1

2

3

4

Unrevealing the Proteolytic Activity of RgpB Gingipain from Computational Simulations

7

8

9

10

Santiago Movilla, Sergio Martí, Maite Roca* and Vicent Moliner*

BioComp Group, Institute of Advanced Materials (INAM), Universidad Jaume I,

12071, Castellón, Spain

14
15
16
17
18
19

*Correspondence should be addressed to:

Maite Roca: mroca@uji.es

Vicent Moliner: moliner@uji.es

23

Abstract: Alzheimer's disease represents one of the most medical concerns for today's population and health services. Its multifactorial inherent nature represents a challenge for its treatment and requires the development of a broad spectrum of drugs. Recently, the cysteine protease gingipain RgpB has been related to neurodegenerative diseases, including Alzheimer's disease, and its inhibition appears to be a promising neuroprotective strategy. Given these features, a computational study that integrates molecular dynamics (MD) simulations with classical and hybrid quantum mechanics/molecular mechanics (QM/MM) potentials was carried out to unravel the atomistic details of RgpB activity. First, a preliminary study based on principal component analysis (PCA), determined the protonation state of the Cys/His catalytic dyad, as well as the crucial role of a flexible loop that favours reactive interactions of the catalytic residues and the peptide in the precatalytic state in its closed conformation. Then, different mechanisms were explored by means of QM/MM MD simulations. The most favorable mechanism consists in two stages. First, an acylation stage that takes place in two steps where, initially, the sulfur atom of C244 residue attacks the carbonylic carbon of the peptide and the proton of C244 residue is transferred to the amino group of the peptide in a concerted manner. Subsequently, the peptide bond is broken and a fragment of the peptide is released. After that, the deacylation stage takes place in a single step where a water molecule attacks the carbonylic carbon of the peptide and a proton of the water is transferred to C244 residue. The free energy barrier of the rate limiting step is in very good agreement with available experimental data. The mechanism exhibits an unusual role of H211 residue compared with other cysteine proteases but a crucial role of the peptide in triggering the catalysis. Notably, the atomic and energetic particularities found represent a significant contribution to the comprehension of the reaction mechanism and a great opportunity for the design of efficient inhibitors of gingipain RgpB.

1. INTRODUCTION

Alzheimer's disease is one of the most challenging conditions for health research worldwide[1,2]. It is a progressive neurodegenerative brain disorder that causes devastating functional, cognitive and behavioral problems.[1-4] Categorized as a multifactorial disease, it has attempted to be treated without success across a broad spectrum of metabolic targets.[5] Thus, in the absence of a successful treatment, comprehension of the reaction mechanism of enzymatic systems related with Alzheimer's disease and proposal of new targets as potential therapeutic strategies is necessary to bring this overwhelming disease under control.[1-5]

For some years now, several reports have been linking oral health problems to cognitive impairment conditions such as Alzheimer's disease. Specifically, infections with *Porphyromonas gingivalis*, the cornerstone pathogen in the development of periodontitis, have been shown to have a significant influence on the formation of amyloid β-peptide (Aβ) plaques, development of dementia, and aggravation of Alzheimer's disease.[6-11] Skillfully, a recent study identified in the brain of Alzheimer's patients a group of enzymes secreted by *Porphyromonas gingivalis*, gingipain proteases, as a direct cause of this neuronal damaging effect. The study concluded that gingipain small-inhibitors have promising neuroprotective brain effects.[11]

Gingipains are enzymes belonging to the CD clan of cysteine peptidases. The CD clan involves a number of cysteine proteases within a wide range of parasitic protozoa. Members of this clan mainly differ from those of all other clans in terms of primary and tertiary structure, the proteins that they hydrolyze and so their metabolic functions.[12] More in depth, two large groups of gingipains can be distinguished, Lys-gingipains (Kgp) and Arg-gingipains (Rgp), according to the residue they recognize at the P1 position to cleave peptides.[13] Although all gingipains are related to a wide range of diseases from gingivitis to cardiovascular problems,[14-17] RgpB shows a stronger correlation with the progression of Alzheimer's disease.[11] RgpB is the gingipain-R encoded by the *rgpb* gene and differs from RgpA and HRgpA forms by the absence of the hemagglutinin/adhesin domains.[18-20] With this in mind, a better understanding of the action of this attractive pharmacological target is necessary in order to design possible efficient inhibitors to treat neurodegenerative diseases.

RgpB activity is extended to the hydrolysis of basically any peptide bond with arginine at the P1 position.[21] Some other residue preferences along peptide chain have been reported, however, none as essential as that of arginine.[13,22] Despite the limited kinetic information available, is well known that RgpB exhibits proteolytic activity in a broad pH range between 6.5 (50%) and 9.5 (100%). Outside this

range it is hypothesized to become structurally unstable.[23] As a result of the aforementioned, RgpB is capable of destroying human connective tissues, cell surface proteins and receptors, cytokines, components of the coagulation and complement cascades, heme and iron-binding proteins, immunoglobulins and proteinase inhibitors.[24–27] This fact enhances the pharmacological importance of understanding in depth the reaction mechanism of RgpB for exploiting medical proposes.[28]

Some features are common and accepted in the mechanisms that have been reported so far for almost all cysteine proteases reactions.[29-42] In general, the reaction proceeds through two main stages. The first stage, the acylation, corresponds to the formation of the acylenzyme through the nucleophilic attack of the cysteine sulfur and the displacement of the peptidic leaving fragment. Afterwards, in a second stage, a base-activated water molecule attacks the C1 of the peptide and the cysteine - peptide bond is broken to give way to the second product release. Despite these common features and other similarities between cysteine proteases, numerous studies have shown a remarkably wide reactivity among them. Disparity between cysteine proteases range from significant structural differences in active sites to small details in the transition states (TSs) of the rate determining steps. In fact, these differences are the cornerstone in the design of potent and selective inhibitors/drugs. Our group's experience[38,40,43–45] and other studies on cysteine proteases[29-42] can be used to focus the attention on those critical points. Some of the most differentiated factors among the mechanisms of cysteine proteases include the protonation state of the Cys/His catalytic dyad, its role into the reaction mechanism, the residues that act as acid/base and whether the bond forming and bond breaking processes are concerted or stepwise. [29-42]

In the reaction mechanism of cysteine proteases, it is generally accepted that a proton transfer takes place between the Cys/His catalytic dyad prior to the formation of the acylenzyme.[29-42] However, in the particular case of RgpB, the peptide is located between these two catalytic residues and direct proton transfer between the catalytic Cys and His residues looks not feasible.[22] Given this, few known mechanisms in proteases are viable to be adapted to RgpB. Furthermore, E-64,[46] a commonly used reference inhibitor of cysteine proteases, shows no inhibitory activity on this cysteine peptidase family, as it does on all others.[47] This fact, together with the atypical spatial disposition of the catalytic dyad residues in RgpB and the peptide, point to crucial differences in the reaction mechanism that must be explored for the drug design process.

Herein, we conducted a computational study that integrates molecular dynamics (MD) simulations with classical and hybrid quantum mechanics/molecular mechanics (QM/MM) potentials, and structural data analysis aimed to understand the activity of the RgpB gingipain protease from

1 *Porphyromonas gingivalis* in atomistic detail. Our results show conformational changes of the

2 precatalytic state depending on the protonation states of the catalytic dyad Cys/His. Additionally, from

3 the most stable and reactive structure, we propose a molecular mechanism by which the proteolysis

4 reaction can proceed, revealing new valuable particularities important for a future rational drug design.

5

6 **2. COMPUTATIONAL DETAILS**

7   **System Set Up** The initial coordinates of the system were obtained from the crystal structure of

8 RgpB from *Porphyromonas gingivalis* in complex with a peptide-like inhibitor (PDB code 1CVR).[22] The

9 inhibitor was replaced by the protein fragment Cys-Ala-Tyr-Arg-Thr-Ser-Pro (acetylated terminals) of

10 human pancreatic ribonuclease (UniProtKB-07998)[48] preserving as many atoms of the crystallized

11 inhibitor as possible. The fragment was chosen based on inhibitor's size, its electrostatic neutrality and

12 reports of being hydrolyzed by RgpB.[23] The coordinates of the missing atoms and hydrogens were added

13 in sterically favorable positions. Protonation states for the titratable residues were selected based on the

14 results provided by the PROPKA3 software.[49] The selected pH value was 7.5 as adjusted experimentally

15 in the activity and inhibition studies.[11,21,22] A total of 16 to 18 Na^+ ions,[50] depending on the protonation

16 states of the Cys/His catalytic dyad, were added in electrostatically optimal positions (those positions

17 where the electrostatic potential reaches a maximum) around the enzyme in order to neutralize the

18 system. Finally, the system was solvated with a 107.5 $Å^3$ cubic box of water molecules (TIP3P)[51] with a

19 minimum distance of 15 Å between any protein atom and the edge of the box. The water molecules from

20 the crystal structure were preserved. The complete system contains ~116 k atoms. All the steps previously

21 described were done using the AmberTools17 tleap package.[52]

22   As mentioned in the introduction section, the peptide is located between the catalytic dyad and

23 no proton transfer is possible between C244 and H211 residues (see Scheme 1 and Figure 1). Moreover,

24 in the absence of experimental data, three initial structures with different protonation states of these two

25 residues were generated and classically simulated to evaluate the stability and reactivity of the

26 precatalytic state because of the possible high impact over the reaction mechanism. Scheme 1 shows the

27 structures of all plausible and chemically reasonable combinations of protonation states of C244 and

28 H211 residues (**R1**, **R2** and **R3**). The first one has both residues neutral (**R1**). The second one corresponds

29 to its protomer with the C244 residue deprotonated (negatively charged) and the H211 residue protonated

30 (**R2**). The last one shows the C244 residue neutral and the H211 residue protonated (**R3**). There is another

31 protonation state, the C244 residue deprotonated and H211 neutral, but it does not have chemical sense

1   for the mechanistic point of view because none of the residues could act as an acid. Thus, this last

2   protonation state was discarded. Frequently, in cysteine proteases the first two protomers (**R1** and **R2**)

3   are in chemical equilibrium[35,38,40–42] and the preference for one over the other can be explored by

4   QM/MM methods. However, as previously mentioned, RgpB has the peptide between these residues and

5   a proton exchange between C244 and H211 is physically inconceivable. The last protonation state (**R3**)

6   is uncommon but was previously proposed[22] for RgpB, so we decided to evaluate it. Scheme 1 also

7   displays some key residues that interact with the peptide, C244 or H211 residues. Figure 1 shows the

8   initial structures of the three different precatalytic states to be studied (**R1**, **R2** and **R3**) with the

9   interactions with the oxyanionic hole residues, G212 and C244, and with H211.
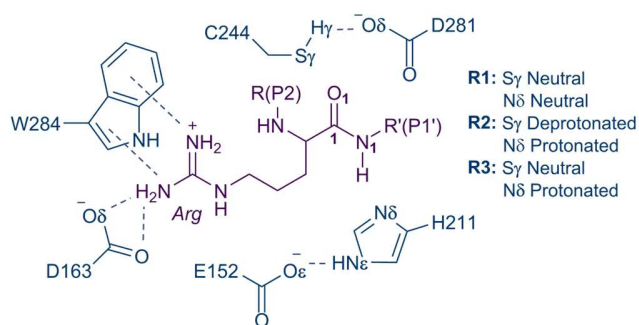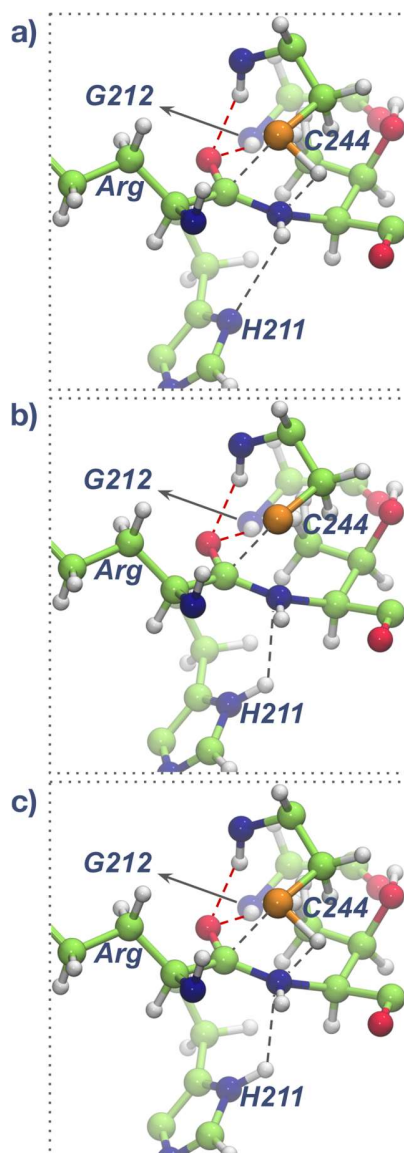

10

11   Scheme 1. General disposition of the active site and protonation states of the proposed initial structures.

12   Interactions between the residues and the peptide are displayed in dashed lines.

13

Figure 1. Initial structures of the considered precatalytic states a) **R1**, b) **R2** and c) **R3**. Adapted and prepared from the crystal structure template, 1CVR.[22] The interactions between the peptide and the backbone of oxyanionic hole residues, C244 and G212, are shown in red dashed lines while the possible bond forming and bond breaking distances are in black dashed lines.

Equilibration of initial reactants states. Optimizations and classical MD simulations were carried out to relax and equilibrate **R1**, **R2** and **R3** initial states in solution using AMBERff14SB[53] and TIP3P[51] force fields to describe the protein and solvent water molecules, respectively. First, a minimization was performed on the solvent molecules, ions and hydrogens using 2500 minimization

1 steps with the conjugate gradient algorithm. Then, a short 2 ns dynamic of the solvent molecules and

2 ions with the constrained position of the heavy atoms of the protein (restraint constant of 300 kcal·mol^{-1}

3 ^{1}·Å^{-2}) was carried out, followed by two energy minimizations, one with the protein backbone restrained

4 and another completely unrestrained. Then, the whole system was heated in four consecutive dynamics.

5 The heating dynamics were: 1) NPT ensemble, 100 K and 1 bar, time step of 0.5 fs; 2) NVT ensemble,

6 200 K, time step of 0.5 fs; 3) NVT ensemble, 300 K, time step of 0.5 fs; and 4) NVT ensemble, 300 K,

7 time step of 2 fs. Subsequently, 50 ns NVT was run at 310 K. Up to this point, restraints were applied to

8 interactions that we considered crucial for the catalysis (E152-H211, C244-Peptide, H211-Peptide,

9 D281-C244). Finally, sampling production dynamics were performed with the system completely

10 unrestrained with the NVT ensemble at 310 K and a time step of 2 fs. In particular, 150 ns of classical

11 MD were followed by 2 ns of QM/MM MD using the semiempirical PM6[54] Hamiltonian for the quantum

12 region and the AMBERff14SB[53] and TIP3P[51] force fields to describe the protein and the water molecules,

13 respectively. The QM/MM frontier was treated using the link atom procedure. Atoms included in the

14 quantum region are shown in Scheme 2. The classical calculations were run in the AMBER GPU[55,56]

15 software version. The cut-off limits for short-range non-bonded interactions were 10 Å and a Particle

16 Mesh Ewald (PME)[57,58] model was used for the long-range interactions. Temperature control was

17 performed using Langevin dynamics[59,60] with a 3 ps^{-1} collision frequency. For all equilibration

18 simulations, SHAKE[61,62] algorithm was used to constrain light atoms and the velocity Verlet[63] algorithm

19 was used to update the velocities. Equilibrium convergence was confirmed by the evaluation of the

20 RMSD of the backbone atoms using the CPPTRAJ package.[64]

21

22

23

24

25

26

Scheme 2. Schematic representation of the region treated quantum mechanically to explore the mechanism of a) acylation stage and b) deacylation stage. Link atoms added to treat QM-MM frontiers are depicted in circles.

Potential of Mean Force (PMF). From QM/MM equilibrated structures, 1D and 2D potential energy surfaces (PES) were calculated through sequential minimizations along selected collective variables that best describe each chemical transformation. Conjugate gradient algorithm was employed for the minimizations using a gradient tolerance of 0.1 kcal mol^{-1} as a convergence criterion. Later, free energy surfaces (FESs) were generated in terms of PMFs[65,66] at 310 K using the structures of the above mentioned 1D and 2D PESs as starting points for each window. Each window had a relaxation time of 5 ps and a sampling time of 25 ps using a time step of 0.5 fs in NVT ensemble. Temperature control was performed using Langevin dynamics[59,60] with a 3 ps^{-1} collision frequency. The umbrella sampling[67] method was used to restrain the reaction coordinates. The force constant used for each window was 580 kcal·mol^{-1}·Å^{-2} and the window width was between 0.05 Å and 0.1 Å

depending on the distinguished reaction coordinate. The number and the width of the windows selected ensure a correct overlapping of windows. The umbrella integration method,[68] as implemented in the QM3 suite,[69] was used to analyze the biased sampling dynamics and to generate the PMFs along selected coordinates. The reaction coordinates chosen for each step, the number of windows and the intervals of the reaction coordinates are described in Table S1 of the Supporting Information. In order to improve the PM6 energy function used to generate the PMFs, an electronic correction has been applied over the PM6-level optimized stationary points. Transition state structures were selected by density peaks clustering processes[70] (RMSD cut 0.7 Å) at the saddle points of the PM6/MM FESs. Subsequently, these structures were optimized and verified by analysis of the Hessian and by tracing down the intrinsic reaction coordinate path (IRC), and the extremes were finally optimized. Single point energy calculations were carried out on these structures at DFT/MM level, employing PBE functional[71–74] and the D3(BJ) dispersion correction[75] with the 6-311+G** basis set, to correct the PM6/MM electronic energy differences. The thermal contributions calculated by the statistical methods at the PM6/MM level were thus preserved. In order to verify the results, the transition states corresponding to the rate limiting steps of both reaction stages (acylation and deacylation) were also optimized using the high level method and verified by inspection of the normal modes. Charges were calculated from electrostatic potentials using grid based method (CHelpG)[76] on the stationary structures localized using PBE+D3(BJ)/MM method with the 6-311+G** basis set. All these calculations were carried out using the Amber14,[77] Gaussian09,[78] fDynamo[79,80] software integrated with the QM3 suite.[69]

3. RESULTS AND DISCUSSION

**Precatalytic State – Protonation States Strongly Affect Structural Stability and Catalytic Power of RgpB.** The first step in our study consists in determining the structure and the protonation state of the catalytic dyad H211 and C244 in the initial enzyme-substrate complex based on classical MD simulations. The analysis of the time evolution of the root mean square deviation (RMSD) graphs of the protein backbone obtained with the three possible systems, **R1**, **R2** and **R3** (see Scheme 1), shows how the three systems reached the equilibrium and none of the systems exceeded, on average, the 2.0 Å threshold of the crystal resolution (see Figure 2). Nevertheless, remarkable differences can be detected. Thus, while the RMSD computed for the whole backbone atoms of the protein are similar in the three systems, the RMSD computed on the atoms of the active site for systems **R2** and **R3** show larger relative

standard deviations than for system **R1** (see Table S2 in the SI)**.** Moreover, the root mean square fluctuation (RMSF) analysis of the protein α-carbon atoms in **R2** shows higher fluctuations in the loop between 147-159 residues (see the inset graphic of the RMSF of **R2** system). The significantly high fluctuations observed in the **R2** system and larger deviations in the RMSD in the active sites of **R2** and **R3** systems, suggest that the system with both neutral residues (**R1**) must be the most stable.

In order to get a deeper insight into these fluctuations, a principal component analysis (PCA) was carried out, revealing two possible conformations, open and closed, of the loop 147-159, containing the catalytic residue E152 that interacts with the H211 residue. The first component PC1 reflects the opening and closing movement of this loop as shown in Figure 3a. Figure 3b shows how while **R3** solely explores open conformations, **R2** varies across a much wider range of values on PC1 and **R1** was restricted to closed conformations during the whole simulation. Conformational changes of the loop between open and closed conformations in **R2** is correlated with the highly fluctuating regions in the RMSF (right panel Figure 2) while the large frequencies of **R1** in the closed conformation and the **R3** in open conformations (Figure 3b) are in agreement with few changes in the loop conformations (low RMSF values).

A detailed structural analysis showed that the position of this loop can affect the interactions between E152, H211 and the peptide (Figure 3a). So, in the open conformations these catalytic residues are not constantly close to the active site and thus the proteolytic activity must be decreased. H211 has been reported to be essential for catalysis, but the only role proposed so far is as an acid at the beginning of the catalysis.[22] This proposed role would force the histidine to be protonated at the beginning of the reaction. However, simulations with protonated histidine showed significantly large fluctuations over some critical distances and an open conformation of the loop (**R3**) or a wide variety of conformations (**R2**), giving rise to inactive structures.

1
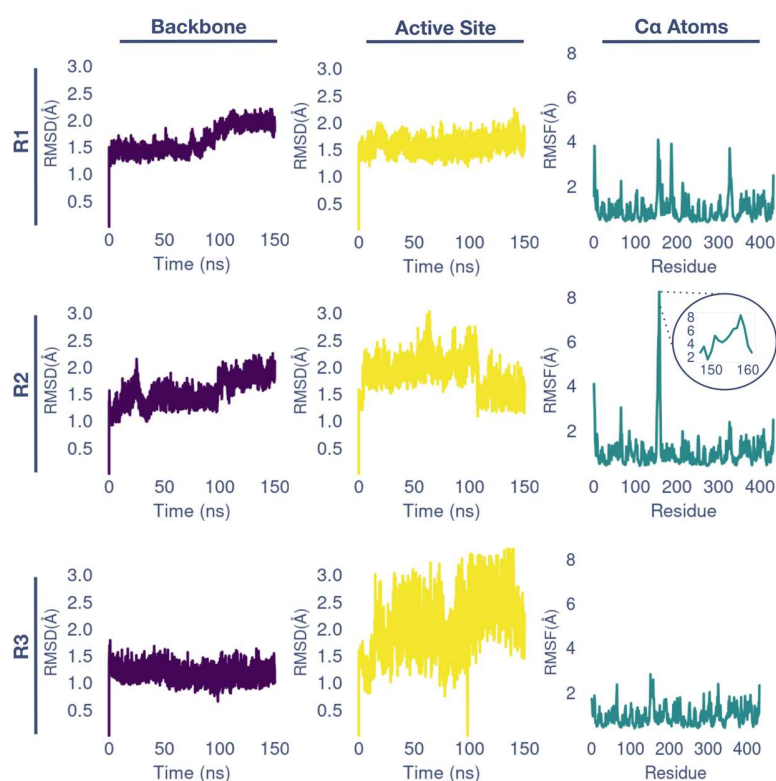
Figure 2. Geometrical analysis of the trajectories obtained for **R1**, **R2** and **R3** systems along the 150 ns of classical MD simulations. Left panels: time evolution of the root mean square deviation (RMSD) of protein backbone atoms (Cα, C, N, O). Center panels: time evolution of the RMSD of the active site atoms (heavy atoms of the E152, H211, C244, D281 residues and peptide). Right panels: root mean square fluctuation (RMSF) analysis of protein α carbon atoms. The inset in the RMSF plot of R2 system shows the values of RMSF of the flexible loop 147-159 residues.

As can be seen by the population analysis of the structures generated along the MD simulations (see Figure 3c), the Nδ:H211 – N1:Pep distance was kept above 6.5 Å for most of the time of the **R2** and **R3** simulations. In contrast, when the H211 residue is neutral (**R1**), the distance between these two nitrogen remained stable at an average of 3.2 Å during the MD simulations of the **R1** system. This suggests that the structural repercussion of opening this loop is stronger than the hydrogen bonds or electrostatic interactions that may exist at the active site when H211 residue is charged. As expected, also the distance between Oε:E152 and Nδ:H211 is strongly influenced by the opening of the loop. The **R2** and **R3** systems show very poor interactions between these two residues while **R1** remains stable in a reactive arrangement. The interaction between the peptide with C244 residue is also affected by the

protonation state of C244 and H211 residues. Particularly, while the **R1** and **R3** systems has an average distance of 3.4 Å between Sγ:C244 and C1:Pep atoms, the **R2** system places the sulfur at an average distance of 5.0 Å due to a rotation of the thiolate group pointing to the opposite side of the peptide. Our predicted reduced catalytic power of the enzyme due to conformational changes induced by the presence of the thiolate ion (**R2** structure) is in agreement with previous hypothesis based on experimental results.[47] An analysis of the hydrogen bond population involving the proton of the thiol group of C244 residue showed a strong interaction with the O1 atom of the electrophilic carbonyl group of the peptide. This interaction locates the sulfur close to the carbon. This also means that the interaction between D281 and C244 residues, that was assumed to be important for the mechanism,[22] is not observed in any of the simulations. In fact, D281 residue was far from C244 residue in all the simulations we performed. Additionally, the absence of a basic residue (like D281) close to the cysteine residue in the active site of other crystal structures[21] of gingipains allow to conclude that this residue is not essential for proteolysis reaction. Previous computational studies[81] also support this structure by showing that in solvent accessible locations the interaction Cys:S-H⋯⁻O-C:Asp is less stable than Wat:O-H⋯⁻O-C:Asp.

Finally, the calculated pK_a for C244 and H211 residues in the open conformation are 6.31 and 9.92, respectively. These values closely match the pK_a experimentally[47] estimated not only for RgpB but also for other related gingipains. Still fascinating is the fact that these values coincide with the pH limits from which catalytic activity decreases dramatically, doubtless because of structural instability. In light of our findings, such structural instability may be related with the open/closed conformations of the loop associated with the Cys/His protonation states. At pH values between 6.5 and 9.5 the enzyme has both residues neutral and it is in closed (active) loop conformation.
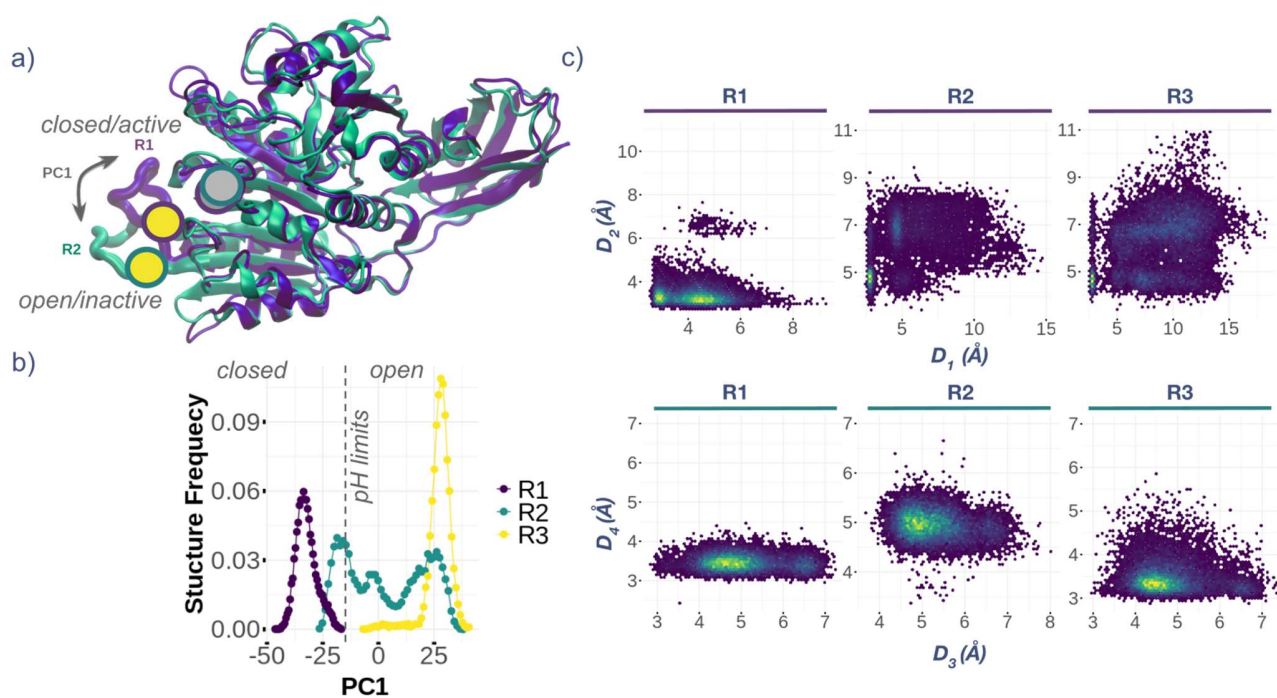
Figure 3. a) Overlay of the protein structures with the open (R2) and close (R1) conformations of the 147-159 loop defined by the PC1. Grey and yellow circles represent H211 and E152 residues, respectively. b) Structure frequency of each reactive structure over the PC1 for the open and closed conformations of the 147-159 loop. c) 2D representations of some distances between Cys/His catalytic dyad, the peptide and other residues. D1: [Nε:H211-Oε:E152], D2: [Nδ:H211-N1:Pep], D3: [Sγ:C244-Oδ:D281], D4:[Sγ:C244-O1:Pep]. For D1 and D3 the distances were considered to the closest oxygen atom of the carboxylic group. Distances are in Å.

In all, our simulations pointed to the H211 and the C244 residues in their neutral states (**R1**) as the most probable precatalytic state. The other two protomers (**R2** and **R3**) proved to be significantly more unstable and thus implausible as starting point structures. This result agrees with the protonation states proposed by Elsässer and coworkers for the cysteine protease legumin, which shows a similar spatial distribution of the catalytic Cys/His dyad and the peptide.[39]

As revealed by analysis of the average structure of R1, the peptide is in between the catalytic Cys/His dyad (see Figure S1). The thiol group of the C244 residue is oriented to the peptide. Moreover, the hydrogen bond interaction between E152 and H211 residues is kept, the flexible loop is in a closed conformation. With regard to the arginine residue of the peptide, a salt bridge interaction is shown with the D163 residue and a π staking interaction with the W284 residue.

15

Acylation Stage Mechanism – An Essential Substrate Assistance Triggers the Proteolytic Reaction. Starting from **R1** system, that has been demonstrated to be the most probably reactant state, several mechanisms were proposed and studied as shown in Scheme 3. All the PMFs are displayed in Figure S2 in the Supporting Information and the free energy profile of the most feasible reaction mechanism is depicted in Figure 4. The stationary point structures of the most plausible mechanisms are depicted in Figure 5. The free energy profiles of the alternative reaction mechanisms are shown in Figure S3 in the Supporting Information. The first mechanism explored involves the proton transfer from Sγ:C244 atom to O1 atom of the peptide (O1:Pep) that appears to take place concomitant with the attack of Sγ:C244 atom on the C1:Pep atom to achieve **I1** intermediate. The associated transition state **TS^{R1-I1}** has a barrier of 24.5 kcal·mol^{-1} (see Figure S3 in the SI) and the intermediate **I1** has an energy of 15.6 kcal·mol^{-1} over **R1**. Starting from **R1** or **I1**, there are alternative paths to reach **I4**. One of them, from **I1**, consists in the proton transfer from O1:Pep atom to Nδ:H211 atom (**I1 → I2**), which would serve as a mediator to transfer the proton to N1:Pep atom in a subsequent step (**I2 → I4**). The free energy barrier to reach **TS^{I1-I2}** transition state is 29.3 kcal·mol^{-1}, while **I2** is found at 28.7 kcal·mol^{-1} above **R1** (see Figure S3 in the SI). This high free energy barrier involving H211 as a mediator of proton transfer reflects a low basicity of Nδ:H211 atom at this point in the reaction. To avoid crossing through this high energy, we studied the possibility of a direct transfer from O1:Pep atom to N1:Pep atom (**I1 → I4**). However, as expected, formation of a four-membered transition state implies a substantial energy penalty, 46.5 kcal·mol^{-1} over **R1** and 30.9 kcal·mol^{-1} over **I1**. A similar free energy barrier, 32.7 kcal·mol^{-1}, was earlier reported for an analogous process in another cysteine protease.[38] Thus, from **I1** intermediate onward there is no viable mechanism for the reaction to proceed.
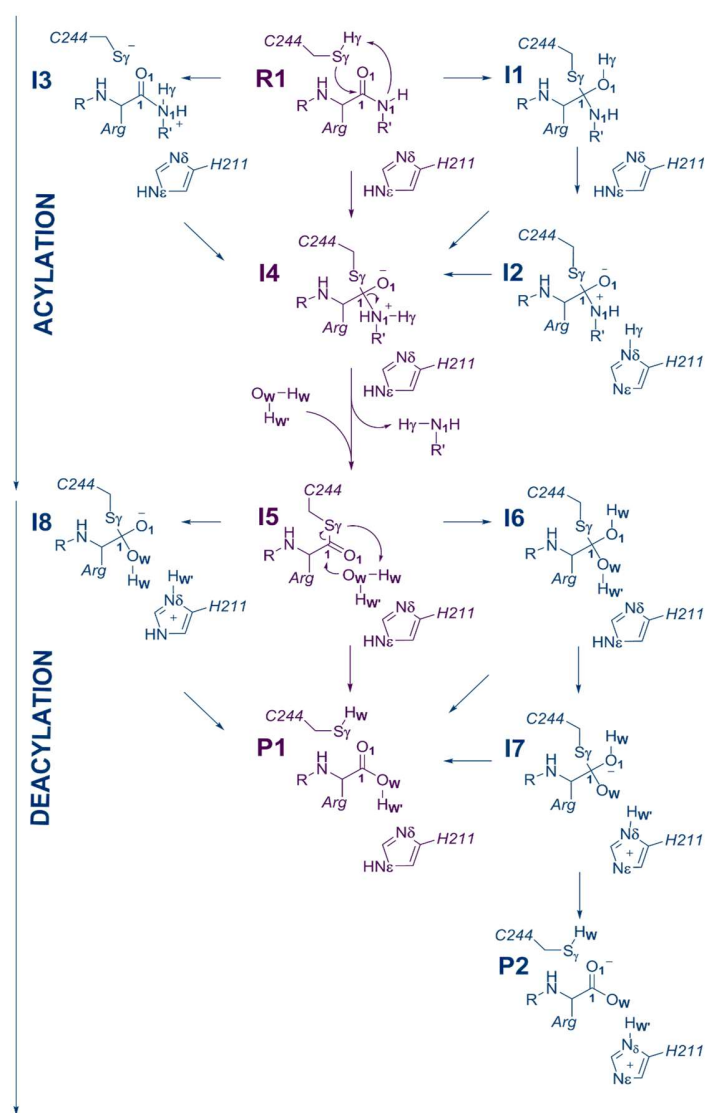
Scheme 3. Explored molecular mechanisms for the proteolysis reaction catalyzed by the RgpB gingipain. The most plausible mechanism, according to the QM/MM free energy profiles, is depicted in purple.

Then, another possible mechanism, starting from **R1**, consists in the proton transfer from Sγ:C244 atom to N1:Pep atom (**R1 → I3**) and subsequently the attack of Sγ:C244 atom on the C1:Pep atom (**I3 → I4**). However, despite our efforts, the **I3** alternative intermediate with the charged Sγ:C244 atom and unbound to the C1:Pep atom could not be localized on any of the computed surfaces. Instead, the concerted reaction mechanism from **R1** to **I4** intermediate appears to be viable, with a free energy barrier of 23.4 kcal·mol^{-1} and the **I4** intermediate 14.2 kcal·mol^{-1} over **R1** (see Figure 4). The nucleophilic attack

of Sγ:C244 atom on the C1:Pep atom is facilitated by the oxyanion hole formed by the backbone nitrogen protons of G212 and C244 residues. The localized transition state (TS^{R1-I4}) shows an advanced state of proton transfer which is confirmed by analysis of the free energy surface (Figure 5 and Figure S2 in the SI). This is also reflected in the increase of the negative charge located on the Sγ:C211 atom (see Table S3 in the SI) and the increase of the positive charge on the Hγ:C211 atom from reactants (**R1**) to transition state (TS^{R1-I4}). On the other hand, the charge on the electrophile C1:Pep atom increases while the charge on N1:Pep atom decreases as the reaction proceeds. The endergonic character of this step is determined by the pseudo stability of **I4** structure, in which the N1:Pep – C1:Pep and the Sγ:C211 – C1:Pep bond distances are significantly longer than the standard values (1.61 Å and 2.04 Å, respectively). In fact, decomposition of **I4** into **I5** intermediate by the N1:Pep – C1:Pep peptide bond breaking occurs through a low barrier of 4.6 kcal·mol^{-1}. Intermediate **I5** is located at -1.7 kcal·mol^{-1} with respect to **R1**. A minimum that is both structurally and energetically similar to that reported by Elsässer and coworkers (Figures 4 and 5).[39] Thus, from all the possible explored mechanisms for the acylation step (**R1** to **I5**), the most favorable reaction path takes place in two steps, through a zwitterionic stable intermediate, **I4**. The 23.4 kcal·mol^{-1} barrier associated with the TS^{R1-I4} transition state is in good agreement with the experimentally data of 22.8 kcal·mol^{-1} reported value.[21] A picture of the rate limiting TS^{R1-I4} optimized at PBE+D3(BJ)/MM level is shown in Figure 6.

The role of the N1:Pep atom as the base has been previously proposed in the similar mechanism of the cysteine protease legumain.[39] Then, this reaction catalyzed by RgpB gingipain can be classified as a Substrate Assisted Catalysis (SAC).[82] This role of the substrate can also explain the inability of compound E-64[46] to inhibit this family of enzymes.[47] In E-64[46], the group that should act as basic/leaving group and the carbonyl group are distant as a consequence of the stereochemistry of the reactive epoxide. Thus, the relative distribution is inappropriate to carry out the attack with the sulfur atom of C244 residue that is not enough acid to be previously deprotonated. Then, the presence of a correctly disposed basic/leaving group must be carefully considered when designing covalent inhibitors for gingipains. In this regard, due to the values of activation and reaction free energies of the **R1**→**I1** step, similar to **R1**→**I4** step, new inhibitors could be designed mimicking this mechanism.
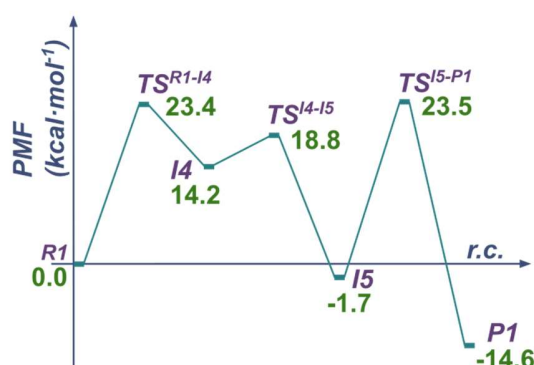
Figure 4. PBE+D3(BJ):PM6/MM free energy profile of the proteolysis catalyzed by the RgpB gingipain.

We also proceed to analyze the interaction energies between the protein and the peptide in this first step (see Figure S4). The results show strong interactions between the peptide and D163 and W284 residues. This result was expected given that these residues are responsible for the arginine selectivity of the RgpB enzyme. D163 and W284 residues strongly interact with the guanidinium group of arginine via a salt bridge and π stacking interactions (see Scheme 1 and Figure S1), respectively, making the enzyme highly selective for arginine-containing peptides. It is also noteworthy that H211 and E152 residues also interact strongly with the peptide (see Scheme 1 and Figure S1). These results reveal an exceptional role of these residues in the electrostatic peptide-enzyme interaction. Further, they are more determinant in the electrostatic stabilization than even the oxyanionic hole residues, C244 and G212 which exhibit lower interaction energies. That means that H211 and E152 residues play an essential role in the structural/electrostatic stability of the precatalytic complex (see Figure S1).

Deacylation Stage – An Uncommon Role of Histidine Residue. Once the acylation step is completed, two water molecules occupies the space generated by the release of the first fragment of the peptide. An analysis of the previous stationary point structures of the mechanism revealed that these two water molecules are able to enter once the first fragment of the peptide has been released. No water molecules were present around the C1:Pep atom prior to the **I5** intermediate. Several mechanisms were proposed for the deacylation stage (see Scheme 3). Based on a previous study,[39] we considered the concerted attack of a water molecule to the C1:Pep atom together with the proton transfer to the O1:Pep atom (**I5 → I6**). Although this process leads to a stable intermediate (-5.1 kcal·mol^{-1}), it has a high activation energy, 32.9 kcal·mol^{-1} (**TS^{I5-I6}**, see Figure S3 in the SI). From **I6**, two different mechanisms

1   can be explored to obtain products. On one hand, the reaction mechanism can take place through a direct

2   proton transfer from O1:Pep atom to Sγ:C244 atom to reach **P1** protomer. The activation free energy

3   (**TS**[I6-P1]) of this step is 14.2 kcal·mol^{-1}. On the other hand, **I6** can evolve into an alternative product with

4   the protonated histidine (**I6 → I7 → P2**). The mechanism results in an almost barrierless process, where

5   the highest free energy barrier corresponds to the step from **I7** to **P2** with an activation energy of 0.8

6   kcal·mol^{-1}. The supposed product obtained by this mechanism has a particular low energy. Fact that can

7   be rationalized due the salt bridge formed between the resulting two charged residues (peptide and H211

8   residue). Anyway, none of these reaction mechanisms are feasible, considering the first high free energy

9   barrier (**TS**[I5-I6]).
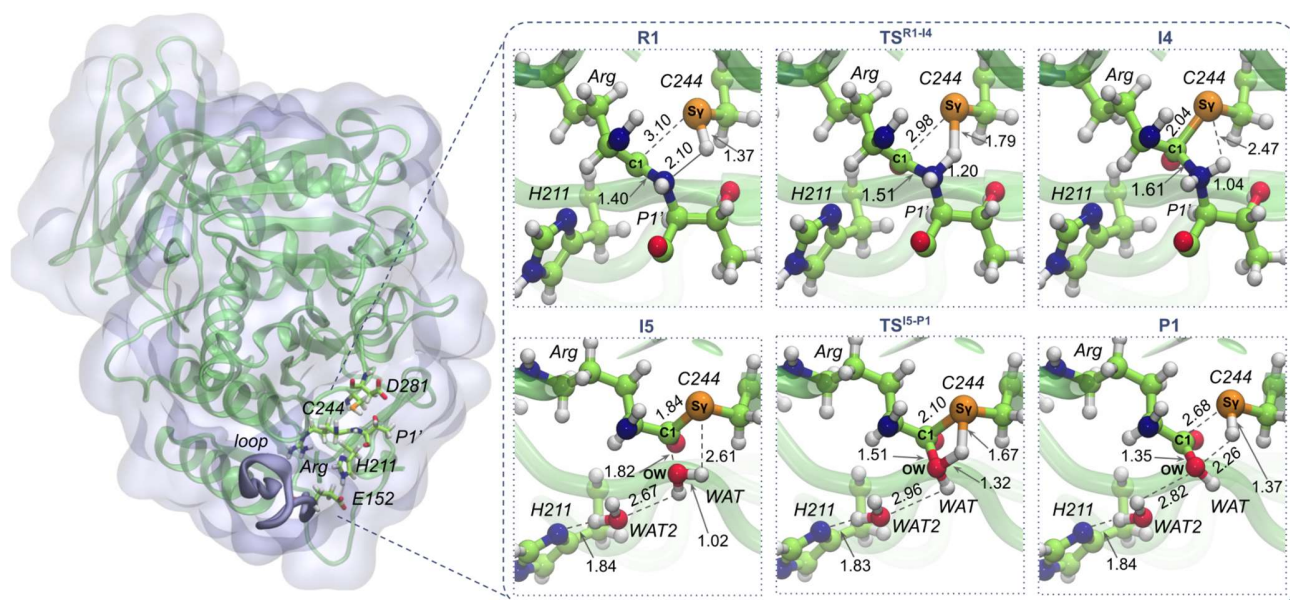
Figure 5. Left panel: Representation of the whole protein where key residues and the peptide are represented as sticks. Right panel: PM6/MM optimized structures of the key states appearing along the most favorable proteolysis reaction mechanism: **R1**, **TS**[R1-I4], **I4**, **I5**, **TS**[I5-P1] and **P1**. Distances are in Å.

16   The role of H211 residue as the base to activate the water molecule to attack the C1:Pep atom to

17 reach **I8** intermediate was also explored. Due to the conformation of the active site, once the H211 residue

18 is protonated by a water molecule cannot transfer the proton to C244 residue to reach **P1**. Instead, the

19 proton from the hydroxylic group bonded to C1:Pep can be transferred to Sγ:C244 concomitant with the

20 proton transfer from H211 to the hydroxylic group. However, it was impossible to locate **I8** where H211

residue assisted as a base. H211 residue plays a distinctive electrostatic/structural role that stands out from the covalent participation that it usually plays in other systems.[38,40] Furthermore, we can observe that the interaction of H211 with the reactive water occurs by means of a bridge of two water molecules. As a matter of fact, this arrangement allows the reactive water to keep its proton oriented towards the $S\gamma$:C244 atom of the cysteine (see **I5** snapshot in Figure 5).

Analysis of the active site in **I5** intermediate (see **I5** in Figure 5) allow us to propose the direct proton transfer from the water molecule to $S\gamma$:C244 atom and the attack of the hydroxylic group to C1:Pep atom. The resulting free energy surface indicates that this reaction proceeds in a concerted manner with a free energy barrier associated with the transition state **TS^{I5-P1}** of 23.5 kcal·mol^{-1}. The atomic charges analysis (see Table S3 in the SI) reveals a late proton transfer in the transition state, while the attack of oxygen atom of water molecule on C1:Pep atom is found to be advanced and the charge on the $S\gamma$:C244 atom slightly becomes more negative at the transition state **TS^{I5-P1}**. Then, the deacylation step will take place through this reaction mechanism. As in the case of the acylation step, the rate limiting **TS^{I5-P1}** was optimized at PBE+D3(BJ)/MM level (see Figure 6). The comparison between the structures localized and characterized using a high level method (PBE+D3(BJ)/MM) and low level method (PM6/MM) shows structural similarities giving rise to the robustness of our conclusions based on DFT/MM corrections over QM/MM MD simulations using the PM6/MM method. A structural analysis of the stationary structures in the deacylation stage reveals the presence of two water molecules in the position of the first fragment of the peptide released where H211 residue disposes them in a reactive orientation.

The deacylation step shows a free energy barrier of 25.2 kcal·mol^{-1} from **I5** which could be dictating the rate limiting step of the full process. Neverthess, considering the energies of this **TS^{I5-P1}** and the first transition state of the acylation process, **TS^{R1-I4}**, relative to reactants state (23.4 and 23.5 kcal·mol^{-1}, respectively, as shown in Figure 4), together with the intrinsic uncertainty of the employed computational method, both steps would be contributing to determine the kinetics of the overall reaction. The relative energy of the product of the reaction, P1, describes the reaction as an exergonic process in thermodynamic terms (-14.6 kcal·mol^{-1}). The deacylation process occurs similarly to some of those studies done in our group and their energies are quite comparable,[38,40] 26.6 kcal·mol^{-1} and 22.8 kcal·mol^{-1}, respectively. Nevertheless, due to the different distribution of the active site, the role that plays the catalytic H211 residue is significantly different.
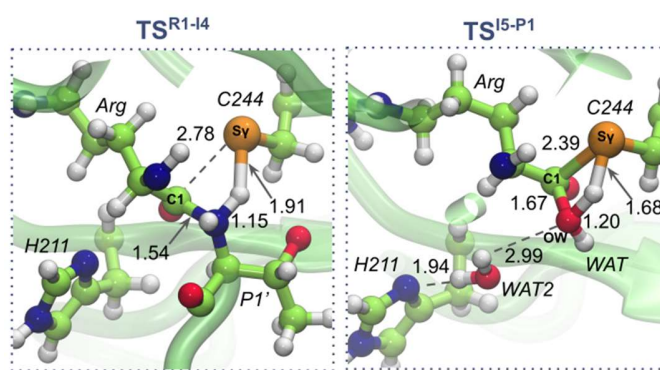
Figure 6. PBE+D3(BJ)/MM optimized structures of the rate limiting transition states of the acylation (**TS^{R1-I4}**, left panel) and deacylation (**TS^{I5-P1}**, right panel) stages in the reaction of proteolysis catalyzed by the RgpB gingipain.

5. CONCLUSIONS

In this study the reaction mechanism of proteolysis of RgpB gingipain was revealed at the atomic level by means of QM/MM MD simulations. Initially, due to the lack of data of the protonation states of the Cys/His catalytic dyad, their roles in the reaction mechanism and their particular arrangement in the active site, three structures with different protonation states (**R1**, **R2** and **R3**) were explored by performing classical MD simulations. From the analysis of the MD simulations, conformational changes of a flexible loop that affects the interactions between the peptide and E152 and H211 residues were observed. While R2 and R3 show a wide range of conformations or an open non-reactive conformation of the flexible loop, R1 is always in a closed and reactive conformation keeping the interactions between E152, H211 residues and the peptide.

Starting from R1, an exhaustive mechanistic study has been performed by the exploration of all the plausible reaction mechanisms catalyzed by RgpB gingipain. The most likely one consists in two steps for the acylation stage and one step for the deacylation stage. In the first step of the acylation takes place the nucleoplilic attack of Sγ:C244 atom to the C1 atom of the peptide and the proton transfer from Sγ:C244 atom to N1 atom of the peptide in a concerted manner. Subsequently, the bond breaking of the peptide bond takes place with the localization and characterization of an I5 intermediate both structurally and energetically similar to the one reported in the study of the protease mechanism of Human Legumain carried out by Elsässer and coworkers.[39] The rate limiting step for the acylation is the first step with an activation free energy of 23.4 kcal·mol^{-1}, very close to the value that can be deduced from previous kinetic reports, 22.8 kcal·mol^{-1}.[21] Finally, the deacylation stage takes place in a single step where a water

molecule attacks the C1 of the peptide and one proton of the water is transferred to the Sγ:C244 atom in a concerted way with a free energy barrier of 23.5 kcal·mol^{-1} relative to R1 reactant state.

It is noteworthy that along this reaction mechanism, promoted by the unusual arrangement of the catalytic dyad in active site, H211 residue does not play a role as a base. Alternatively, it plays an essential role in the orientation of the two water molecules located in the active site after the first fragment of the peptide is release in the acylation step. This reaction mechanism also suggests the crucial role played by the substrate in promoting enzyme reactivity. This role should be taken into account for future drug designing of inhibitors for RgpB for the treatment of Alzheimer's disease and other human conditions.

ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge.

Details to perform the PMFs; standard deviations of RMSD values for R1, R2 and R3; snapshot of the active site of RgpB protease in R1 state; all the PMFs computed; free energy profile of the alternative reaction mechanisms; main average interaction energies between peptide and the protein residues; CHelpG charges of the key atoms calculated on the localized PBE+D3(BJ)/MM stationary points; quantum region TS Cartesian coordinates computed using PBE-D3(BJ)/MM.

Data and Software Availability

Amber14 Package can be purchased from ambermd.org/GetAmber.php
QM3 Suite is freely available via a public GitHub repository github.com/sergio-marti/qm3
AmberTools17 that can be get from ambermd.org/AmberTools.php
Gaussian 09 D01 can be purchased from gaussian.com
fDYNAMO v2.2 can be freely downloaded from www.pdynamo.org/downloads

6. REFERENCES

(1)   Sharma, P.; Srivastava, P.; Seth, A.; Tripathi, P. N.; Banerjee, A. G.; Shrivastava, S. K. Comprehensive Review of Mechanisms of Pathogenesis Involved in Alzheimer's Disease and Potential Therapeutic Strategies. *Prog. Neurobiol.* **2019**, *174*, 53–89.

(2)   De Strooper, B.; Karran, E. The Cellular Phase of Alzheimer's Disease. *Cell* **2016**, *164*, 603–615.

(3)   Mirakhur, A.; Craig, D.; Hart, D. J.; Mcllroy, S. P.; Passmore, A. P. Behavioural and Psychological Syndromes in Alzheimer's Disease. *Int. J. Geriatr. Psychiatry* **2004**, *19*, 1035–1039.

(4)   Kumar, A.; Singh, A.; Ekavali. A Review on Alzheimer's Disease Pathophysiology and Its Management: An Update. *Pharmacol. Reports* **2015**, *67*, 195–203.

(5)   Huang, Y.; Mucke, L. Alzheimer Mechanisms and Therapeutic Strategies. *Cell* **2012**, *148*, 1204–1222.

(6)   Kaye, E. K.; Valencia, A.; Baba, N.; Spiro, A.; Dietrich, T.; Garcia, R. I. Tooth Loss and Periodontal Disease Predict Poor Cognitive Function in Older Men. *J. Am. Geriatr. Soc.* **2010**, *58*, 713–718.

(7)   Gatz, M.; Mortimer, J. A.; Fratiglioni, L.; Johansson, B.; Berg, S.; Reynolds, C. A.; Pedersen, N. L. Potentially Modifiable Risk Factors for Dementia in Identical Twins. *Alzheimer's Dement.* **2006**, *2*, 110–117.

(8)   Stein, P. S.; Desrosiers, M.; Donegan, S. J.; Yepes, J. F.; Kryscio, R. J. Tooth Loss, Dementia and Neuropathology in the Nun Study. *J. Am. Dent. Assoc.* **2007**, *138*, 1314–1322.

(9)   Kamer, A. R.; Pirraglia, E.; Tsui, W.; Rusinek, H.; Vallabhajosula, S.; Mosconi, L.; Yi, L.; McHugh, P.; Craig, R. G.; Svetcov, S.; Linker, R.; Shi, C.; Glodzik, L.; Williams, S.; Corby, P.; Saxena, D.; de Leon, M. J. Periodontal Disease Associates with Higher Brain Amyloid Load in Normal Elderly. *Neurobiol. Aging* **2015**, *36*, 627–633.

(10)   Noble, J. M.; Borrell, L. N.; Papapanou, P. N.; Elkind, M. S. V; Scarmeas, N.; Wright, C. B.

Periodontitis Is Associated with Cognitive Impairment among Older Adults: Analysis of NHANES-III. *J. Neurol. Neurosurg. Psychiatry* **2009**, *80*, 1206–1211.

(11) Dominy, S. S.; Lynch, C.; Ermini, F.; Benedyk, M.; Marczyk, A.; Konradi, A.; Nguyen, M.; Haditsch, U.; Raha, D.; Griffin, C.; Holsinger, L. J.; Arastu-Kapur, S.; Kaba, S.; Lee, A.; Ryder, M. I.; Potempa, B.; Mydel, P.; Hellvard, A.; Adamowicz, K.; Hasturk, H.; Walker, G. D.; Reynolds, E. C.; Faull, R. L. M.; Curtis, M. A.; Dragunow, M.; Potempa, J. Porphyromonas Gingivalis in Alzheimer's Disease Brains: Evidence for Disease Causation and Treatment with Small-Molecule Inhibitors. *Sci. Adv.* **2019**, *5*, eaau3333.

(12) Mottram, J. C.; Helms, M. J.; Coombs, G. H.; Sajid, M. Clan CD Cysteine Peptidases of Parasitic Protozoa. *Trends Parasitol.* **2003**, *19*, 182–187.

(13) Mayrand, D.; Holt, S. C. Biology of Asaccharolytic Black-Pigmented Bacteroides Species. *Microbiol. Rev.* **1988**, *52*, 134–152.

(14) Mahendra, J.; Mahendra, L.; Kurian, V. M.; Jaishankar, K.; Mythilli, R. Prevalence of Periodontal Pathogens in Coronary Atherosclerotic Plaque of Patients Undergoing Coronary Artery Bypass Graft Surgery. *J. Maxillofac. Oral Surg.* **2009**, *8*, 108–113.

(15) Katz, J.; Chegini, N.; Shiverick, K. T.; Lamont, R. J. Localization of P. Gingivalis in Preterm Delivery Placenta. *J. Dent. Res.* **2009**, *88*, 575–578.

(16) Ishikawa, M.; Yoshida, K.; Okamura, H.; Ochiai, K.; Takamura, H.; Fujiwara, N.; Ozaki, K. Oral Porphyromonas Gingivalis Translocates to the Liver and Regulates Hepatic Glycogen Synthesis through the Akt/GSK-3β Signaling Pathway. *Biochim. Biophys. Acta - Mol. Basis Dis.* **2013**, *1832*, 2035–2043.

(17) Mougeot, J.-L. C.; Stevens, C. B.; Paster, B. J.; Brennan, M. T.; Lockhart, P. B.; Mougeot, F. K. B. Porphyromonas Gingivalis Is the Most Abundant Species Detected in Coronary and Femoral Arteries. *J. Oral Microbiol.* **2017**, *9*, 1281562.

(18) Potempa, J.; Travis, J. Porphyromonas Gingivalis Proteinases in Periodontitis, a Review. *Acta Biochim. Pol.* **1996**, *43*, 455–465.

(19) Potempa, J.; Pavloff, N.; Travis, J. Porphyromonas Gingivalis: A Proteinase/Gene Accounting Audit. *Trends in Microbiology*. Elsevier Current Trends November 1, 1995, pp 430–434.

(20) Curtis, M. A.; Kuramitsu, H. K.; Lantz, M.; Macrina, F. L.; Nakayama, K.; Potempa, J.; Reynolds, E. C.; Aduse-Opoku, J. Molecular Genetics and Nomenclature of Proteases of

Porphyromonas Gingivalis. *J. Periodontal Res.* **1999**, *34*, 464–472.

(21) Ally, N.; Whisstock, J. C.; Sleprawska-Lupa, M.; Potempa, J.; Le Bonniec, B. F.; Travis, J.; Pike, R. N. Characterization of the Specificity of Arginine-Specific Gingipains from Porphyromonas Gingivalis Reveals Active Site Differences between Different Forms of the Enzymes. *Biochemistry* **2003**, *42*, 11693–11700.

(22) Eichinger, A.; Beisel, H. G.; Jacob, U.; Huber, R.; Medrano, F. J.; Banbula, A.; Potempa, J.; Travis, J.; Bode, W. Crystal Structure of Gingipain R: An Arg-Specific Bacterial Cysteine Proteinase with a Caspase-like Fold. *EMBO J.* **1999**, *18*, 5453–5462.

(23) Potempa, J.; Mikolajczyk-Pawlinska, J.; Brassell, D.; Nelson, D.; Thøgersen, I. B.; Enghild, J. J.; Travis, J. Comparative Properties of Two Cysteine Proteinases (Gingipains R), the Products of Two Related but Individual Genes of Porphyromonas Gingivalis. *J. Biol. Chem.* **1998**, *273*, 21648–21657.

(24) Bedi, G. S. Purification and Characterization of Lysine- and Arginine-Specific Gingivain Proteases from Porphyromonas Gingivalis. *Prep. Biochem.* **1994**, *24*, 251–261.

(25) Fujimura, S.; Shibata, Y.; Nakamura, T. Purification and Partial Characterization of a Lysine-Specific Protease of Porphyromonas Gingivalis. *FEMS Microbiol. Lett.* **1993**, *113*, 133–137.

(26) Hinode, D.; Hayashi, H.; Nakamura, R. Purification and Characterization of Three Types of Proteases from Culture Supernatants of Porphyromonas Gingivalis. *Infect. Immun.* **1991**, *59*, 3060–3068.

(27) Bleeg, H. S. Non-Specific Cleavage of Collagen by Proteinases in the Presence of Sodium Dodecyl Sulfate. *Eur. J. Oral Sci.* **1990**, *98*, 235–241.

(28) Griffen, A. L.; Becker, M. R.; Lyons, S. R.; Moeschberger, M. L.; Leys, E. J. Prevalence of Porphyromonas Gingivalisand Periodontal Health Status. *J. Clin. Microbiol.* **1998**, *36*, 3239–3242.

(29) Harrison, M. J.; Burton, N. A.; Hillier, I. H. Catalytic Mechanism of the Enzyme Papain: Predictions with a Hybrid Quantum Mechanical/Molecular Mechanical Potential. *J. Am. Chem. Soc.* **1997**, *119*, 12285–12291.

(30) Wei, D.; Huang, X.; Liu, J.; Tang, M.; Zhan, C. G. Reaction Pathway and Free Energy Profile for Papain-Catalyzed Hydrolysis of N-Acetyl-Phe-Gly 4-Nitroanilide. *Biochemistry* **2013**, *52*, 5145–5154.

(31) Arad, D.; Langridge, R.; Kollman, P. A. A Simulation of the Sulfur Attack in the Catalytic Pathway of Papain Using Molecular Mechanics and Semiempirical Quantum Mechanics. *J. Am. Chem. Soc.* **1990**, *112*, 491–502.

(32) Welsh, W. J.; Lin, Y. Discussion of the Catalytic Pathway of Cysteine Proteases Based on AM1 Calculations. *J. Mol. Struct. THEOCHEM* **1997**, *401*, 315–326.

(33) Sulpizi, M.; Rothlisberger, U.; Carloni, P. Molecular Dynamics Studies of Caspase-3. *Biophys. J.* **2003**, *84*, 2207–2215.

(34) Sulpizi, M.; Laio, A.; VandeVondele, J.; Cattaneo, A.; Rothlisberger, U.; Carloni, P. Reaction Mechanism of Caspases: Insights from QM/MM Car-Parrinello Simulations. *Proteins Struct. Funct. Genet.* **2003**, *52*, 212–224.

(35) Miscione, G. Pietro; Calvaresi, M.; Bottoni, A. Computational Evidence for the Catalytic Mechanism of Caspase-7. A DFT Investigation. *J. Phys. Chem. B* **2010**, *114*, 4637–4645.

(36) Angelides, K. J.; Fink, A. L. Mechanism of Thiol Protease Catalysis: Detection and Stabilization of a Tetrahedral Intermediate in Papain Catalysis. *Biochemistry* **1979**, *18*, 2363–2369.

(37) Ŝtrajbl, M.; Florián, J.; Warshel, A. Ab Initio Evaluation of the Free Energy Surfaces for the General Base/Acid Catalyzed Thiolysis of Formamide and the Hydrolysis of Methyl Thiolformate: A Reference Solution Reaction for Studies of Cysteine Proteases. *J. Phys. Chem. B* **2001**, *105*, 4471–4484.

(38) Arafet, K.; Ferrer, S.; Moliner, V. Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease. *ACS Catal.* **2017**, *7*, 1207–1215.

(39) Elsässer, B.; Zauner, F. B.; Messner, J.; Soh, W. T.; Dall, E.; Brandstetter, H. Distinct Roles of Catalytic Cysteine and Histidine in the Protease and Ligase Mechanisms of Human Legumain As Revealed by DFT-Based QM/MM Simulations. *ACS Catal.* **2017**, *7*, 5585–5593.

(40) Świderek, K.; Moliner, V. Revealing the Molecular Mechanisms of Proteolysis of SARS-CoV-2 Mproby QM/MM Computational Methods. *Chem. Sci.* **2020**, *11*, 10626–10630.

(41) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, *10*, 12544–12554.

(42) Ramos-Guzmán, C. A.; Zinovjev, K.; Tuñón, I. Modeling Caspase-1 Inhibition: Implications for Catalytic Mechanism and Drug Design. *Eur. J. Med. Chem.* **2019**, *169*, 159–167.

(43) Arafet, K.; Ferrer, S.; González, F. V.; Moliner, V. Quantum Mechanics/Molecular Mechanics

Studies of the Mechanism of Cysteine Protease Inhibition by Peptidyl-2,3-Epoxyketones. *Phys. Chem. Chem. Phys.* **2017**, *19*, 12740–12748.

(44) Arafet, K.; Świderek, K.; Moliner, V. Computational Study of the Michaelis Complex Formation and the Effect on the Reaction Mechanism of Cruzain Cysteine Protease. *ACS Omega* **2018**, *3*, 18613–18622.

(45) Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V.; Świderek, K.; Moliner, V. Mechanism of Inhibition of SARS-CoV-2 M pro by N3 Peptidyl Michael Acceptor Explained by QM/MM Simulations and Design of New Derivatives with Tunable Chemical Reactivity. *Chem. Sci.* **2021**, *12*, 1433–1444.

(46) Hanada, K.; Tamai, M.; Yamagishi, M.; Ohmura, S.; Sawada, J.; Tanaka, I. Isolation and Characterization of E–64, a New Thiol Protease Inhibitor. *Agric. Biol. Chem.* **1978**, *42*, 523–528.

(47) Rangarajan, M.; Smith, S. J. M.; U, S.; Curtis, M. A. Biochemical Characterization of the Arginine-Specific Proteases of Porphyromonas Gingivalis W50 Suggests a Common Precursor. *Biochem. J.* **1997**, *323*, 701–709.

(48) Bateman, A.; Martin, M.-J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bursteinas, B.; Bye-A-Jee, H.; Coetzee, R.; Cukura, A.; Da Silva, A.; Denny, P.; Dogan, T.; Ebenezer, T.; Fan, J.; Castro, L. G.; Garmiri, P.; Georghiou, G.; Gonzales, L.; Hatton-Ellis, E.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Jokinen, P.; Joshi, V.; Jyothi, D.; Lock, A.; Lopez, R.; Luciani, A.; Luo, J.; Lussi, Y.; MacDougall, A.; Madeira, F.; Mahmoudy, M.; Menchi, M.; Mishra, A.; Moulang, K.; Nightingale, A.; Oliveira, C. S.; Pundir, S.; Qi, G.; Raj, S.; Rice, D.; Lopez, M. R.; Saidi, R.; Sampson, J.; Sawford, T.; Speretta, E.; Turner, E.; Tyagi, N.; Vasudev, P.; Volynkin, V.; Warner, K.; Watkins, X.; Zaru, R.; Zellner, H.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Blatter, M.-C.; Bolleman, J.; Boutet, E.; Breuza, L.; Casals-Casas, C.; de Castro, E.; Echioukh, K. C.; Coudert, E.; Cuche, B.; Doche, M.; Dornevil, D.; Estreicher, A.; Famiglietti, M. L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Keller, G.; Kerhornou, A.; Lara, V.; Le Mercier, P.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T. B.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pourcel, L.; Pozzato, M.; Pruess, M.; Rivoire, C.; Sigrist, C.; Sonesson, K.; Stutz, A.; Sundaram, S.; Tognolli, M.; Verbregue, L.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Garavelli, J. S.; Huang, H.; Laiho,

K.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Q.; Wang, Y.; Yeh, L.-S.; Zhang, J.; Ruch, P.; Teodoro, D. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489.

(49)   Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(50)   Li, P.; Song, L. F.; Merz, K. M. Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883–895.

(51)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(52)   Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210.

(53)   Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(54)   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.

(55)   Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

(56)   Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

(57)   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(58)   Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N ·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(59)   Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of Trajectories in Canonical Molecular-Dynamics Simulations: Observation, Explanation, and Exploitation. *J. Chem. Phys.* **2004**, *120*, 6363–6374.

(60) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. Bad Seeds Sprout Perilous Dynamics: Stochastic Thermostat Induced Trajectory Synchronization in Biomolecules. *J. Chem. Theory Comput.* **2009**, *5*, 1624–1631.

(61) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(62) Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J. Comput. Chem.* **2001**, *22*, 501–508.

(63) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98–103.

(64) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(65) Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275–282.

(66) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **1995**, *16*, 1339–1350.

(67) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(68) Kästner, J.; Thiel, W. Bridging the Gap between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: "Umbrella Integration." *J. Chem. Phys.* **2005**, *123*, 144104.

(69) Martí, S. QMCube (QM3): An All-Purpose Suite for Multiscale QM/MM Calculations. *J. Comput. Chem.* **2021**, *42*, 447–457.

(70) Rodriguez, A.; Laio, A. Clustering by Fast Search and Find of Density Peaks. *Science (80-.).* **2014**, *344*, 1492–1496.

(71) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(72) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An

Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.

(73) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(74) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21+G Basis Set for First-Row Elements, Li-F. *J. Comput. Chem.* **1983**, *4*, 294–301.

(75) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(76) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.

(77) Case Tom Darden Thomas E Cheatham III Carlos Simmerling Adrian Roitberg Junmei Wang, D. A.; Götz SDSC, A. W.; István Kolossváry Budapest, U.; Shaw Francesco Paesani Jian Liu Xiongwu Wu Thomas Steinbrecher, D. *Amber 14 Reference Manual Principal Contributors to the Current Codes*.

(78) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 09, Revision A.02. *Gaussian, Inc. Wallingford CT* **2009**.

(79) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The Dynamo Library for

Molecular Simulations Using Hybrid Quantum Mechanical and Molecular Mechanical Potentials. *J. Comput. Chem.* **2000**, *21*, 1088–1100.

(80) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*; Cambridge University Press, 1999.

(81) Mazmanian, K.; Sargsyan, K.; Grauffel, C.; Dudev, T.; Lim, C. Preferred Hydrogen-Bonding Partners of Cysteine: Implications for Regulating Cys Functions. *J. Phys. Chem. B* **2016**, *120*, 10288–10296.

(82) Dall'Acqua, W.; Carter, P. Substrate-Assisted Catalysis: Molecular Basis and Biological Significance. *Protein Sci.* **2008**, *9*, 1–9.

Table of Contents