

# Origin of Selectivity in Protein Hydrolysis by Zr(IV)-Containing Metal Oxides as Artificial Proteases.

Albert Solé-Daura,<sup>†</sup> Antonio Rodríguez-Forteá,<sup>†</sup> Josep M. Poblet,<sup>†</sup> David Robinson,<sup>‡§</sup> Jonathan D. Hirst,<sup>‡</sup> Jorge J. Carbó<sup>†\*</sup>

<sup>†</sup> Department de Química Física i Inorgànica, Universitat Rovira i Virgili, Marcel·lí Domingo 1, 43007 Tarragona, Spain.

<sup>‡</sup> School of Chemistry, University of Nottingham, Nottingham, NG7 2RD, United Kingdom.

<sup>§</sup> Department of Chemistry and Forensics, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, United Kingdom.

**KEYWORDS:** Zirconium, peptide hydrolysis, selectivity, artificial proteases, polyoxometalate, molecular dynamics, DFT, mechanistic study.

**ABSTRACT:** The origin of selectivity in protein hydrolysis promoted by Zr(IV)-substituted polyoxometalates (POMs) has been investigated through a variety of computational techniques. Initially, we analyzed the reaction mechanism for the observed hydrolysis at the Asn44-Arg45 site in hen egg-white lysozyme protein (HEWL) by Zr-substituted Lindqvist anion  $[W_5O_{18}Zr(H_2O)(OH)]^{3-}$  (**ZrL**), using cluster models obtained from molecular dynamics (MD) simulations, quantum mechanics / molecular mechanics (QM/MM) calculations and metadynamics simulations. The mechanism characterization shows that the overall activity is governed by the energy cost to reach the transition state for C—N bond cleavage from reactants, resulting in a calculated, overall free-energy barrier of 121 kJ mol<sup>-1</sup>, which is in excellent agreement with the values derived from experimental rate constants (113–134 kJ mol<sup>-1</sup>). In addition, QM/MM metadynamics simulations on the early stages of the mechanism revealed the formation of an exergonic, non-covalent POM···protein complex at the protein surface stabilized by positively charged amino acids that are maintained during Zr coordination to amide oxygen. For non-reactive, related sites containing Arg (Asn13—Arg14, Arg45—Asn46, and Arg21—Gly22) we found very similar overall barriers within the cluster model approach (124, 124 and 120 kJ mol<sup>-1</sup>, respectively); however their non-bonding POM···protein interactions along the simulated coordination of Zr to the amide oxygen are significantly weaker than those for the reactive Asn44-Arg45 site. Thus, for HEWL protein the selectivity is governed by an *enzyme-like recognition* of the **ZrL** at the cleavage site that results in an overall acceleration of the reaction rate compared to other sites. Conversely, for human serum albumin (HSA) the observed selectivity is not directed by non-bonding POM···protein interactions, but it is controlled by protein secondary structure. Calculations on several Arg-Leu sites placed in positive patches showed that peptide bonds in  $\alpha$ -helix structure have higher overall free-energy barriers, while for the active Arg14-Leu15 site in a random coil region, the C—N cleavage is facilitated by the extended conformation of the protein chain. All in all, this study has identified and evaluated two complementary factors controlling the selectivity in peptide hydrolysis promoted by transition metal-substituted POMs; hydrolysis is disfavored at  $\alpha$ -helical regions of protein, and then, specific positively charged patches can trap the POM via electrostatic-type POM···protein interactions and accelerate the reaction.

## 1. INTRODUCTION

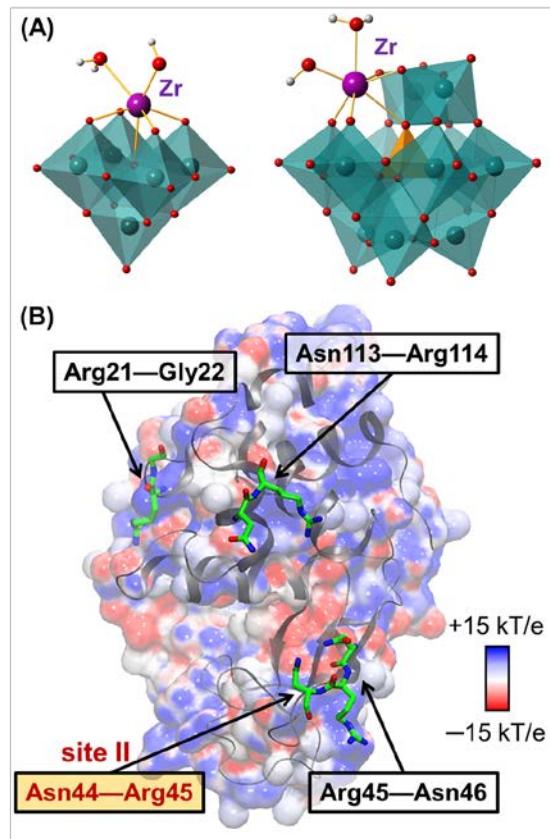
The selective hydrolysis of peptide bonds is a process of particular interest, finding application in the fields of biochemistry and biomedicine, including proteomics, protein engineering or the digestion of pathogenic proteins.<sup>1</sup> Nevertheless, peptide bonds in proteins are highly kinetically inert, with an estimated half-life of 600 years at 25 °C and neutral pH,<sup>2</sup> which makes mandatory the use of catalysts to achieve their hydrolysis. In nature, there is a family of hydrolytic enzymes called hydrolases or proteases that accelerate this reaction effectively. However, they are usually expensive, provide low selectivity and can

only operate in a narrow range of pH and temperature.<sup>1</sup> To overcome these issues, many research groups have devoted themselves to the development of artificial metalloproteases.<sup>3</sup> Among them, the group of Parac-Vogt probed the hydrolytic activity of Zr(IV)-, Ce(IV) and Hf(IV)-substituted polyoxometalates (POMs) towards dipeptides and oligopeptides,<sup>4,5</sup> and more importantly, towards peptide bonds in proteins in a highly selective manner.<sup>6–12</sup>

Zr-substituted POMs (see some structures in Figure 1A) were found to hydrolyze horse heart myoglobin (HHM)<sup>9</sup> and hemoglobin (Hb)<sup>12</sup> only at peptide bonds of Asp—X

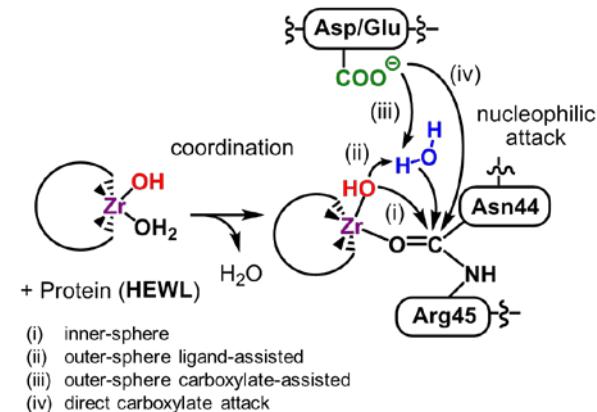
topology, where X is any other amino acid. In contrast, other proteins such as human serum albumin (HSA)<sup>7</sup> or hen egg-white lysozyme (HEWL)<sup>6</sup> were found to be selectively hydrolyzed at chemically different sites. Specifically, HEWL is hydrolyzed in the presence of Zr-POMs at only two peptide bonds located between Trp28 and Val29 and between Asn44 and Arg45, labeled site I and site II, respectively.<sup>6</sup> For this protein, previous molecular dynamics (MD) simulations<sup>14</sup> and crystallographic studies<sup>15–17</sup> identified two positively charged regions on its surface located at the vicinity of the cleavage sites where POMs can establish persistent interactions. Later, a combined docking and MD study described similar findings for HSA,<sup>18</sup> for which four binding sites were identified nearby the four chemically distinct cleavage sites. Therefore, it was hypothesized that these positively charged patches could be related to the observed selectivity. However, more recent MD studies with HEWL showed that other regions of the protein exhibit persistent interactions with POMs, besides those at the vicinity of the cleavage sites.<sup>19</sup> Thus, although these simulations provide insight into the nature of the non-covalent POM–protein interactions, they cannot unequivocally explain the origin of selectivity in peptide hydrolysis. For HSA the same selectivity was observed for different POM structures such as Lindqvist, Keggin and Wells-Dawson-type anions,<sup>7</sup> indicating that selectivity does not depend on the nature of the POM. Moreover, previous simulations revealed that although the features of the POM can influence their binding affinity to proteins, the positively charged regions of the protein that participate in these interactions are the same regardless the POM structure.<sup>14,19</sup>

Scheme 1 compiles different reaction mechanisms analyzed computationally for the peptide hydrolysis by Zr-substituted POMs,<sup>5,12,20</sup> and by other transition metal-based systems.<sup>21–29</sup> Once the Zr<sup>IV</sup> ion coordinates to the amide oxygen of the peptide bond, the mechanisms can be classified as follows: the inner-sphere attack of the hydroxo ligand of Zr (i); the outer-sphere nucleophilic attack of an external water molecule assisted by the ZrOH moiety (ii) or by a carboxylate group of a neighboring residue (iii); and the direct attack of a carboxylate group nearby (iv). Earlier studies by Pierloot and co-workers on the hydrolysis of dipeptides found that the most likely pathway resembles mechanism (iii) in Scheme 1, but with a nucleophilic attack assisted by the C-terminus carboxylate group acting as a Brønsted base.<sup>5</sup> Prabhakar and co-workers analyzed mechanisms (i) and (ii) for the hydrolysis of four different peptide bonds in HSA by ZrK using static quantum mechanics / molecular mechanics (QM/MM) calculations in the full molecular system.<sup>20</sup> The authors concluded that the inner-sphere nucleophilic attack of the hydroxo ligand of Zr (i) is somewhat preferred over the outer-sphere mechanism (ii). Although this study shed some light on the reaction mechanism, the reasons for the experimentally observed selectivity remained unexplored.



**Figure 1.** (A) 3D-representation of hydrolytically active Zr-containing metal oxide clusters:  $[Zr(H_2O)(OH)W_5O_8]^{3-}$  (**ZrL**; left) and  $[Zr(H_2O)(OH)PW_{11}O_{39}]^{3-}$  (**ZrK**; right) with Lindqvist and Keggin structures, respectively. (B) Molecular electrostatic potential (MEP) of hen egg-white lysozyme (HEWL;  $q = 8+$ ) showing in green sticks the amino acids belonging to the four analyzed peptide bonds. The experimentally hydrolyzed site II (Asn44–Arg45) is highlighted with a red label. Blue coloring represents positively charged regions on HEWL surface, while red represents the negatively charged ones. The MEP was obtained by means of the PDB2PQR web-based software.<sup>13</sup>

### Scheme 1. Proposed mechanisms for peptide bond hydrolysis by Zr-substituted POMs.



Very recently, a static DFT study using model tripeptides discussed the origin of selectivity for specific Asp–X bonds in Hb.<sup>12</sup> The proposed mechanism involves the direct nucleophilic attack of the carboxylate side chain of

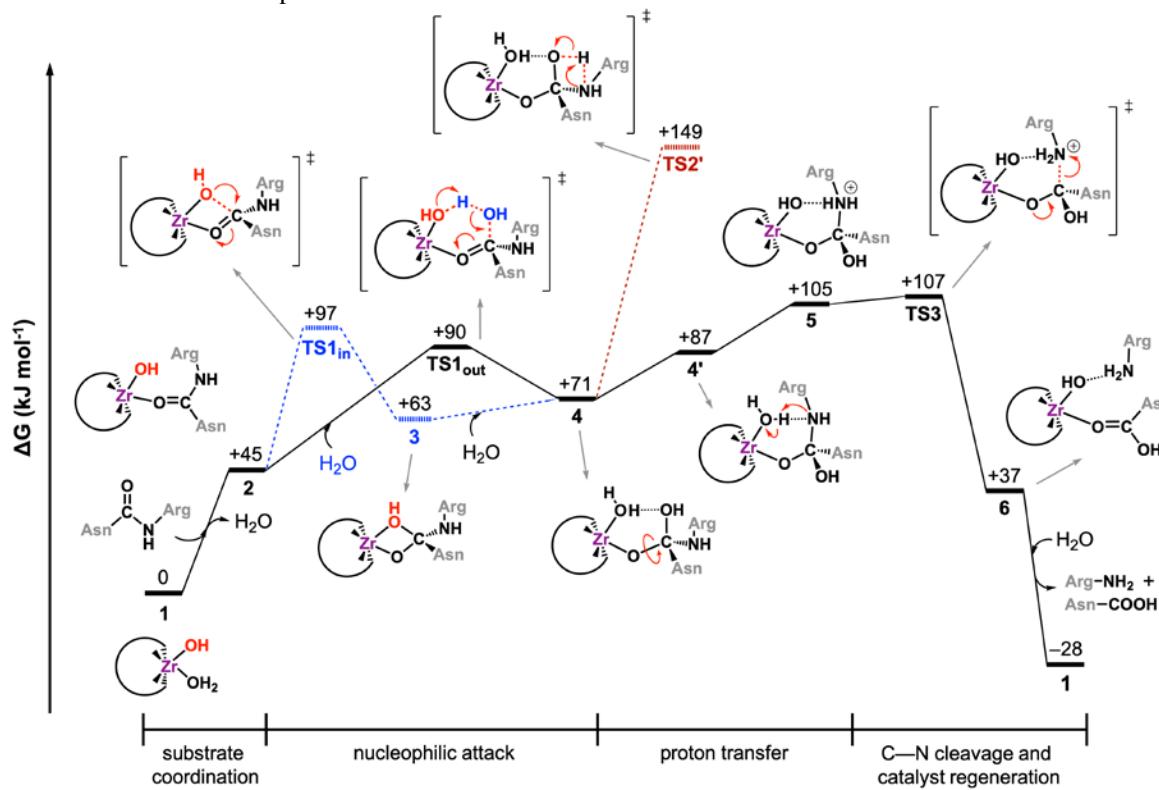
Asp residue on the amide carbon with formation of a cyclic anhydride (mechanism *iv*), which shows a lower activation energy than the structurally related Glu-X bond.<sup>12</sup> Although these results nicely explain the preference for hydrolyzing Asp-X sites, experimentally not all the Asp-X bonds in Hb are hydrolyzed, and therefore, the authors recognize that positive patches on protein surface might also influence the selectivity.<sup>12</sup> Moreover, as noted above for HSA and HEWL the selective hydrolysis was observed at non-carboxylate containing sites, even though Asp-X bonds are also present. Thus, we suspect that dynamic protein arrangements around the negatively charged polyoxometalate cluster occurring along the reaction mechanism are crucial to explain the observed selectivity. Here, we select the model HEWL protein and the Zr-substituted Lindqvist anion [ $\text{Zr}(\text{H}_2\text{O})(\text{OH})\text{W}_5\text{O}_{18}]^{3-}$  (**ZrL**, in Figure 1A) as catalyst, and compare the observed hydrolysis of the Asn44—Arg45 peptide bond with those of the structurally analogous Asn113—Arg114, Arg45—Asn46, and Arg21—Gly22 (Figure 1B). We employ a combination of detailed DFT calculations in cluster models with MD and hybrid QM/MM metadynamics simulations to evaluate the effect of protein structure and

POM…protein non-covalent interactions on the reaction mechanism and the selectivity. Also, we have analyzed the origin of selectivity in another protein, namely Human Serum Albumin (HSA), which is also cleaved at chemically different sites in the presence of Zr-substituted POMs.<sup>7</sup>

## 2. RESULTS AND DISCUSSION

### 2.1. Characterization of the reaction mechanism.

Initially, we studied the reaction mechanism for the hydrolysis of the Asn44—Arg45 peptide bond combining static DFT calculations within the *cluster model* approach,<sup>30</sup> MD simulations, QM/MM calculations on the full protein model, and metadynamics simulations. Then we compared the results for this site with other sites that showed resistance to hydrolysis experimentally (Asn113—Arg114, Arg45—Asn46, and Arg21—Gly22). Figure 2 shows the free-energy profile computed for the hydrolysis of Asn44—Arg45 peptide bond by **ZrL** via mechanisms (*i*) and (*ii*) (Scheme 1) and Figure 3 represents the main transition state structures.



**Figure 2.** Free-energy profile ( $\text{kJ mol}^{-1}$ ) for the hydrolysis of Asn44—Arg45 peptide bond by **ZrL** using a cluster model obtained from an MD trajectory. Black solid lines and blue dashed lines represent two energetically-accessible mechanisms that converge in species 4, whereas red dashed lines represent an alternative mechanism for the proton transfer proposed in ref. 20 that is higher in energy.

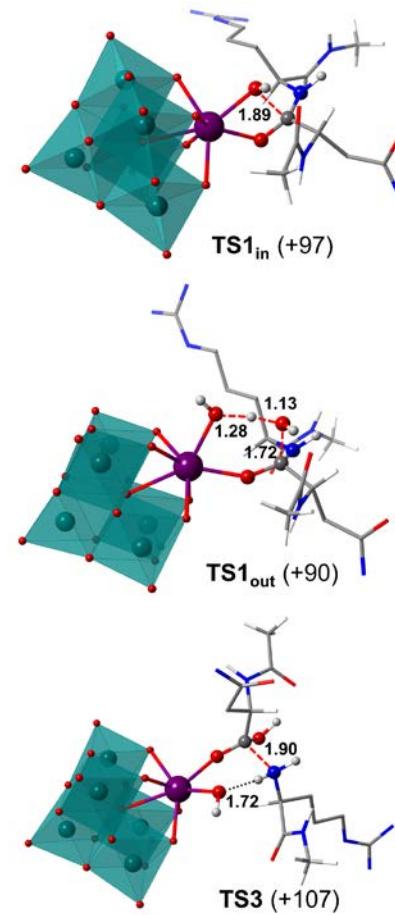
The reaction mechanisms can be divided into four main steps: the coordination of **ZrL** to the amide oxygen of the peptide bond, the nucleophilic attack of the Zr-hydroxo group (inner-sphere, *i*) or an external water (assisted outer-sphere, *ii*) on the amide carbon, the proton-transfer to the amide nitrogen, and the final C—N bond cleavage

to yield the products and regenerate the catalyst. The first coordination step was computed to be endergonic by 45  $\text{kJ mol}^{-1}$ . Note that the geometry for the cluster model was generated from restricted MD simulations fixing the Zr-oxygen bond (*vide infra*). Also, it has to be pointed out that the profile of Figure 2 has been constructed using a

single MD snapshot and therefore, it is meant to provide a qualitative description of the mechanism that will be complemented in the next section with conformational analyses. The coordinated Zr<sup>IV</sup> ion acts as a Lewis acid activating the amide carbon for the subsequent nucleophilic attack which can occur through different pathways (Scheme 1). Our calculations indicate that the outer-sphere nucleophilic attack (**TS1<sub>out</sub>**) is slightly more favorable than the inner-sphere one (**TS1<sub>in</sub>**) [ $\Delta G^\ddagger$  ( $2 \rightarrow \text{TS}$ ) of 45 vs 52 kJ mol<sup>-1</sup>]. This is ascribed to the less strained six-membered ring in **TS1<sub>out</sub>**. Nevertheless, the small free-energy difference suggests that both mechanisms could be operative. This result differs slightly from that found for the hydrolysis of HSA by [PW<sub>n</sub>O<sub>39</sub>Zr(OH)]<sup>4-</sup> where the inner-sphere mechanism was energetically preferred.<sup>20</sup> In fact, our calculations initially pointed towards inner-sphere mechanism, but introducing the entropy corrections to the ideal gas thermochemistry analysis of standard DFT codes resulted in a slight preference of the associative process of the outer-sphere mechanism (*ii*) (see Computational Details). The inner-sphere mechanism leads to the four-membered ring intermediate **3** that lies 63 kJ mol<sup>-1</sup> above the reactants, while the outer-sphere one results in intermediate **4**, which is 8 kJ mol<sup>-1</sup> less stable. The interconversion between **3** and **4** requires the incorporation or release of an aqua ligand in the coordination sphere of Zr, which is expected to be fast due to the lability of water coordination.<sup>31</sup>

From **4**, we propose a step-wise process that involves an intramolecular rearrangement via bond rotation to yield **4'**, a proton transfer from the aqua ligand to the amide nitrogen to give **5**, and final C—N bond cleavage (see Figure 2). This process goes uphill in energy until it reaches the transition state for the C—N bond cleavage, in line with that proposed for hydrolysis of dipeptides by Zr-containing POMs.<sup>12</sup> The proton transfer (see Figure S1 for further details) results in the thermodynamically and kinetically unstable intermediate **5** from which the C—N cleavage proceeds through a very small free-energy barrier of 2 kJ mol<sup>-1</sup>. Thus, the existence of intermediate **5** might depend on the specific conditions transforming the step-wise hydrogen transfer then C-N bond cleavage in a concerted process. Also, we cannot discard that the interconversion between **3** and **4'** could occur through a step-wise mechanism involving ligand rotation followed by water coordination (see Figures S2 and S3). Finally, transition state **TS3** yields intermediate **6** with the release of the amine product and the Zr-coordinated carboxylic acid, which then is exchanged by a water molecule recovering the catalyst and providing the thermodynamic driving force of the reaction (-28 kJ mol<sup>-1</sup>). Additionally, we analyzed direct hydrogen transfer from the carboxylic acid group in **4** which is concomitant with the C—N bond cleavage (red dashed lines in Figure 2), as proposed computationally for the hydrolysis of HSA protein by Zr-substituted POM.<sup>20</sup> However, the corresponding four-membered ring transition state (**TS2'** in Figure 2; geometry displayed in Figure S4) is significantly higher in energy (149 kJ mol<sup>-1</sup> vs. 107 kJ mol<sup>-1</sup> for **TS3**).

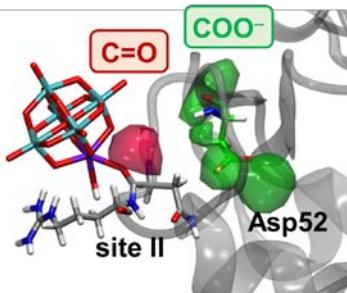
The overall activity of the hydrolysis reaction is governed by the energy cost to reach the transition state for C—N bond cleavage from the reactants (**1** + protein → **TS3**) since the reaction proceeds uphill in energy along the process (see Figure 2). Thus, it is not crucial for the understanding of the activity and selectivity to determine the preference for the inner-sphere pathway (*i*) or the outer-sphere (*ii*), both converging in intermediate **4** before reaching **TS3**. Nevertheless, we have analyzed in more detail the nucleophilic attack of both mechanisms on the solvated, full protein model using a QM/MM approach in conjunction with the replica-path method.<sup>32</sup> The potential energy curves (see Figure S5) suggest that both mechanisms (*i*) and (*ii*) are possible within the protein environment and neither of them is particularly favored over the other, since the large radius of the Zr atom makes the strained four-membered ring transition state to be close in energy to the six-membered ring TS of outer-sphere.



**Figure 3.** B3LYP-optimized geometries of key transition states in the reaction profile of Figure 2. Selected distances are shown in Å, and relative free energies are shown in kJ mol<sup>-1</sup>. Hydrogen atoms in the side chains are omitted for clarity.

We have also evaluated the feasibility of the carboxylate-assisted mechanisms (*iii*) and (*iv*) by analyzing the distribution of aspartate and glutamate residues around site II along the MD simulation on the coordinated

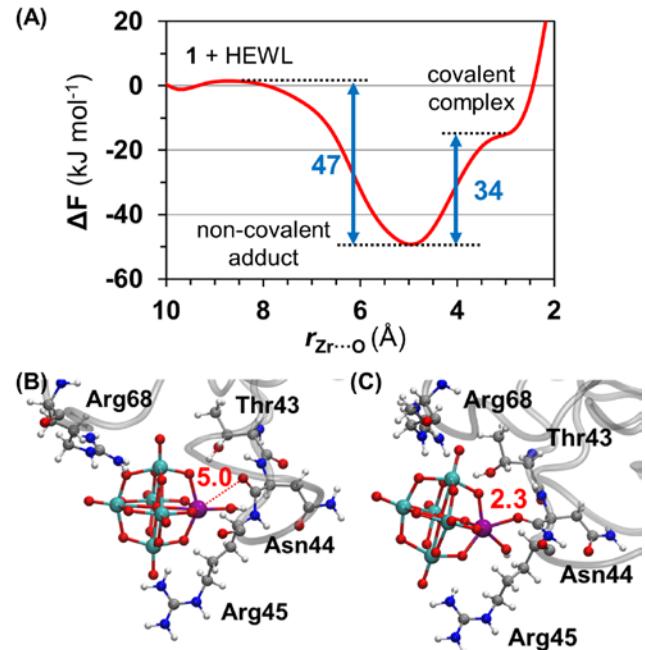
POM—protein complex. Distance analysis identifies Asp52 as the closest carboxylate-containing residue to the Asn44—Arg45 peptide bond (see Figure S6). Thus, we analyzed the density of water molecules surrounding the  $\text{COO}^-$  moiety (green density) and the carbonyl group of the peptide bond (red density) as shown in Figure 4. The gap between both densities suggests that the probability of finding one water molecule close enough from both groups to react via mechanism *iii* (Scheme 1) is low. Even so, it is possible to find snapshots of the dynamics in which a water molecule is simultaneously hydrogen-bonded to the carboxylate of Asp52 and the amide oxygen of Asn44. From one of these snapshots we generated a new cluster model with three residues (Asn44, Arg45 and Asp52, see Figure S7) in order to evaluate the energy cost of the nucleophilic attack of an external water molecule assisted by the carboxylate group. The free-energy barrier associated to this process ( $105 \text{ kJ mol}^{-1}$ ) is higher than those for the nucleophilic attacks to amide carbon associated to mechanisms *(i)* and *(ii)* ( $90$  and  $97 \text{ kJ mol}^{-1}$ ). Overall, the carboxylate-assisted mechanism *(iii)* is unlikely because of the higher energy barrier and the low occurrence (0.4 %) of conformations with short enough distances between the amide carbonyl and the carboxylate. On the other hand, the MD trajectory does not provide any geometry suitable for modeling a TS geometry for the direct attack of the carboxylate group, and therefore, we concluded that mechanism *(iv)* is also not likely to occur.



**Figure 4.** Representation of the volumetric density (isovalue 0.3) of water molecules within 3.5 Å from the carbonyl group of site II (red isosurface) and from the carboxylate group of Asp52 (green isosurface). Densities were computed taking as a reference both the C and O atoms of the protein groups and all the solvent atoms; and were averaged over 100 ns of MD simulation from which data were sampled every 4 ps. See Figures S8 and S9 for further details on these interactions.

Finally, we analyze in more detail the initial coordination of the POM to the protein. We suspect that this step could be very sensitive to the rearrangements of the protein chain and the environment of the cleavage site to accommodate the POM moiety;<sup>12</sup> and consequently it could influence the selectivity. Since the cluster model does not reflect these interactions, we performed QM/MM simulations on the real, solvated system. To observe such a coordination of the POM to the protein in an affordable simulation time, we accelerated the dynamics by means of the metadynamics technique, defining the Zr...O distance as the collective variable. Figure 5 displays

the calculated free-energy landscape along the coordination path, from the POM in solution to the formation of the covalent bond between the Zr and the amide oxygen. First, we observed the formation of a minimum corresponding to a non-covalent POM-protein complex at the protein surface that is exergonic by almost  $50 \text{ kJ mol}^{-1}$ . At an atomistic level, the POM...protein interactions resemble those previously characterized with MD simulations,<sup>14,19</sup> in which the oxygens of the POM framework form hydrogen bonds with positively charged and polar amino acids, namely, Arg45, Arg68 and Thr43. Noticeably, Parac-Vogt et al. postulated that interaction energy should be close to  $42 \text{ kJ mol}^{-1}$  on the basis of DFT calculations and experimental rate constants,<sup>12</sup> although so far, there was no direct proof for that. From this minimum, the coordination of the Zr center to the amide oxygen is estimated to be endergonic by ca.  $34 \text{ kJ mol}^{-1}$ .



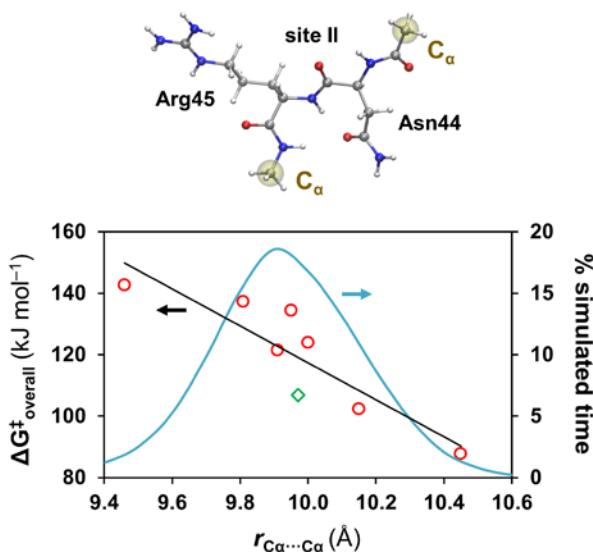
**Figure 5.** (A) Free-energy landscape along the coordination process of  $\text{ZrL}$  to site II (Asn44—Arg45) of HEWL; obtained from QM/MM metadynamics on the fully solvated real system. The reaction coordinate was described using the distance between the Zr center and the O atom of the peptide bond as a single collective variable. (B and C) Representative snapshots of the non-covalent adduct and covalent complex, respectively. Zr...O distances (in Å) displayed in red.

Note that the shallow minima such as that for the covalent complex cannot be clearly appreciated in the reconstructed free-energy surface due to the large width of the added Gaussian functions (1 Å) that we used to sample in the most efficient and reliable manner a wide range of Zr...O distance. During the simulation, we did not observe the coordination of any QM water molecule to the Zr. However, if we add the free-energy change of a water decoordination from **1** (+4  $\text{kJ mol}^{-1}$  at DFT level), the metadynamics-derived coordination energy becomes  $38 \text{ kJ mol}^{-1}$ , very close to that derived from the cluster model ( $45 \text{ kJ mol}^{-1}$ ). This indicates that the strength of non-

bonding POM–protein interactions does not vary significantly along the reaction coordinate. Thus, we can assume that the overall free-energy barrier is properly described by the energy difference between the reactants and  $\text{TS}_3$  in the cluster model. Finally, we must point out that the covalent complex lies below the free-energy level of the POM in solution.

## 2.2. Influence of protein conformation on reactivity.

Considering that the highest-energy TS corresponds to a bond cleavage in the protein main chain, we analyze whether the conformational variability of the protein may influence the reaction kinetics, as observed in other biological systems.<sup>33–37</sup> To assess this point, we analyzed the free-energy barriers associated to the rate-determining process ( $\Delta G^\ddagger_{\text{overall}}$ ) of a set snapshots taken from different stages of the MD simulation (Table S1). Indeed, the conformation of the protein was found to play an important role in its reactivity since the array of computed barriers is rather broad, comprising values from 88 to 143 kJ mol<sup>-1</sup>. Thus, the reaction kinetics should not only be studied on a single protein conformation but on a thermodynamic ensemble.



**Figure 6.** Overall free-energy barrier for the hydrolysis of Asn44–Arg45 as a function of the distance between alpha carbons,  $r_{\text{C}\alpha\cdots\text{C}\alpha}$  (highlighted in yellow in the balls-and-sticks representation) using different snapshots of the MD simulation. The green diamond corresponds to the snapshot used for computing the profile of Figure 2. Values were fitted to a linear regression model ( $r^2 = 0.78$ ). The abundance of each  $r_{\text{C}\alpha\cdots\text{C}\alpha}$  value is expressed as the % of simulated time (blue line).

Interestingly, we found a correlation between the free energy barriers and the distance between capping alpha carbons ( $r_{\text{C}\alpha\cdots\text{C}\alpha}$ ) that is represented in Figure 6. We observed that the longer the  $\text{C}_\alpha\cdots\text{C}_\alpha$  distance, the lower the barrier and thus, the faster hydrolysis. We attributed this trend to the fact that longer distances for a given protein chain secondary structure may increase the strain incurred to the peptide bond to facilitate the cleavage of the

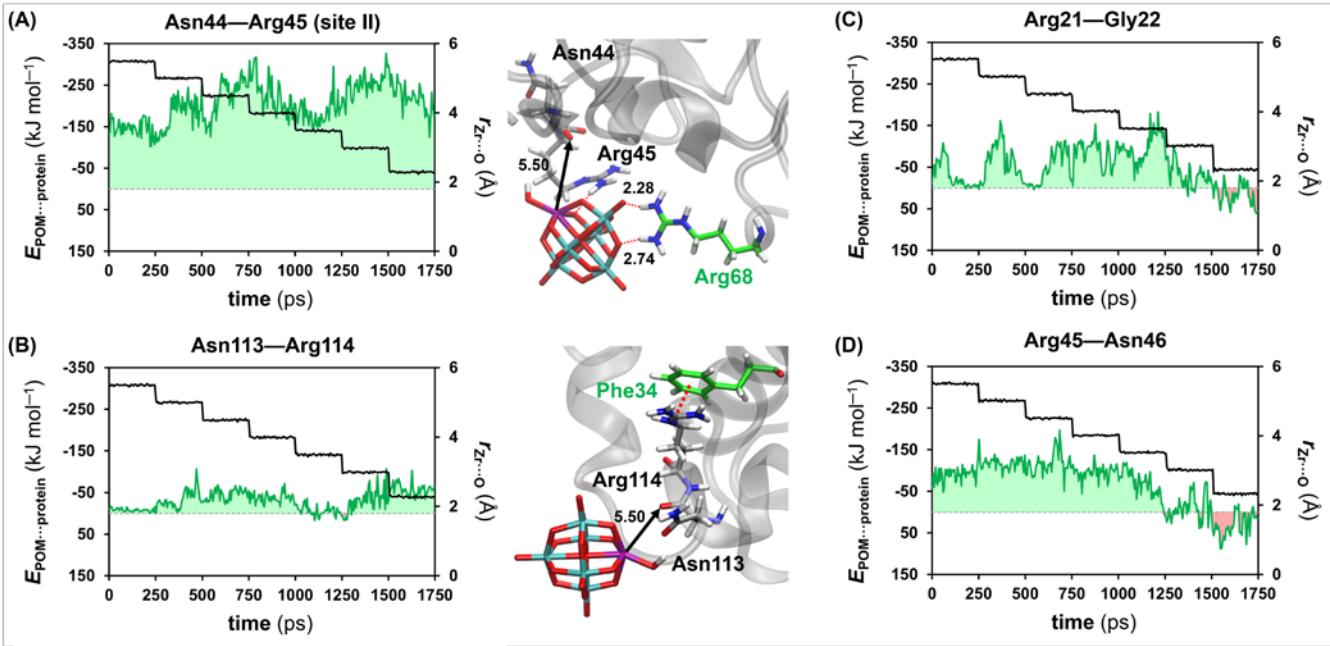
C–N bond. Figure 6 also displays the distribution of  $\text{C}_\alpha\cdots\text{C}_\alpha$  distances computed over 100 ns of MD simulation with the POM coordinated to the cleavage site. The abundance of each  $\text{C}_\alpha\cdots\text{C}_\alpha$  distance value is expressed as the percentage of simulated time that is reported. The  $\text{C}_\alpha\cdots\text{C}_\alpha$  distances show an almost symmetric unimodal distribution centered at 9.90 Å, which is the most likely distance accounting for the 18.6 % of the simulated time. As the dependence of  $\Delta G^\ddagger_{\text{overall}}$  on  $r_{\text{C}\alpha\cdots\text{C}\alpha}$  can be fitted to a linear regression, we could determine the free-energy barrier for each conformation ( $\text{C}_\alpha\cdots\text{C}_\alpha$  distance) using the linear equation (Table S2). Then, the average free-energy barrier can be estimated from this set of values weighted with the abundance of each conformation. This approach leads to an average free-energy barrier of 121 kJ mol<sup>-1</sup> for the hydrolysis of Asn44–Arg45. Importantly, this value lies within the range of values derived from experimental rate constants (113–134 kJ mol<sup>-1</sup>; Table S3),<sup>4,7,9,10</sup> and it is equal to that computed at the most likely  $\text{C}_\alpha\cdots\text{C}_\alpha$  distance (see snapshot 4 in Table S1).

## 2.3. Origin of selectivity in peptide hydrolysis.

Initially, we analyzed how the energy barrier for the peptide bond cleavage varies at different protein sites and whether this effect governs the selectivity. Table 1 compares the computed, overall free-energy barriers ( $\mathbf{1} + \text{protein} \rightarrow \text{TS}_3$ ) at hydrolyzed site II (Asn44–Arg45) with those at structurally-related, non-hydrolyzed sites (Asn113–Arg114, Arg21–Gly22, and Arg45–Asn46). For each site, we selected a cluster model derived from the most likely protein conformation, which is obtained from the analysis of the  $\text{C}_\alpha\cdots\text{C}_\alpha$  distance abundance along the simulation. In all cases, the  $\text{C}_\alpha\cdots\text{C}_\alpha$  distances describe unimodal distributions analogous to that depicted in Figure 6 that are rather symmetric around the maximum (see Figure S10), justifying the use of a single conformation as representative of the distribution. Note that the selected peptide bonds are embedded in different regions of the protein with varied motifs of the secondary structure ( $\beta$ -sheet,  $\alpha$ -helix, and random coil), and consequently, the  $\text{C}_\alpha\cdots\text{C}_\alpha$  distances cannot be compared. All four peptide bonds exhibited very similar  $\Delta G^\ddagger_{\text{overall}}$  values within a range of 4 kJ mol<sup>-1</sup>. This indicated that there is no intrinsic preference for hydrolyzing site II over other sites and therefore, we hypothesized that the origin of selectivity might lie in the enzyme-like recognition of  $\text{ZrL}$  anion at the protein region of the hydrolyzed peptide bond.

**Table 1. Comparison of Free Energy Barriers (kJ mol<sup>-1</sup>) for the Hydrolysis of Several Peptide Bonds in HEWL.**

cleavage site	motif	$\Delta G^\ddagger_{\text{overall}}$
Asn44–Arg45 (site II)	$\beta$ -sheet	121
Asn113–Arg114	$\alpha$ -helix	124
Arg21–Gly22	random coil	124
Arg45–Asn46	$\beta$ -sheet	120



**Figure 7.** Evolution of the POM···protein non-covalent interaction potential in  $\text{kJ mol}^{-1}$  (green lines) and the  $\text{Zr}_{\text{POM}}\cdots\text{O}_{\text{protein}}$  distance in Å (black lines) along simulated coordination processes to four different sites: hydrolyzed site II Asn44–Arg45 (A), and the non-hydrolyzed sites Asn113–Arg114 (B), Arg21–Gly22 (C) and Arg45–Asn46 (D). Representative, selected 1750 ps runs of restricted MD simulations. Typical snapshots of  $\text{ZrL}\cdots\text{protein}$  interaction for simulations A and B.

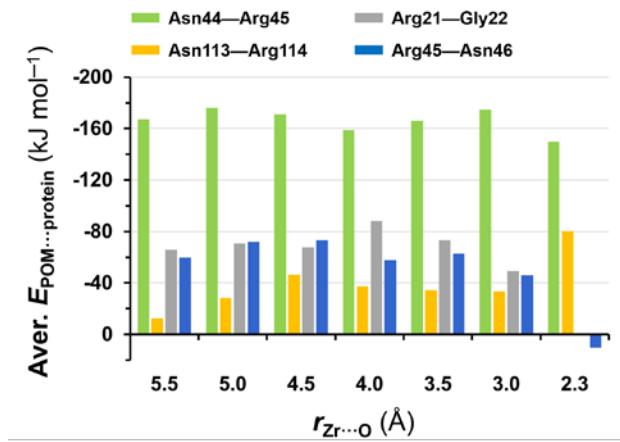
Next, we analyzed the effect of protein arrangements around the polyoxometalate on the selectivity by comparing the non-covalent POM···protein interactions at the four different sites highlighted in Figure 1B. These interactions should be present from the early stages of the reaction, that is the Zr ion coordination to the amide oxygen. Thus, we performed constrained MD simulations, in which the  $\text{Zr}_{\text{POM}}\cdots\text{O}_{\text{protein}}$  distance ( $r_{\text{Zr}\cdots\text{O}}$ ) was initially set to the non-coordination length of 5.50 Å and decreased successively every 250 ps of simulation until 2.33 Å. This latter value corresponds to the bond equilibrium distance predicted by DFT (see Computational Details for details). The procedure was repeated along five parallel runs for each site to obtain a more reliable sampling.

Figure 7 displays the time evolution of the POM···protein interaction ( $E_{\text{POM}\cdots\text{protein}}$ ) and the  $\text{Zr}\cdots\text{O}$  distance for a selected run of each site (see Figures S11–S14 for the complete set of runs); and Figure 8 collects the  $E_{\text{POM}\cdots\text{protein}}$  energies averaged over the five independent MD runs at different  $\text{Zr}_{\text{POM}}\cdots\text{O}_{\text{protein}}$  distances. The four selected sites contain positively charged arginine residues which can interact with the negatively charged POM framework. Consequently, Figures 7 and 8 show attractive POM···protein interactions along the simulated coordination process in all the cases. However, we observed remarkable differences between the sites which can explain the experimental selectivity. Most importantly, the interaction energies at the hydrolyzed site (Asn44–Arg45 in Figure 7A) are significantly stronger than for the other sites. Visual analysis of the trajectories revealed that during the coordination  $\text{ZrL}$  interacts simultaneously with Arg45 and with Arg68 located nearby, as illustrated by the

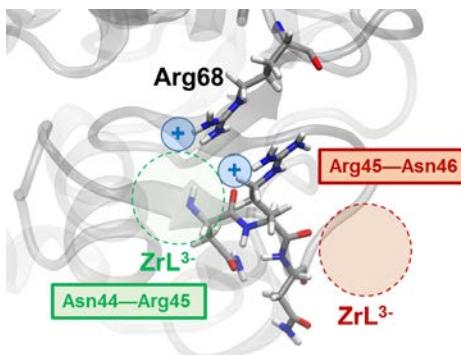
snapshot of Figure 7A, and in agreement with the QM/MM trajectory (see Figure 5). Figure S15 provides the averaged, non-covalent interaction energies of  $\text{ZrL}$  with individual amino acids, confirming the simultaneous interaction of Arg45 and Arg68. Additional hydrogen bonding interactions with polar residues Thr43, Thr51 and Tyr53 were also characterized, being analogous to those of previous MD simulation on noncovalent adducts.<sup>14,18,19</sup> Moreover, close contacts of the guanidine groups of Arg45 and Arg68 with Zr-substituted POMs had been actually observed in crystal X-ray structures of non-covalent complexes.<sup>15–17</sup>

Moving from simulations at site Asn44–Arg45 to the contiguous Arg45–Asn46 site, we only observed a significant stabilizing interaction between the  $\text{ZrL}$  anion and Arg45 that vanishes as the zirconium-oxygen distance shortens (Figure S18). As a result, the overall non-covalent interaction is significantly weaker for the Arg45–Asn46 site, becoming even slightly repulsive at coordination  $\text{Zr}\cdots\text{O}$  distance (Figures 7D and 8). The coordination to either of these consecutive sites places the POM at different regions of the protein causing a differentiated pattern of non-covalent interactions (see Figure 9). For the Asn113–Arg114 site, which is chemically equivalent to the hydrolyzed Asn44–Arg45 site, the stabilizing interactions are even weaker than in Arg45–Asn46 site (Figures 7B and 8). Besides the lack of efficient cooperative effects at non-reactive sites, POM···protein interactions are often hampered by neighboring amino acids that compete with the POM for interacting with the side chain of arginines. In this case, the positively charged guanidine group of Arg114 is trapped in a cation–π interaction with the phenyl ring of

Phe34 (snapshot in Figure 7B, and SI for details). Similarly, for the Arg21-Gly22 site, weak **ZrL**...protein interaction was observed during the coordination (Figures 7C and 8). Additional details and discussion on the forces governing the binding of the POM to the protein can be found in the SI (Figure S21 and derived discussion).



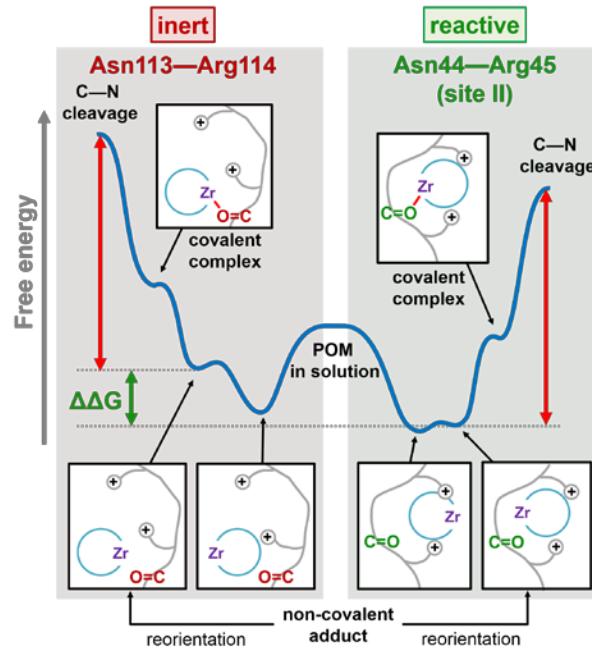
**Figure 8.** Average non-covalent interaction energies between the **ZrL** POM and the protein ( $E_{\text{POM...protein}}$ ) as function of the  $\text{Zr}_{\text{POM}}\cdots\text{O}_{\text{protein}}$  distance during the coordination of **ZrL** to Asn44—Arg45, Asn113—Arg114, Arg21—Gly22 and Arg45—Asn46 (green, yellow, gray and blue bars, respectively). Data were sampled every 2 ps and averaged over the 5 runs.



**Figure 9.** Representation of the preference of **ZrL** to coordinate Asn44—Arg45 rather than Arg45—Asn46 peptide bond.

To sum up, previous simulations have shown that POM anions can form persistent non-covalent interactions at different cationic patches of the HEWL protein surface,<sup>14,19</sup> however, only specific peptide bonds of these sites are hydrolyzed. We propose that the observed selectivity in HEWL protein is governed by non-covalent interactions occurring during the POM approach to the protein and the transition-metal coordination to the amide oxygen. In the selectively hydrolyzed site (Asn44-Arg45), the side chains of two arginines (Arg45 and Arg68) clamp the POM during the coordination process, and presumably, along the whole reaction profile. This effect compensates to some extent the endothermicity of Zr coordination and prevents the inverse process that releases the Zr-substituted POM to the solvent. Combining the insight

gained from DFT, metadynamics and MD simulations, we can propose schematic free-energy profiles for the reaction at the inert and the reactive peptide bonds (Figure 10). Note the resemblance between the energy profile in Figure 10 (right side) and that of Figure 5 obtained with QM/MM metadynamics simulations, despite in the latter the computational cost limitation did not allow us to identify shallow minima associated to non-covalent POM...protein interactions. Although the barriers required for carbon-nitrogen bond breaking from non-covalent adducts are similar in both cases, in the reactive peptide bond the POM catalyst forms strong interactions with specific amino acids of the protein, shifting the energy profile down and favoring the irreversibility of the specific POM...protein interaction. In contrast, coordination to non-reactive sites shows weaker stabilizing POM...protein interactions shifting the free-energy profile up and disfavoring POM coordination and overall hydrolysis reaction. Further supporting the specific recognition of the POM as the source of selectivity in HEWL, experimental studies with Hf-containing POMs<sup>16</sup> revealed hydrolysis patterns at pH 5 that differ from those observed at pH 7.4.<sup>6</sup> At pH 5, the histidine residues of the protein might be found in their protonated, cationic form according to the pK<sub>a</sub> value of 6.5, which is expected to alter the charge distribution at the protein surface and in turn, the recognition patterns of the protein towards anionic POM proteases.

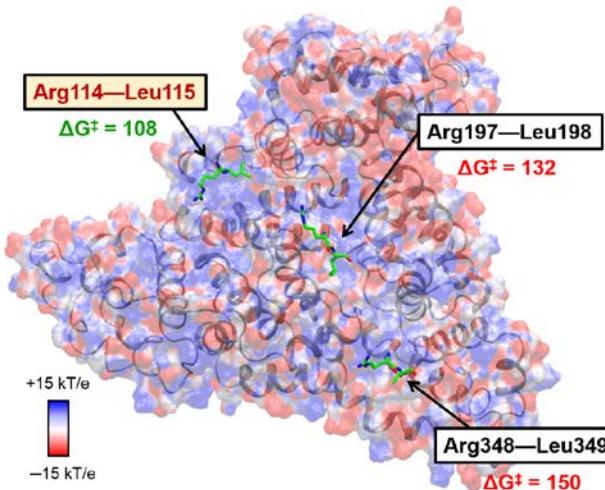


**Figure 10.** Pictorial representation of the origin of selectivity in peptide hydrolysis in HEWL by Zr-substituted POMs.

## 2.4. Origin of selectivity in human serum albumin.

To evaluate whether the factors governing the selectivity in HEWL protein could be also extended to explain the observed selectivity in other proteins, we next examined

the case of HSA, which is also selectively hydrolyzed at chemically distinct sites in the presence of Zr-containing POMs.<sup>7</sup> This protein is cleaved at Arg114-Leu115, while other peptide bonds of the same topology, also located on positively charged patches of the protein surface showed resistance against hydrolysis: Arg197-Leu198, Arg218-Leu219 and Arg348-Leu349 (see Figure 11). Thus, we performed a new set of constrained MD simulations to compare the magnitude of the POM...protein non-bonding interactions during the coordination of **ZrL** to these sites. For this study we excluded Arg218-Leu219 motif because the amide oxygen is not surface-accessible, being buried within the protein structure.



**Figure 11.** Molecular electrostatic potential (MEP) of human serum albumin (HSA;  $q = 14-$ ) showing in green sticks the amino acids belonging to the three analyzed peptide bonds. The experimentally hydrolyzed site (Arg114—Leu115) is highlighted with a red label and overall free-energy barriers ( $\Delta G^\ddagger$ ) for each peptide bond are shown in  $\text{kJ mol}^{-1}$ . Blue coloring represents positively charged regions on HEWL surface, while red represents the negatively charged ones. The MEP was obtained by means of the PDB2PQR web-based software.<sup>33</sup>

As shown in Figure S22, the non-bonding POM...protein interactions observed during the coordination of **ZrL** to the different sites are all highly attractive in agreement with the cationic nature of the protein regions. Even the strongest interaction corresponds to the non-reactive Arg348-Leu349 site, indicating that in this case the selectivity is not controlled by the POM...protein interactions. On the other hand, the free-energy barriers, calculated as in previous sections, revealed a significantly lower energy demand for the reactive site ( $\Delta G_{\text{overall}}^\ddagger = 108 \text{ kJ mol}^{-1}$  for Arg114-Leu115 versus 132 and 150  $\text{kJ mol}^{-1}$  for Arg197-Leu198 and Arg348-Leu349, respectively). Interestingly, the reactive site Arg114-Leu115 is located in a random coil region of protein secondary structure, the inert sites are found in  $\alpha$ -helix regions. In fact, a similar trend was experimentally observed in the tryptic hydrolysis of apolipoprotein A-II, for which the authors suggested that peptide bonds in  $\alpha$ -helix structure have a slower rate of hydrolysis because of their structure.<sup>38</sup> The nature of the calculated rate-determining transition state provides an

explanation to the observed trends. The C—N bond cleavage is easier in a random coil where the protein chain is more extended and with less conformational restrictions compared to the  $\alpha$ -helical region. This structural effect directs the hydrolysis towards the Arg114—Leu115 peptide bond, compensating the larger non-bonding interactions of other cationic sites such as Arg348—Leu349. Note that it has been experimentally demonstrated that HSA protein has a high conformational stability at temperatures below 70°C,<sup>39</sup> while hydrolysis experiments were performed at 60°C. Therefore, in contrast with HEWL protein, the enzyme-like recognition of the POM by the environment of HSA does not determine the selectivity. Instead, the selectivity is governed by the secondary structure of the protein chain, facilitating the C—N bond cleavage step in regions with extended main chain.

## CONCLUSIONS

For the first time, the origin of the selectivity in protein hydrolysis promoted by transition-metals has been unraveled, using a computational-based protocol that combines several modeling tools. We show that for Zr(IV)-containing polyoxometalate acting as an artificial protease, the non-bonding POM-protein interactions and the secondary structure of protein chain control the observed selectivity in HEWL and HSA proteins, respectively. Using DFT calculations on cluster models derived from MD simulations and complemented with QM/MM calculations, we propose a reaction mechanism consisting of four main steps: i) coordination of the Zr center to the amide oxygen; ii) nucleophilic attack of either the hydroxo ligand or of Zr or an external water molecule assisted by the ZrOH group as general base; iii) protonation of the amide nitrogen and iv) C—N bond cleavage. The overall activity is governed by the energy cost to reach the transition state for C—N bond cleavage, the highest energy point, from reactants. The computed overall free-energy barrier of the hydrolysis is very sensitive to the protein conformation but its weighted average for HEWL (121  $\text{kJ mol}^{-1}$ ) agrees very well with the values derived from the experimental rate constants (113–134  $\text{kJ mol}^{-1}$ ). QM/MM metadynamics simulations on the early stages of the mechanism for active Asn44—Arg45 cleavage site in HEWL show the formation of an exergonic, non-covalent POM...protein complex at protein surface stabilized by positively charged amino acids. These simulations also indicated that the non-bonding interactions are maintained during Zr coordination to amide oxygen, and presumably, along the reaction coordinate.

Using  $[\text{W}_5\text{O}_{18}\text{Zr}(\text{H}_2\text{O})(\text{OH})]^{3-}$  (**ZrL**) as an artificial protease, the origin of selectivity in HEWL protein can be ascribed to an *enzyme-like* recognition process, in which the cooperative effect of positively charged and polar amino acids traps the POM more strongly at a specific site, Asn44—Arg45. These strong POM...protein interactions compensate the energy cost of POM reorientation and coordination to the amide oxygen, shifting the energy

profile down and favoring the irreversibility of the specific POM–protein interaction. For the selective hydrolysis of HSA protein, we identified a different scenario by comparing chemically equivalent sites (Arg-Leu) at different positive patches of the protein. In this case, the strength of non-bonding POM–protein interactions does not direct the selectivity, but it is the secondary structure of protein chain that prevents hydrolysis at  $\alpha$ -helical regions because of higher overall free-energy barriers for the C–N bond cleavage. Moreover, the selectivity rules undisclosed for Lindqvist-type anion ZrL might apply for other POMs since the same selectivity have been observed for POM proteases of different structure.<sup>7,9</sup>

Overall, we have identified and evaluated two key, complementary factors influencing the selectivity in peptide hydrolysis promoted by Lewis-acid, transition-metal substituted polyoxometalates. The secondary structure of peptide chain disfavors hydrolysis at the  $\alpha$ -helical regions, and then, strong electrostatic-type POM–protein interactions at positive patches can accelerate the rate of hydrolysis of specific peptide bonds. Finally, we note that the new insight gained on the origin of selectivity might be used to derive guidelines for the design of novel catalysts for protein engineering.

## COMPUTATIONAL DETAILS

**DFT calculations** were performed at the B<sub>3</sub>LYP level<sup>40</sup> and using Gaussian09 rev. Ao2 quantum chemistry package.<sup>41</sup> Ti and W centers were described using the LANL2DZ basis set with the corresponding pseudopotential,<sup>42</sup> while the 6-31G(d,p) basis set<sup>43</sup> was used for the remaining atoms. Solvent effects of water ( $\epsilon = 78.3553$ ) were included in the geometry optimizations and energy calculations by means of the IEF-PCM continuum solvent model<sup>44</sup> as implemented in Gaussian09. To describe the cleavage site within a *cluster model* approach,<sup>30</sup> the C<sub>a</sub>CO—Asn44(ZrL)—Arg45—NH $\text{C}_\alpha$  fragment was taken from the protein structure and the alpha carbons at the limits of the cluster were capped with hydrogen atoms. Also, the position of these alpha carbons was constrained during the geometry optimizations in order to reproduce the strain imposed by the main chain of the protein into the peptide bond. Frequency calculations confirmed the nature of the stationary points on the potential energy surface for all minima and transition state structures. For the latter, the unique imaginary frequency is associated with the normal mode of vibration connecting reactants and products. The standard-state correction of +8 kJ mol<sup>-1</sup> (from ideal gas at 1 atm to 1 mol L<sup>-1</sup> at 298.15 K) was applied to the free energy of all the species except water molecules. For them, the standard state corresponds to a higher concentration as water acts as solvent in this reaction. Thus, according to the water density of 0.997 g cm<sup>-3</sup>, the correction becomes +18 kJ mol<sup>-1</sup> for the free energy of water molecules, as previously assumed by other authors.<sup>45</sup>

**Classical MD simulations** were performed with the GROMACS 4.5.4 code<sup>46</sup> following the methodology described in previous studies devoted to analyze the interaction between POMs and proteins.<sup>14,19</sup> See Supporting Information for technical details of the classical simulations. For constrained simulations, a fictitious bond between the Zr center of the POM and the O atom of the protein was defined with a force constant of  $7.5 \times 10^5$  kJ mol<sup>-1</sup> nm<sup>-2</sup>. Also, angle parameters involving the Zr—O bond were

described with a force constant of  $7.5 \times 10^3$  kJ mol<sup>-1</sup> degree<sup>-2</sup>, with equilibrium values that correspond to those in the DFT-optimized geometry of the POM—GlyGly complex, calculated at the same level of theory defined above. In all cases, simulations were carried out in a periodic simulation box of dimensions  $75.7 \times 78.4 \times 79.8$  Å<sup>3</sup> filled with one HEWL protein ( $q = 8+$ ), one ZrL anion ( $q = 3-$ ), five Cl<sup>-</sup> anions to neutralize the charge and a number of water molecules close to 14650, with slight variation from one simulation to another.

**Static, hybrid QM/MM calculations** were performed with the Qchem(4.2)<sup>47</sup>–CHARMM(41b2)<sup>48</sup> interface, describing the QM region at same level of theory as in static DFT calculations, whereas the low layer atoms were described with the CHARMM36 force field.<sup>49</sup> The system was generated with CHARMM-GUI and contains one HEWL—ZrL complex obtained from previous constrained simulations, 8834 water molecules, 24 K<sup>+</sup> cations and 29 Cl<sup>-</sup> anions embedded in a periodic truncated-octahedral box with dimensions of  $a = b = c = 74.0$  Å and  $\alpha = \beta = \gamma = 109.47^\circ$ . The high-level region (represented in Figure S28) includes, besides the POM, the main chain of the dipeptide to be cleaved (from N<sub>Asn44</sub> to C<sub>Arg45</sub>). In calculations related to the outer-sphere mechanism, one solvent water molecule was also included in the QM subsystem. The four QM/MM borders crossing chemical bonds were treated with the single link atom (SLA) approach<sup>50</sup> and the EXGR method<sup>51</sup> to exclude QM/MM electrostatic interactions that involve MM host groups. The through-space partition was described by the electronic embedding scheme, in which the set of MM Gaussian-delocalized charges<sup>52</sup> with a finite width of 1.5 Å polarize the QM wave function. Replica-path<sup>32</sup> calculations were carried out with 16 images. Those corresponding to the reactants and the products were obtained from QM/MM optimizations and those in between were initially generated by linear interpolation of Cartesian coordinates. For each replica, we initially performed 80 steps of QM/MM steepest descent (SD) minimization followed by 80 additional steps using the Average-Basis Newton Raphson (ABNR) algorithm. Geometry optimizations along the replica path were carried freezing the position of all atoms further than 6 Å from the QM region, and constraining the distance between adjacent replicas with a force constant ( $k_{\text{RMS}}$ ) of  $2.5 \times 10^6$  kJ mol<sup>-1</sup> Å<sup>-2</sup> that induces an energy penalty through a Hooke's Law-like expression. To control the linearity of the reaction path, we used a force constant ( $k_{\text{angle}}$ ) of 836 kJ mol<sup>-1</sup>. This applies an angle energetic penalty upon deviation from linearity through the law of cosines, which permits a maximum value for  $\cos(\Theta)$  of 0.975 before applying any penalty (where  $\Theta$  is the angle defined by replicas  $i$ ,  $i \pm 1$  and  $i \pm 2$  on the potential energy surface). Also, an additional energy penalty ( $k_{\text{max}} = 2.5 \times 10^6$  kJ mol<sup>-1</sup> Å<sup>-2</sup>) is applied if one point moves further than 0.1 Å from their neighbors during the replica-path minimization. For more details on the constraints used in replica-path calculations, see ref. 32.

**Multiscale QM/MM simulations and metadynamics** were carried out with the CP2K code<sup>53</sup> (version 6.1) using the Quick-Step<sup>54</sup> module to compute interactions within the QM subsystem and the FIST module as implemented in CP2K as the MM driver. The QM subsystem is represented in Figure S29 and consists of 538 atoms, including the POM, several residues of the protein (Thr43, Asn44, Arg45 and the side chain of Arg68) and all the water molecules within a distance of 7 Å from the QM atoms of the POM and the protein, which account for 148 solvent molecules. The size of the QM tetragonal box is 35.0 × 35.0 × 43.0 Å<sup>3</sup> and bears an overall charge of -1. The rest of the system

was described at the MM level and was included in an orthorhombic 3D-periodic supercell of  $75.7 \times 78.4 \times 79.1 \text{ \AA}^3$ . The QM region was treated at the BLYP level<sup>55</sup> using a dual Gaussian and plane waves (GPW) method to describe the wave function with atom-centered Gaussian functions and the electron density by an auxiliary plane waves basis set. Core electrons were replaced by Goedecker-Teter-Hutter (GTH) pseudopotentials<sup>56</sup> and valence electrons were described with a double- $\zeta$  basis set supplemented with polarization functions (DZVP-MOLOPT-SRGTH);<sup>57</sup> and expanded to a plane wave basis set with an energy cutoff of 420 Ry. The MM region was treated by the AMBER99SB force field<sup>58</sup> and the TIP3P model<sup>59</sup> was adopted for classical water molecules. The interaction between the QM and the MM region was treated by the electrostatic embedding scheme, in which the set of MM charges polarizes the QM electron density. This interaction was computed through the Gaussian expansion in the electrostatic potential method (GEEP),<sup>60</sup> using six Gaussian functions and a spherical cutoff of 16 Å. Thus, MM charges are not represented as discrete point charges but as Gaussian functions, in order to prevent the electron spill-out problem. The QM/MM boundaries crossing chemical bonds were treated with the link atom approach to saturate the QM moieties with hydrogen atoms. Non-bonding electrostatic interactions within the MM region were calculated using the smooth particle mesh Ewald (SPME)<sup>61</sup> method with a cutoff of 9 Å in the real space. Simulations were performed at 298 K with NVT conditions, starting from random velocities. We used a time step of 0.5 fs to integrate Newton equations of motion, and the electronic wave function was optimized at every MD step to a tolerance of  $5 \times 10^{-5}$  Hartree within the Born-Oppenheimer MD scheme. The temperature of the system was controlled by a Nose-Hoover thermostat<sup>62</sup> with a relaxation time of 50 fs. The metadynamics technique<sup>63</sup> was used for studying the coordination process between the POM and the protein, since the time scale affordable for QM/MM simulations does not allow the observation of rare events such as thermally-activated processes. To this end, we used a single collective variable (distance between the Zr center of the POM and the amide oxygen of the protein cleavage site) and we deposited Gaussian-like hills of 1.25 kJ mol<sup>-1</sup> height and 1 Å of perpendicular width to the free-energy surface along the reaction coordinate every 50 steps (25 fs) of simulation. As shown in Figure S30, after ca. 25000 steps (~12.5 ps), the system has explored the distance range from 2.3 to 8 Å several times, and after that, the POM rides to the solvent bulk and therefore the simulation was stopped after 36160 steps (18 ps). For this reason, we can only analyze the computed free-energy landscape until ca. 8 Å, as done in Figure 4. The initial configuration was obtained from a classical MD simulation (*vide supra*) and before applying any biasing potential, the system was equilibrated for 2.3 ps of QM/MM MD simulation under NVT conditions.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: XXXXX  
 Transition state structures, energy profiles, additional analyses, snapshots and discussion about MD simulations, representations of the systems studied at QM/MM level, additional computational details and DFT-optimized Cartesian coordinates of the most representative structures (PDF).

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [j.carbo@urv.cat](mailto:j.carbo@urv.cat)

### ORCID

Josep M. Poblet: oooo-0002-4533-0623

Jorge J. Carbó: oooo-0002-3945-6721

Antonio Rodríguez-Fortea: oooo-0001-5884-5629

Albert Solé-Daura: oooo-0002-3781-3107

Jonathan D. Hirst: oooo-0002-2726-0983

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

We thank the Spanish Ministry of Science (PGC2018-100780-B-100, CTQ2017-87269-P and UNRVI5-EE-3935), the Generalitat de Catalunya (2017SGR629) and the URV for generous support. J.M.P. also thanks the ICREA foundation for an ICREA ACADEMIA award. We are grateful for access to the University of Nottingham high performance computer.

## REFERENCES

- See this review and references therein: Grant, K. B.; Kassai, M. *Curr. Org. Chem.* Major advances in the hydrolysis of peptides and proteins by metal ions and complexes. **2006**, *10*, 1035–1049.
- Radzicka, A.; Wolfenden, R. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *J. Am. Chem. Soc.* **1996**, *118*, 6105–6109.
- See for instance: Wezynfeld, N. E.; Frączyk, T.; Bal, W. Metal assisted peptide bond hydrolysis: Chemistry, biotechnology and toxicological implications. *Coord. Chem. Rev.* **2016**, *327–328*, 166–187.
- Absillis G.; Parac-Vogt, T. N. Peptide Bond Hydrolysis Catalyzed by the Wells-Dawson  $Zr(\alpha_2P_2W_{17}O_{61})_2$  Polyoxometalate. *Inorg. Chem.* **2012**, *51*, 9902–9910.
- Ly, H. G. T.; Mihaylov, T.; Absillis, G.; Pierloot, K.; Parac-Vogt, T. N. Reactivity of Dimeric Tetraazirconium(IV) Wells-Dawson Polyoxometalate toward Dipeptide Hydrolysis Studied by a Combined Experimental and Density Functional Theory Approach. *Inorg. Chem.* **2015**, *54*, 11477–11492.
- Stroobants, K.; Moelants, E.; Ly, H. G. T.; Proost, P.; Bartik, K.; Parac-Vogt, T. N. Polyoxometalates as a Novel Class of Artificial Proteases: Selective Hydrolysis of Lysozyme under Physiological pH and Temperature Promoted by a Cerium(IV) Keggin-Type Polyoxometalate. *Chem. Eur. J.* **2013**, *19*, 2848–2858.
- Stroobants, K.; Absillis, G.; Moelants, E.; Proost P.; Parac-Vogt, T. N. Regioselective Hydrolysis of Human Serum Albumin by  $Zr^{IV}$ -Substituted Polyoxotungstates at the Interface of Positively Charged Protein Surface Patches and Negatively Charged Amino Acid Residues. *Chem. Eur. J.* **2014**, *20*, 3894–3897.
- Stroobants, K.; Goovaerts, V.; Absillis, G.; Bruylants, G.; Moelants, E.; Proost P.; Parac-Vogt, T. N. Molecular Origin of the Hydrolytic Activity and Fixed Regioselectivity of a  $Zr^{IV}$ -Substituted Polyoxotungstate as Artificial Protease. *Chem. Eur. J.* **2014**, *20*, 9567–9577.
- Ly, H. G. T.; Absillis, G.; Janssens, R.; Proost, P.; Parac-Vogt, T. N. Highly Amino Acid Selective Hydrolysis of Myoglobin at Aspartate Residues as Promoted by Zirconium (IV)-

- Substituted Polyoxometalates. *Angew. Chem. Int. Ed.* **2015**, *54*, 7391–7394.
10. Sap, A.; Van Tichelen, L.; Mortier, A.; Proost, P.; Parac-Vogt, T. N. Tuning the Selectivity and Reactivity of Metal-Substituted Polyoxometalates as Artificial Proteases by Varying the Nature of the Embedded Lewis Acid Metal Ion. *Eur. J. Inorg. Chem.* **2016**, *32*, 5098–5105.
  11. Quanten, T.; Mayaer, T. D.; Shestakova, P.; Parac-Vogt, T. N. Selectivity and reactivity of Zr<sup>IV</sup> and Ce<sup>IV</sup> substituted Keggin type polyoxometalates toward cytochrome c in surfactant solutions. *Front. Chem.* **2018**, *6*, 372.
  12. Ly, H. G. T.; Mihaylov, T. T.; Proost, P.; Pierloot, K.; Harvey, J. N.; Parac-Vogt, T. N. Chemical Mimics of Aspartate-Directed Proteases: Predictive and Strictly Specific Hydrolysis of a Globular Protein at Asp-X Sequence Promoted by Polyoxometalate Complexes Rationalized by a Combined Experimental and Theoretical Approach. *Chem. Eur. J.* **2019**, *25*, 14370–14381.
  13. a) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522–W525; b) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667; c) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to micro-tubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037–10041.
  14. Solé-Daura, A.; Goovaerts, V.; Stroobants, K.; Absillis, G.; Jiménez-Lozano, P.; Poblet, J. M.; Hirst, J. D.; Parac-Vogt, T. N.; Carbó, J. J. Probing polyoxometalate–protein interactions using molecular dynamics simulations. *Chem. Eur. J.* **2016**, *22*, 15280–15289.
  15. Sap, A.; De Zitter, E.; Van Meervelt, L.; Parac-Vogt, T. N. Structural Characterization of the Complex between Hen Egg-White Lysozyme and Zr<sup>IV</sup>-Substituted Keggin Polyoxometalate as Artificial Protease. *Chem. Eur. J.* **2015**, *21*, 11692–11695.
  16. Vandebroek, L.; De Zitter, E.; Ly, H. G. T.; Conić, D.; Mihaylov, T.; Sap, A.; Proost, P.; Pierloot, K.; Van Meervelt, L.; Parac-Vogt, T. N. Protein-Assisted Formation and Stabilization of Catalytically Active Polyoxometalate Species. *Chem. Eur. J.* **2018**, *24*, 10099–10108.
  17. Vandebroek, L.; Mampaey, Y.; Antonyuk, S.; Van Meervelt, L.; Parac-Vogt, T. N. Noncovalent Complexes Formed between Metal-Substituted Polyoxometalates and Hen Egg White Lysozyme. *Eur. J. Inorg. Chem.* **2019**, *3*, 506–511.
  18. Paul, T. J.; Parac-Vogt, T. N.; Quiñonero, D.; Prabhakar, R. Investigating Polyoxometalate–Protein Interactions at Chemically Distinct Binding Sites. *J. Phys. Chem. B* **2018**, *122*, 7219–7232.
  19. Solé-Daura, A.; Poblet, J. M.; Carbó, J. J. Structure-Activity Relationships for the Affinity of Chaotropic Polyoxometalate Anions towards Proteins. *Chem. Eur. J.* **2020**, *26*, 5799–5809.
  20. Jayasinghe-Arachchige, V. M.; Hu, Q.; Sharma, G.; Paul, T. J.; Lundberg, M.; Quinonero, D.; Parac-Vogt, T. N.; Prabhakar, R. Hydrolysis of Chemically Distinct Sites of Human Serum Albumin by Polyoxometalate: A Hybrid QM/MM (ONIOM) Study. *J. Comp. Chem.* **2019**, *40*, 51–61.
  21. a) Zhang, T.; Zhu, X.; Prabhakar, R. Mechanistic insights into metal (Pd<sup>2+</sup>, Co<sup>2+</sup>, and Zn<sup>2+</sup>)–β-cyclodextrin catalyzed peptide hydrolysis: A QM/MM approach. *J. Phys. Chem. B* **2014**, *118*, 4106–4114; b) Zhang, T.; Ozbil, M.; Barman, A.; Paul, T. J.; Bora, R. P.; Prabhakar, R. Theoretical insights into the functioning of metallopeptidases and their synthetic analogues. *Acc. Chem. Res.* **2015**, *48*, 192–200; c) Zhang, T.; Sharma, G.; Paul, T. J.; Hoffmann, Z.; Prabhakar, R. Effects of ligand environment in Zr (IV) assisted peptide hydrolysis. *J. Chem. Inf. Model.* **2017**, *57*, 1079–1088.
  22. Pelmenschikov, V.; Blomberg, M. R. A.; Siegbahn, P. E. M. A theoretical study of the mechanism for peptide hydrolysis by thermolysin. *J. Biol. Inorg. Chem.* **2002**, *7*, 284–298.
  23. Bora, R. P.; Barman, A.; Zhu, X.; Ozbil, M.; Prabhakar, R. Which one among aspartyl protease, metallopeptidase, and artificial metallopeptidase is the most efficient catalyst in peptide hydrolysis? *J. Phys. Chem. B* **2010**, *114*, 10860–10875.
  24. Blumberger, J.; Lamoureux, G.; Klein, M. L. Peptide hydrolysis in thermolysin: Ab initio QM/MM investigation of the Glu143-assisted water addition mechanism. *J. Chem. Theory Comput.* **2007**, *3*, 1837–1850.
  25. Brás, N. F.; Fernandes, P. A.; Ramos, M. J. QM/MM studies on the β-galactosidase catalytic mechanism: Hydrolysis and transglycosylation reactions. *ACS Catal.* **2014**, *4*, 2587–2597.
  26. Díaz, N.; Suárez D. Peptide hydrolysis catalyzed by matrix metalloproteinase 2: a computational study. *J. Phys. Chem. B* **2008**, *112*, 8412–8424.
  27. Paul, T. J.; Barman, A.; Ozbil, M.; Bora, R. P.; Zhang, T.; Sharma, G.; Hoffmann, Z.; Prabhakar, R. Mechanisms of peptide hydrolysis by aspartyl and metalloproteases. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24790–24801.
  28. Leopoldini, M.; Russo, N.; Toscano, M. Which one among Zn (II), Co (II), Mn (II), and Fe (II) is the most efficient ion for the methionine aminopeptidase catalyzed reaction? *J. Am. Chem. Soc.* **2007**, *129*, 7776–7784.
  29. Alberto, M. E.; Leopoldini, M.; Russo, N. Can human prolidase enzyme use different metals for full catalytic activity? *Inorg. Chem.* **2011**, *50*, 3394–3403.
  30. The cluster model approach consists of taking a fragment from a biological system to perform calculations constraining the relative position of the capping groups during the geometry optimizations to mimic the strain in the real system (see also Computational Details section). For an example of its successful application, see: Himo F; Siegbahn P. E. M. Quantum chemical studies of radical-containing enzymes. *Chem. Rev.* **2003**, *103*, 2421–2456.
  31. a) Jiménez-Lozano, P.; Solé-Daura, A.; Wipff, G.; Poblet, J. M.; Chaumont, A.; Carbó, J. J. Assembly Mechanism of Zr-Containing and Other TM-Containing Polyoxometalates. *Inorg. Chem.* **2017**, *56*, 4248–4156; b) Jiménez-Lozano, P.; Carbó, J. J.; Chaumont, A.; Poblet, J. M.; Rodríguez-Fortea, A.; Wipff, G. Nature of Zr-Monosubstituted Monomeric and Dimeric Polyoxometalates in Water Solution at Different pH Conditions: Static Density Functional Theory Calculations and Dynamic Simulations. *Inorg. Chem.* **2014**, *53*, 778–786.
  32. Woodcock, H. L.; Hodošček, M.; Sherwood, P.; Lee, Y. S.; Schaefer III, H. F.; Brooks, B. R. Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor. Chem. Acc.* **2003**, *109*, 140–148.
  33. Ferrer, S.; Tuñón, I.; Martí, S.; Moliner, V.; Garcia-Viloca, M.; González-Lafont, Á.; Lluch, J. M. A theoretical analysis of rate constants and kinetic isotope effects corresponding to different reactant valleys in lactate dehydrogenase. *J. Am. Chem. Soc.* **2006**, *128*, 16851–16863.
  34. Hu, P.; Zhang, Y. Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: An ab initio QM/MM-FE study with multiple initial structures. *J. Am. Chem. Soc.* **2006**, *128*, 1272–1278.

35. Lodola, A.; Mor, M.; Zurek, J.; Tarzia, G.; Piomelli, D.; Harvey, J. N.; Mulholland, A. J. Conformational Effects in Enzyme Catalysis: Reaction via a High Energy Conformation in Fatty Acid Amide Hydrolase. *Biophys. J.* **2007**, *92*, L20–L22.
36. Benkovic, S. J.; Hammes, G. G.; Hammes-Schiffer, S. Free-energy landscape of enzyme catalysis. *Biochemistry* **2008**, *47*, 3317–3321.
37. a) Ribeiro, A. J.; Ramos, M. J.; Fernandes, P. A. The catalytic mechanism of HIV-1 integrase for DNA 3'-end processing established by QM/MM calculations. *J. Am. Chem. Soc.* **2012**, *134*, 13436–13447; b) Ribeiro, A. J. M.; Santos-Martins, D.; Russo, N.; Ramos, M. J.; Fernandes, P. A. Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1Protease. *ACS Catal.* **2015**, *5*, 5617–5626.
38. Massey, J. B.; Hickson-Bick, D. L. M.; Gotto, Jr., A. M.; Pownall, H. J. Kinetics of tryptic hydrolysis as a probe of the structure of human plasma apolipoprotein A-II. *Biochim. Biophys. Acta* **1989**, *999*, 121–127.
39. Picó, G. A. Thermodynamic features of thermal unfolding of human serum albumin. *Int. J. Biol. Macromol.* **1997**, *20*, 63–73.
40. a) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789; b) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652; c) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
41. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
42. Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 299–310.
43. a) Franci, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665; b) 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; c) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret. Chim. Acta* **1973**, *28*, 213–222.
44. Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model. *J. Chem. Phys.* **1997**, *107*, 3032–3041.
45. See for instance: Mihaylov, T. T.; Ly, H. G. T.; Pierloot, K.; Parac-Vogt, T. N. Molecular Insight from DFT Computations and Kinetic Measurements into the Steric Factors Influencing Peptide Bond Hydrolysis Catalyzed by a Dimeric Zr(IV)-Substituted Keggin Type Polyoxometalate. *Inorg. Chem.* **2016**, *55*, 9316–9328.
46. a) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447; b) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718; c) Berendsen, H. J. C.; Van Der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
47. Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock III, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio Jr., R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; William A. Goddard III, Mark S. Gordon, G. III M. S.; Hehre, W. J.; Klamt, A.; Schaefer III, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W. Z.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.
48. Brooks, B. R.; Brooks, C. L. 3rd; Mackerell, A. D. Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
49. a) MacKerell, A. D., Jr., Feig, M. and Brooks, III, C. L. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–699; b)

- MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **1998**, *102*, 3586–3616.
50. a) Vasilyev, V. V. Tetrahedral intermediate formation in the acylation step of acetylcholinesterases. A combined quantum chemical and molecular mechanical model. *J. Mol. Struct. (Theochem)* **1994**, *110*, 129–141; b) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. A combined quantum mechanical/molecular mechanical model of the potential energy surface of ester hydrolysis by the enzyme phospholipase A<sub>2</sub>. *J. Chem. Soc. Perkin Trans. 2* **1991**, *2*, 225–231.
51. Singh, U. C.; Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH<sub>3</sub>Cl + Cl<sup>-</sup> exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* **1986**, *7*, 718–730.
52. Das, D.; Eurenus, K. P.; Billings, E. M. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.* **2002**, *117*, 10534–10547.
53. Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: atomistic simulations of condensed matter systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 15–25.
54. VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comp. Phys. Com.* **2005**, *167*, 103–128.
55. Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
56. a) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710; b) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641–3662.
57. Schaefer A, Huber C, Ahlrichs R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1994**; *100*, 5829–5835.
58. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712–725.
59. 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.
60. Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J. Chem. Theory Comput.* **2006**, *2*, 1370–1378.
61. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald potential. *J. Chem. Phys.* **1995**, *103*, 8577–85923.
62. a) Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519; b) Nosé S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Physics* **1984**, *52*, 255–268.
63. Laio, A.; Parrinello, M. Escaping free-energy minima. *PNAS* **2002**, *99*, 12562–12566.

