JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

The Catalytic Machinery of Rhomboid Proteases: Combined MD and QM Simulations

Neta Uritsky, Michael Shokhen,* and Amnon Albeck*

The Julius Spokojny Bioorganic Chemistry Laboratory, Department of Chemistry, Bar Ilan University, Ramat Gan 52900, Israel

Supporting Information

ABSTRACT: Rhomboid proteases are a ubiquitous family of intramembrane serine proteases in prokaryotic and eukaryotic organisms that cleave membrane proteins in their transmembrane region. Their catalytic activity is centered at a His−Ser catalytic dyad. We applied molecular dynamics and quantum mechanics calculations in order to clarify the protonation state of the catalytic residues of *E. coli* GlpG rhomboid protease and how it is affected by the immersion of the enzyme in the membrane. We identified (Nε)H150^{dpr}_H254^{dpr}_S201^{pr} as the protonation (and H150 tautomeric) state of free GlpG in both lipid-solubilized and membrane environments. We used our MD-QM/SCRF(VS) computational protocol to rationalize and predict the trend of pK_a change caused by the decrease of water exposure of the active site of GlpG due to ligand binding. The catalytic diad of lipid-solubilized GlpG exists as an H254(+)_S201(−) ion pair at the Michaelis complex stage, with Ser201 ready for nucleophilic attack on the substrate. Therefore, deprotonation of S201 does not contribute to the activation barrier of covalent tetrahedral complex formation. In contrast, both catalytic residues, H254 and S201, are neutral in the Michaelis complex of GlpG in the membrane. Therefore, S201 deprotonation by H254 general base catalysis should contribute to the activation barrier of the covalent tetrahedral complex formation.

INTRODUCTION

Rhomboid proteases (EC 3.4.21.105) are a ubiquitous family of intramembrane serine proteases in prokaryotic and eukaryotic organisms that usually cleave membrane proteins in their transmembrane (TM) or juxtamembrane region.[1−3] The biological functions of this widely conserved protease family range from growth factor signaling to the regulation of mitochondrial dynamics and to parasitic invasion.[4] Recent progress in the reconstitution of rhomboid proteases' proteolytic activity from purified recombinant proteins *in vitro* enabled mechanistic questions to be addressed and inhibitors to be tested.[5,6] Despite biochemical advances, structures of GlpG rhomboids from two bacteria only have been solved by X-ray crystallography, namely, two from *H. influenza*[7,8] and, from *E. coli*, six structures of the free enzyme[9−15] and three complexes with covalent inhibitors.[14−17] The active site of rhomboid proteases is membrane-immersed about 10 Å below the presumed membrane surface but located in a cavity partially opened to periplasmic water.[18]

Interpretation of the catalytic machinery of rhomboid proteases is based on the available crystal structures, analysis of conserved residues, and reactivity of wild-type vs mutant enzymes. The current consensus is that the catalytic activity is centered at the His254−Ser201 catalytic dyad,[5,7−22] rather than the more common Asp−His−Ser catalytic triad in water solvated serine proteases.[23] Three molecular dynamics simulations of rhomboid proteases were published. One was applied to the structural analysis of the wild-type and a mutant *E. coli* GlpG free enzyme, immersed in different lipid bilayers simulating a native-like membrane environment.[24] Another study combined molecular dynamics of free wild-type GlpG and experimental enzyme activity of wild-type and mutant GlpG, identifying the importance of the side chains of Q189 and S185 in internal water retention, essential for peptide bond hydrolysis.[25] The third study is a coarse-grained molecular dynamics simulation of the gating of the transmembrane domain of the substrate Spitz into GlpG rhomboid protease.[26] All these MD simulations ignore the importance of the protonation state of catalytic residues and proton exchange to the mechanism of rhomboids.

The native environment of rhomboid proteases is the lipid bilayer of the cell membrane. However, all 3D structures of rhomboids were determined for detergent- or lipid-bicelle-solubilized enzymes. This may lead to structural artifacts since the hydrophobic parts of the rhomboid are screened from water by only a thin annulus of lipids and detergent molecules.[15] Nevertheless, experimental proteolytic activity and substrate specificity *in vitro* of detergent-solubilized rhomboids indicates that the membrane environment does not have a critical effect on intramembrane proteolysis.[5,6] On the other hand, most potent small molecule inhibitors of serine proteases are inefficient for rhomboids *in vivo*,[19] but this could be due to impermeability through the bacterial outer membrane. It was demonstrated recently that series of monocyclic β-lactams can inhibit mammalian and bacterial rhomboids, and they are effective both *in vitro* and *in vivo*.[27] The present work addresses the identification of the catalytic machinery of rhomboid proteases and how membrane immersion affects it.

METHODS

All molecular graphics, editing, and MD simulations were conducted using the YASARA Structure software package.[28] Quantum mechanical calculations were performed using the

Received: May 13, 2012
Published: September 18, 2012

   dx.doi.org/10.1021/ct3003767 | J. Chem. Theory Comput. 2012, 8, 4663−4671

semiempirical PM6 Hamiltonian[29] implemented in MOPAC2009.[30]

MD Simulation. The 2xov.pdb crystal structure of free GlpG was used as the initial point of the modeling. All cocrystallized lipids were deleted. The standard YASARA scripts were used for MD simulations of the protein in membrane and the lipid-solubilized protein in water. In the former—GlpG in membrane—the periodic rectangular simulation cell was extended on each side of the protein in the membrane plane (*XZ* plane) by 30 Å and by 20 Å on each side of the protein (filled by solvent water molecules) along the third axis (*Y* axis; Figure 1a,b). YASARA builds a membrane

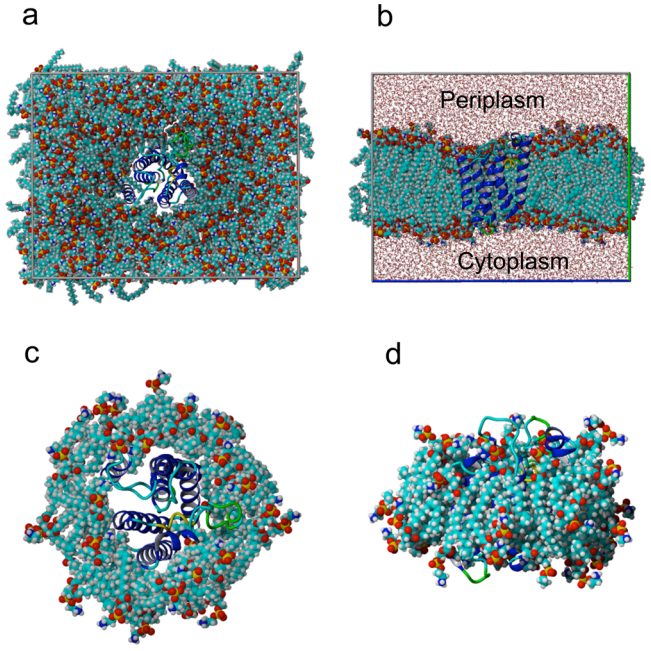

Figure 1. (a) Frontal view from the periplasm of free GlpG in the membrane in the MD simulation cell. Solvent water molecules are hidden. (b) A lateral view of GlpG in the membrane. The protein can be observed by using a clipping plane that slices the lipid membrane and solvent water molecules, showing only the back half of the environment. (c) Frontal view from periplasm of free lipid-solubilized GlpG. The simulation cell and solvent water molecules are hidden. (d) A lateral view of lipid-solubilized GlpG. The simulation cell and solvent water molecules are hidden.

(consisting of 1-palmytoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine, POPE, in this work) of the required size and runs a 250 ps equilibration simulation, during which the membrane is artificially stabilized while it adapts to the protein. The lipid-solubilized GlpG was simulated as a protein surrounded by an annulus of POPE lipids removed no more than 1.5 Å (minimal lipid−protein interatomic distance) from the molecular surface of GlpG (Figure 1c,d). The system was immersed in the simulation cell, which was extended 15 Å on each side of the protein−lipid system.

In both simulated environments, the AMBER03 force field[31] was assigned with a 7.86 Å van der Waals interaction cutoff, and long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) algorithm.[32] The TIP3P water molecules model[33] was used for explicit solvent simulation. All simulated systems were neutralized at pH 7.4 by counterions using 0.9% NaCl (concentration in mass percent, a physiological solution). Solvent density was 0.993 g/mL (water at 310 K). YASARA predicted pK_a values for Asp, Glu, His, and Lys residues using a method that has been calibrated on experimental pK_a measurements.[34] The protonation states of H150, H254, and S201 were predefined in every MD simulation (see Results and Discussion). The MD simulation sampled from the canonical ensemble (NVT). The constant simulation temperature of 310 K was set by assigning random velocity vectors to all atoms using a Maxwell−Boltzmann distribution and a Berendsen thermostat, so that the resulting kinetic energy matched the specified temperature. The time average macroscopic temperature was kept at the requested value by rescaling the atom velocities.[35] All possible combinations of protonation states of the three active site residues H150, H254, and S201 required 24 different MD runs—12 for lipid-solubilized GlpG (Table 1 and Tables S1 and S2 in the Supporting Information) and 12 for GlpG in membrane (Table 2 and Tables S3 and S4 in the Supporting Information). The MD simulation time was ∼15 ns for solubilized GlpG and ∼30 ns for GlpG in membrane. Snapshots were collected every 10 ps. The convergence of MD simulations to equilibrium was controlled by the convergence of the potential energy and by a geometrical criterion of the amplitude of fluctuations of Cα atoms around the equilibrium position. The convergence of Cα atoms to a horizontal asymptote was estimated as RMSD[(RMSD(Cα)$_i$], where RMSD(Cα)$_i$ is the root-mean-square deviation of Cα atoms in the MD simulation i-snapshot from the initial protein structure. The convergence characteristics are presented in the third and the last columns in Tables 1 and 2. The convergence of S201 Oγ−Nε H254 and S201 Oγ−Nε H150 interatomic distances is graphically illustrated in the Supporting Information (Figures S2 and S3).

MD-QM/SCRF(VS) Approach. We have modified our original QM/SCRF(VS)[36−39] approach by combining QM calculations with molecular dynamics simulations, so this approach is termed MD-QM/SCRF(VS). The water solvated enzyme in QM/SCRF(VS) is treated as a two layer system. The inner layer is the enzyme active-site molecular cluster, explicitly calculated by quantum mechanics. The outer layer includes the rest of the protein, the solvent water molecules, and all other molecules from the protein environment—such as lipids in this work. The outer layer is accounted for by QM calculations as a uniform dielectric continuum, "virtual solvent" (VS), which is characterized by a dielectric constant, ε. The molecular cluster includes the catalytic residues and all residues interacting with them by H bonds, π stacking, and residues identified as conserved in sequence analysis and located in the enzyme active site.[4,19] All water molecules forming H-bonds with catalytic residues were also included. Thus, around 270 atoms were included in the GlpG active site molecular clusters used for QM calculations. Figures S4 and S5 in the Supporting Information present two examples of molecular clusters simulating the active site of free GlpG in the H150^{dpr}_H254^{dpr}_S201^{pr} lipid-solubilized protonation state and in the membrane, respectively.

We applied a sampling algorithm in MD-QM/SCRF(VS), collecting 200 snapshots from the end time interval (see third column in Tables 1 and 2) of each of the MD trajectories. We have considered in total 24 different combinations of potonation states of H150, H254, and S201−12 for the membrane environment and 12 for lipid-solubilized GlpG (Tables 1, 2, and S1−S4). Each of the 24 variants of the QM molecular clusters was generated by a specialized automated

dx.doi.org/10.1021/ct3003767 | J. Chem. Theory Comput. 2012, 8, 4663−4671

Table 1. Free Lipid-Solubilized GlpG[a]

protonation state[b]	no. of protons	simulation period (ns)[c]	Oγ S201 Nε H254[d]	Oγ S201 Nε H150[d]	RMSD Cα[e]
H150^{pr}_H254^{pr}_S201^{pr}	**3**	**6.00−14.02**	**2.85**	**9.19**	**0.09**
		0.0008	*0.10*	*0.68*	
(Nδ)H150^{dpr}_H254^{pr}_S201^{pr}	2	1.23−14.30	2.96	2.92	0.13
		0.0008	*0.23*	*0.26*	
(Nε)H150^{dpr}_H254^{pr}_S201^{pr}	2	17.3−22.2	6.07	4.21	0.09
		0.0008	*0.72*	*0.67*	
H150^{pr}_H254^{dpr}_S201^{pr}	**2**	**2.60−17.05**	**2.77**	**9.28**	**0.13**
		0.0007	*0.10*	*0.68*	
H150^{pr}_H254^{pr}_S201^{dpr}	2	5.50−13.99	2.74	2.73	0.11
		0.0008	*0.08*	*0.08*	
(Nδ)H150^{dpr}_H254^{pr}_S201^{dpr}	1	6.90−17.82	3.09	7.29	0.10
		0.0008	*0.48*	*0.53*	
(Nε)H150^{dpr}_H254^{pr}_S201^{dpr}	1	12.00−17.49	2.76	2.87	0.09
		0.0008	*0.09*	*0.14*	
(Nδ)H150^{dpr}_H254^{dpr}_S201^{pr}	1	12.90−21.16	3.05	4.49	0.10
		0.0008	*0.62*	*0.49*	
(Nε)H150^{dpr}_H254^{dpr}_S201^{pr}	**1**	**12.71−16.57**	**2.85**	**6.87**	**0.14**
		0.0008	*0.19*	*0.50*	
H150^{pr}_H254^{dpr}_S201^{dpr}	1	9.24−13.99	9.78	2.78	0.11
		0.0008	*0.52*	*0.17*	
(Nδ)H150^{dpr}_H254^{dpr}_S201^{dpr}	0	9.30−15.58	8.19	7.25	0.13
		0.0008	*0.61*	*0.84*	
(Nε)H150^{dpr}_H254^{dpr}_S201^{dpr}	0	11.80−17.73	*5.16*	*2.90*	*0.12*
			0.36	*0.14*	
experimental values[f]			**2.67**	**5.21**	
			0.16	*0.61*	

[a]Averaged values of active-site interatomic distances (Å). [b]The protonation state for which the calculated distances fit the experimental data is boldfaced. [c]The MD simulation period column also presents the values of RMSD(E)/{abs[AVERAGE(E)]} characterizing the convergence of total energy, E, calculated on the MD trajectory snapshots at this period. RMSD is the standard deviation. [d]RMSDs in Å of the corresponding interatomic distances are presented in italic. [e]The convergence of the protein 3D structure during MD simulation measured on positions of the Cα atoms as RMSD[(RMSD(Cα)$_i$], where RMSD(Cα)$_i$ is a root-mean-square deviation of the Cα atoms in MD simulated i-snapshot from the initial protein structure. [f]Experimental interatomic distances averaged over 2ic8, 3b45, 2o7l, and 2xov pdb files of free GlpG crystal structures.

script for cutting the molecular cluster form the corresponding MD simulation trajectory with a predefined protonation state. Subsequently, open valences in the QM clusters were capped by hydrogen atoms.

The 200 molecular clusters of every protonation state were then transferred to the next automated procedure that calculated over the 200 snapshots their averaged heats of formation, $\langle H_f(\varepsilon)\rangle$, in semiempirical PM6 Hamiltonian.[29] The $\langle H_f(\varepsilon)\rangle$ of each of the 12 different protonation states of GlpG in the membrane and the analogous 12 variants of solubilized GlpG were calculated at 10 different dielectric constants, ε, simulating molecular cluster solvation in a virtual solvent by COSMO,[40] implemented in MOPAC2009. Thus, we have conducted 24 × 200 × 10 = 48 000 QM calculations in total of molecular clusters containing ∼270 atoms. This task is obviously unrealistic for *ab initio* or high level DFT QM calculations, applying also a continuum solvation model. Therefore, we applied a semiempirical QM method.

pK_a Calculations. The pK_a values at every dielectric constant of the virtual solvent were calculated as

$$\langle \Delta H_f(\varepsilon)\rangle = \langle H_f(\varepsilon)^{deprot}\rangle - \langle H_f(\varepsilon)^{prot}\rangle \quad (1)$$

$$pK_a(\varepsilon)^{theor} = [\langle \Delta H_f(\varepsilon)\rangle + C]/2.303RT \quad (2)$$

$\langle H_f(\varepsilon)^{deprot}\rangle$ and $\langle H_f(\varepsilon)^{prot}\rangle$ are the averaged heats of formation of deprotonated and protonated states, calculated over the corresponding $H_f(\varepsilon)$ values on the sampled snapshots. When referring to unprotonated (neutral) H150, at each dielectric constant we selected the $\langle \Delta H_f(\varepsilon)\rangle$ value of the more stable tautomer between the Nε and Nδ proton positions (see details in the Supporting Information).

$$C = H^+(g) + G_{aq}(H^+) \quad (3)$$

$H^+(g)$ = 367.5 kcal/mol is the experimental gas-phase heat of proton formation,[41] and $G_{aq}(H^+)$ is the experimental absolute free energy of proton hydration, which varies from −252 to −264 kcal/mol.[42,43] We used the value of C = 104.5 kcal/mol.

The calculated pK_a^{theor} values of histidine and serine were improved by eqs 4 and 5, using calibration coefficients we have obtained by linear regression of p$K_a(\varepsilon)^{theor}$ vs experimental pK_a, calculated on small molecules of nitrogen heterocycles and alcohols, respectively (Tables S5 and S6 in the Supporting Information):

$$pK_a(\varepsilon)^{improved}(His) = 1.0219pK_a^{theor} - 0.3544 \quad (4)$$

$$pK_a(\varepsilon)^{improved}(Ser) = 0.3325pK_a^{theor} + 6.6353 \quad (5)$$

The improved p$K_a(\varepsilon)$ values are presented in Tables 3−5.

RESULTS AND DISCUSSION

Knowledge of the protonation state of catalytic residues at every elementary step along the reaction coordinate is key to the understanding of enzyme catalytic mechanisms. The protonation state of the catalytic residues in the free enzyme

dx.doi.org/10.1021/ct3003767 | J. Chem. Theory Comput. 2012, 8, 4663−4671

Table 2. Free GlpG in Membrane[a]

protonation state[b]	no. of protons	simulation period (ns)[c]	Oγ S201 Nε H254[d]	Oγ S201 Nε H150[d]	RMSD Cα[e]
H150^{pr}_H254^{pr}_S201^{pr}	**3**	**5.00−30.33**	**3.00**	**9.83**	**0.10**
		0.0012	*0.25*	*0.68*	
(Nδ)H150^{dpr}_H254^{pr}_S201^{pr}	**2**	**8.50−32.45**	**2.96**	**5.03**	**0.08**
		0.0012	*0.17*	*0.37*	
(Nε)H150^{dpr}_H254^{pr}_S201^{pr}	2	12.5−30.1	7.66	3.81	0.08
		0.0011	*0.34*	*0.42*	
H150^{pr}_H254^{dpr}_S201^{pr}	2	6.00−31.48	3.73	3.98	0.09
		0.0012	*0.88*	*1.02*	
H150^{pr}_H254^{pr}_S201^{dpr}	2	12.00−31.23	2.79	2.75	0.12
		0.0012	*0.13*	*0.15*	
(Nδ)H150^{dpr}_H254^{pr}_S201^{dpr}	1	13.00−30.05	4.21	6.56	0.12
		0.0014	*0.28*	*0.43*	
(Nε)H150^{dpr}_H254^{pr}_S201^{dpr}	1	25.5−34.8	4.80	8.68	0.07
		0.0011	*0.16*	*0.47*	
(Nδ)H150^{dpr}_H254^{dpr}_S201^{pr}	1	10.00−33.33	3.95	8.66	0.10
		0.0011	*0.83*	*1.00*	
(Nε)H150^{dpr}_H254^{dpr}_S201^{pr}	**1**	**15.5−30.0**	**2.85**	**5.84**	**0.07**
		0.0011	*0.16*	*0.51*	
H150^{pr}_H254^{dpr}_S201^{dpr}	1	8.00−36.025	4.95	2.71	0.08
		0.0011	*0.31*	*0.08*	
(Nδ)H150^{dpr}_H254^{dpr}_S201^{dpr}	0	25.0−35.3	5.29	4.90	0.10
		0.0011	*0.22*	*0.16*	
(Nε)H150^{dpr}_H254^{dpr}_S201^{dpr}	0	18.5−27.4	4.80	4.74	0.06
		0.0010	*0.15*	*0.29*	

[a]Averaged values of active-site interatomic distances (Å). [b]The protonation state for which the calculated distances fit the experimental data is boldfaced. [c]The MD simulation period column also presents the values of RMSD(E)/{abs[AVERAGE(E)]} characterizing the convergence of total energy, E, calculated on the MD trajectory snapshots at this period. RMSD is the standard deviation. [d]RMSDs in Å of the corresponding interatomic distances are presented in italic. [e]The convergence of the protein 3D structure during MD simulation measured on positions of the Cα atoms as RMSD[(RMSD(Cα)$_i$], where RMSD(Cα)$_i$ is a root-mean-square deviation of the Cα atoms in MD simulated i-snapshot from the initial protein structure.

as a starting point on the reaction coordinate predetermines the character of pK_a changes caused by ligand (either a substrate or an inhibitor) binding.[36−39] We have estimated the protonation state of the catalytic residues in the lipid-solubilized and membrane-immersed free GlpG by molecular modeling. Our analysis is based on a combination of molecular dynamics (MD) simulations followed by quantum mechanics (QM) calculations applying our MD-QM/SCRF(VS) protocol. We have selected *E. coli* GlpG as a target since most of the experimental atomic details of rhomboids have been collected for this enzyme.

Identification of the Protonation State by Geometry. The 2xov.pdb crystal structure,[14] of 1.7 Å resolution, was used as a starting structure in our MD simulations of *E. coli* free GlpG. Phosphatidylethanolamine constitutes ∼70% of the *E. coli* membrane.[6] Therefore, we used 1-palmytoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine (POPE) lipids for the construction and MD simulations of cell membrane and lipid annulus of solubilized rhomboid. We performed all-atom MD simulations at 37 °C. The system was immersed in a periodic parallelepiped cell filled with water molecules and Na^+ and Cl^- ions at 0.9% concentration for the neutralization of excessive coulomb charges on the protein and for simulation of physiological salt conditions. The membrane environment was simulated by systems comprised of one protein molecule, ∼300 lipid molecules, ∼20 000 water molecules, and counterions, with a total of about 101 000 atoms (Figure 1a,b). Lipid-solubilized free enzyme was simulated by a system containing a protein molecule surrounded by an annulus of POPE molecules removed no more than 1.5 Å (minimal lipid−protein interatomic distance) from the molecular surface of GlpG, which amounts to about 30−40 lipids, and ∼24 000 water molecules and counterions. This system contains ∼80 000 atoms (Figure 1c,d).

The crucial importance of S201, H254, and H150 for GlpG catalysis was demonstrated by alanine point mutations, resulting in a complete loss of enzymatic activity.[44] Whereas S201 and H254 form the catalytic diad, structural analysis of GlpG complexes with covalent inhibitors reveals that H150 stabilizes the oxyanion in the tetrahedral complex by hydrogen bonding.[16] We examined all possible protonation states of these active site residues in the free enzyme. Unprotonated (neutral) H150 was considered in two tautomeric forms—with a proton on either Nε or Nδ. In free GlpG, the single proton on H254 is located on Nδ as indicated by the crystal structure of the GlpG complex with an isocoumarin inhibitor, in which N254 Nε forms a covalent bond with the inhibitor.[14] Thus, 12 protonation states in total were considered: one with three protons, four with two protons, five with one proton, and two with zero ionizable protons (Tables 1 and 2). We have analyzed how the changes in the protonation state of these three residues influence the geometry of the active site in free GlpG. The data are collected in Table 1 (for lipid-solubilized GlpG) and in Table 2 (for GlpG in membrane). Tables 1 and 2 demonstrate that the interatomic distances in the pairs S201/H254 and S201/H150 are very sensitive to variations of the protonation state in both the lipid and the membrane environments of GlpG. Therefore, these distances were used as key geometric

Table 3. MD-QM/SCRF(VS) Calculated pK_a of H150, H254, and S201 in Three Single-Proton States: H150^{pr}_H254^{dpr}_S201^{dpr}, H150^{dpr}_H254^{pr}_S201^{dpr}, and H150^{dpr}_H254^{dpr}_S201^{pr}[a]

	p$K_a(\varepsilon)$		
ε	H150	H254	S201
lipid-solubilized GlpG			
1	0.4	**25.0**	17.1
2	2.7	**19.1**	15.3
3	3.3	**15.8**	14.2
4	3.7	**13.9**	13.7
6	4.3	11.8	**13.0**
8	4.7	10.8	**12.7**
10	4.9	10.1	**12.5**
20	5.4	8.7	**12.0**
30	5.6	8.2	**11.9**
40	5.8	8.0	**11.8**
GlpG in membrane			
1	−6.7	6.5	**13.0**
2	−8.8	−0.5	**12.2**
3	−9.8	−1.0	**11.9**
4	−10.4	−0.4	**11.7**
6	−11.1	0.3	**11.5**
8	−11.5	0.7	**11.5**
10	−11.7	0.9	**11.4**
20	−12.3	1.5	**11.3**
30	−12.5	1.7	**11.3**
40	−12.6	1.8	**11.2**

[a]Calculated at 310 K. The boldfaced values indicate the preferred position of the proton in the single-proton state at each ε (at pH = 7.4).

parameters for the comparison of the MD simulated and the corresponding experimental crystal structure values for the identification of the protonation state of the active site of free GlpG. The last row in Table 1 contains the averaged values of the corresponding experimental interatomic distances collected from several crystal structures of free GlpG.

The average experimental value of the S201 Oγ−Nε H254 distance (2.67 Å) from crystal structures of free GlpG indicates a strong H-bond interaction between these atoms (Table 1). Conversely, the very large average experimental S201 Oγ−Nε H150 distance of 5.21 Å means that H150 cannot activate the S201 nucleophile, in agreement with its function as the oxyanion hole.[16] In our analysis, we accepted protonation states that satisfy the following experimental structural criteria: (1) H-bond distance for S201 Oγ−Nε H254 no longer than 3.0 Å and (2) the relation number (RN) between S201 Oγ−Nε H254 and S201 Oγ−Nε H150 distances should be close to the experimental value of RN = 0.5 (the former distance is considerably smaller than the latter). Three protonation states among the 12 MD simulated lipid-solubilized GlpG models satisfy these criteria: H150^{pr}_H254^{pr}_S201^{pr}, RN = 0.3; H150^{pr}_H254^{dpr}_S201^{pr}, RN = 0.3; and (Nε)-H150^{dpr}_H254^{dpr}_S201^{pr}, RN = 0.4 (Table 1). Analogously, three MD simulated GlpG in the membrane environment satisfy both conditions: H150^{pr}_H254^{pr}_S201^{pr}, RN = 0.3; (Nδ)H150^{dpr}_H254^{pr}_S201^{pr}, RN = 0.6; and (Nε)-H150^{dpr}_H254^{dpr}_S201^{pr}, RN = 0.5 (Table 2).

Clearly, the geometrical criteria alone cannot unequivocally identify the protonation state of the catalytic residues of free GlpG. Therefore, we have applied a complementary energetic

Table 4. MD-QM/SCRF(VS) Calculated pK_a Values of Catalytic Residues in the Two Ionizable Protons States[a]

	H150^{dpr}_H254^{pr}_S201^{pr}		H150^{pr}_H254^{dpr}_S201^{pr}		H150^{pr}_H254^{pr}_S201^{dpr}	
	p$K_a(\varepsilon)$					
ε	S201	H254	S201	H150	H254	H150
lipid-solubilized GlpG						
1	3.8	−15.7	**7.7**	−28.6	−3.5	−28.2
2	5.7	−10.2	**8.6**	−17.9	−3.4	−19.9
3	6.7	−7.4	**9.0**	−12.9	−3.5	−15.9
4	**7.4**	−5.5	**9.2**	−9.9	−3.5	−13.6
6	**8.1**	−3.4	**9.5**	−6.6	−3.5	−11.1
8	**8.5**	−2.2	**9.6**	−4.8	−3.5	−9.7
10	**8.7**	−1.5	**9.7**	−3.6	−3.6	−8.8
20	**9.2**	0.0	**9.9**	−1.2	−3.6	−6.9
30	**9.4**	0.5	**9.9**	−0.3	−3.6	−6.2
40	**9.5**	0.8	**10.0**	0.1	−3.6	−5.9
GlpG in membrane						
1	6.1	−14.7	7.0	−25.0	4.2	−9.0
2	**9.4**	−9.2	**9.5**	−17.2	7.3	−1.0
3	**10.1**	−6.6	**10.7**	−13.3	**8.7**	0.0
4	**10.2**	−5.0	**11.5**	−10.9	**9.6**	−0.3
6	**10.4**	−3.3	**12.5**	−8.3	**10.7**	−0.7
8	**10.5**	−2.3	**13.0**	−6.8	**11.2**	−0.9
10	**10.6**	−1.7	**13.3**	−5.8	**11.6**	−1.0
20	**10.7**	−0.4	**14.1**	−3.8	**12.4**	−1.4
30	**10.7**	0.1	**14.3**	−3.0	**12.7**	−1.5
40	**10.8**	0.3	**14.5**	−2.7	**12.8**	−1.6

[a]Calculated at 310 K. The boldfaced values indicate the preferred position of the proton at each ε (at pH = 7.4).

Table 5. MD-QM/SCRF(VS) Calculated pK_a of H150, H254, and S201 in the Fully Protonated State: H150^{pr}_H254^{pr}_S201^{pr}[a]

	p$K_a(\varepsilon)$		
ε	H150	H254	S201
lipid-solubilized GlpG			
1	−54.7	−41.9	−4.8
2	−32.0	−24.3	1.8
3	−22.0	−16.5	4.7
4	−16.6	−12.2	6.4
6	−10.6	−7.4	**8.2**
8	−7.4	−4.8	**9.2**
10	−5.3	−3.2	**9.8**
20	−1.0	0.1	**11.1**
30	0.5	1.3	**11.5**
40	1.2	1.9	**11.8**
GlpG in membrane			
1	−52.8	−42.5	−8.2
2	−31.3	−23.4	−0.5
3	−21.7	−15.0	3.0
4	−16.1	−10.2	5.1
6	−10.0	−5.0	7.3
8	−6.8	−2.3	**8.6**
10	−4.7	−0.6	**9.4**
20	−0.4	3.0	**11.0**
30	1.1	4.3	**11.6**
40	1.9	4.9	**11.9**

[a]Calculated at 310 K. The boldfaced values indicate the preferred position of the proton at each ε (at pH = 7.4).

dx.doi.org/10.1021/ct3003767 | J. Chem. Theory Comput. 2012, 8, 4663−4671

criterion—the quantum mechanical pK_a calculation as discussed below.

We have also analyzed interatomic distances between the catalytic H254 and N251 and D243 considered in the literature to play a mechanistic role in catalysis. Both the averaged experimental and the MD calculated (in this work) distances between the Cα atoms of D243 and H254 indicate that D243 is too far from H254 to influence its pK_a (Tables S1 and S3 in the Supporting Information). D243A mutation had only a mild effect on the enzymatic activity of GlpG.[44] Therefore, D243 should not be considered a catalytic residue in rhomboid protease. In good agreement with the crystal structures, our MD simulations demonstrated that the side chain of N251 is turned toward the lipid layer both in the membrane and in lipid-solubilized GlpG. The speculation that N251 may assist H254 in activating S201 through a H-bonding network and thus could be considered as a catalytic residue[9] is refuted by an N251A mutant that exhibited only a modest decrease in enzymatic activity.[44]

A protein substrate of rhomboid enters the active site laterally from the membrane into the gate between helices 2 and 5.[19] This gate is closed in the free enzyme by attractive interactions between side chains of residues in these helices. Point mutations of three pairs of aromatic and hydrophobic residues, F153/W236, W157/F232, and Y160/L229 bound by van der Waals dispersive interactions along helices 2 and 5, demonstrated their key role in the gate opening and therefore in the proteolytic activity of rhomboid protease.[44] The distances between the Cα atoms of the corresponding pairs are presented in Tables S2 and S4 in the Supporting Information. These distances exhibit only minor sensitivity to the variations in the protonation state of the active site residues in both lipid-solubilized (Table S2) and membrane-immersed free GlpG (Table S4).

pK_a Prediction. The X-ray crystal structures of free GlpG do not show protons. Moreover, the structure of the lipid-solubilized and membrane-immersed protein may be somewhat different from the structure of the crystallized protein. MD simulations provide a common approach for structure analysis of noncrystallized proteins in different environments. Unfortunately, classical force fields like Amber03, used in this work, cannot account for changes in atomic partial charges caused by MD simulated thermal atomic movements. Moreover, the applied variant of classical MD does not allow protonation–deprotonation reactions that could occur in a real protein as a result of conformational changes. We worked around the problem through MD simulations of all possible protonation states of the catalytic residues in question—H150, H254, and S201 (Table 1 and 2). We then applied MD-QM/SCRF(VS) to quantum mechanically calculate pK_a values on the corresponding molecular clusters simulating the enzyme active site. We have calculated pK_a's of H150, H254, and S201 in all variants of protonation states with one, two, and three ionizable protons. The results are collected in Tables 3, 4, and 5 (see also Supporting Information for details of pK_a calculations). The model states with two and three ionizable protons are irrelevant since at every value of dielectric constant ε, no more than one residue can be protonated simultaneously at physiological pH = 7.4 (boldfaced pK_a values in Tables 4 and 5). Table 3 presents all possible variants of a single ionizable proton state in both environments—lipid-solubilized and in a membrane. A change in the protonation state of any catalytic residue causes considerable conformational change of the protein structure (Tables 1 and 2) and consequently its interaction with its environment. The different environments of lipid-solubilized GlpG and GlpG immersed in a membrane result in a dramatic pK_a difference of H254 and H150 (Table 3).

For GlpG in a membrane, a single ionizable proton is located on S201 at all ε values. In contrast, there are two alternative positions of a single proton in lipid-solubilized GlpG, either on S201 or on H254, depending on ε. We demonstrate below that free GlpG is characterized by large values of ε. This will select the protonation state in which the ionizable proton is on S201 in both environments—membrane and lipid-solubilized. Two possible such states correspond to two tautomers of H150 (Table 2). The pK_a values presented in Table 3 were calculated for the most stable H150 tautomer, that is, (Nε)-H150^{dpr}_H254^{dpr}_S201^{pr} for both environments (Tables S8 and S28, Supporting Information). Concluding the pK_a analysis, one can observe that the QM predicted protonation state for free GlpG in a membrane and lipid-solubilized GlpG falls in the allowed set previously filtered by geometry. The combined selection of protonation states by geometry and pK_a analysis finally gave only one possible protonation state of free GlpG in both environments—(Nε)H150^{dpr}_H254^{dpr}_S201^{pr}.

GlpG Catalysis. The boldfaced values in Table 3 indicate the preferred position of the proton in the single-proton state at each ε. The highest probability for the residence of a single proton is on S201 for GlpG in a membrane. In lipid-solubilized GlpG, there is an alteration in the preferred proton position: the ionizable proton is on S201 at $\varepsilon \geq 6$ and on H254 if $\varepsilon < 6$. This alteration stems from the fact that H254 pK_a becomes larger than S201 pK_a when ε decreases (Table 3). The data in Table 3 present a trend of pK_a increase with decreasing ε of the residues with a preferred proton position in both environments, in a membrane and lipid-solubilized. This trend can be explained by analysis of the interactions that contribute to the pK_a value. We previously demonstrated that the value of the dielectric constant used in the continuum solvation model is proportional to the degree of water exposure of active site residues.[37] The active site of the free enzyme is much more water exposed (ε is larger) than in the bound state with a ligand. We have derived in frames of QM/SCRF(VS) an expression of pK_a as a function of two components: (1) the gas phase free energy of deprotonation in a molecular cluster simulating explicitly all atoms in the enzyme active site and (2) the difference in free energy of solvation between the deprotonated and the protonated states of the active site molecular cluster in VS. Water solvation dominates the effect of the VS on the active site.[37] Ligand binding reduces the degree of water solvation of the catalytic residues in the enzyme active site, and consequently ε of the virtual solvent reduces, too. This effect is formulated as a simple rule (derived from our equation) that rationalizes and predicts the trend in pK_a change of catalytic residues caused by ligand binding (decrease of ε):[37−39]

1. pK_a increases when $Q^2(B) - Q^2(A) > 0$.
2. pK_a decreases if $Q^2(B) - Q^2(A) < 0$.

$Q(A)$ and $Q(B)$ are the total charges of all atoms constituting the active site when the target ionizable functional group is in its acidic A or basic B form. We will mark this pK_a trend-predicting rule as TP-rule in the following discussion.

Ligand binding, which corresponds in our model to the decrease of ε, increases the pK_a of both S201 and H254 (Table 3) for lipid-solubilized GlpG in agreement with the TP-rule:

$$(\text{A form})\text{H254}(0)^{\text{dpr}}\text{S201}(0)^{\text{pr}} \rightarrow (\text{B form})\text{H254}(0)^{\text{dpr}}\text{S201}(-1)^{\text{dpr}} \quad \text{with } Q^2(\text{B}) - Q^2(\text{A}) = 1 - 0 = 1$$

$$(\text{A form})\text{H254}(+1)^{\text{pr}}\text{S201}(-1)^{\text{dpr}} \rightarrow (\text{B form})\text{H254}(0)^{\text{dpr}}\text{S201}(-1)^{\text{dpr}} \quad \text{with } Q^2(\text{B}) - Q^2(\text{A}) = 1 - 0 = 1$$

At values of $\varepsilon \leq 6$, simulating the progressive screening of the GlpG active site from water, the H254 pK_a increases faster than that of S201 (Table 3) and thus changes the character of the single-proton state from H150$^{\text{dpr}}$_H254$^{\text{dpr}}$_S201$^{\text{pr}}$ in free GlpG to H150$^{\text{dpr}}$_H254$^{\text{pr}}$_S201$^{\text{dpr}}$ in the Michaelis complex. Thus, the active site configuration of the Michaelis complex of lipid-solubilized GlpG is the H254(+)−S201(−) ion pair. This is different from molecular modeling interpretation of the mechanism of the "classical" Asp−His−Ser catalytic triad of soluble serine proteases, in which the pK_a of the catalytic His in the Michaelis complex is smaller than that of Ser.[23,45−50] General base catalysis is the initial step in water-soluble serine proteases, leading to the formation of the tetrahedral intermediate as the rate-determining step. For instance, the QM/MM calculated free energy of activation of 17.8 kcal/mol in trypsin refers to the concerted proton transfer from Ser195 to His57 and the nucleophilic attack of Ser195 on the substrate.[49] The calculated pK_a value of H254 for GlpG in the membrane is 6−13 units lower than the pK_a of S201 (depending on ε, Table 3). Therefore, the general base catalysis of H254 in the membrane environment should contribute about 8−18 kcal/mol to the barrier of S201 activation in the first catalytic step ($\Delta G \sim 1.4\Delta$pK_a at 310 K). Consequently, the activation barrier for the acylation step is smaller in solubilized GlpG than in the membrane. In other words, substrate hydrolysis should be faster for lipid-solubilized GlpG than in the membrane. The latter conclusion should be considered with caution as a preliminary speculation that needs further computational studies and validation by corresponding experiments.

The experimentally measured pK_a is a macroscopic value reflecting the free energy change of the whole system caused by deprotonation of the target residue. We stress that not only proton abstraction from S201 contributes to the pK_a change in the Michaelis complex formation. Energy of conformational changes caused by deprotonation and by substrate gating, as well as the accompanying reorganization of the lipid and water molecules around the protein, also contribute to the pK_a difference of catalytic residues in the Michaelis complex from the free enzyme. We should note that the TP-rule is a practical tool for interpretation of the pK_a trend caused by changes in water solvation of the enzyme active site but not for quantitative pK_a prediction. The TP-rule does not account for the conformational energy changes accompanying deprotonation. Therefore, the trend prediction by the TP-rule is relevant only if the change of solvation energy dominates over conformational energy. For example, the TP-rule cannot be applied for the prediction of the pK_a dependence on ε for H254 in the membrane or of the pK_a of H150 in lipid-solubilized GlpG (Table 3).

The Role of Y205. A tyrosine residue was suggested as a third member of a catalytic triad in rhomboid protease, based on the close packing of the aromatic rings of H254 and Y205 in *E. coli* GlpG[9] and H169 and Y120 in *Haemophilus influenzae* hiGlpG.[7] $\pi-\pi$ interactions between the aromatic rings could orient the catalytic His general base in its active position.[7,9] The structural proposal is supported by the observation that the detergent-solubilized GlpG Y205A mutant is inactive, but it exhibits low level activity in living cells (natural membranes).[44] We propose an alternative or additional catalytic role for Y205 in GlpG, functioning to increase the pK_a of H254. Indeed, Y205 and H254 form a $\pi-\pi$ stacking pair in free GlpG. Face-to-face and edge-to-face stacking of aromatic rings to His is found in many proteins. It was demonstrated that such conformations elevate the pK_a[51] and H-bonding ability[52] of histidine.

CONCLUSIONS

We applied molecular dynamics (MD) and quantum mechanics (QM) calculations in order to clarify the catalytic machinery of rhomboid proteases and how it is affected by the immersion of the enzyme in membrane. We showed that in both membrane and lipid-solubilized environments the S201/H254 and S201/H150 interatomic distances of GlpG are most sensitive to variations of the protonation state of the active site residues. Therefore, they were used as geometrical parameters for the comparison of the MD simulated structures with the corresponding experimental values for the identification of the protonation state of the active site of free GlpG. Several protonation states satisfy the geometrical criterion. Thus, as a complementary filtration of the protonation states we have applied an energetic criterion—quantum mechanical pK_a calculation and analysis. This combined analysis yielded (Nε)H150$^{\text{dpr}}$_H254$^{\text{dpr}}$_S201$^{\text{pr}}$ as the protonation state of the active site residues of free GlpG in both environments. We used our MD-QM/SCRF(VS) computational protocol and the TP-rule to rationalize and predict the trend of pK_a change caused by the decrease of water exposure of the active site of GlpG due to ligand binding. The predicted His254 pK_a is larger than that of S201 in the Michaelis complex of lipid-solubilized GlpG, forming the H254(+)_S201(−) ion pair with Ser201 ready for the nucleophilic attack on the substrate. It means that deprotonation of S201 in the Michaelis complex of lipid-solubilized GlpG does not contribute to the activation barrier of the covalent tetrahedral complex formation. The suggested pK_a value of H254 in the Michaelis complex of GlpG in the membrane is considerably lower than the pK_a of S201. Therefore, S201 deprotonation by H254 (general base catalysis) in the membrane environment should contribute to the activation barrier of covalent complex formation. In other words, substrate hydrolysis should be faster for lipid-solubilized GlpG than in the membrane. This latter conclusion is a preliminary speculation that needs further computational study and experimental validation.

ASSOCIATED CONTENT

Supporting Information

MD simulations: tables of interatomic distances, figures of distances along MD trajectories, and figures of QM molecular clusters. pK_a calculations: tables and correlation graphs of pK_a calculations of small molecule nitrogen heterocycles and alcohols, tables of QM calculated heats of formations, and tables of pK_a calculations of the active site residues at various protonation states. This information is available free of charge via the Internet at http://pubs.acs.org

AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.shokhen@biu.ac.il; amnon.albeck@biu.ac.il.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was supported by the "Marcus Center for Pharmaceutical and Medicinal Chemistry" at Bar Ilan University.

REFERENCES

(1) Freeman, M. Rhomboid proteases and their biological functions. *Annu. Rev. Genet.* **2008**, *42*, 191−210.

(2) Urban, S. Rhomboid proteins: conserved membrane proteases with divergent biological functions. *Genes Dev.* **2006**, *20*, 3054−3068.

(3) Urban, S.; Lee, J .R.; Freeman, M. *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **2001**, *107*, 173−182.

(4) Urban, S.; Dickey, S. W. The rhomboid protease family: a decade of progress on function and mechanism. *Gen. Biol.* **2011**, *12*, 231.

(5) Lemberg, M. K.; Menendez, J.; Misik, A.; Garcia, M.; Koth, C. M.; Freeman, M. Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO J.* **2005**, *24*, 464−472.

(6) Urban, S.; Wolfe, M. S. Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 1883−1888.

(7) Lemieux, M. J.; Fischer, S. J.; Cherney, M. M.; Bateman, K. S.; James, M. N. G. The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 750−754.

(8) Brooks, C. L.; Lazareno-Saez, C.; Lamoureux, J. S.; Mak, M. W.; Lemieux, M. J. Insights into substrate gating in H. influenzae rhomboid. *J. Mol. Biol.* **2011**, *407*, 687−697.

(9) Wang, Y.; Zhang, Y.; Ha, Y. Crystal structure of a rhomboid family intramembrane protease. *Nature* **2006**, *444*, 179−183.

(10) Wu, Z.; Yan, N.; Feng, L.; Yan, H.; Gu, L.; Shi, Y. Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1084−1091.

(11) Bibi, E.; Fass, D.; Ben-Shem, A. Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 462−466.

(12) Ha, Y. Open-cap conformation of intramembrane protease GlpG. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2098−2102.

(13) Wang, Y.; Maegawa, S.; Akiyama, Y.; Ha, Y. The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J. Mol. Biol.* **2007**, *374*, 1104−1113.

(14) Vinothkumar, K. R.; Strisovsky, K.; Andreeva, A.; Christova, Y.; Verhelst, S.; Freeman, M. The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J.* **2010**, *29*, 3797−809.

(15) Vinothkumar, K. R. Structure of rhomboid proteases in a lipid environment. *J. Mol. Biol.* **2011**, *407*, 232−247.

(16) Xue, Y.; Ha, Y. Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J. Biol. Chem.* **2012**, *287*, 3099−3107.

(17) Xue, Y.; Chowdhury, S.; Liu, X.; Akiyama, Y.; Ellman, J.; Ha, Y. Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites. *Biochemistry* **2012**, *51*, 3723−3731.

(18) Wu, Z.; Yan, N.; Feng, L.; Oberstein, A.; Yan, H.; Baker, R. P.; Gu, L.; Jeffrey, P. D.; Urban, S.; Shi, Y. Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1084−1091.

(19) Urban, S. Taking the plunge: integrating structural, enzymatic and computational insights into a unified model for membrane-immersed rhomboid proteolysis. *Biochem. J.* **2010**, *425*, 501−512.

(20) Erez, E.; Fass, D.; Bibi, E. How intramembrane proteases bury reactions in membrane. *Nature* **2009**, *459*, 371−378.

(21) Ha, Y. Structure and mechanism of intramembrane protease. *Semin. Cell Dev. Biol.* **2009**, *20*, 240−250.

(22) Wolfe, M. S. Intramembrane proteolysis. *Chem. Rev.* **2009**, *109*, 1599−1612.

(23) Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **2002**, *102*, 4501−4524.

(24) Bondar, A. N.; del Val, C.; White, S. H. Rhomboid dynamics and lipid interactions. *Structure* **2009**, *17*, 395−405.

(25) Zhou, Y.; Moin, S. M.; Urban, S.; Zhang, Y. An Internal Water-Retention Site in the Rhomboid Intramembrane Protease GlpG Ensures Catalytic Efficiency. *Structure* **2012**, *20*, 1255−1263.

(26) Reddy, T; Rainey, J. K. Multifaceted substrate capture scheme of a rhomboid protease. *J. Phys. Chem. B* **2012**, *116*, 8942−8954.

(27) Pierrat, O. A.; Strisovsky, K.; Christova, Y.; Large, J.; Ansell, K.; Bouloc, N.; Smiljanic, E.; Freeman, M. Monocyclic β-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases. *ACS Chem. Biol.* **2011**, *6*, 325−335.

(28) YASARA Structure. http://www.yasara.com (accessed Sept. 2012).

(29) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.

(30) Stewart, J. J. P. *MOPAC2009*; Computational Chemistry: Colorado Springs, CO, 2008. http://OpenMOPAC.net (accessed Sept. 2012).

(31) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

(32) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys. B* **1995**, *103*, 8577−8593.

(33) Miyamoto, S.; Kollman, P. A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(34) Krieger, E.; Nielsen, J. E.; Spronk, C. A.; Vriend, G. Fast empirical pKa prediction by Ewald summation. *J. Mol. Graphics Modell.* **2006**, *25*, 481−486.

(35) Krieger, E.; Darden, T.; Nabuurs, S.; Finkelstein, A.; Vriend, G. Making optimal use of empirical energy functions: force field parameterization in crystal space. *Proteins* **2004**, *57*, 678−683.

(36) Shokhen, M.; Khazanov, N.; Albeck, A. The cooperative effect between active site ionized groups and water desolvation controls the alteration of acid/base catalysis in serine proteases. *ChemBioChem* **2007**, *8*, 1416−1421.

(37) Shokhen, M.; Khazanov, N.; Albeck, A. Screening of the active site from water by the incoming ligand triggers catalysis and inhibition in serine proteases. *Proteins* **2008**, *70*, 1578−1587.

(38) Shokhen, M.; Khazanov, N.; Albeck, A. Challenging a paradigm: Theoretical calculations of the protonation state of the Cys25-His159 catalytic diad in free papain. *Proteins* **2009**, *77*, 916−926.

(39) Shokhen, M.; Khazanov, N.; Albeck, A. The mechanism of papain inhibition by peptidyl aldehydes. *Proteins* **2011**, *79*, 975−985.

(40) Klamt, A.; Schüümann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.

(41) *CRC Handbook of Chemistry and Physics*, 78th ed.; Lide, D., Ed.; CRC Press: Boca Raton, FL, 1997.

(42) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. Calculation of absolute and relative acidities of substituted imidazoles in aqueous solvent. *J. Phys. Chem. A* **1997**, *101*, 10075−10081.

(43) Daqing Gao, D.; Svoronos, P.; Wong, P. K.; Maddalena, D.; Hwang, J.; Walker, H. pKa of acetate in water: a computational study. *J. Phys. Chem. A* **2005**, *109*, 10776−10785.
(44) Baker, R. P.; Young, K.; Feng, L.; Shi, Y.; Urban, S. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8257−8262.
(45) Warshel, A.; Russel, S. Theoretical correlation of structure and energetic in the catalytic reaction of trypsin. *J. Am. Chem. Soc.* **1986**, *108*, 6569−6579.
(46) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J. K. How do serine proteases really work? *Biochemistry* **1989**, *28*, 3629−3637.
(47) Bentzien, J.; Muller, R. P.; Florian, J.; Warshel, A. Hybrid ab initio QM/ molecular mechanics calculations of free energy surfaces for enzymatic reactions: the nucleophilic attack in subtilisin. *J. Phys. Chem. B.* **1998**, *102*, 2293−2301.
(48) Topf, M.; Varnai, P.; Richards, W. G. Ab initio QM/MM dynamics simulation of the tetrahedral intermediate of serine proteases: insights into the active site hydrogen-bonding network. *J. Am. Chem. Soc.* **2002**, *124*, 14780−14788.
(49) Kato, S.; Ishida, T. Theoretical perspectives on the reaction mechanism of serine proteases: the reaction free energy profiles of the acylation process. *J. Am. Chem. Soc.* **2003**, *125*, 12035−12048.
(50) Ishida, T.; Kato, S. Role of Asp102 in the catalytic relay system of serine proteases: a theoretical study. *J. Am. Chem. Soc.* **2004**, *126*, 7111−7118.
(51) Loewnthal, R.; Sanco, J.; Fersht, A. Histidine-aromatic interactions in barnase. Elevation of histidine pKa and contribution to protein stability. *J. Mol. Biol.* **1992**, *224*, 759−770.
(52) Mignon, P.; Loverix, S.; Greelings, P. Interplay between π-π interactions and H-bonding ability of aromatic nitrogen bases. *Chem. Phys. Lett.* **2005**, *401*, 40−46.