Accelerated Publications

Computational Prediction of Structure, Substrate Binding Mode, Mechanism, and Rate for a Malaria Protease with a Novel Type of Active Site†

Sinisa Bjelic and Johan Åqvist*

Department of Cell and Molecular Biology, Uppsala University Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden

Received August 13, 2004; Revised Manuscript Received October 8, 2004

ABSTRACT: The histo-aspartic protease (HAP) from the malaria parasite *P. falciparum* is one of several new promising targets for drug intervention. The enzyme possesses a novel type of active site, but its 3D structure and mechanism of action are still unknown. Here we use a combination of homology modeling, automated docking searches, and molecular dynamics/reaction free energy profile simulations to predict the enzyme structure, conformation of bound substrate, catalytic mechanism, and rate of the peptide cleavage reaction. We find that the computational tools are sufficiently reliable both for identifying substrate binding modes and for distinguishing between different possible reaction mechanisms. It is found that the favored pathway only involves direct participation by the catalytic aspartate, with the neighboring histidine providing critical stabilization (by a factor of ∼10000) along the reaction. The calculated catalytic rate constant of about 0.1 s^{-1} for a hexapeptide substrate derived from the α chain of human hemoglobin is in excellent agreement with experimental kinetic data for a similar peptide fragment.

Genome sequencing projects are producing vast amounts of information in the form of genes with both known and unknown functions. The functional characterization of such data for different organisms is progressing at an increasing rate, and major structural genomic projects are undertaken to further elucidate the detailed mechanisms of action of the proteins involved. The specific targeting of drugs against enzymes is often dependent on knowledge of their 3D structure and catalytic reaction mechanism, and since the structure determination process can still be fairly slow, many pharmaceutically interesting targets will be awaiting structural characterization for some time. In this work we are concerned with theoretically deriving as much information as possible from the amino acid sequence of a malarial enzyme with unknown structure and mechanism using state-of-the-art computational methods. This so-called histo-aspartic protease (HAP)[1] is one of several proteases involved in the hemoglobin degradation pathway that is essential for the survival and reproduction of the most virulent malaria parasite, *Plasmodium falciparum* (*1*).

The erythrocytic cycle of the parasite starts with the invasion of erythrocytes in the host organism, then proceeds by growth and maturation, and ends with the release of new parasites from ruptured blood cells. The proteases involved both in the invasion and rupture of erythrocytes and in the degradation of hemoglobin have emerged as highly promising drug targets for therapies against malaria (*2*). Hemoglobin is a major nutrient source during the intraerythrocytic life stage of *P. falciparum* and is initially processed by four different but homologous proteases, plasmepsins (Plm) I, II, and IV and HAP, located in the acidic food vacuole of the parasite. The plasmepsins are aspartic proteases with the catalytic aspartic residues situated in an active site cleft formed by two domains. The catalytic mechanism of peptide bond cleavage in aspartic proteases is generally considered to involve the activation of a water molecule by one of the aspartates accompanied by nucleophilic attack of the water/hydroxide on the carbonyl carbon of the substrate. The transient tetrahedral intermediate formed in this process is believed to be stabilized by the second aspartate, and subsequent proton transfer to the leaving peptide nitrogen leads to the final bond breaking (*3*). The HAP enzyme, on the other hand, among several substitutions has a histidine at the position of one of the catalytic aspartates, thus presenting a novel type of protease active site (*4*).

While crystallographic structures are available for inhibitor complexes of plasmepsins II (*5*−*7*) and IV (*8*), no structural information has yet been reported for HAP or Plm I. The HAP enzyme has about 60% sequence identity to Plm I, Plm II, and Plm IV, with notable substitutions also in the β-hairpin loop ("flap") that closes with bound substrate, in addition to the active site Asp34 → His mutation. It is generally believed that HAP would follow an aspartic protease-like catalytic mechanism in view of the high sequence identity to the plasmepsins. However, the Tyr77 → Ser mutation in the

† Support from the Swedish Research Council (VR) and Swedish Foundation for Strategic Research (SSF/Rapid) is gratefully acknowledged.

* Corresponding author. Phone: +46 18 471 4109. Fax: +46 18 53 69 71. E-mail: aqvist@xray.bmc.uu.se.

[1] Abbreviations: HAP, histo-aspartic protease; Plm, plasmepsin; MD, molecular dynamics; EVB, empirical valence bond; rmsd, root mean square deviation.

10.1021/bi048252q CCC: $27.50
Published on Web 10/28/2004

active site flap could conceivably place this residue, together with His34 and Asp214, in a configuration that would allow a serine protease mechanism mediated by such a "catalytic triad". The enzyme has also been found to be weakly inhibited by the serine protease inhibitor PMSF, unlike the plasmepsins (*1*). On the other hand, the low pH of the parasite's acidic food vacuole would seem to imply that His34 is protonated, which is not compatible with a serine protease type of mechanism.

In the present work our main objective is to clarify the catalytic action of HAP, and it consists of three main parts: homology modeling of HAP based on its sequence similarity to Plm II, automated docking of a peptide substrate into the modeled structure of the enzyme, and calculations of the catalytic mechanism by molecular dynamics (MD), empirical valence bond (EVB), and free energy perturbation simulations. Having constructed a homology model, the MD simulations are performed within a solvated sphere (diameter 44 Å) centered on the active site of HAP. While this procedure involves restraining the outer parts of the enzyme, it is justified by the high degree of sequence identity between HAP and Plm II, and the resulting backbone deviations for the freely moving parts within the simulation sphere are also found to be small (see below). It can further be noted that the average deviation of core backbone atoms for experimentally determined structures with the same homology is 0.83 Å (*9*).

The reaction mechanism of the HAP model is investigated with the EVB method (*10*, *11*) that treats chemical reactions in terms of a number of resonance structures or valence bond states. The efficiency of the method lies in the possibility to calibrate the potential energy surface of the uncatalyzed reaction in solution against available experimental data or high level ab initio calculations. By carrying out simulations of the same reaction in the active site of the solvated enzyme using the earlier calibrated potential energy surface, the effect of the environment can be examined by comparing the solution and enzyme reactions as well as different related enzymes. In general, for base-catalyzed peptide bond hydrolysis in solution, as well as in the serine proteases, the rate-limiting chemical step is considered to be associated with the first high-energy tetrahedral intermediate, i.e., leading to the acyl-enzyme in serine proteases. Also in the case of aspartic proteases, the limiting chemical step in the favored mechanistic proposals is associated with the tetrahedral intermediate (*12*, *13*). Nevertheless, the activation barrier involved in the breakdown of the tetrahedral intermediate is believed to be approximately of the same height as for its formation, and we therefore evaluate the energetics of the total peptide hydrolysis reaction catalyzed by HAP for different substrate binding modes and protonation states of the active site histidine residue.

A previous model of the HAP structure was reported in ref *4*, but it was only discussed in a qualitative way regarding the position of His34, and no attempts to use the model for docking or elucidation of the enzyme mechanism were reported. In fact, the present work is the first to our knowledge where the substrate binding mode, catalytic mechanism, and reaction energetics of an enzyme are calculated on a microscopic level starting directly from the amino acid sequence.

MATERIALS AND METHODS

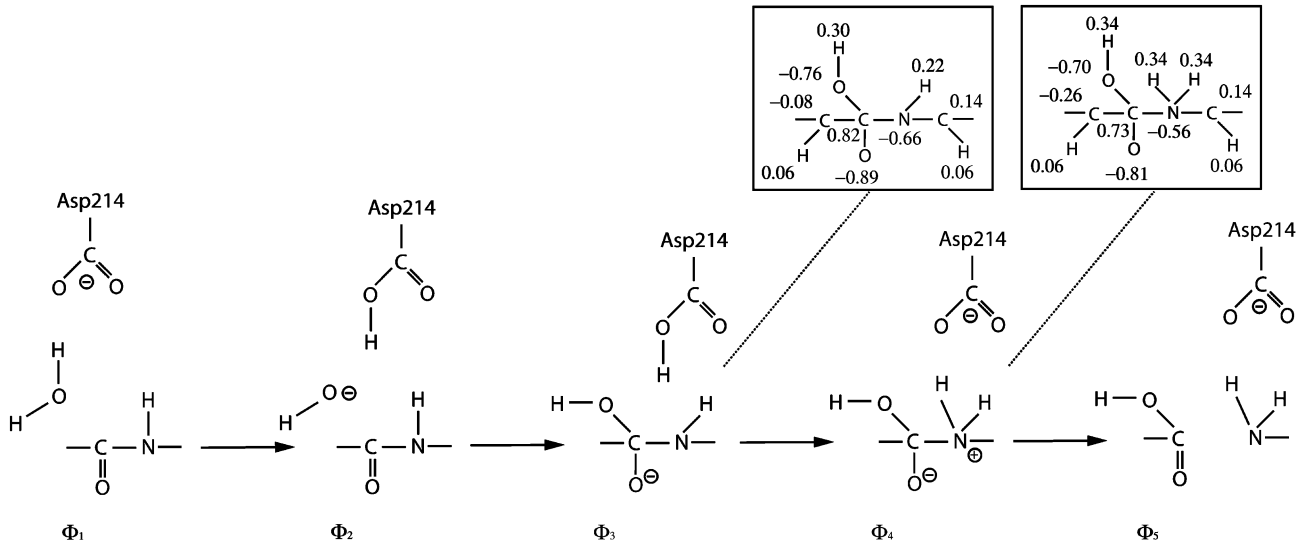
The HAP homology modeling made use of Plm II structures with PDB codes 1LEE, 1LF2, 1LF3, and 1LF4 (*6*, *7*) and employed the automated SWISS-MODEL server (*14*). MD simulations both for refinement of the initial model and for subsequent reaction profile calculations were carried out with the program Q (*15*) using the OPLS-AA force field (*16*). Since no major deviations in the overall protein fold and backbone conformation are expected with the present level of sequence identity (~60%), the MD simulations were done with a solvated spherical system of radius 22 Å centered on the active site Asp214 $O_{\delta 2}$ atom. Atoms within this sphere moved freely while protein atoms outside the sphere were restrained by a 200 kcal/(mol Å^2) force constant to their initial positions. The water surface was subjected to radial and polarization restraints according to our version of the SCAAS model (*15*, *17*). Long-range electrostatic interactions beyond 10 Å were treated with the local reaction field method (*18*) while short-ranged Lennard-Jones forces were truncated with a 10 Å cutoff. Prior to equilibration the system was slowly heated from 1 to 300 K with an initial increase of the MD time step from 0.01 fs to a final step size of 1 fs.

Docking was done with the AutoDock3 program (*19*) with a six-residue flexible peptide substrate (sequence: RMFLSF), after stripping the final MD equilibrated model of its water molecules [except for the crystallographically observed catalytic water in Plm II (*7*)]. Grid maps were calculated with 50 × 92 × 62 points and 0.375 Å spacing centered close to the active site residue Asp214. Two runs were performed with 10 million (Lamarckian genetic algorithm) generations, 10 million energy evaluations, a population size of 250, and 150 runs as recommended for blind docking (*20*). The resulting 300 conformations of the ligand were clustered with a root mean square deviation (rmsd), for all atoms in the docked ligand, with a tolerance of 4 Å, which was found to be optimal due to the large flexibility of the ligand. Six docked conformations belonging to the two top-ranking, most populated, clusters were used for subsequent EVB/MD simulations.

The reaction free energy surface was described by the EVB model (*10*, *11*) in terms of five basic resonance structures (Scheme 1) and iteratively fitted to available energetic data for the uncatalyzed (reference) reaction in water. Free energy profiles in aqueous solution both for the stepwise and for the concerted formation of the tetrahedral intermediate, i.e., the attack of a water molecule/hydroxide ion on the substrate Phe−Leu peptide bond with an Asp side chain abstracting a proton from the nucleophile, were initially determined. The resonance or valence bond structures used in the stepwise formation of the tetrahedral intermediate are $\Phi_1 \rightarrow \Phi_2 \rightarrow \Phi_3$ (Scheme 1). Similarly, the concerted pathway is represented by $\Phi_1 \rightarrow \Phi_3$ (i.e., along the "diagonal" on the surface spanned by the two reaction coordinates of the stepwise scheme). Each resonance structure is associated with an analytical potential energy function (force field), describing the bonded and nonbonded interactions of the reacting fragments and the surrounding enzyme/solvent system.

A key feature of the EVB method is the possibility to calibrate the potential which involves determining the gas-phase energy differences, $\Delta\alpha_{ij}$, as well as off-diagonal matrix elements, H_{ij}, between pairs of VB states (*10*, *11*). Here, we

Scheme 1: EVB Resonance Structures (Φ_i) Used for Modeling the HAP Reaction Mechanism



use data from refs *21*−*23* to obtain reliable estimates of the energetics associated with forming the tetrahedral intermediate through attack of a water molecule on the carbonyl carbon of the substrate with an aspartate side chain acting as a general base. The free energy difference in water between states Φ_1 and Φ_2 (proton transfer) is 15 kcal/mol from the pK_a difference between water and an Asp side chain, with an intervening barrier of about 18 kcal/mol (*24*, *25*). The cost of forming the tetrahedral intermediate by attack of OH^-, i.e., the free energy difference between Φ_2 and Φ_3, can be estimated to be 12 kcal/mol (*21*). The corresponding activation free energy for this step is obtained both from experiments (*22*, *23*) and from calculations (*21*) as approximately 18 kcal/mol when making the correction for bringing the reactants from a 1 M standard state into a solvent cage with the same 55 M concentration as surrounding water (*10*). The free energy surface for this type of reaction in solution has been found to be relatively flat in the transition state region (*21*), which implies that the stepwise and concerted pathways have similar energetics and that the above calibration can be used for both of these alternatives. The overall barrier for reaching the first tetrahedral intermediate is thus estimated here to be 33 kcal/mol.

The formed tetrahedral intermediate is a short-lived species, and the breakdown proceeds through the two steps in which the neutral Asp donates a proton to the nitrogen of the scissile peptide bond, which then spontaneously dissociates to an amine and a carboxylic acid moiety. As previously, we investigated the possibility that the reaction is either concerted or stepwise, proceeding through $\Phi_3 \rightarrow \Phi_5$ or $\Phi_3 \rightarrow \Phi_4 \rightarrow \Phi_5$, respectively (Scheme 1). The reaction free energy for the proton transfer from the protonated aspartate to the nitrogen of the tetrahedral intermediate, i.e., from Φ_3 to Φ_4, is about −5 kcal/mol and corresponds to a pK_a difference of 3.5 units between the aspartate side chain (pK_a = 4.5) and protonated nitrogen of the tetrahedral intermediate (pK_a = 8) (*21*). The activation barrier of ∼2 kcal/mol can be obtained as previously from Eigen's data (*24*, *25*), taking into account that it costs 2.4 kcal/mol to bring the reactants into the same solvent cage due to the entropic contribution, $\Delta G_{cage} = -RT \ln(1/55)$ (*10*). The proton transfer to the tetrahedral intermediate is followed by the breakdown of the scissile bond, $\Phi_4 \rightarrow \Phi_5$, with an activation barrier of 3 kcal/mol and a reaction free energy difference of −12 kcal/mol (*21*). The overall activation and reaction free energy for concerted breakdown of the tetrahedral intermediate are thus 2 and −17 kcal/mol, respectively, adding together the corresponding energies in the stepwise mechanism. The solution reaction was calibrated to reproduce the relevant free energy differences associated with the reaction steps above, yielding a total of 11 points on the free energy surface to be fitted. The calibration is supported also by experiments on the kinetics of the *N*-arylcarbamates (*26*), taking into account that the pK_a of the tetrahedral intermediate nitrogen is 8 (*21*) while the general acid (Asp214) has the pK_a about 4 units lower and that the pK_a of the leaving group is high (∼9). Under these conditions the rate-controlling step is the nitrogen protonation while the C−N bond dissociation is fast, thus implying a lower activation barrier (*26*). It can also be noted that proton transfer to anilide acetals, which are more representative of the reaction studied here, and their subsequent cleavage is believed to proceed similarly to that of the *N*-arylcarbamates (*27*).

The force field parameters used here were the standard OPLS-AA ones (*16*) except for the charges of the tetrahedral intermediates, Φ_3 and Φ_4, that were derived by ab initio HF/6-31G** calculations. These charges are given in the insets to Scheme 1. To fit the 11 points on the overall EVB reaction energy surface in water, we have 10 relative free energies to calibrate. This fitting is accomplished with the gas-phase energy differences, $\Delta\alpha_{ij}$, together with the off-diagonal matrix elements, H_{ij}, between VB states (*10*, *11*), and the resulting values of these parameters are given in Table 1. It should be noted that the 10 EVB parameters thus provide a unique solution to the fitting problem and that, once the uncatalyzed solution surface is determined, there are no adjustable parameters in the EVB model.

Reaction free energy profiles for different pathways were calculated with the free energy perturbation/umbrella sampling method (*10*, *11*) following 250−300 ps of MD equilibration of the enzyme−substrate complexes. The free energy perturbation calculations each comprised 41 discrete steps with a 4 ps trajectory generated at each step (e.g., the stepwise path from reactants to products consists of four

Table 1: Reaction Free Energy Surface Parametrization Constants[a]

$\Phi_1 \rightarrow \Phi_2 \rightarrow \Phi_3$[b]					$\Phi_3 \rightarrow \Phi_4 \rightarrow \Phi_5$				
H_{12}[c]	α_2[d]	H_{23}	α_3	H_{13}	H_{34}	α_4	H_{45}	α_5	H_{35}
38.5	46.8	81.4	191.6	116.1	78.9	190.5	60.8	−53.1	127.3

[a] The reaction surface for the water reaction is determined explicitly by a set of constant parameters (in kcal/mol). [b] Φ_i represents the resonance structures used in the EVB method. [c] H_{ij} is the off diagonal matrix element between the corresponding diabatic surfaces. [d] α_i is the gas-phase energy difference between the resonance structures Φ_{i+1} and Φ_0 at infinite separation. These energy constants are added to the molecular mechanics potential function; i.e., $H_{ii} = V_{ii} + \alpha_i$. Morse potentials for bonds within the reacting fragments are taken from ref *37*.

separate simulations of this kind). In practice, simulations were initiated from the tetrahedral intermediate and carried out toward reactants and products. After calibration of the reaction surface for the peptide, the attacking water molecule, and an aspartate ion in water, the energetics of the same process in the solvated active site environment of the enzyme was evaluated by the same procedure without changing the earlier calibration (*10*, *11*). By carrying out multiple simulations and by using somewhat different initial structures, the precision of the calculated free energy values could be estimated to be about ±1.5 kcal/mol, which is satisfactory in view of the fact that we are dealing with computationally derived initial structures.

RESULTS AND DISCUSSION

The initial homology model of HAP was generated by an automated protein modeling server (*14*) against four different structures of Plm II (PDB codes 1LEE, 1LF2, 1LF3, and 1LF4) (*6*, *7*). This model was further refined by molecular dynamics (MD) simulations after positioning Asp214 as in Plm II and Plm IV, optimizing both the neutral and protonated forms of the His34 side chain, and solvating the system. With His34 charged, one water molecule is located approximately at the position of the "catalytic" water molecule observed in 1LF4 (the other Plm II complexes all have an inhibitor hydroxyl group at the same location) (*5*−*7*). The final equilibrated MD structure with His34 protonated has all-atom and C_α rms deviations for atoms within 16 Å from Asp214 of only 0.91 and 0.55 Å, respectively, relative to the starting model, illustrating its robustness. Already at this stage did it become clear that no stable structure emerges with a putative catalytic triad, involving the unprotonated His34, in a reasonable configuration for catalysis. Moreover, the essential oxyanion hole found in serine proteases has no counterpart in our generated HAP models, and we therefore conclude that there is no structural basis for a serine protease type of mechanism. Furthermore, as mentioned above, the pH of the parasite vacuoles is acidic and the pH−rate profile for HAP also shows that the enzyme becomes inactive in the pH region where His34 would be neutral (*1*), which strongly suggests that this residue should be protonated for catalysis.

The resulting enzyme structure with His34 protonated was then used for automated docking simulations (*19*) where a flexible six amino acid peptide substrate (Arg-Met-Phe-Leu-Ser-Phe), corresponding to residues 31−36 of the human hemoglobin α chain, was docked with the (rigid) enzyme.

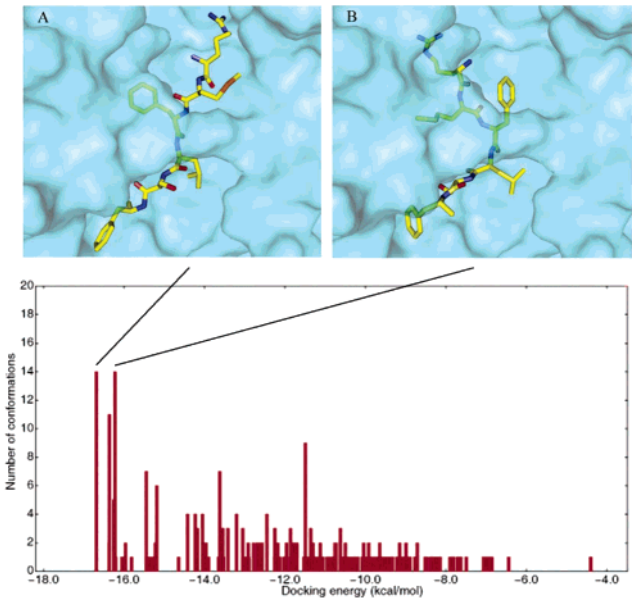


FIGURE 1: Results from automated docking (*19*) calculations where each bar (lower panel) represents the energy and population of a conformational cluster. Representative conformations of the two most populated clusters, denoted A and B, are also shown along with the semitransparent enzyme surface (upper panel).

A relatively large grid (19 × 34 × 23 Å) was used for these searches, and no strong bias regarding the location of the substrate binding was thus imposed. These simulations yielded 300 conformations of the enzyme−substrate complex that were clustered and ranked according to their predicted binding score (Figure 1).

Remarkably, two unique low-energy solutions (A and B) appear from the docking calculations, both of which place precisely the Phe−Leu scissile bond at attacking distance from the catalytic water molecule observed in aspartic proteases (Figure 1). These solutions do not only have the lowest docking energy but also are the most populated. The clustering of the docked conformations with an rmsd criterion of 4.0 Å gives many different clusters, and most of them are populated with just one or two conformations, indicating that the use of a lower rmsd criterion is not meaningful. On the other hand, clustering with a larger rmsd of 4.5 Å assigns different conformations to the same cluster; i.e., the side chains of the substrate do not bind uniquely into the explicit subsites. Hence, the 4.0 Å criterion seems appropriate, and although many clusters are obtained, there are clearly a few of them that stand out as particularly significant. A closer inspection of the cluster with the lowest docking energy reveals the substrate bound with Phe33 and Leu34 in the S1 and S1′ subsites, respectively (Figure 1A). This is in accordance with the expected mode of binding in Plm II and is supported by both experimental studies (*5*−*7*, *28*) and our earlier theoretical calculations on inhibitor binding to Plm II (*29*). It is thus quite encouraging that AutoDock manages to dock a rather large substrate consisting of six amino acids into a vaguely defined active site, obtaining a seemingly accurate solution, and also assigns it the lowest docking energy.

The second most populated low-energy cluster gives a solution where the S1 subsite is occupied by Met32 (Figure 1B), instead of by Phe33 as in conformation A. Although this alternative binding conformation (B) has a slightly higher

docking energy (<0.2 kcal/mol) than two other intervening clusters, with docking energies between A and B, it is more populated (the less populated intervening clusters also, in fact, have structures similar to either A or B). The interesting aspect with this binding mode is that it resembles the conformation of the corresponding amino acids in native hemoglobin. Compared to the corresponding sequence in the crystal structure of the hemoglobin α chain (*30*), the four middle amino acids of the substrate have an rmsd of approximately 1 Å for the C_α atoms. The fact that Plm II initially cleaves native hemoglobin at the α33−34 position may suggest that Met32 in that case could be positioned in the S1 subsite, depending on whether the local tertiary structure of hemoglobin is retained in the complex.

Since we find two possible low-energy binding modes for the substrate, both of them were considered in subsequent computer simulations of the catalytic reaction pathways. Initial calculations on the enzyme−substrate complex were performed for six distinct substrate conformations, three within the conformation A cluster and three within the B cluster. Since the major deviation between conformers belonging to the same cluster is the orientation of the substrate side chains, especially at the ends of the peptide, the initial criterion for choosing a particular conformer is their optimal positioning for reaction. In principle, conformers belonging to the same cluster should converge after extensive simulation to the same average structure, but this could require very long calculations since our clustering criterion is rather loose. Several of the initially examined conformations were also found to yield less effficient catalysis and could therefore be discarded, thus enabling us to focus on the most active ones.

Modeling of the catalytic mechanism of HAP was done with the empirical valence bond (EVB) method in combination with MD free energy perturbation simulations (*10*, *11*). Data from high level ab initio calculations including solvation effects (*21*), as well as experiments (*22*, *23*, *26*), were used for calibration of the acid−base-catalyzed reaction potential energy surfaces in aqueous solution as described above. This procedure involves (iteratively) fitting a set of valence bond resonance structures to the reference data and mixing these states to obtain the ground-state surface, from which the reaction free energy surface is constructed through the free energy perturbation/umbrella sampling technique (*10*, *11*). The modulation of this solution surface, which accurately reproduces experimental data, by the enzyme can then be directly quantified by "moving" the reacting groups from solution into the HAP active site. In determining the most favorable catalytic mechanism, we considered both stepwise and concerted pathways for hydrolysis of the Phe−Leu peptide bond by a water molecule assisted by general acid−base catalysis from Asp214 alone. We also examined a possible pathway where His34 carries out the general acid function and protonates the transient tetrahedral intermediate to effect its decomposition into products. Furthermore, the case with His34 in its unprotonated form was considered as a control in order to further assess the role of this residue in the reaction.

The free energy surfaces from EVB simulations of both the solution and enzyme reactions are summarized in Figure 2, where the enzyme−substrate complex has the substrate bound in conformation A. The acetate-"catalyzed" process in water (Figure 2, upper surface) corresponds to a very slow reaction with $k_{ace} \sim 10^{-12}$ s^{-1} ($\Delta G^{\ddagger}_{ace} = 33$ kcal/mol). This rate is, in fact, somewhat lower than that estimated for neutral water hydrolysis of a peptide bond (*31*) (and for water hydrolysis through an OH^- mechanism at pH 7), which is in agreement with no detection of catalysis by acetate for the solution reaction (*31*). From the lower free energy surface of Figure 2 it can immediately be seen that HAP exerts a large catalytic effect on peptide cleavage with an overall activation barrier of only 17−20 kcal/mol. This corresponds to a rate enhancement by a factor of 10^{11} and an absolute value of $k_{cat} \sim 0.1$ s^{-1}, estimated from standard transition state theory, which is in very good agreement with the experimental measurements for an octapeptide substrate containing the same sequence (*1*). It can further be seen that a stepwise hydrolysis pathway appears more favorable than the concerted path. Figure 3 shows a representative MD structure of the HAP active site with the bound tetrahedral intermediate, where the characteristic H-bond between the catalytic Asp and the Thr (or Ser) residue in position $i + 3$, observed in most aspartic proteases, can also be seen. In fact, the MD simulations predict that the H-bond network involving the Asp214-Ser-Ala-Thr sequence in HAP preserves the very same H-bonding pattern as found in the active sites of aspartic proteases (*3*).

The role of His34 was first examined by considering the possibility that this residue could be neutral in the active site microenvironment (*1*). As can be seen from the dashed curve of Figure 2, this has a significant destabilizing effect on catalysis and is predicted to slow the rate by about a factor of 10000 ($\Delta\Delta G^{\ddagger}_{cat} = 5.3$ kcal/mol). The reason for this is a strong stabilizing interaction between His34 and the water/hydroxide nucleophile as well as the developing negative charge on the substrate carbonyl oxygen (Figure 3). A similar deleterious effect would thus be predicted for mutation of His34 to a nonpolar amino acid. As discussed above, one might imagine that a neutral histidine could instead serve as the catalytic base (*1*). However, such a proposal is at variance with the pH−rate profile for HAP which has a typical bell-shaped optimum around pH 5.7, where both Asp214 and His34 would be charged, and the rate falls off at higher pH values with a characteristic pK_a matching that of a histidine residue (*1*). Furthermore, it is clear that such a mechanism would inevitably lead to a destabilization of the nucleophile by the negative charge on Asp214 since the two (negative) charges will be close to each other.

A more reasonable mechanistic alternative could be that the charged form of His34 directly participates in the catalytic reaction and effects decomposition of the transient intermediate by protonating the leaving group nitrogen. However, in our structural model His34 is not optimally positioned for this task (Figure 3), and this is reflected by the calculated free energy profile for this mechanism which is estimated here to have an activation barrier of at least 25 kcal/mol (gray arrow in Figure 2). Hence, we conclude that the most favorable mechanism is that Asp214 performs both catalytic acid−base functions and that His34 stabilizes this pathway through a strong interaction with its positive charge.

We also evaluated free energy profiles for the above mechanism with the substrate bound in conformation B. Surprisingly, the reaction surface for this conformation is very similar to that for conformation A, with an overall

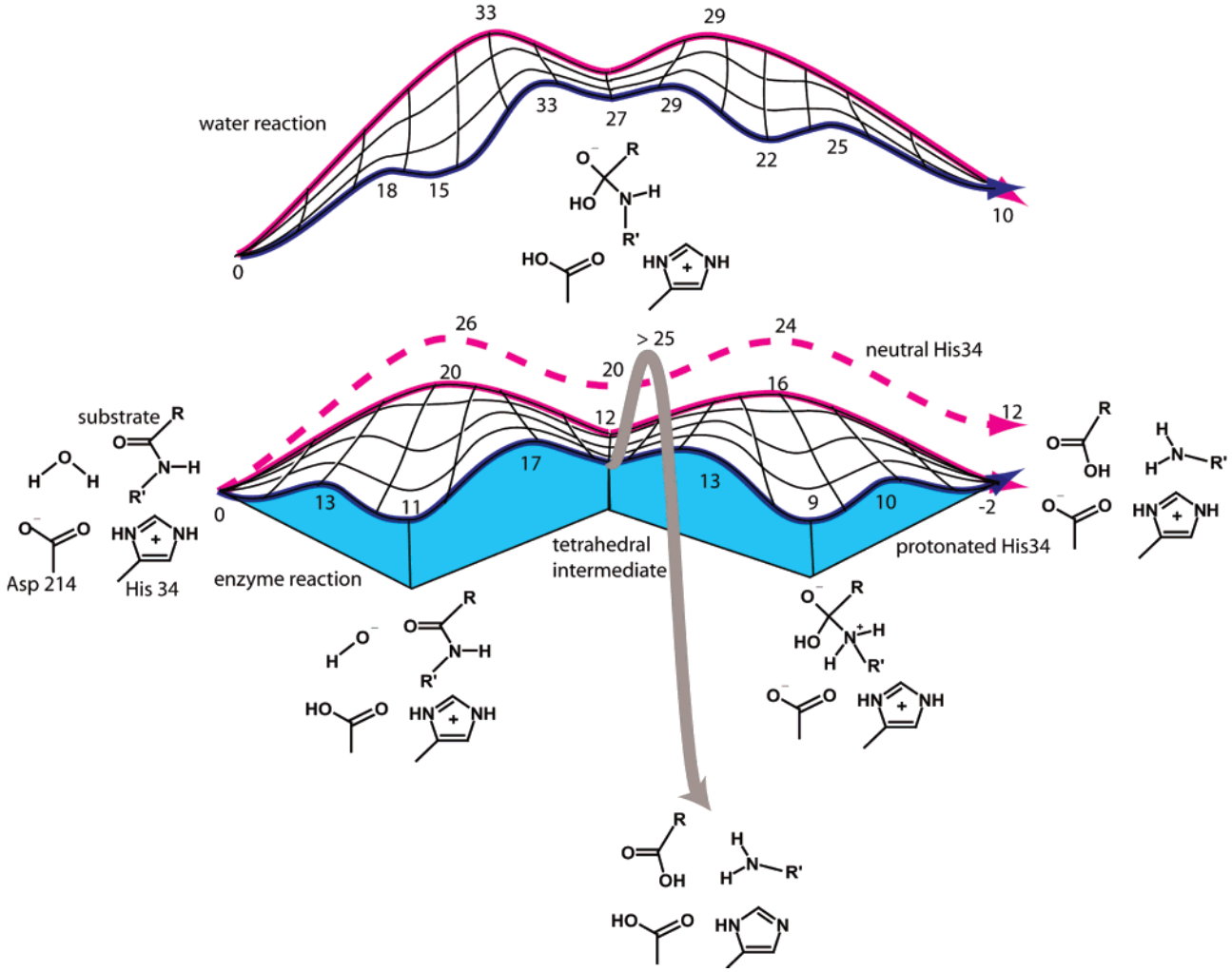


FIGURE 2: Calculated free energy surfaces (kcal/mol) for the peptide hydrolysis reaction in aqueous solution, calibrated as described in the text (upper surface), and in the solvated HAP enzyme (lower surface). The front (blue) and back (pink) paths on the surfaces correspond to stepwise and concerted reactions, respectively, and relevant chemical structures are indicated. The dashed (concerted) profile denotes the case with a neutral His34, while the gray arrow denotes a path where the charged His34 protonates the tetrahedral intermediate instead of Asp214.

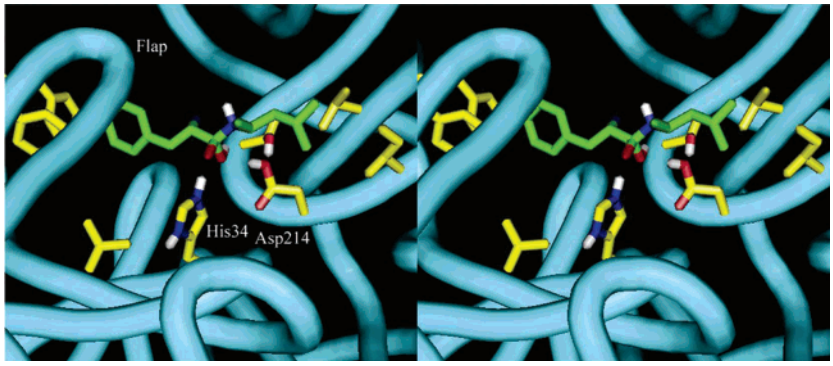


FIGURE 3: Stereoview of the active site of HAP, with the central part of the bound substrate, in the transient tetrahedral intermediate state of the catalytic reaction. In addition to His34 and Asp214, some hydrophobic residues of the active site together with Thr217 (interacting with Asp214) are also shown.

barrier of about 19 kcal/mol (data not shown). It is thus not possible to distinguish which is the most reactive conformation on the basis of calculated k_{cat} values alone, but it will rather be the relative binding affinity of the two conformers that determines the active substrate binding mode. In this respect it can be noted that the docking results point to conformation A as having the higher affinity, although the energy difference is small and the estimate from the empirical scoring function is probably not very accurate. It is still difficult to obtain accurate estimates of binding free energies for large ligands, such as these, from direct microscopic simulations, and to further address this issue, we first examined two additional scoring functions (*32*, *33*). Both of these again give the result that the two conformations have similar affinities for HAP. However, empirical scoring functions usually do not treat intramolecular ligand strain in

an optimal way, and we therefore carried out more accurate force field minimizations of the enzyme−substrate complex for the two conformations. This showed that conformation B has a significantly higher internal (substrate) energy than A by about 12 kcal/mol. If one adds solvation effects to the isolated substrate conformations with the generalized Born treatment (*34*), this energy difference drops to about 8 kcal/mol, but the prediction is clearly that conformation A is the most stable in solution and that it costs a nonnegligible amount of energy to convert a short peptide substrate to conformation B.

Experiments also seem to indicate that HAP is more active on extended substrate conformations as it shows a higher activity against a hemoglobin peptide fragment as well as acid-denatured globin compared to the native human hemoglobin (*1*). The latter experiments are not entirely unambiguous, since a combination of HAP and Plm II was found to cleave the native protein faster than either enzyme alone (*1*), but the general picture emerging is that HAP may act downstream of Plm I and Plm II in the hemoglobin degradation process (*1*).

Regarding inhibition of HAP, only one weakly active compound has so far been reported (*35*), while for Plm I and Plm II, a number of very potent inhibitors have been synthesized (*5*, *28*, *36*). With inhibition constants in the nanomolar region, some of these compounds have also shown promising efficacy on infected red blood cells (*36*). On the other hand, a significant effect on parasite growth has also, surprisingly, been detected for a newly synthesized compound that showed no inhibition of Plm I or Plm II, which may suggest an interaction with HAP or Plm IV instead (*36*). At any rate, it is clear that structural and mechanistic information on the intriguing HAP enzyme can contribute decisively to such inhibitor design efforts. We have shown here how a combination of state-of-the-art computational methods, including homology modeling, automated docking, and reaction simulations, can essentially yield a complete picture of the function of a new enzyme, provided that there is sufficient sequence homology to known structures. Most parts of this computational scheme also lend themselves to a high degree of automation which, in principle, can allow the exploration of a multitude of structural and mechanistic alternatives.

REFERENCES

1. Banerjee, R., Liu, J., Beatty, W., Pelosof, L., Klemba, M., and Goldberg, D. E. (2002) Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine, *Proc. Natl. Acad. Sci. U.S.A. 99*, 990−995.
2. Rosenthal, P. J. (1998) Proteases of malaria parasites: New targets for chemotherapy, *Emerging Infect. Dis. 4*, 49−57.
3. Dunn, B. M. (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases, *Chem. Rev. 102*, 4431−4458.
4. 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*, 149−154.
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*, 10034−10039.
6. 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. D 58*, 2001−2008.
7. 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*, 173−181.
8. Asojo, O. A., Gulnik, S., Afonina, E., Randad, R., and Silva, A. (2004) Crystal structure of plasmepsin IV from *P. Falciparum* in complex with pepstatin A, manuscript in preparation.
9. Chothia, C., and Lesk, A. M. (1986) The relation between the divergence of sequence and structure in proteins, *EMBO J. 5*, 823−826.
10. Warshel, A. (1991) *Computer modeling of chemical reactions in enzymes and solutions*, John Wiley & Sons, New York.
11. Aqvist, J., and Warshel, A. (1993) Simulation of enzyme-reactions using valence-bond force-fields and other hybrid quantum-classical approaches, *Chem. Rev. 93*, 2523−2544.
12. Davies, D. R. (1990) The structure and function of the aspartic proteinases, *Annu. Rev. Biophys. Biophys. Chem. 19*, 189−215.
13. Hyland, L. J., Tomaszek, T. A., and Meek, T. D. (1991) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism, *Biochemistry 30*, 8454−8463.
14. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res. 31*, 3381−3385.
15. Marelius, J., Kolmodin, K., Feierberg, I., and Aqvist, J. (1998) Q: A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems, *J. Mol. Graphics 16*, 213−225.
16. Jorgensen, W. L., Maxwell, D. S., and TiradoRives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc. 118*, 11225−11236.
17. King, G., and Warshel, A. (1989) A surface constrained all-atom solvent model for effective simulations of polar solutions, *J. Chem. Phys. 91*, 3647−3661.
18. 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*, 3100−3107.
19. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem. 19*, 1639−1662.
20. Hetenyi, C., and van der Spoel, D. (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site, *Protein Sci. 11*, 1729−1737.
21. Strajbl, M., Florian, J., and Warshel, A. (2000) Ab initio evaluation of the potential surface for general base-catalyzed methanolysis of formamide: A reference solution reaction for studies of serine proteases, *J. Am. Chem. Soc. 122*, 5354−5366.
22. Deraniyagala, S. A., Adediran, S. A., and Pratt, R. F. (1995) Beta-secondary and solvent deuterium kinetic isotope effects and the mechanisms of base-catalyzed and acid-catalyzed hydrolysis of penicillanic acid, *J. Org. Chem. 60*, 1619−1625.
23. Slebocka-Tilk, H., Neverov, A. A., and Brown, R. S. (2003) Proton inventory study of the base-catalyzed hydrolysis of formamide. Consideration of the nucleophilic and general base mechanisms, *J. Am. Chem. Soc. 125*, 1851−1858.
24. Eigen, M. (1964) Proton transfer, acid−base catalysis, and enzymatic hydrolysis. Part I: Elementary processes, *Angew. Chem., Int. Ed. 3*, 1−72.
25. Åqvist, J. (1997) in *Computational approaches to biochemical reactivity* (Náray-Szabó, G., and Warshel, A., Eds.) pp 341−362, Kluwer Academic Publishers, Dordrecht, The Netherlands.
26. Johnson, S. L., and Morrison, D. L. (1972) Kinetics and mechanism of decarboxylation of *N*-arylcarbamates. Evidence for kinetically important zwitterionic carbamic acid species of short lifetime, *J. Am. Chem. Soc. 94*, 1323−1334.
27. McClelland, R. A., and Patel, G. (1981) Kinetics and mechanism of the hydrolysis of anilide acetals, *J. Am. Chem. Soc. 103*, 6908−6911.
28. 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*, 1428−1440.
29. 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*, 3723−3733.
30. Luisi, B., and Shibayama, N. (1989) Structure of hemoglobin in the deoxy quaternary state with ligand bound at the alpha-hemes, *J. Mol. Biol. 206*, 723−736.
31. Radzicka, A., and Wolfenden, R. (1996) Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases, *J. Am. Chem. Soc. 118*, 6105−6109.
32. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., and Mee, R. P. (1997) Empirical scoring functions. 1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, *J. Comput. Aided Mol. Des. 11*, 425−445.
33. Wang, R. X., Lai, L. H., and Wang, S. M. (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction, *J. Comput.-Aided Mol. Des. 16*, 11−26.
34. Qiu, D., Shenkin, P. S., Hollinger, F. P., and Still, W. C. (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii, *J. Phys. Chem. A 101*, 3005−3014.
35. 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*, 8459−8464.
36. 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*, 110−122.
37. Vagedes, P., Rabenstein, B., Åqvist, J., Marelius, J. and Knapp, E.-W. (2000) The deacylation step of acetylcholine esterase: computer simulation studies, *J. Am. Chem. Soc. 122*, 12254−12262.

BI048252Q