Different enzyme conformations induce different mechanistic traits in HIV-1 protease

RESEARCH ARTICLE

Different Enzyme Conformations Induce Different Mechanistic Traits on HIV-1 Protease

João T. S. Coimbra,[a] Rui P. P. Neves,[a] Ana V. Cunha,[b,c] Maria J. Ramos,[a] and Pedro A. Fernandes*[a]

[a] Dr. J. T. S. Coimbra, Dr. R. P. P. Neves, Prof. Dr. M. J. Ramos, Prof. Dr. P. A. Fernandes
     LAQV/REQUIMTE, Departamento de Química e Bioquímica
     Faculdade de Ciências Universidade do Porto
     Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal
     E-mail: pafernan@fc.up.pt
[b] Prof. Dr. A. V. Cunha
     Scientific Computing Group
     Oak Ridge National Laboratory
     1 Bethel Valley Rd, Oak Ridge, TN 37831-6373 USA
[c] Prof. Dr. A. V. Cunha (current address)
     Department of Chemistry
     University of Antwerp
     Groenenborgerlaan 171, 2020 Antwerp, Belgium

   Supporting information for this article is given via a link at the end of the document.

Abstract: The influence of the dynamical flexibility of enzymes in reaction mechanisms is a cornerstone in biological sciences. In this study we aim to: (1) study the convergence of the activation free energy, using the first step of the reaction catalysed by HIV-1 protease as a case study; and (2) provide further evidence for a mechanistic divergence in this enzyme, as two different reaction pathways were observed to contribute to this step. Here we used quantum mechanics/molecular mechanics molecular dynamics simulations, on four different initial conformations, which led to different barriers in a previous study. Despite the sampling, the four activation free energies still spanned a range of 5.0 kcal·mol^{-1}. Furthermore, the present simulations did confirm the occurrence of an unusual mechanistic divergence, with two different mechanistic pathways displaying equivalent barriers. An active site water molecule was proposed to influence the mechanistic pathway.

Introduction

The human immune deficiency virus (HIV) is the pathogen responsible for the acquired immunodeficiency syndrome (AIDS). By the end of 2019, it was estimated that 37.7 million people were living with HIV, and despite the several progresses in the last decades, HIV continues to be a major public health issue.[1] One of the most alarming issues nowadays has been the increase in the levels of HIV resistance to antiretroviral drugs, which increases the risk of these drugs becoming partly or fully inactive.[2]

The HIV protease (HIV-PR) is essential for the virus maturation, hydrolysing the viral Gag and Gag-Pol precursor polyproteins during the maturation of the infectious viral particle, and is thus an important target for the treatment of HIV. Currently, there are several approved drugs that target the HIV-PR.[3] One of the most potent FDA-approved HIV-PR inhibitors, Darunavir, binds to wild-type HIV-1 PR with a potency of less than 5 pM. However, in a recent study by Henes *et al.*, an HIV-1 PR variant, with mutations in and outside the active site was identified.

Furthermore, it was shown that the potency to Darunavir is reduced in up to 150,000-fold.[4] This is an excellent example of the implications of drug resistance to antiretroviral therapy that can impact the ongoing efforts to end the AIDS epidemic by 2030.
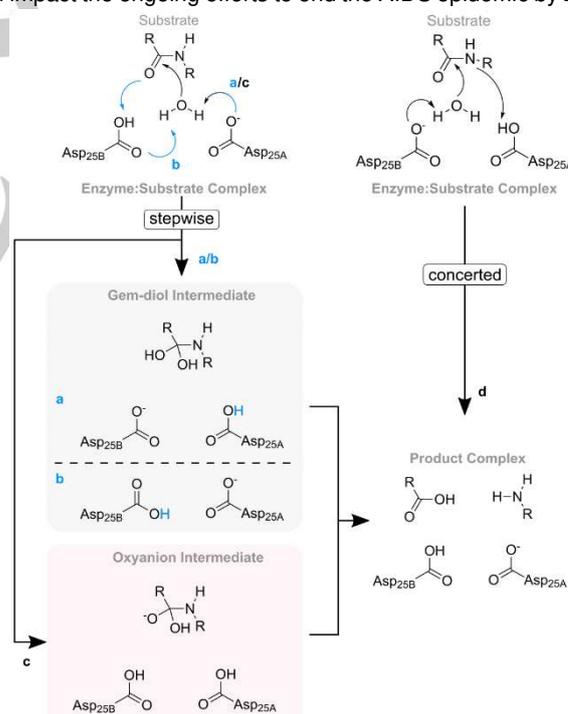
Scheme 1. Most relevant mechanistic proposals for the hydrolysis of peptides by HIV-1 PR: the proposed stepwise (**a**, **b**, and **c**) and concerted (**d**) general acid-base reaction mechanism. In this work we explored the a and b pathways (highlighted in blue) up to the gem-diol intermediate state.

The HIV-1 PR is an aspartic protease consisting of a homodimer with 99 amino acid residues per monomer. Despite the numerous studies presented in the literature, its catalytic mechanism is still under debate.[5] The HIV-1 PR uses two catalytic aspartic acid residues to hydrolyse a peptide bond

through an acid-base mechanism. Important features about the reaction have been reported by solution enzyme kinetics and NMR studies,[5-6] revealing important information about the protonation state of the two Asp residues and the role of a lytic water molecule in the reaction. They also revealed information regarding and the formation of a tetrahedral amide hydrate intermediate along the reaction, which collapses to the product after a step comprising the protonation of the nitrogen of the C-N scissile bond by a catalytic Asp.

Due to the presence of multiple basic oxygen atoms in the active site of HIV-1 PR, and the lack of structural information on the reactant Michaelis complex, tetrahedral intermediate and products, several proposals for the reaction mechanism have surfaced in the literature.[5, 7] Recent computational studies supported a gem-diol tetrahedral intermediate,[7a] in line with previous calculations by Carloni and coworkers that also found the gem-diol intermediate to be more stable than the charged oxyanion.[8] Still, a recent crystallographic study on a protease:peptidomimetic inhibitor complex, proposed the tetrahedral intermediate to be an oxyanion instead.[5] More recently, a one-step concerted mechanism for the enzyme was proposed based on Quantum Mechanics/Molecular Mechanics (QM/MM) studies.[7b] A summary of the most relevant mechanistic proposals, based in theory and experiment, is depicted in Scheme 1.[7a]

Despite the continuous efforts to decipher the reaction pathway for this enzyme, there is still ongoing debate. This study focuses on the recent hypotheses by Calixto et al.,[9] where the authors have discussed the possibility of mechanistic divergence, depending on the instantaneous enzyme conformation. This study was performed with the adiabatic QM/MM ONIOM (Our own N-layered Integrated molecular Orbital and molecular Mechanics) methodology using 19 different enzyme:substrate conformations, which were collected from an equilibrated *NPT* ensemble generated *via* a classical MD simulation. The authors observed that the conformational dynamics of the enzyme influenced both the barrier heights and the reaction pathway of the first step of the reaction mechanism – the nucleophilic attack by a water molecule to the carbonyl carbon of the substrate's scissile bond. They reported a spread on the activation barriers of 14.9 kcal·mol^{-1} at the M06-2X/6-311++G(2d,2p):ff99SB//B3LYP/6-31G(d):ff99SB level of theory. They also observed that the first step of the reaction could proceed through either a one-aspartate or a two-aspartate mechanistic pathway (see Scheme 1, mechanisms a or b), with similar barriers. These two possible routes were shown to be dependent on multiple structural factors, including the interaction with a nearby water molecule in the active site and with the Thr26B residue.

The Potential Energy Surface (PES) of an enzymatic reaction is dependent on the specific protein conformation.[10] Conformational fluctuations span a wide array of timescales that can be studied with different computational methods. An excellent study of Carloni and co-workers [11] focused on the role of large timescale conformational fluctuations on the reaction by HIV-1 PR, movements that are beyond the possibilities of a study at the QM/MM level. Here we focus on faster, sub-nanosecond timescale movements, complementing the study of Carloni and co-workers by covering a faster timescale.

In the case of HIV-1 protease, 36.8 kcal·mol^{-1} variations have been reported among calculated activation free energies starting from conformations separated by 2 ns.[12] This highlights that the choice of the enzyme:substrate conformation has a relevant influence in single-structure QM/MM ONIOM calculation results. This aspect can be overcome by a proper sampling of the conformational space in the study of enzymatic reactions, either through multi-PES QM/MM or QM/MM MD methods.[13] Nevertheless, the use of the reactant conformations taken from good-quality X-ray structures can be a solution for this limitation, as the X-ray conformation is not a "single conformation" but instead an average over all conformations in the crystal.[13a, 14] However, the availability of very high resolution X-Ray structures is often scarce. From a more technical point of view, the large energy fluctuations in the barrier could also derive in part from MM energy contributions that often result from conformational changes and fluctuations taking place during geometry optimizations. This can be in part accounted for by freezing the outer MM region that is most distant from the region where the reaction takes place.[15]

A typical alternative is to perform QM/MM MD simulations, which provide a wider sampling along the reaction coordinate. This is however quite challenging, as these simulations require a very large number of computations of the system's energy and its gradient. These simulations are thus often performed with lower-level QM methods, such as wavefunction semi-empirical methods or SCC-DFTB, which tend to yield less accurate energy barriers. Hence, these results are often refined by single-point energy calculations with higher levels of theory, or by re-parameterization of semi-empirical methods for a particular system.[13a]

Since QM/MM MD methods are often computationally very demanding, adiabatic methods that resort to the QM cluster model or adiabatic QM/MM protocols are thus frequently and successfully employed.[16] These often require multiple conformations of the system under study, *i.e.* multi-PES QM/MM, to circumvent the limited sampling problem of the very complex potential energy surface of enzymes.[17] Still, if conformational sampling effects are not substantial and if the limitations of the modelling are grasped and carefully inspected, one can derive important conclusions and correct reaction mechanisms with limited-sampling approaches and rationalize the impact of key structural aspects to the energy barriers of the studied reaction, something for which adiabatic methods are very well suited.[9, 12]

We were interested in exploring the first step of the reaction mechanism of peptide hydrolysis by HIV-1 PR through QM/MM MD umbrella sampling simulations, to confirm if the mechanistic divergence found in the work published by Calixto *et al.* back in 2019 could be further supported by extensive conformational sampling. We have selected four structures from that work, specifically those that led to the highest and lowest barriers for the first step of the reaction for each of the two different mechanistic pathways that have been observed – the one-aspartate and two-aspartate pathways (Scheme 1). The two extreme starting points were chosen to evaluate how well the convergence of the free energy barrier for the first step of the reaction would be accomplished, and the two mechanistic scenarios were chosen to determine whether the reaction followed a one-aspartate or a two-aspartate mechanism within a robust computational protocol, and to analyse some of the structural features that influence this step.

The results highlighted technical aspects that are broadly applicable in the study of enzyme reaction mechanisms with hybrid QM/MM approaches, and underlined important features of the catalysis by HIV-1 PR. They also expanded the understanding of aspartic proteases, something that can have implications in the

studies of other retroviruses such as the human T-cell lymphotropic virus type 1 (HTLV-1), and on future endeavours on drug design campaigns.[18]

Results and Discussion

This study focuses on the first step of the catalytic mechanism of HIV-1 PR. There are multiple proposals in the literature regarding this catalytic mechanism. One of the most recent proposals was published by Calixto et al.,[9] where the authors observed two probable pathways for the reaction to take place. Using adiabatic QM/MM methods, they also observed large fluctuations on the energy barrier of the first step, which was independent of the mechanistic pathway for the reaction. We decided to expand on this study and evaluate the convergence of the free energy profile for the first step, and the convergence or divergence of the mechanistic pathway. We also discuss the role of a water molecule buried at the reaction site, which we found to influence the mechanism of the enzyme.
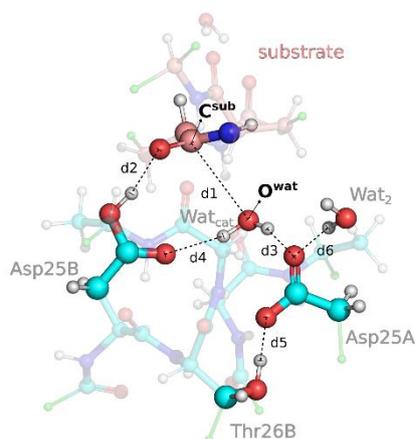

Figure 1. Atoms included in the QM region used in this study, shown as ball-and-sticks with the link atoms in green. The most important atoms for the reaction are highlighted (*i.e.*, shown without transparency). Important distances and residues discussed along the work are also labelled.

To understand the free energy profile dependence on the starting conformation, we selected four starting structures from the work by Calixto et al..[9] With these four starting conformations, the authors have reported an energy span among the four activation barriers (zero-point corrected total energies, $\Delta E_0^{\ddagger}$) of 14.9 kcal·mol^{-1}, and two possible pathways for the first step: i) one in which the deprotonation of the hydrolytic water molecule and the protonation of the peptide oxygen rely on the same aspartate residue (Scheme 1 b); and ii) the other where both aspartates have a catalytic role (Scheme 1 a) – one protonates the peptide oxygen and the other deprotonates the hydrolytic water molecule. The reported activations barriers and mechanistic pathways are presented in Table S1, for each of the four starting structures and as reported by Calixto et al. These initial conformations will be hereby referred to as "one Asp E^{low}", "one Asp E^{high}", "two Asp E^{low}", and "two Asp E^{high}". We explored the same step with QM/MM MD simulations to evaluate how well the four free energy profiles converged to a similar barrier of activation, and whether this could be achieved within the ps-timescale typically used in this kind of simulations. The sampling of the QM region was performed at the PBE/DZVP level of theory.

The QM region is represented in Figure 1, together with a depiction of important distances and residues that are discussed in this work. The dangling bonds between the QM and MM region are also highlighted, and these have been capped with hydrogen atoms.

In Figure 2, we represent the free energy profiles for the first step by HIV-1 PR, using the four distinct starting conformations of the enzyme:substrate complex. We can see the change in the free energy profiles as we decrease the interatomic distance between the oxygen atom (O_{wat}) of the catalytic water molecule and the carbon atom (C_{sub}) of the substrate's scissile bond (d1 in Figure 1). Figure 2 shows that the free energy profile derived from each of the four simulations that started from different initial structures was similar, and that the free energy barriers varied between 17.5-22.0 kcal·mol^{-1}. It is interesting to see that the spread of the barriers was much larger than the statistical error within each umbrella sampling simulation (*ca.* 0.1 kcal·mol^{-1}), showing that the QM/MM MD results still depend significantly on the starting conformation, even with 10 ps sampling/window, implying that longer timescale sampling is needed to bring down this uncertainty to within 1-2 kcal·mol^{-1}, a level where it is no longer a relevant contributor for the overall accuracy of the result.
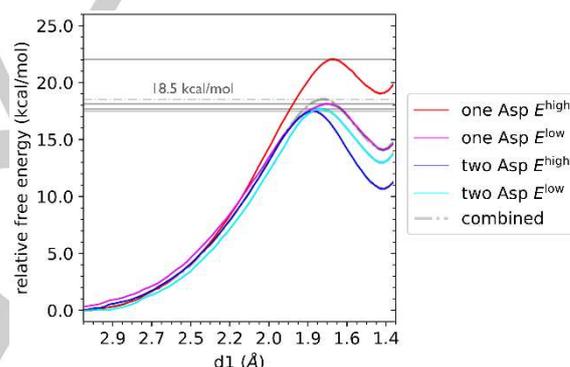

Figure 2. Free energy profiles for the nucleophilic attack to the carbonyl carbon atom of the substrate's scissile bond by a water molecule. The d1 distance, representing the distance between the nucleophilic oxygen atom of the catalytic water molecule (O_{wat}) and the carbonyl's carbon atom of the scissile bond on the substrate (C_{sub}), was used as a reaction coordinate. The free energy profiles of the four starting configurations (one Asp E^{high}, one Asp E^{low}, two Asp E^{high}, two Asp E^{low}) are depicted. In grey lines we show the maximum of each free energy profile. The dashed grey profile represents the free energy profile of all four starting conformations combined and the maximum value of this profile is depicted. For this analysis we have discarded the first 2 ps of each umbrella sampling window.

Our lowest value was close to the lowest value reported by Calixto et al.[9] of 17.3 kcal·mol^{-1}, but our upper limit was lower in *ca.* 10 kcal·mol^{-1} when compared with the highest barrier reported in that same work (32.2 kcal·mol^{-1}). This was expected because Calixto et al. reported instantaneous barriers, whose values are intrinsically more spread, and we report averaged barriers, where the instantaneous barrier variations are averaged out up to the ps timescale. The spread of instantaneous barriers is considered to be physically realistic, and not an artifact of the simulation, and is in line with the observations made in single-molecule experiments.[12, 19] When compared to the value obtained by Krzemińska et al.[7a] of 8.5
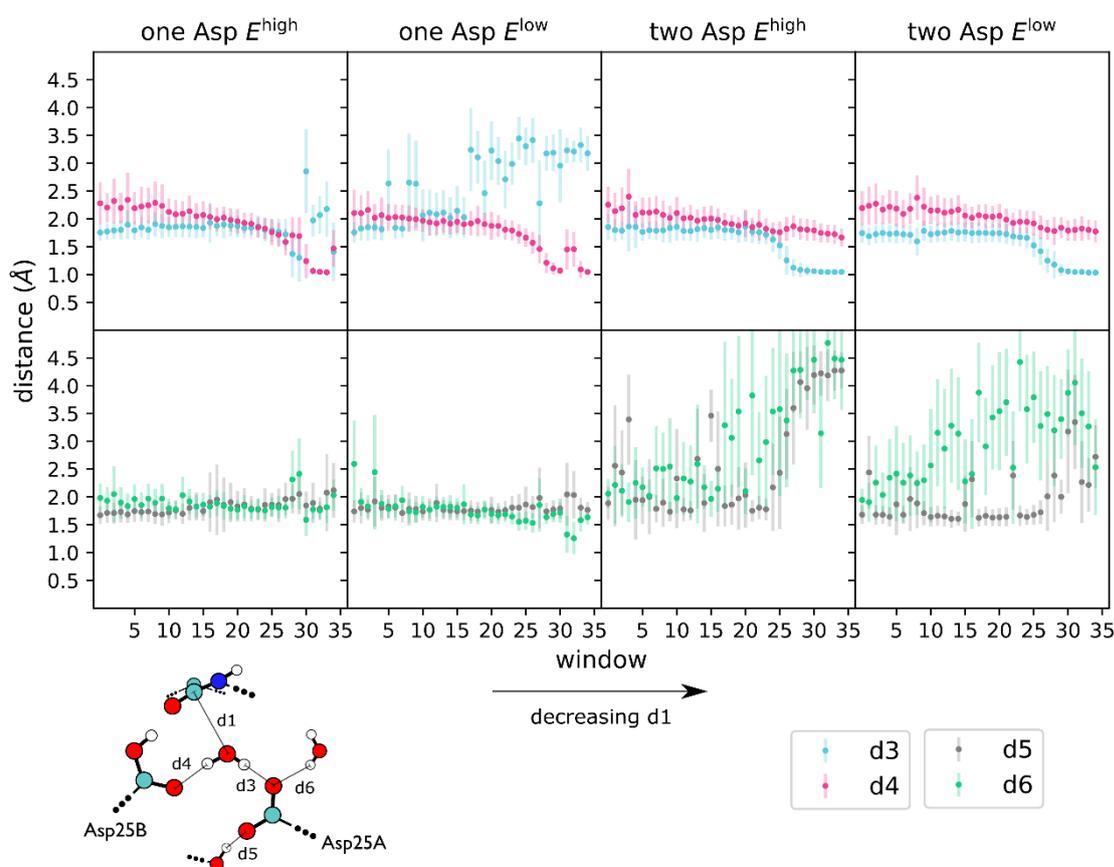
Figure 3. Selected distances – d3 (cyan), d4 (purple), d5 (grey), and d6 (green) were averaged and plotted for each umbrella sampling window and for each of the four starting configurations (one Asp E^{high}, one Asp E^{low}, two Asp E^{high}, two Asp E^{low}). In the bottom left corner, we show a representation of the selected distances. Error bars represent the standard deviation.

kcal·mol[-1], our lowest barrier was *ca.* 9 kcal·mol[-1] higher. Other computational works have also observed an activation barrier of 18.1, 16.26, and 16.5 kcal·mol[-1] for this first step (the last value is an apparent barrier, *i.e.* an average of 39 k_{cat} values, and includes a correction of 4.6 kcal·mol[-1] for the probability of having the catalytic water bound to the active site, which is a rare event).[12, 20] There are inherent differences to these works that can explain the discrepancy between all these results, those including different levels of theory, the size of the QM layer, or even conformational differences at the active site. Nevertheless, our free energy barriers are in good agreement with most values in the literature.

To the best of our knowledge, few works have addressed the convergence of the activation energies of enzyme reaction mechanisms using QM/MM MD simulations on different starting conformations. For class A β-lactamases, Chudyk *et al.*[21] observed a good convergence (standard deviation ranging from 0.1 to 2.8 kcal·mol[-1]) when they repeated the QM/MM MD simulations three times and for a total of 20 ps per window. Since we used different initial conformations of the enzyme:substrate complex and a lower simulation time for each window, and we have deliberately chosen initial structures that led to a very large difference in the adiabatic mapping barriers, we obtained larger free energy differences between the four structures. Recently, our group has also highlighted the impact of QM/MM MD simulations on the barrier heights of the first step catalysed by α-amylase.[22] This study also used four starting conformations (including the X-

ray structure) that have previously shown a large spread in adiabatic activation energies. They observed that most free energy profiles converged within 1-2 kcal·mol[-1] with 10 ps/window, even though free energy barriers could still differ within 5 kcal·mol[-1]. Hence, our differences in free energy barriers should be mostly explained by the differences among the starting conformations. In addition, we hypothesize that the fact that both enzymes relied on a water molecule from the bulk to perform catalysis, may also hamper the rapid convergence of the free energy barriers for different starting conformations.

The statistical convergence of the free energy profiles for the MD simulations that started from each of the four tested conformations was also analysed (Figure S1). For this evaluation, we have divided the production simulations in blocks of 2 ps, where we followed the evolution of the free energy profiles. Even though the barrier coming from one of the MD simulations (two Asp E^{high}) could still be decreasing, the rest of the profiles seem well converged within the simulated 10 ps per window. It also seems that the starting structures that in the previous study led to high energy barriers, generated larger energy fluctuations in this analysis. The results show that the intrinsic differences on the four starting conformations require extensive simulations to reach a satisfactory convergence of the free energy profiles, *i.e.*, to bring the spread of activation barriers within a 1-2 kcal·mol[-1] difference.

We should recall that the starting configurations were obtained from an MD simulation where the distance between the catalytic hydrogen atom of Asp25B and the carbonyl oxygen of

4

the substrate was constrained. This was done to ensure that a proper sampling of the less abundant catalytic conformation was produced. Hence, we cannot conclude on the impact of more abundant configurations outside this distance threshold and their impact on the convergence of the free energy profile for the first step of the reaction, even though it is expected that they would take more time to converge as they were less extensively pre-organized for the reaction to take place. Furthermore, since these starting conformations were spaced at the nanosecond time scale, they should not possess large conformational fluctuations in relation to the enzyme's folding. Nevertheless, Piana *et al.*[23] have already observed the impact of more substantial conformational fluctuations on the reaction by HIV-1 PR, where they obtained prohibitive configurations for the reaction to take place. Hence, there are inherent limitations that QM/MM MD simulations cannot easily circumvent, and that often require a more weighted and multi-variate approach, especially if these involve large and long-scale conformational fluctuations.[11]

Influence of the starting conformation on the reaction pathway. We also looked at the reaction pathway for the first step of the reaction catalysed by HIV-1 PR. It is remarkable that, despite the extensive relaxation that the QM/MM MD simulations provided, the simulations that started from conformations where the one-aspartate mechanism was previously observed in previous adiabatic-mapping calculations, led to that same one-aspartate mechanism. In the same line, the simulations that started from conformations where the two-aspartate mechanism was previously observed in adiabatic-mapping calculations, also led to that same two-aspartate mechanism (Figure 3). Additionally, the free energy profiles of both mechanisms led to statistically equivalent barriers. The results strongly support the previously observed mechanistic divergence in the reaction mechanism, and support that the origin of the phenomenon is the region of the conformational space where the enzyme is located, at the moment of the reaction.

Specifically, the evolution into the one-aspartate or two-aspartate pathways has been previously related to a contribution of multiple key interactions in the active site, including the hydrogen bond distance between Asp25A and both Wat$_2$ and Thr26B, and the difference in the distance between the hydrolytic water molecule and both catalytic Asp25 residues. The first two distances (d6 and d5 here) have been proposed to influence the pK_a of Asp25A, with longer distances causing an increase of its pK_a and thus favouring the deprotonation of the catalytic water by Asp25A and therefore the two-aspartate mechanism. The authors observed that these two distances influenced, but were not fully responsible, for the outcome of the mechanism. A good and clear correlation between the active site interactions and the outcome of the reaction mechanism, only emerged when the distances d3 and d4 were also considered, in addition to d5 and d6.[9]

Our results also suggested that the described network of hydrogen bonds, defined by the distances in Figure 3, indeed had a role in defining the reaction pathway. If we focus on the two Asp E^{high} and two Asp E^{low} simulations, we can see this trend – both structures led to the two-aspartate mechanism and displayed longer and unstable d5 and d6 distances, which supports that the lowering of the Asp25A pK_a due to the anchoring of the Asp25A base by Thr26B and Wat$_2$, disfavours the two-aspartate mechanism.

These results indicated that the mechanistic pathway does diverge between a one- and two-aspartate mechanism and

support the substantial influence of instantaneous enzyme conformations for this reaction.

Additionally, to assess the influence of a different representation of the active site water molecule (Wat$_2$) in the reaction, we placed it in the MM layer instead of the QM layer. We have then simulated all four initial conformations with this new QM layer description. The presence of Wat$_2$ in the MM layer, rather than the QM layer, still showed similar differences between the free energy barriers of the four simulations that started from different initial structures, even though free energy barriers decreased in *ca.* 2.0 kcal·mol^{-1} (15.1-20.3 kcal·mol^{-1}) relatively to the systems with a larger QM region (Figure S2). Interestingly, these new simulations again highlighted the role of Wat$_2$ (and possibly Thr26B) for defining the reaction pathway. When Wat$_2$ was omitted from the QM region, we observed a higher destabilization of d5 and d6 along the reaction for three of the starting structures – one Asp E^{low}, two Asp E^{low}, and two Asp E^{high} (Figure S3). In these circumstances, the strength of the hydrogen bonds it provides may decrease due to the lack of Wat$_2$ polarization, and the one-aspartate mechanism with the lower activation energy, one Asp E^{low}, becomes a two-aspartate mechanism.
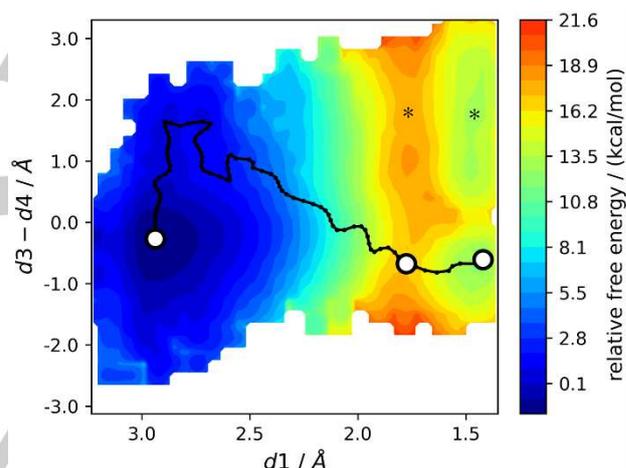

Figure 4. 2D profile for the nucleophilic attack by a water molecule to the carbonyl carbon atom of the substrate's scissile bond. Two sets of distances were used for building these profiles: d1, the reaction coordinate; and the difference between d3 and d4, which dictates the occurrence of a one-aspartate vs. a two-aspartate mechanism, *i.e.*, while negative values will indicate that the proton from the catalytic water molecule, Wat$_{cat}$, is transferred to Asp25A (two-aspartate mechanism), positive values will mean that the proton is transferred to Asp25B (one-aspartate mechanism). This analysis contemplates all four starting structures. The minimum free energy path between the reagent and the product is shown. In addition, the most likely stationary points of the alternative path (*i.e.*, one-aspartate mechanism) are also highlighted with an asterisk.

2D free energy profiles. A 2D profile (Figure 4) was calculated by accumulating the data from the four MD simulations. This helps to understand better the divergence between the one-aspartate and two-aspartate mechanisms. We have plotted the reaction coordinate (d1), against the difference between the d3 and d4 distances (Figure 1). Negative values of the difference between d3 and d4 will imply the progression through a two-aspartate mechanism, positive values will indicate a one-aspartate mechanism. A large spread for the d3-d4 difference is present throughout the reaction progression (Figure 4), which indicates that there is no overall prevalence for a one-aspartate

or a two-aspartate mechanism. The barrier along the minimum free energy path favours the two-aspartate mechanism (by only *ca.* 1 kcal·mol⁻¹, relative to the one-aspartate pathway whose most likely stationary points are marked with an asterisk in Figure 4). In addition, the intermediate states of both mechanisms are rapidly interconvertible at room temperature (barrier of *ca.* 3 kcal·mol⁻¹), and coexistent.

Conclusion

The conformational diversity of enzyme:substrate complexes was shown in the past to impact the barrier of the reaction steps. In the case of HIV-1 protease it further influences the mechanistic pathway. Using distinct initial conformations of this complex in QM/MM MD simulations, we observed free energy barriers that spanned a difference of *ca.* 5.0 kcal·mol⁻¹. These starting structures had previously observed an energy difference of 14.9 kcal·mol⁻¹, with adiabatic QM/MM methods. Nevertheless, just 10 ps timescale sampling per umbrella window provided a significant decrease on the previously observed energy differences obtained with adiabatic mapping studies on the same initial conformations. Furthermore, the obtained results were qualitatively similar between the two approaches, QM/MM MD *vs.* adiabatic mapping using a multi-conformation strategy, which support their use in the study of enzyme reaction mechanisms.

Regarding the different mechanistic proposals that were observed for HIV-1 protease – the one-aspartate and two-aspartate mechanism – a QM/MM MD protocol confirmed the existence of a rare mechanistic divergence for the studied reaction. We showed that a nearby water molecule and Thr26B, influenced the mechanistic route for this step. Low distances to these two residues, along the reaction, were shown to favour a one-aspartate mechanism, whether larger distances were shown to favour the two-aspartate mechanism instead. This agrees with what has been proposed previously for this enzyme using multi-PES QM/MM scans.

Computational Methods

System preparation. The starting structures for the QM/MM MD simulations were selected from the work by Calixto *et al.* from 2019,[9] originally obtained from an equilibrated *NPT* ensemble generated through conventional MD simulations. The modelling of the enzyme-substrate complex was obtained using the 4HVP PDB structure.[24] In the conventional MD simulation, the distance between the catalytic hydrogen atom of Asp25B and the carbonyl oxygen atom of the substrate was also constrained, to ensure a proper sampling of the less abundant catalytic conformation.[12] The enzyme residues were described with the Amber ff99SB force field and the system was solvated with TIP3P water molecules. The full protocol details can be found in the original reference.[9]

We have selected four structures among the 19 that have been previously characterized through QM/MM adiabatic mapping methods. More specifically, we have chosen the four structures for which Calixto *et al.* obtained the lowest and highest activation barrier for the first step of the reaction in each of the two mechanistic pathways observed, also called the one-aspartate and two-aspartate mechanisms (a and b in Scheme 1).

QM/MM MD simulations and umbrella sampling. The QM/MM calculations were performed with the CP2K 7.1 software,[25] using the QM module QuickStep [26] and MM module FIST. The QM region was described at the DFT level with the PBE functional. The PBE functional

has been shown to provide a good description of the stability of the active site of HIV-1 protease and acceptable energies for the enzyme:intermediate complex.[27] In addition, it is not as CPU-demanding as hybrid and hybrid-meta density functionals, due to the lack of Hartree-Fock exchange, which facilitates the QM/MM MD simulations, themselves extremely CPU-demanding. We employed the dual basis set of Gaussian plane-waves (GPW) formalism, and a double-ζ valence polarized (DZVP) basis set. The plane wave was expanded up to a density cut-off of 300 Ry and used in conjunction with the GTH pseudopotential of Goedecker *et al.*[28] to describe the core electrons. The rest of the system was treated classically, using the same MM parameters as the ones used for the conventional MD simulations and a non-bonded cut-off of 10 Å. The optimization of the two regions was performed separately based on the IMOMM method and within the electrostatic embedding scheme.

The QM region was essentially the same as in the work by Calixto *et al.*,[9] except for the substrate for which we have increased the atoms treated at the QM level, making a total of 96 atoms (not including link atoms). This was made to include an important interaction formed with a structural water molecule in the active site. We have also performed QM/MM MD simulations without including Wat₂ (Figure 1) in the QM layer, *i.e.*, this water molecule was only described at the MM level. We were particularly interested in evaluating if treating this water molecule at the MM level, would influence the one-aspartate vs. two-aspartate mechanistic pathway.

All QM/MM MD simulations were performed in the *NVT* ensemble using an integration time-step of 0.5 fs, assuming a constant volume of the equilibrated *NPT* classical ensemble within the simulated time. We have used the velocity rescaling thermostat to equilibrate the temperature of the system at 298 K.[29] The initial conformations were first energy minimized with CP2K, before an MD equilibration of 0.5 ps, and subsequently we sampled the reaction coordinate of interest. In this case, we have used a simple reaction coordinate to avoid biases in the reaction pathway, defined as the distance between the nucleophilic oxygen atom of the catalytic water molecule (O_{wat}), and the carbonyl's carbon atom of the scissile bond on the substrate (C_{sub}). This distance (d1 in Figure 1) was decreased from 3.0 Å to 1.4 Å at a constant rate of -0.0004 Å/fs during 4 ps and using a harmonic restraint of 250 kcal·mol⁻¹·Å⁻².

Then, for each of the starting conformations, a total of 35 umbrella sampling windows, spaced every 0.05 Å along the reaction coordinate, were used to sample the free energy profile for the reaction. We used a constant harmonic potential at each window of 250 kcal·mol⁻¹·Å⁻². Each window was simulated for 10 ps, making a total of 0.35 ns of accumulated simulated time per starting conformation. The free energy profiles have been obtained through the weighted histogram analysis method (WHAM)[30] combined with the umbrella sampling approach. Visual inspection of the system and analysis of important distances was performed with the VMD software (version 1.9.4).[31]

2D free energy profiles. The individual probabilities for each conformation from the umbrella sampling simulations, p_i, were retrieved from the general WHAM equation,

$$p_i = \frac{1}{\sum_j N_j e^{\left(F_j - U_{ij}^{bias}\right)/k_B T}}$$

where the constants F_j correspond to the reference energy of the jth window retrieved from the final converged free energy profiles after the WHAM analysis, N_j corresponds to the number of conformations in the jth window, and U_{ij}^{bias} corresponds to the bias potential energy of the conformation i in the jth window. The 2D free energy plots were generated after grouping the data from selected distances along $\sqrt[3]{\sum_j N_j}$ bins per axis, using a weight of p_i per datapoint. Free energies were then computed from the total probabilities, p, of each bin, as

$$\Delta F = -k_B T ln(p)$$

Conflicts of Interest

The authors declare no conflict of interest.

Keywords: Enzyme Catalysis • Aspartic Proteases • Mechanistic Divergence • Quantum Mechanics/Molecular Mechanics • Molecular Dynamics

[1] Geneva, Switzerland: World Health Organization, "HIV/AIDS", can be found under https://www.who.int/news-room/fact-sheets/detail/hiv-aids, **2019** (accessed July 2021).

[2] HIV Drug Resistance Report 2019, *Geneva, Switzerland: World Health Organization*, **2019** (WHO/CDS/HIV/19.21)

[3] D. Triki, M. Kermarrec, B. Visseaux, D. Descamps, D. Flatters, A. C. Camproux, L. Regad, *J. Biomol. Struct. Dyn.* **2020**, *38*, 5014-5026.

[4] M. Henes, G. J. Lockbaum, K. Kosovrasti, F. Leidner, G. S. Nachum, E. A. Nalivaika, S. K. Lee, E. Spielvogel, S. T. Zhou, R. Swanstrom, D. N. A. Bolon, N. K. Yilmaz, C. A. Schiffer, *Acs Chem. Biol.* **2019**, *14*, 2441-2452.

[5] M. Kumar, K. Mandal, M. P. Blakeley, T. Wymore, S. B. H. Kent, J. M. Louis, A. Das, A. Kovalevsky, *Acs Omega* **2020**, *5*, 11605-11617.

[6] a) L. J. Hyland, T. A. Tomaszek, G. D. Roberts, S. A. Carr, V. W. Magaard, H. L. Bryan, S. A. Fakhoury, M. L. Moore, M. D. Minnich, J. S. Culp, R. L. Desjarlais, T. D. Meek, *Biochemistry-Us* **1991**, *30*, 8441-8453; b) L. J. Hyland, T. A. Tomaszek, T. D. Meek, *Biochemistry-Us* **1991**, *30*, 8454-8463; c) D. B. Northrop, *Accounts Chem. Res.* **2001**, *34*, 790-797; d) D. R. Kipp, J. S. Hirschi, A. Wakata, H. Goldstein, V. L. Schramm, *P. Natl. Acad. Sci. USA* **2012**, *109*, 6543-6548; e) V. Y. Torbeev, S. B. H. Kent, *Org. Biomol. Chem.* **2012**, *10*, 5887-5891; f) E. J. Rodriguez, T. S. Angeles, T. D. Meek, *Biochemistry-Us* **1993**, *32*, 12380-12385.

[7] a) A. Krzeminska, V. Moliner, K. Swiderek, *J. Am. Chem. Soc.* **2016**, *138*, 16283-16298; b) Z. K. Sanusi, M. M. Lawal, P. L. Gupta, T. Govender, S. Baijnath, T. Naicker, G. E. M. Maguire, B. Honarparvar, A. E. Roitberg, H. G. Kruger, *J. Biomol. Struct. Dyn.* **2020**, *40*, 1736-1747.

[8] V. Carnevale, S. Raugei, S. Piana, P. Carloni, *Comput. Phys. Commun.* **2008**, *179*, 120-123.

[9] A. R. Calixto, M. J. Ramos, P. A. Fernandes, *Chem. Sci.* **2019**, *10*, 7212-7221.

[10] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. F. Ke, F. Y. Liu, H. B. Li, L. N. Ding, K. Morokuma, *Chem. Rev.* **2015**, *115*, 5678-5796.

[11] V. Carnevale, S. Raugei, C. Micheletti, P. Carloni, *J. Am. Chem. Soc.* **2006**, *128*, 9766-9772.

[12] A. J. M. Ribeiro, D. Santos-Martins, N. Russo, M. J. Rarnos, P. A. Fernandes, *Acs Catal.* **2015**, *5*, 5617-5626.

[13] a) R. Lonsdale, J. N. Harvey, A. J. Mulholland, *Chem. Soc. Rev.* **2012**, *41*, 3025-3038; b) S. Ahmadi, L. B. Herrera, M. Chehelamirani, J. Hostas, S. Jalife, D. R. Salahub, *Int. J. Quantum Chem.* **2018**, *118*, e25558.

[14] S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Bras, N. M. F. S. A. Cerqueira, P. A. Fernandes, M. J. Ramos, *Wires Comput. Mol. Sci.* **2017**, *7*, e1281.

[15] A. R. Calixto, M. J. Ramos, P. A. Fernandes, *J. Chem. Theory Comput.* **2017**, *13*, 5486-5495.

[16] a) A. C. C. Barbosa, R. P. P. Neves, S. F. Sousa, M. J. Ramos, P. A. Fernandes, *Acs Catal.* **2018**, *8*, 9298-9311; b) R. P. P. Neves, P. A. Fernandes, M. J. Ramos, *P. Natl. Acad. Sci. USA* **2017**, *114*, E4724-E4733; c) D. S. Gesto, N. M. F. S. A. Cerqueira, P. A. Fernandes, M. J. Ramos, *J. Am. Chem. Soc.* **2013**, *135*, 7146-7158; d) M. Prejano, T. Marino, N. Russo, *Front. Chem.* **2018**, *6*, 249.

[17] U. Ryde, *J. Chem. Theory Comput.* **2017**, *13*, 5745-5752.

[18] G. J. Lockbaum, M. Henes, N. Talledge, L. N. Rusere, K. Kosovrasti, E. A. Nalivaika, M. Somasundaran, A. Ali, L. M. Mansky, N. K. Yilmaz, C. A. Schiffer, *Acs Chem. Biol.* **2021**, *16*, 529-538.

[19] a) Q. F. Xue, E. S. Yeung, *Nature* **1995**, *373*, 681-683; b) R. D. Smiley, G. G. Hammes, *Chem. Rev.* **2006**, *106*, 3080-3094.

[20] a) S. Piana, D. Bucher, P. Carloni, U. Rothlisberger, *J. Phys. Chem. B* **2004**, *108*, 11139-11149; b) N. Okimoto, T. Tsukui, M. Hata, T. Hoshino, M. Tsuda, *J. Am. Chem. Soc.* **1999**, *121*, 7349-7354.

[21] E. I. Chudyk, M. A. L. Limb, C. Jones, J. Spencer, M. W. van der Kamp, A. J. Mulholland, *Chem. Commun.* **2014**, *50*, 14736-14739.

[22] R. P. P. Neves, A. V. Cunha, P. A. Fernandes, M. J. Ramos, *Chemphyschem* **2022**, e202200159.

[23] S. Piana, P. Carloni, M. Parrinello, *J. Mol. Biol.* **2002**, *319*, 567-583.

[24] M. Miller, J. Schneider, B. K. Sathyanarayana, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, S. B. H. Kent, A. Wlodawer, *Science* **1989**, *246*, 1149-1152.

[25] T. D. Kuhne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schutt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borstnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Muller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Gloss, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, J. Hutter, *J. Chem. Phys.* **2020**, *152*, 194103.

[26] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter, *Comput. Phys. Commun.* **2005**, *167*, 103-128.

[27] J. Garrec, P. Sautet, P. Fleurat-Lessard, *J. Phys. Chem. B* **2011**, *115*, 8545-8558.

[28] S. Goedecker, M. Teter, J. Hutter, *Phys. Rev. B* **1996**, *54*, 1703-1710.

[29] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.

[30] A. Grossfield, version 2.0.9.1 ed., http://membrane.urmc.rochester.edu/wordpress/?page_id=126.

[31] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph. Model.* **1996**, *14*, 33-38.

Entry for the Table of Contents
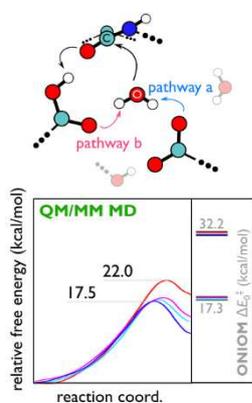
Using multiple conformations (previously characterized with adiabatic methods) and quantum mechanics/molecular mechanics molecular dynamics simulations, we assessed the convergence of activation free energies and the mechanistic divergence of the first step catalysed by HIV-1 protease. The mechanistic divergence for this step was still present and our barriers spanned a range of 5 kcal·mol⁻¹.