Chemical Physics 561 (2022) 111602

Contents lists available at ScienceDirect

Chemical Physics

journal homepage: www.elsevier.com/locate/chemphys

Conformational analysis and water dynamics: a molecular dynamics study on the survival of a β-lactoglobulin peptide in the archaeological record

Beatriz Fonseca [a,*], Colin L. Freeman [b], Matthew J. Collins [a,c]

[a] The Globe Institute, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen K, Denmark
[b] Department of Materials Science and Engineering, The University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom
[c] McDonald Institute for Archaeological Research, University of Cambridge, CB2 3ER Cambridge, United Kingdom

ARTICLE INFO

Keywords:
Protein hydrolysis
β-lactoglobulin
Water dynamics
Protein survival

ABSTRACT

One of the greatest successes of the application of Palaeoproteomics to Archaeology is its use by a number of authors to track evidence of dairying practice, both in terms of its origin and the selection of animal species. To this end, the whey protein β-lactoglobulin entrapped in pottery and dental calculus is widely studied *because* it is so frequently recovered but why is it differentially preserved? Hydrolysis plays a big part in the breakdown of proteins. Therefore, it is essential to explore the role of water in degradation to uncover some of the patterns linked to protein survival. One approach to understand the hydrolytic process is to examine the molecular behaviour of this protein and in particular of the peptide most commonly recovered: T_{125}PEVDXEALEK$_{135}$. In this study, we use Molecular Dynamics, with the Amber14SB forcefield and the SPC/E water model in Gromacs 2020, to first explore the dynamics of this peptide in bulk water. Despite the difficulties in describing reactive processes with classical methods, we were able to identify geometric arrangements between water and protein residues which are similar to the ones described in the literature for protein hydrolysis. These arrangements helped to identify potential sites for hydrolysis along the bovine β-lactoglobulin T_{125}PEVDDEALEK$_{135}$ amino acid chain.

1. Introduction

Increasingly ancient biomolecules, such as DNA and proteins, recovered from archaeological contexts are providing new insights into the lifeways of people in the past. In the context of proteins, undoubtedly the most widely reported discoveries are those surrounding the understanding of the advent of dairying [1–5]. There is evidence that some of the milk proteins may get trapped into limescale in archaeological ceramics or dental calculus [1,3–5]. However, it is unclear why in so many contexts, milk rather than other food proteins are being detected. Furthermore, when milk is detected, it is the whey protein β-lactoglobulin (BLG) which is most commonly recovered [2,5]. This finding is even more peculiar given that early populations were unable to digest lactose, another major component of the whey fraction of milk [3].

In particular, an eleven amino acid long tryptic peptide T_{125}PEVDXEALEK$_{135}$ is more frequently recovered from archaeological materials than any other BLG fragment [2]. This peptide is unique in two ways. First, a mutation in the amino acid on position 6 of the sequence allows for species differentiation [2] between cow (Asp$_{130}$), sheep (Asn$_{130}$) and goat (Lys$_{130}$). Second, such sequence is rich in acidic residues, such as aspartates and glutamates, and comes from a flexible region of the BLG protein which consists of a turn going into a α-helix. These flexible regions are not only often those with the greatest sequence variation, but they are also more exposed and more likely to undergo hydrolysis [6,7]. Therefore, the widespread persistence of this peptide in the archaeological record is unexpected and may be the result of binding to mineral surfaces, as has been previously suggested as a survival mechanism for a peptide in ostrich eggshell [7].

To understand the mechanism of the unlikely preservation of the bovine BLG peptide T_{125}PEVDDEALEK$_{135}$, this study first seeks to explore its stability in solution, in order to understand its behaviour without the presence of a mineral surface. We then aim to establish whether it is as prone to non-enzymatic hydrolysis as we might expect.

Although reactive processes are more suitably described by quantum mechanical methods, the length of the peptide and a lack of previous knowledge as to where in the chain reaction might occur render such

Abbreviations: BLG, β-lactoglobulin; metaD, Metadynamics; MD, Molecular Dynamics; QM/MM, Quantum Mechanics / Molecular Mechanics; CPMD, Car-Parrinello Molecular Dynamics.

* Corresponding author.
E-mail address: bia@palaeome.org (B. Fonseca).

https://doi.org/10.1016/j.chemphys.2022.111602
Received 8 February 2022; Received in revised form 30 May 2022; Accepted 1 June 2022
Available online 6 June 2022

methods computationally prohibitive. Hybrid methods such as QM/MM and CPMD have been used to study reactions in biological systems [8–10] However, such methods are still very limited in the size of the QM region.

In order to resolve this conundrum, we have explored an alternative view to investigate the ease of hydrolysis along a peptide chain. We draw on the literature in hydrolysis and use Molecular Dynamics to explore how the susceptibility of a peptide bond to undergo hydrolysis changes as a consequence of peptide configuration, the presence of cations and charge. Even though it is difficult to tackle such a problem from a classical point of view, the literature in hydrolysis of proteins is extensive enough to allow us to investigate the hydrolytic process from different fronts.

1.1. Hydrolysis of peptide sequences

The hydrolysis of proteins at different pHs has been extensively studied by several authors [11–30]. Although it is well known that the peptide bond is more prone to hydrolysis in more acidic and alkaline pHs [18,22,23,25], there has been some effort in understanding the non-enzymatic peptide bond breakage in neutral water due to its importance for biological processes [11–13,26].

In general, theoretical studies focus on unveiling the reaction mechanism [11,13,18,25,26,29]. In these studies, formamide is usually employed as a model compound [11,18,25,29], since its simplicity is ideal for quantum mechanics based methods. There is no consensus on whether the reaction occurs in a stepwise or concerted fashion. However, several authors agree that a water assisted mechanism, with or without the formation of an intermediate, is more energetically favourable [11,18,23,25,29–31]. In more than one study, a key step seems to involve the attack of one or more water molecules on the carbonyl group of the peptide bond [11,18,31]. For acid catalyzed hydrolysis, the O-protonation is more preferential than the *N*-protonation [16,18,31]. However, Antonczak et al. [18] have suggested that a similar mechanism is likely to occur in bulk neutral water, as the carbonyl oxygen hydrogen bonds with water molecules. Despite not taking solvent effects into account, the authors have found that the protonation of the oxygen has a catalytic effect in both acidic and neutral conditions [18].

Gorb et al. [11] have taken the study of the hydrolysis of formamide one step further by taking into consideration the effect of interactions with the bulk. They have unveiled a preferential water assisted mechanism that also occurs through attack on the carbonyl backbone bond, leading to the formation of an amino-*gem*-diol intermediate in a concerted fashion. This result agrees with previous findings that have determined that the addition of a hydroxyl group to the amide bond is the rate limiting step for the neutral hydrolysis of model amide compounds [29,31]. The transition state structure involves two water molecules, where a water assisted proton transfer occurs between the hydroxyl and the amino group. Scheme 1 shows the reaction pathway for this mechanism. Pan et al. [26] concluded that this mechanism is more likely to occur in both neutral and acidic pHs by performing a more sophisticated study using *N*-methyl acetyl acrylamide (*N*-MAA) as a model compound and including solvent effects. The authors also concluded that the degradation of proteins by hydrolysis is a complicated process, with many degrees of freedom involved, since at least five reaction coordinates are necessary to describe peptide bond breakage. These reaction coordinates go beyond the more commonly used carbonyl carbon to water oxygen distance [11,18,29,31] and relate to not only protein structure but also solvent dynamics.

Since the susceptibility of a given peptide bond to undergo hydrolysis is subjected to factors relating to protein structure and dynamics of the solvent shell [26], other authors have focused their efforts in understanding how protein sequence affects reaction mechanisms and rates [17,21,22,30,32]. Specifically, there is an agreement towards the role of the carboxylic acid side chain of aspartates and glutamates in lowering the energy barrier for reaction to occur. Friedrich et al. [21] observed spontaneous cleavage of the adjacent peptide bond to the C-terminal side of glutamate and glutamine residues, although the reaction was more pronounced when glutamate was present. Such an effect occurred both at neutral and acidic pHs. The presence of Ala, Leu, Pro, Val or Arg residues on the *N*-terminal side of the glutamate residue increased the incidence of cleavage. The authors hypothesize a mechanism in which the carboxylic group of glutamate promotes a nucleophilic attack on the carbonyl group of the peptide backbone. Similarly, there has been evidence of the spontaneous peptide cleavage at the C-terminal of Asp and Asn residues [30,32–34]. Moreover, Wang et al. [32] have found this reaction occurs at the pH range of 5.0 to 7.4. Two competing mechanisms have been suggested. The first involves the attack of the carboxylic group on the adjacent C-terminal peptide bond, forming a reactive Asp anhydride intermediate [30,32]. The second is through the formation of a succinimide intermediate and has been broadly studied in literature [33–36].

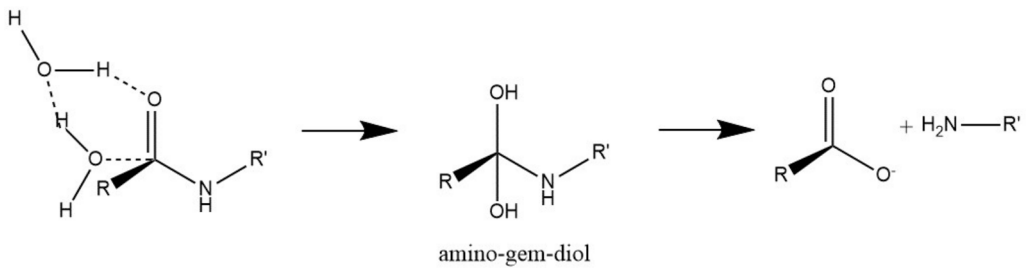
In this study, we use Molecular Dynamics to understand the dynamics of the bovine β-lactoglobulin T_{125}PEVDDEALEK$_{135}$ amino acid chain in solution and gain insight into the mechanism for its survival in the archaeological record. To do this, we first explore the conformational space of the peptide in solution to determine stable peptide configurations. We then use these configurations to identify geometric arrangements between water and protein residues which are similar to the ones described in literature and are therefore consistent with a greater likelihood for hydrolysis. Such arrangements may help identify potential sites for hydrolysis along the bovine β-lactoglobulin T_{125}PEVDDEALEK$_{135}$ amino acid chain.

2. Methods

2.1. Simulation details

The dynamics of the bovine β-lactoglobulin eleven amino acid long peptide T_{125}PEVDDEALEK$_{135}$ was investigated in neutral water. The initial geometry for the peptide was cut from the whole bovine β-lactoglobulin (PDB id: 1CJ5 [37]) between positions 125 and 135.

Although it would be valuable to also investigate the dynamics of T_{125}PEVDDEALEK$_{135}$ at different pH conditions, we have chosen to run the simulations at pH 7 due to the simplicity of the system. The peptide



Scheme 1. Concerted hydrolysis with the formation of an amino-*gem*-diol intermediate [11,26].

then had amine groups protonated and carboxylic groups deprotonated, with the *N*-terminal kept as NH_3^+ and the C-terminal kept as CO_2^-. It is important to note that different protonation states of charged groups along the amino acid chain will likely alter the preferential stable conformations and therefore the peptide bonds susceptible to hydrolysis found in this study. However, such comparison represents a large body of work beyond the scope of this paper.

All simulations in this study were performed using the GROMACS 2020.3 [38–40] software. The peptide was described by the Amber14SB [41] forcefield and solvated in a SPC/E [42,43] cubic water box containing 5676 water molecules. The net charge of the system was neutralised with Na^+ ions. The system was first equilibrated for 1 ns in an NVT ensemble with a reference temperature of 300 K. In all simulations, the temperature was controlled by a Nosé-Hoover thermostat with a time coupling constant of 0.1 ps. This was followed by a 1 ns equilibration in the NPT ensemble with a reference pressure of 1 bar. The pressure was controlled by a Parrinello-Rahman barostat with a time constant coupling of 2 ps. A production MD in the NVT ensemble at 300 K was then performed, with a time step of 0.5 fs. For all calculations, the Particle Mesh Ewald (PME) with a real space cut-off of 1.0 nm, Fourier grid spacing of 0.12 nm and a 4th order spline was used for long range electrostatic interactions. The Lennard-Jones 12–6 interactions had a cut-off distance of 1.0 nm. Periodic boundary conditions were implemented in all directions.

To enhance sampling further, metadynamics [44,45] (metaD) was performed with the PLUMED [46,47] v2.6 plug-in for GROMACS. Gaussian hills were added at a regular interval of 0.5 ps with a 1 kJ/mol height and a width of 0.35. Since it was observed in previous simulations that the peptide structure tended to switch between a folded and an unfolded arrangement, the radius of gyration (R_g) was chosen as a collective variable. Runs had the total length of 30 ns.

2.2. Conformational analysis

A distinct conformation is defined as a maximum variation of 0.2 nm in the root mean square deviation (RMSD) of the backbone atoms (C Cα N O), maintained for at least 5 ns during a given molecular dynamics (MD) simulation. This provided meaningful statistics to be able to more confidently explore water dynamics around the structure and aided identification of potential sites for hydrolysis. A standard 50 ns molecular dynamics (MD) simulation was first performed in order to sample different geometries. These geometries were used as starting conformations for further 20 ns standard MD simulations and for metadynamics. The minima found in metadynamics simulations were also taken forward as starting geometries for standard MD simulations. Sampled arrangements that did not fulfill the requirements to be considered a stable conformation in this study were used as starting points for further replica simulations. Otherwise, if a given conformation was stable, its contribution to the total average potential energy of the system was calculated. Assuming independent states, the average energy of the ensemble can be defined as:

$$\langle E \rangle = \frac{\sum_i E_i e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} \tag{1}$$

where $exp(-\beta E_i)$ is the Boltzmann weight of a certain state i with probability $p_i \propto e^{-\beta E_i}$, and $\beta = 1/k_B T$. When there were no more significant changes to the value of $\langle E \rangle$, it was considered that the conformational space had been sufficiently sampled. A workflow for the simulations is displayed in Fig. 1. The average number of intramolecular hydrogen bonds for each residue was calculated in order to better understand differences in structure.

2.3. Water dynamics and determining potential sites for hydrolysis

To be able to identify sites that may be more likely to undergo hydrolysis, we have used three different approaches based on existing literature. Firstly, to address the possibility of the carboxyl groups engaging in hydrolysis, the radial distribution function (RDF) for cations and water were calculated to generate their distribution around carboxyls. In cases where it was necessary to obtain the coordination number, the first maximum of the RDF was smoothed with a second order polynomial Savitzky-Golay filter and integrated.

Secondly, water dynamics and water availability around the peptide chain were investigated, since there is a consensus that water assisted mechanisms are more likely to occur in neutral water [11,12,18,26]. Hydrogen bond time autocorrelation functions were computed for hydrogen bonds between water molecules and atoms within the peptide bond to determine water residence time around sites of interest for

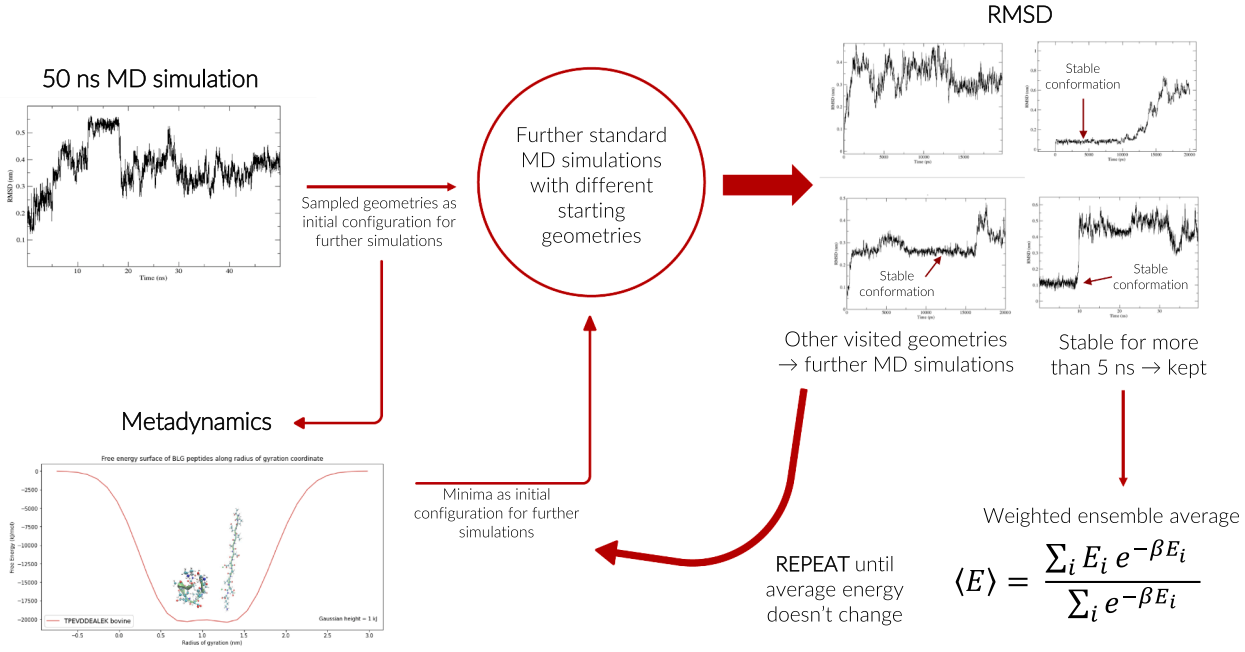


Fig. 1. Schematic drawing of workflow of simulations.

hydrolysis. The autocorrelation function is given by [48,49]:

$$C_{HB}(\tau) = \frac{\langle s_i(t_0)s_i(t)\rangle}{\langle s_i^2(t_0)\rangle} \quad (2)$$

where $s_i(t)$ is the existence function of a given hydrogen bond *i*. If $s_i(t) = 1$, a hydrogen bond with donor–acceptor distance of maximum 0.35 nm and a hydrogen-donor–acceptor angle of 30° exists at time t. Otherwise, $s_i(t) = 0$. Hydrogen bond lifetimes were obtained by fitting the autocorrelation functions with a third order exponential.

Finally, we focused our efforts in identifying water-peptide arrangements along the T_{125}PEVDDEALEK$_{135}$ amino acid chain that were favorable for hydrolysis. Geometric parameters linked to reported reactant structures were calculated and compared to published values in literature [11,18,26] (Table 1).

All analyses were performed with standard GROMACS analysis tools and with self-written scripts in Python 3.

3. Results and discussion

3.1. Exploring conformational space for the bovine BLG peptide

Preliminary simulations have shown that the T_{125}PEVDDEALEK$_{135}$ peptide is very flexible. Such observation is not unexpected as the sequence has an overall charge of $-4e$ and is relatively short. Therefore, it was necessary to explore its conformational space prior to any analysis of water dynamics and identification of possible sites for hydrolysis. Metadynamics simulations revealed that the potential energy surface of the bovine β-lactoglobulin T_{125}PEVDDEALEK$_{135}$ peptide along the radius of gyration coordinate has two distinct minima (Fig. 2) relating to a folded and an unfolded conformation. The energy barrier between these two minima is very small, consistent with the ease the peptide changes between folded and unfolded states during a simulation.

In total, five stable peptide arrangements were found (Fig. 3). These consist of three folded (structures 1, 2 and 3) and two unfolded peptide conformations (structures 4 and 5). Such structures are stabilised by intramolecular hydrogen bonds and, in some cases, the proximity of a Na^+ ion to negatively charged residues along the peptide chain. Table 2 lists the average number of intramolecular hydrogen bonds per residue in each conformation. Analysing the three folded structures, we observe that structure 3 has on average a total of 10.67 intramolecular hydrogen bonds, whilst structures 1 and 2 have on average a total of 7.20 and 6.26 intramolecular hydrogen bonds, respectively. The higher number of intramolecular hydrogen bonds in structure 3 is because it lacks any interaction with Na^+, so it depends solely on interactions between the positively charged side chain of the lysine residue and other negatively charged residues in the middle of the chain - in particular the aspartates - to stay in its folded state. The *N*-terminal threonine then interacts with the peptide bond between lysine and the glutamate in position 10 to close the loop. On the other hand, structures 1 and 2 depend partially on the interaction with the cations in solution to maintain their conformation. The Na^+ ions not only act as a bridge to close the loop, but they also provide a positive charge that helps to stabilize the repulsive forces caused by the proximity of aspartate and glutamate residues due to folding of the structure. This is reflected in a lower average number of intramolecular hydrogen bonds for these structures.

A similar effect is observed between the unfolded structures 4 and 5. Structure 4 depends on intramolecular hydrogen bonds to form the turn towards the *N*-terminal and the helix towards the C-terminal, which translates into an average of 12.19 hydrogen bonds throughout the chain. In comparison, structure 5 has only 3.22 intramolecular bonds overall as it relies on four Na^+ ions to maintain the conformation in an unfolded state, due to electrostatic interactions between the cations and all carboxyl groups present in the peptide chain. Such carboxyl groups belong to not only aspartates and glutamates, but also to the C-terminus lysine. In general, the higher the number of Na^+ ions contributing to stabilising a given structure, the less intramolecular hydrogen bonds are required. Therefore, conformations such as structures 1, 2 and especially 5 could be an indication of how the peptide binds to the mineral surface, where the cations present in the mineral would replace the role of Na^+.

Table 1
Geometric parameters reported in literature for the reactant structure for the hydrolysis of amides. Relevant groups and atoms are represented as follows: C═O peptide bond carbonyl, NH peptide bond amine, O_w water oxygen, H_w water hydrogen.

SOURCE	Model	Geometric parameter	Value
Gorb et al.[11]	Formamide - reactant structure	distance C═O–H_w	1.788 Å /1.805 Å
		distance NH–O_w	1.821 Å / 2.028 Å
Pan et al.[18]	*N*-MAA – reactant structure	distance O═C–O_w	(2.07, +∞) Å
		distance N–H_w	(1.79, +∞) Å
Antonczak et al.[26]	Formamide – pre-reactive hydrogen bonded complex	distance O═C–O_w	2.749 Å
		distance N–H_w	2.109 Å

3.2. Influence of the cation on water availability and limiting reaction pathways

Another effect caused by the proximity of cations to a given site along the peptide chain is the limited water availability around those sites. Fig. 4A shows the first maximum of the RDF of both water hydrogens (H_w) and Na^+ ions to the carboxyl oxygens in structure 1 (see Figure S1 for RDFs). The higher the value of the RDF for water, the lower the value of the equivalent RDF for Na^+ ions. For example, for the aspartate residue in positions 5 and 6 in the peptide, it is evident that there is high probability to find water around them. For the same residues, however, the probability of finding Na^+ ions in their vicinity is effectively zero. This suggests that the cations limit the amount of water available to engage in hydrolysis at the site they are closest to. This effect is visible in all structures that rely on Na^+ ions to remain stable, as can be seen by the negative correlation between the coordination number of carboxyl groups in respect to water and Na^+ ions in Fig. 4B.

Previous studies have shown the participation of the carboxylic group in aspartates and glutamates in the spontaneous hydrolysis of the peptide bond on their C-terminal [21,22,32]. By interaction with such groups, the cations hinder their possibility in facilitating the reaction and limit possible reaction mechanisms for the hydrolysis of T_{125}PEVDDEALEK$_{135}$. The effects produced by interaction with the cations in solution may thus be an indication of how the mineral surface enhances preservation of the β-lactoglobulin peptide. The high density of positive charge due to Ca^{2+} ions at a calcite surface probably also limits the interaction of negatively charged residues with water and the peptide chain.

3.3. Water dynamics around peptide bonds

Peptide bonds are the primary site to undergo hydrolysis along a peptide chain and therefore we analysed the behaviour of water molecules around these sites. Fig. 5 shows the hydrogen bond autocorrelation function between each peptide bond in the T_{125}PEVDDEALEK$_{135}$ peptide and water, for each of the five structures previously discussed. It is evident that differences in conformation affect the hydrogen bond lifetimes around these sites. The water availability around a certain peptide bond will change from structure to structure, according to the geometry it adopts influenced by intramolecular hydrogen bonds, the proximity of Na^+ ions and possible side chain steric effects.

In most cases, the hydrogen bond lifetime seems to decay quite rapidly and reach zero before 164 ± 60 ps on average. Such hydrogen bonds are not long lasting and therefore are unlikely to engage in

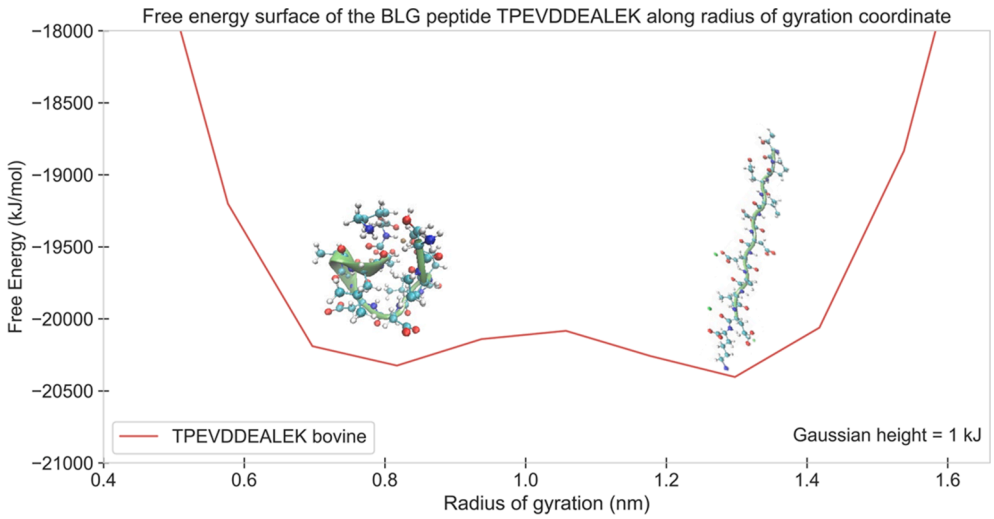


Fig. 2. Free energy surface of the bovine β-lactoglobulin T_{125}PEVDDEALEK$_{135}$ peptide obtained by performing metadynamics along the radius of gyration coordinate. Peptide conformations for each of the minima are represented near their respective minimum. Given the small barrier, the plot has been amplified to focus on the minima.

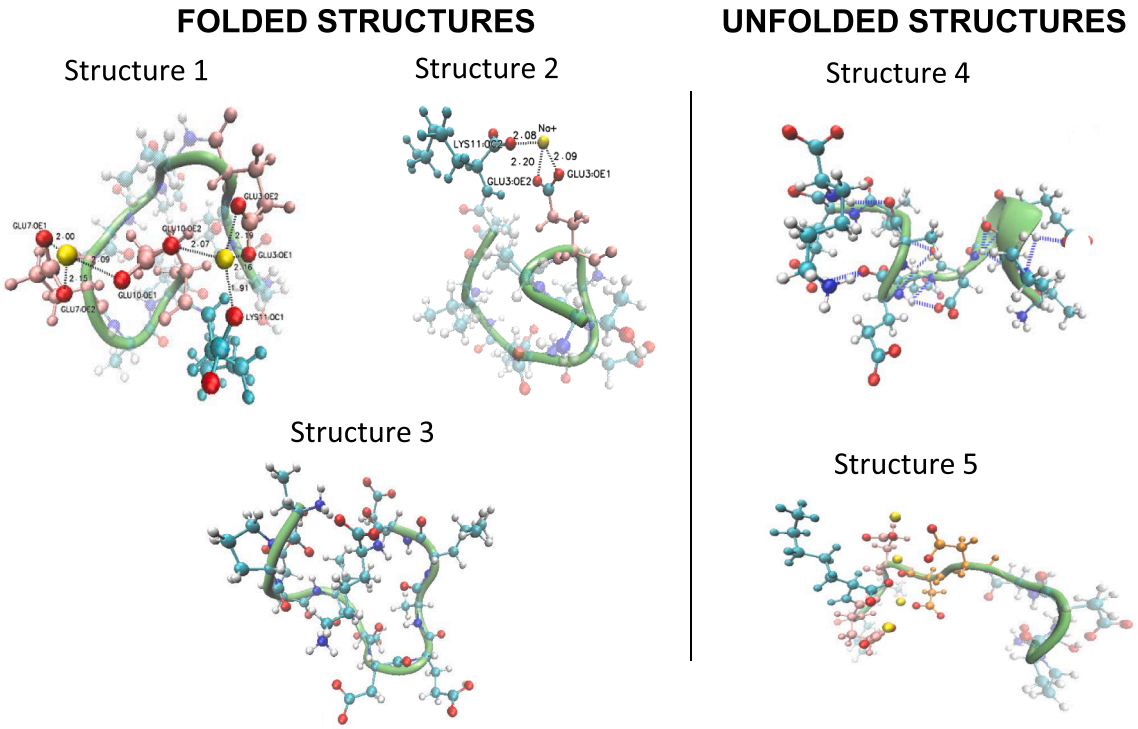


Fig. 3. Stable conformations of the TPEVDDEALEK bovine β-lactoglobulin peptide in solution, found in replica and metadynamics simulations. Structural Na^+ ions are represented as yellow spheres. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hydrolysis. In contrast, long lasting hydrogen bonds with atoms in the peptide bonds between Pro-Glu3 and Asp6-Glu7 for structure 2, and Ala-Leu and Leu-Glu10 for structure 5, are present. These last for almost a nanosecond or more, which suggests they are in an energetically favorable position around the peptide and are unlikely to engage in hydrolysis. Visual inspection of such conformations revealed these are structural water molecules that help stabilise each of the geometric arrangements. A third category of hydrogen bond can be identified with the decay function reaching zero between 408 ± 54 ps and 616 ± 70 ps. Such water molecules might be available for hydrolysis.

3.4. Potential sites for hydrolysis

Several authors agree that for hydrolysis to take place, a water assisted mechanism is preferential as a second water molecule aids with the proton transfer during the reaction, which reduces the energy barrier for the process to occur [11,18,26,31]. Gorb et al. [11] have reported a reactant arrangement between formamide and water molecules for hydrolysis at neutral pH. As listed in Table 1, this reactant structure has an optimised distance between the carbonyl oxygen (C═O) and the hydrogen of water (H_w) of approximately 1.8 Å. Additionally, distances of 1.8 Å to 2.0 Å are reported between the amine hydrogen (NH) and the oxygen of water (O_w). The distance between carbonyl oxygen and water

Table 2
Average number of intramolecular hydrogen bonds per residue.

Residue	Structure 1	Structure 2	Structure 3	Structure 4	Structure 5
THREONINE 1	0.03 ± 0.01	2.35 ± 0.04	2.57 ± 0.03	3.49 ± 0.02	1.10 ± 0.03
PROLINE 2	0.03 ± 0.01	0.00 ± 0.00	0.19 ± 0.01	0.08 ± 0.01	0.05 ± 0.01
GLUTAMATE 3	0.96 ± 0.01	0.48 ± 0.02	1.12 ± 0.01	1.04 ± 0.01	0.62 ± 0.02
VALINE 4	0.00 ± 0.00	0.02 ± 0.01	1.15 ± 0.02	0.94 ± 0.01	0.53 ± 0.02
ASPARTATE 5	0.03 ± 0.01	0.73 ± 0.03	1.04 ± 0.02	2.56 ± 0.02	0.36 ± 0.02
ASPARTATE 6	1.01 ± 0.04	1.70 ± 0.03	0.73 ± 0.02	0.76 ± 0.01	0.37 ± 0.02
GLUTAMATE 7	1.62 ± 0.04	0.09 ± 0.01	0.13 ± 0.01	0.62 ± 0.02	0.09 ± 0.01
ALANINE 8	0.99 ± 0.03	0.08 ± 0.01	1.16 ± 0.02	0.78 ± 0.01	0.01 ± 0.00
LEUCINE 9	0.74 ± 0.02	0.17 ± 0.01	0.12 ± 0.01	0.83 ± 0.01	0.00 ± 0.00
GLUTAMATE 10	0.53 ± 0.02	0.12 ± 0.01	0.11 ± 0.01	0.57 ± 0.01	0.00 ± 0.00
LYSINE 11	1.26 ± 0.03	0.52 ± 0.02	2.35 ± 0.03	0.52 ± 0.02	0.09 ± 0.01
Total per structure	7.20 ± 0.31	6.26 ± 0.07	10.67 ± 0.06	12.19 ± 0.05	3.22 ± 0.05
# of Na^+ ions	2	1	0	0	4

may relate not only to the hydration of the carbonyl bond, a key step for the stepwise formation of an amino-*gem*-diol intermediate [11], but also to a potential pre-protonation of the oxygen atom. Although O-protonation is more commonly described in acid hydrolysis, Antonczak et al. [18] have hypothesised that a similar mechanism may occur in neutral hydrolysis due to interactions with the bulk. The RDF between the carbonyl oxygen and the water hydrogen was calculated for each peptide bond in each of the five structures. The first maximum gives a value of 1.8 Å, in accordance with the value published by Gorb et al [11]. This peak was integrated to give the amount of water molecules around the carbonyl. Fig. 6A shows the coordination number of each carbonyl oxygen plotted against the lifetime of its hydrogen bonds with water. The plot was split into quadrants in order to better visualize the peptide bonds that may be more likely to undergo hydrolysis. A horizontal threshold of one has been set, based on the fact that the carbonyl oxygen needs to hydrogen bond with at least one water molecule for hydrolysis to occur [11,18]. Additionally, the average lifetime of a hydrogen bond on the carbonyl oxygen was calculated in order to set a vertical threshold. Extremely fast or slow lifetimes were not included to not skew the average (Figure S2). This gave a value of approximately 39 ± 4 ps, where any hydrogen bond that breaks sooner than that might not be held around the carbonyl for long enough to react. The upper right quadrant of the plot then represents the peptide bonds which are more likely to hydrolyse. Such peptide bonds include the sequence Glu7-Ala-Leu-Glu10-Lys for folded structures. Two peptide bonds located towards the *N*-terminal also appear in this quadrant and they both belong to unfolded structures: Pro-Glu3 for structure 4, and Val-Asp5 for structure 5. For structure 4, the peptide bond Asp6-Glu7 in the middle of the chain also appears. Although the unfolded structures 4 and 5 rely on intramolecular hydrogen bonding and interaction with cations, respectively, to maintain their conformation, the susceptibility of these specific bonds suggest they are in an exposed position in the chain. Conversely, the lower left quadrant represents bonds that are very unlikely to react, because they may have no hydrogen bonds between water and the carbonyl oxygen and, if they do, they do not last long enough to react. This quadrant is mainly populated by Thr-Pro peptide bonds. It has been previously shown (Table 2) that the threonine residue preferentially hydrogen bonds with the rest of the amino acid chain. That, allied with the hydrophobicity of the proline residue, probably works in tandem to limit the amount of water in their proximity.

To complete the picture of the reactant structure of Gorb et al. [11], the interaction between the amine hydrogen and water oxygen was also investigated. Fig. 6B shows the average distance between the amine hydrogen NH and the water oxygen O_w for each peptide in each structure, along with their associated error bars. The peptide bond between threonine and proline is excluded from this analysis, given that no hydrogen is present due to the cyclic nature of the proline residue. Although amine hydrogens in all peptide bonds are within error of the optimal reported distance (Table 1), albeit in different structures, only one peptide bond has a combination of C═O–H_w and NH–O_w distances that match the geometric parameters reported in literature. This bond is the Ala-Leu peptide bond in the folded structure 3.

Although most peptide bonds do not have the exact combination of geometric parameters as reported by Gorb et al. [11], it is important to note that such parameters were derived for formamide and that the dynamics involved in an amino acid chain like T_{125}PEVDDEALEK$_{135}$ in solution are far more complicated. Nonetheless, based on these characteristics, it is likely that the Ala-Leu bond is a good candidate for the hydrolysis of a folded T_{125}PEVDDEALEK$_{135}$ structure.

As for the reactive structures reported by Antonczak et al. [18] and Pan et al. [26], they relate to the mechanism which involves the simultaneous attack of the carbonyl carbon and the amine nitrogen by water [26,31]. In the case of Antonczak et al. [18], however, the structure is in fact of a pre-reactive hydrogen bonded complex, whereas Pan et al. [26] have optimised the geometry of a reactant. Typical values for the interactions involved in these structures are listed in Table 1. The average minimum distance between carbonyl carbon and water oxygens, and between amine nitrogen and water hydrogen were calculated for each of the peptide bonds for every structure found for the T_{125}PEVDDEALEK$_{135}$ peptide. All O═C–O_W distances fall between 3.0 and 4.0 Å, with most interactions in the interval of 3.4 to 3.7 Å (Fig. 7A). As for the N–H_w distance, values fall between 2.7 and 3.7 Å (Fig. 7B). In every case, the calculated distances are larger than the reported values in literature [18,26]. However, it is not possible to conclude that none of the peptide bonds are prone to undergoing hydrolysis through the aforementioned mechanism as the reactant structures in literature were obtained using smaller model compounds. Because all calculated values are approximately 1 Å larger than literature values, it may be that such elongation is caused by steric effects due to size and composition of the amino acid chain. There is, however, no evidence to say that such elongation hinders hydrolysis.

Because the carbonyl group is a primary site for attack [11,18,26,29,31], the results obtained solely by analysing water dynamics around this group were also considered individually. Assuming that all peptide bonds are within distance to suffer a nucleophilic attack on the carbonyl carbon, the limiting factor thus becomes the hydrogen bond dynamics with the carbonyl oxygen (see Scheme 1). The previously discussed results suggest that the β-lactoglobulin T_{125}PEVDDEALEK$_{135}$ peptide most likely preferentially hydrolyses from the C-terminal inwards, towards the middle of the chain, rather than from the *N*-terminal. However, two peptide bonds towards the *N*-terminal stand out: Pro-Glu3 (structure 4) and Val-Asp5 (structure 5). These may also be candidates for hydrolysis when it comes to unfolded structures. It has been shown that a mechanism involving the carboxyl group of glutamate is enhanced by a proline residue in its *N*-terminal side, as is the case in the T_{125}PEVDDEALEK$_{135}$ sequence. Therefore, the combination of favorable water dynamics around the carbonyl and possible effect of the carboxylic group makes the peptide bond between Pro-Glu3 also likely to spontaneously hydrolyse in neutral water, if the peptide assumes a suitable conformation.

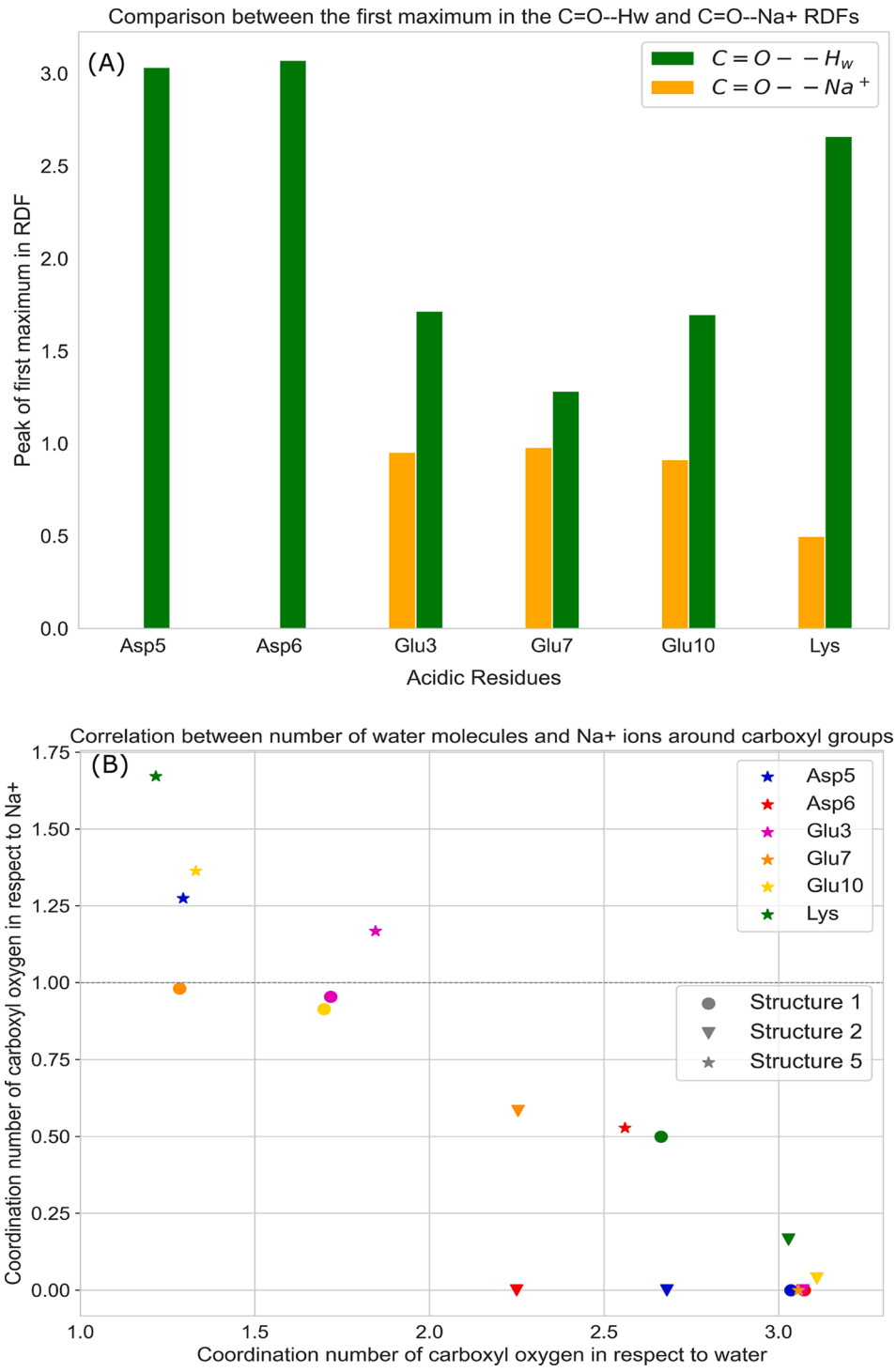


Fig. 4. Correlation between cation proximity and water availability around charged residues. (A) First maximum of the radial distribution functions of water and sodium ions around the carboxyl oxygens in negatively charged residues along the peptide chain for structure 1. (B) Correlation between number of water molecules and sodium ions around carboxylic groups. Only structures that have an active participation of the cation in stabilizing the geometry were included in this analysis. (*single column*).

4. Conclusion

The dynamics of the bovine β-lactoglobulin T_{125}PEVDDEALEK$_{135}$ peptide was investigated in solution at neutral pH. Five stable geometric arrangements were found, three of which rely on interaction with cations to maintain their conformation. Such conformations may be an indication of how the peptide binds to a mineral surface. Moreover, the interaction with cations also limits water availability around certain sites and prevents carboxylic groups in acidic residues from engaging in hydrolysis.

Several mechanisms for hydrolysis are published in literature with differing reactant structures. Some of the most common reported structures have however been selected in an effort to identify potential sites for hydrolysis. The combination of parameters published by Gorb et al. [11] has only been found once in the T_{125}PEVDDEALEK$_{135}$ chain: in the Ala-Leu peptide bond of a folded state, where this bond is more exposed in the loop region of the conformation. As for the parameters published by Pan et al. [26] and Antonczak et al. [18], no optimal combination was found in any of the structures. However, calculated values are approximately 1 Å longer than values in literature. Such elongation might be the result of steric effects caused by amino acid side chains.

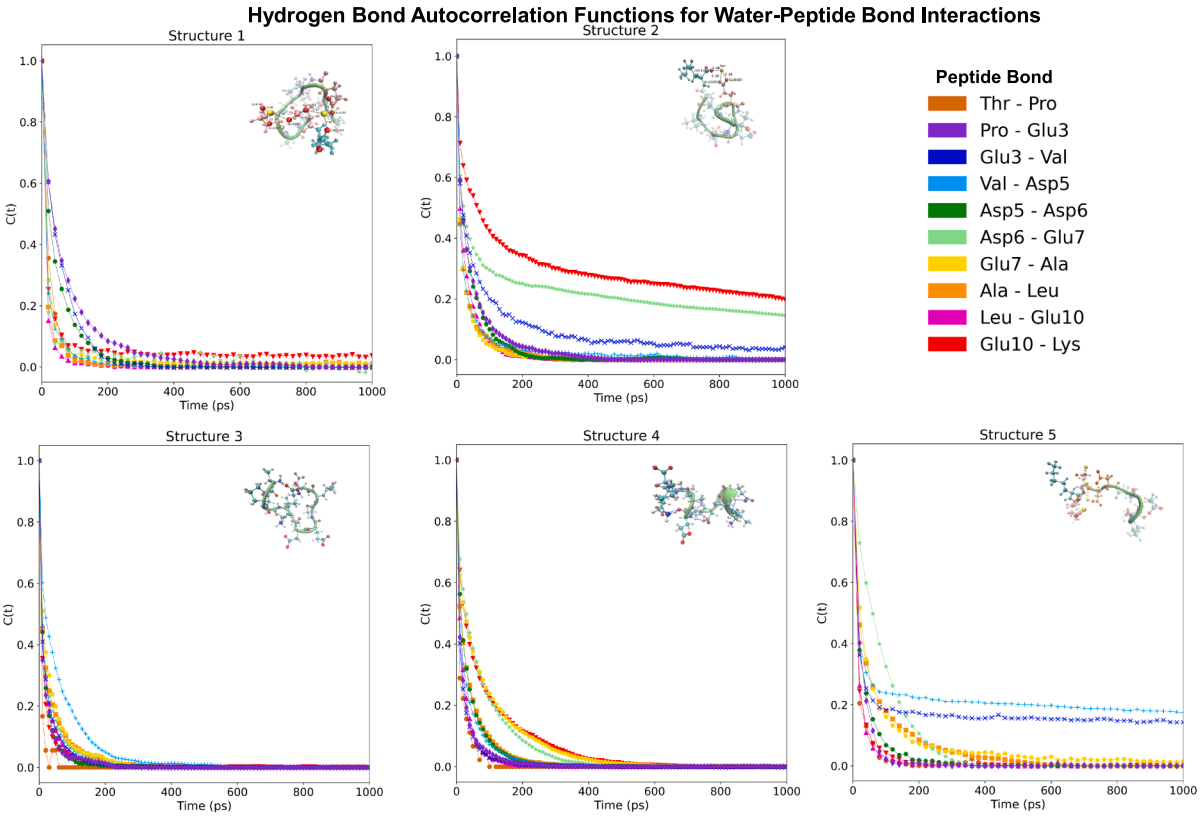


Fig. 5. Hydrogen bond autocorrelation functions (C(t)) for water-peptide bond interactions.

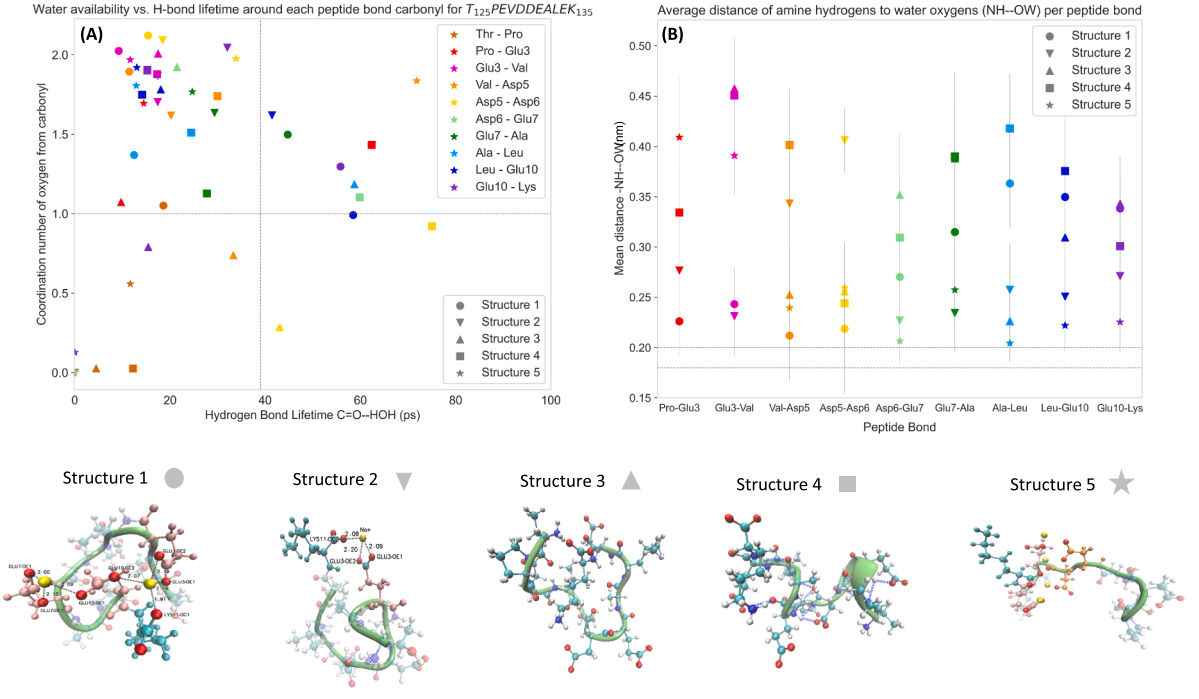


Fig. 6. Geometric parameters investigated according to the reactant structure reported by Gorb et al. [11] (A) Hydrogen bond lifetime vs. coordination number for the carbonyl oxygen in each peptide bond. (B) Average distance of amine hydrogens to water oxygens.

Since the carbonyl group is the primary site for attack, individual results involving solely the carbonyl backbone bond were considered. These suggest that hydrolysis may preferentially occur from the C-terminal inwards for folded structures. To emphasise this further, it was found that the Thr-Pro bond on the *N*-terminal is the one most unlikely to hydrolyse in all investigated scenarios. Nonetheless, two peptide

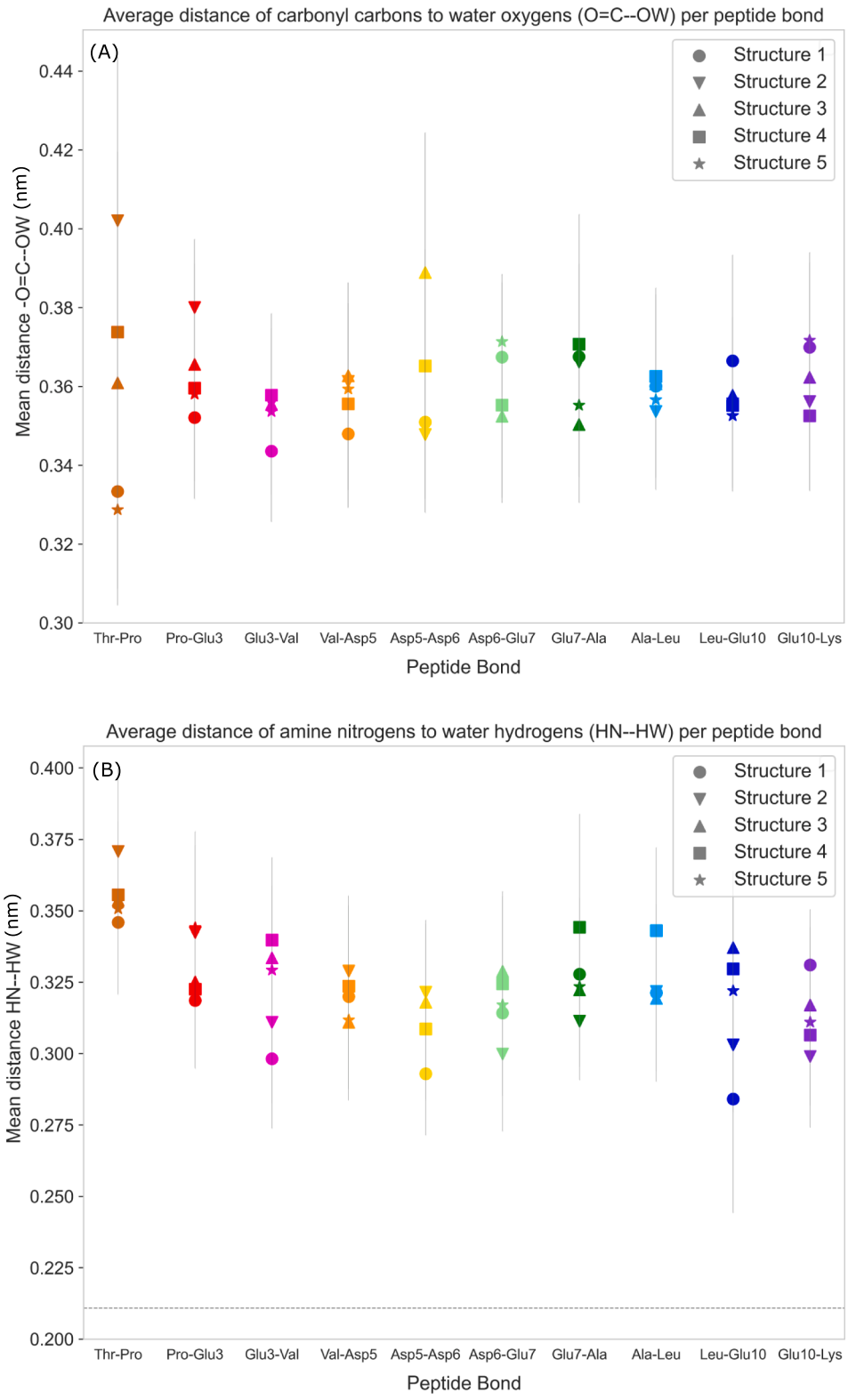


Fig. 7. Average distance of carbonyl carbons to water oxygen (A) and of amine nitrogens to water hydrogens (B) for each peptide bond in each of the five structures found.

bonds stand out towards the *N*-terminal side of the chain, since they are readily available to water molecules in unfolded structures. These are Pro-Glu3 and Val-Asp5 and they may be candidates to reaction mechanisms involving the carboxylic groups in glutamates and aspartates.

Although it is not possible to determine with great accuracy the sites for hydrolysis along an amino acid chain with Molecular Dynamics, an approximate picture of more susceptible bonds was obtained, based on comparison to the published literature and investigations into water dynamics in different peptide conformations. Table 3 highlights the peptide bonds most susceptible to hydrolysis in each of the stable conformations. There is a clear difference between the exposed bonds in unfolded and folded states. Whilst folded structures seem to have higher

Table 3
Summary of peptide bonds susceptible to hydrolysis in each of the T_{125}PEVDDEALEK$_{135}$ conformations.

Type of Strcuture	Conformation	Number of Na^+ Ions	Vulnerable Peptide Bonds
FOLDED	Structure 1	1	Glu7-Ala Leu-Glu10-Lys
	Structure 2	2	Leu-Glu10
	Structure 3	0	Ala-Leu
UNFOLDED	Structure 4	0	Pro-Glu3, Asp6-Glu7
	Structure 5	4	Val-Asp5

vulnerability in bonds from the middle of the chain towards the C-terminal, unfolded structures seem to be more vulnerable towards the *N*-terminal of the peptide chain. A correlation between the number of structural cations and vulnerable sites is, however, not clear. To get a more complete picture in the survival of β-lactoglobulin peptides, it is paramount to compare the results obtained here with those in the presence of a mineral. This will be the focus of further studies into the preservation of β-lactoglobulin peptides in the archaeological record.

CRediT authorship contribution statement

Beatriz Fonseca: Methodology, Investigation, Formal analysis, Validation, Conceptualization, Writing – original draft, Visualization. **Colin L. Freeman:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Matthew J. Collins:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Beatriz Fonseca is currently funded by the Danish National Research Fund DNRF128 and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801199. Colin Freeman and Matthew Collins have also received support from the Danish National Research Fund DNRF128. In addition, Colin Freeman would like to thank funding from an EPSRC Programme Grant (Grant EP/R018820/1). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemphys.2022.111602.

References

[1] S. Wilkin, A. Ventresca Miller, R. Fernandes, R. Spengler, W.-T. Taylor, D.R. Brown, D. Reich, D.J. Kennett, B.J. Culleton, L. Kunz, C. Fortes, A. Kitova, P. Kuznetsov, A. Epimakhov, V.F. Zaibert, A.K. Outram, E. Kitov, A. Khokhlov, D. Anthony, N. Boivin, Dairying enabled Early Bronze Age Yamnaya steppe expansions, Nature 598 (7882) (2021) 629–633, https://doi.org/10.1038/s41586-021-03798-4.
[2] S. Wilkin, A. Ventresca Miller, W.T.T. Taylor, B.K. Miller, R.W. Hagan, M. Bleasdale, A. Scott, S. Gankhuyg, A. Ramsøe, S. Uliziibayar, C. Trachsel, P. Nanni, J. Grossmann, L. Orlando, M. Horton, P.W. Stockhammer, E. Myagmar, N. Boivin, C. Warinner, J. Hendy, Dairy pastoralism sustained eastern Eurasian steppe populations for 5,000 years, Nat Ecol Evol. 4 (3) (2020) 346–355.
[3] S. Charlton, A. Ramsøe, M. Collins, O.E. Craig, R. Fischer, M. Alexander, C. F. Speller, New insights into Neolithic milk consumption through proteomic analysis of dental calculus, Archaeol Anthropol Sci. 11 (11) (2019) 6183–6196.
[4] C. Warinner, J. Hendy, C. Speller, E. Cappellini, R. Fischer, C. Trachsel, J. Arneborg, N. Lynnerup, O.E. Craig, D.M. Swallow, A. Fotakis, R.J. Christensen, J. V. Olsen, A. Liebert, N. Montalva, S. Fiddyment, S. Charlton, M. Mackie, A. Canci, A. Bouwman, F. Rühli, M.T.P. Gilbert, M.J. Collins, Direct evidence of milk consumption from ancient human dental calculus, Sci Rep. 4 (1) (2015), https://doi.org/10.1038/srep07104.
[5] J. Hendy, C. Warinner, A. Bouwman, M.J. Collins, S. Fiddyment, R. Fischer, R. Hagan, C.A. Hofman, M. Holst, E. Chaves, L. Klaus, G. Larson, M. Mackie, K. McGrath, A.Z. Mundorff, A. Radini, H. Rao, C. Trachsel, I.M. Velsko, C.F. Speller, Proteomic evidence of dietary sources in ancient dental calculus, Proc Biol Sci. 285 (1883) (2018) 20180977, https://doi.org/10.1098/rspb.2018.0977.
[6] Nair B. Personal Communication. 2021.
[7] B. Demarchi, S. Hall, T. Roncal-Herrero, C.L. Freeman, J. Woolley, M.K. Crisp, J. Wilson, A. Fotakis, R. Fischer, B.M. Kessler, R. Rakownikow Jersie-Christensen, J.V. Olsen, J. Haile, J. Thomas, C.W. Marean, J. Parkington, S. Presslee, J. Lee-Thorp, P. Ditchfield, J.F. Hamilton, M.W. Ward, C.M. Wang, M.D. Shaw, T. Harrison, M. Domínguez-Rodrigo, R.DE. MacPhee, A. Kwekason, M. Ecker, L. Kolska Horwitz, M. Chazan, R. Kröger, J. Thomas-Oates, J.H. Harding, E. Cappellini, K. Penkman, M.J. Collins, Protein sequences bound to mineral surfaces persist into deep time, Elife. 5 (2016), https://doi.org/10.7554/eLife.17092.
[8] M. Dal Peraro, P. Ruggerone, S. Raugei, F.L. Gervasio, P. Carloni, Investigating biological systems using first principles Car-Parrinello molecular dynamics simulations, Curr Opin Struct Biol. 17 (2) (2007) 149–156.
[9] J. Kang, Y. Hagiwara, M. Tateno, Biological applications of hybrid quantum mechanics/molecular mechanics calculation, J Biomed Biotechnol. 2012 (2012) 1–11.
[10] Q. Cui, Perspective: Quantum mechanical methods in biochemistry and biophysics, J Chem Phys. 145 (14) (2016) 140901, https://doi.org/10.1063/1.4964410.
[11] L. Gorb, A. Asensio, I. Tuñón, M.F. Ruiz-López, The mechanism of formamide hydrolysis in water from ab initio calculations and simulations, Chemistry. 11 (22) (2005) 6743–6753.
[12] A. Radzicka, R. Wolfenden, Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases, J Am Chem Soc. 118 (26) (1996) 6105–6109.
[13] D. Kahne, W.C. Still, Hydrolysis of a peptide bond in neutral water, J Am Chem Soc. 110 (22) (1988) 7529–7534.
[14] J.I. Mujika, J.M. Mercero, X. Lopez, Water-promoted hydrolysis of a highly twisted amide: rate acceleration caused by the twist of the amide bond, J Am Chem Soc. 127 (2005) 4445–4453.
[15] B. Pan, M.S. Ricci, B.L. Trout, Molecular mechanism of acid-catalyzed hydrolysis of peptide bonds using a model compound, J Phys Chem B. 114 (2010) 4389–4399.
[16] B. Wang, Z. Cao, Mechanism of acid-catalyzed hydrolysis of formamide from cluster-continuum model calculations: concerted versus stepwise pathway, J Phys Chem A. 114 (49) (2010) 12918–12927.
[17] C. Oliyai, R.T. Borchardt, Chemical pathways of peptide degradation. VI. Effect of the primary sequence on the pathways of degradation of aspartyl residues in model hexapeptides, Pharm Res. 11 (1994) 751–758.
[18] S. Antonczak, M.F. Ruiz-Lopez, J.L. Rivail, Ab Initio Analysis of Water-Assisted Reaction Mechanisms in Amide Hydrolysis, J Am Chem Soc. 116 (9) (1994) 3912–3921.
[19] D. Cheshmedzhieva, S. Ilieva, B. Hadjieva, B. Galabov, The mechanism of alkaline hydrolysis of amides: a comparative computational and experimental study of the hydrolysis of N -methylacetamide, N -methylbenzamide, and acetanilide, J Phys Org Chem. 22 (6) (2009) 619–631.
[20] A.C. Tsipis, D.N. Gkarmpounis, C.E. Kefalidis, E.M. Papamichael, L.G. Theodorou, Modeling the cysteamine catalyzed cysteine proteinases using DFT: mechanistic insights into the hydrolysis of acetyl- p -nitroanilide, New J Chem. 37 (12) (2013) 4061, https://doi.org/10.1039/c3nj00769c.
[21] M.G. Friedrich, Z. Wang, K.L. Schey, R.J.W. Truscott, Spontaneous Cleavage at Glu and Gln Residues in Long-Lived Proteins, ACS Chem Biol. 16 (11) (2021) 2244–2254, https://doi.org/10.1021/acschembio.1c00379.
[22] T.M. Lauer, G.P.F. Wood, D. Farkas, H.A. Sathish, H.S. Samra, B.L. Trout, Molecular Investigation of the Mechanism of Non-Enzymatic Hydrolysis of Proteins and the Predictive Algorithm for Susceptibility, Biochemistry 55 (23) (2016) 3315–3328.
[23] Y.i. Sun, M. Frenkel-Pinter, C.L. Liotta, M.A. Grover, The pH dependent mechanisms of non-enzymatic peptide bond cleavage reactions, Phys Chem Chem Phys. 22 (1) (2020) 107–113.
[24] R.M. Smith, D.E. Hansen, The pH-Rate Profile for the Hydrolysis of a Peptide Bond, J Am Chem Soc. 120 (35) (1998) 8910–8913.
[25] J.P. Krug, P.L.A. Popelier, R.F.W. Bader, Theoretical study of neutral and of acid and base-promoted hydrolysis of formamide, J Phys Chem. 96 (19) (1992) 7604–7616.
[26] B. Pan, M.S. Ricci, B.L. Trout, A molecular mechanism of hydrolysis of peptide bonds at neutral pH using a model compound, J Phys Chem B. 115 (19) (2011) 5958–5970.
[27] R.B. Martin, Free energies and equilibria of peptide bond hydrolysis and formation, Biopolymers: Original Research on Biomolecules. 45 (5) (1998) 351–353.
[28] K. Marshall-Bowman, S. Ohara, D.A. Sverjensky, R.M. Hazen, H.J. Cleaves, Catalytic peptide hydrolysis by mineral surface: Implications for prebiotic chemistry, Geochim Cosmochim Acta. 74 (20) (2010) 5852–5861.
[29] M. Cascella, S. Raugei, P. Carloni, Formamide hydrolysis investigated by multiple-steering ab initio molecular dynamics, J Phys Chem B. 108 (1) (2004) 369–375.

[30] W. Sang-aroon, V. Amornkitbamrung, V. Ruangpornvisuti, A density functional theory study on peptide bond cleavage at aspartic residues: direct vs cyclic intermediate hydrolysis, J Mol Model. 19 (12) (2013) 5501–5513.
[31] D. Zahn, On the role of water in amide hydrolysis, European J Org Chem. 2004 (19) (2004) 4020–4023.
[32] Z. Wang, M.G. Friedrich, R.J.W. Truscott, K.L. Schey, Cleavage C-terminal to Asp leads to covalent crosslinking of long-lived human proteins, Biochim Biophys Acta: Proteins Proteomics. 1867 (9) (2019) 831–839.
[33] T. Nakayoshi, K. Kato, S. Fukuyoshi, O. Takahashi, E. Kurimoto, A. Oda, Molecular Mechanisms of Succinimide Formation from Aspartic Acid Residues Catalyzed by Two Water Molecules in the Aqueous Phase, Int J Mol Sci. 22 (2) (2021) 509, https://doi.org/10.3390/ijms22020509.
[34] C.E. Voorter, W.A. de Haard-Hoekman, P.J. van den Oetelaar, H. Bloemendal, W. W. de Jong, Spontaneous peptide bond cleavage in aging alpha-crystallin through a succinimide intermediate, J Biol Chem. 263 (35) (1988) 19020–19023.
[35] Geiger T, Clarke S. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. J Biol Chem. 1987;262: 785–794.
[36] R.C. Stephenson, S. Clarke, Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins, J Biol Chem. 264 (11) (1989) 6164–6170.
[37] K. Kuwata, M. Hoshino, V. Forge, S. Era, C.A. Batt, Y. Goto, Solution structure and dynamics of bovine beta-lactoglobulin A, Protein Sci. 8 (1999) 2541–2545.
[38] Lindahl, Abraham, Hess, Spoel V der. GROMACS 2020 Source code. 2020. doi: 10.5281/zenodo.3562495.
[39] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1-2 (2015) 19–25.
[40] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, Comput Phys Commun. 91 (1-3) (1995) 43–56.
[41] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J Chem Theory Comput. 11 (8) (2015) 3696–3713.
[42] P. Mark, L. Nilsson, Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K, J Phys Chem A. 105 (43) (2001) 9954–9960.
[43] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, J Phys Chem. 91 (24) (1987) 6269–6271.
[44] A. Laio, M. Parrinello, Escaping free-energy minima, Proc Natl Acad Sci U S A. 99 (20) (2002) 12562–12566.
[45] G. Bussi, A. Laio, Using metadynamics to explore complex free-energy landscapes, Nature Reviews Physics. 2 (4) (2020) 200–212.
[46] G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird, Comput Phys Commun. 185 (2) (2014) 604–613.
[47] PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. Nat Methods. 2019;16: 670–673.
[48] A. Luzar, D. Chandler, Hydrogen-bond kinetics in liquid water, Nature 379 (6560) (1996) 55–57.
[49] A. Luzar, Resolving the hydrogen bond dynamics conundrum, J Chem Phys. 113 (23) (2000) 10663–10675.