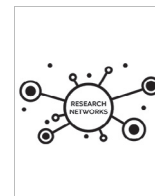




ELSEVIER

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

Early stages of misfolding of PAP248–286 at two different pH values: An insight from molecular dynamics simulations

Nikhil Agrawal ^{a,b,*}, Emilio Parisini ^{a,c,*}^a Latvian Institute of Organic Synthesis, Aizkraukles 21, LV, Riga 1006, Latvia^b College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa^c Department of Chemistry “G. Ciamician”, University of Bologna, Bologna, Italy

ARTICLE INFO

Article history:

Received 7 June 2022

Received in revised form 26 August 2022

Accepted 26 August 2022

Available online 3 September 2022

Keywords:

HIV

PAP248–286

MD simulations

amyloid

ABSTRACT

PAP248–286 peptides, which are highly abundant in human semen, aggregate and form amyloid fibrils that enhance HIV infection. Previous experimental studies have shown that the infection-promoting activity of PAP248–286 begins to increase well before amyloid formation takes place and that pH plays a key role in the enhancement of PAP248–286-related infection. Hence, understanding the early stages of misfolding of the PAP248–286 peptide is crucial. To this end, we have performed 60 independent MD simulations for a total of 24 μ s at two different pH values (4.2 and 7.2). Our data shows that early stages of misfolding of the PAP248–286 peptide is a multistage process and that the first step of the process is a transition from an “I-shaped” structure to a “U-shaped” structure. We further observed that the structure of PAP248–286 at the two different pH values shows significantly different features. At pH 4.2, the peptide has less intra-molecular H-bonds and a reduced α -helical content than at pH 7.2. Moreover, differences in intra-peptide residues contacts are also observed at the two pH values. Finally, free energy landscape analysis shows that there are more local minima in the energy surface of the peptide at pH 7.2 than at pH 4.2. Overall, the present study elucidates the early stages of misfolding of the PAP248–286 peptide at the atomic level, thus possibly opening new avenues in structure-based drug discovery against HIV infection.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

HIV has infected more than 75 million people since its inception and, so far, has claimed the lives of more than 36 million people worldwide [1,2]. Human semen is one of the most important vectors for HIV transmission and women are at high risk of HIV through unprotected vaginal intercourse [3–5]. The one very crucial natural factor that has been identified to play a key role in HIV transmission is fragments of the prostatic acidic phosphatase (PAP) protein [6]. The region of PAP from amino acid residue number 248 to 286 (PAP248–286) is naturally present in high quantity in the human semen. Studies have reported that the PAP248–286 peptide could be responsible for an increase in the rate of infection of HIV by up to 400,000-fold [7,8]. PAP248–286 peptides aggregate and form amyloid fibrils termed Semen-derived enhancers of viral

infection (SEVI). The highly cationic nature of the SEVI fibrils facilitates their binding to the anionic membranes of HIV virions and cells, which leads to an increase in virion cell attachment and fusion [9–11]. In the recent times, several different approaches have been taken to reduce the PAP248–286-mediated rate of HIV infection [12]. These include (i) inhibition of the transformation of PAP248–286 monomers into amyloids, (ii) hindering of the cleavage of the PAP248–286 peptide from the PAP protein, (iii) perturbation of the structure of the SEVI fibrils, and (iv) neutralization of the overall charge of the SEVI fibrils. Out of the above-mentioned approaches, inhibition of early stages of misfolding of PAP248–286 monomers could be very promising, as it has been reported to similar amyloidogenic peptides that are related to Alzheimer's and Parkinson's diseases [13,14]. While the biological activity of PAP248–286 is mostly reported to its aggregated amyloid form, studies have shown that the monomeric form of PAP248–286 is also biologically active [6,15].

Previous studies using *in vitro* and *in vivo* experiments reported that PAP248–286 shows HIV viral enhancement activity mainly at the acidic vaginal environment (pH \sim 4.2) [7,16]. Brender *et al.*

* Corresponding authors at: Latvian Institute of Organic Synthesis, Aizkraukles 21, LV, Riga 1006, Latvia.

E-mail addresses: nikhil.08oct@gmail.com (N. Agrawal), emilio.parisini@osi.lv (E. Parisini).

[16] suggested that the high activity of PAP248–286 in the acidic vaginal environment may be due to its two histidine residues, which are likely to be charged at the acidic pH \sim 4.2, but not at the neutral pH \sim 7.2. Interestingly, other than showing HIV infection enhancement at acidic pH, the PAP248–286 peptide has also been reported to display antimicrobial activity through bacterial agglutination at neutral pH [15,17].

Several experimental studies, mostly using the NMR technique, have been performed to understand the monomeric structure of the full-length PAP248–286 peptide or short regions thereof in different environments. It has been reported that in aqueous and SDS micelles environments the structure of the PAP248–286 peptide is highly disordered and can be divided into three parts: i) a N-terminal disordered region, ii) a stable central part, which forms an α -helix, iii) a C-terminal flexible region, which adopts a 3_{10} / α -helix conformation [18–23]. In trifluoroethanol (TFE), PAP248–286 features a largely α -helical conformation. Indeed, in 30 % TFE, almost all the central part of the peptide adopts a helical conformation while in 50 % TEF the whole peptide, with the exception of its C-terminal end, displays a α -helical conformation. The conformation of PAP248–286 in 50 % TEF solution is similar to the PAP248–286 conformation in the crystal structure of the full-length PAP protein (PDB id: 1CVI) [18].

Misfolding and aggregation of proteins to form amyloids is a well-known phenomenon, particularly but not exclusively in the context of neurodegenerative diseases such as Alzheimer's disease (AD) [24–26] and Parkinson's disease (PD) [27]. In fact, there are almost 50 other human diseases that are related to the formation of amyloid [28]. Understanding the early stages of misfolding of amyloid-forming proteins may help identify amino acid residues that play a key role in the process, and aid the design of potential aggregation inhibitors [29]. Molecular dynamics (MD) simulations are widely used to investigate the misfolding and aggregation process of amyloid proteins. MD simulations provide an atomic-level insight into the transformation of α -helix-rich structures into a disordered or β -rich structures that is currently not attainable by experimental techniques [30]. In recent times, several MD simulations studies have been conducted on the PAP248–286 peptide; however, all these studies were done to investigate the binding of small molecules to PAP248–286 [31,32] or its interaction with biological membrane lipid [15]. To the best of our knowledge, no MD simulation study has been conducted so far to investigate the early stages of misfolding of PAP248–286 at the acidic pH characteristic of the vaginal environment and at neutral pH. The overall aim of the present study is to investigate and compare the mechanism of the early stages of misfolding of PAP248–286 at two different pH values. To this end, we have performed 60 independent all-atom MD simulations of PAP248–286 at pH 4.2 and at pH 7.2 in explicit water environment.

2. Methods

In the present study, we used the NMR structure of PAP248–286 (PDB id: 2L77 and BMRB id: 17346) [18,19] for our molecular dynamics simulations. This PAP248–286 peptide structure closely resembles the conformation of the same amino acid residue portion in the crystal structure of the full-length PAP (PDB id: 1CVI) [33]. Residue numbering of PDB id: 2L77 starts from 1 and ends at 39. At pH 4.2, His3 and His23 of PAP248–286 are protonated, leading to an overall charge of +8 on the peptide, compared to a +6 charge at pH 7.2 (Fig. 1A and B). We used the H⁺⁺ server [34] to add hydrogen atoms to PAP248–286 at pH 4.2 to account for the acidic vaginal environment and at pH 7.2 to represent the physiological environment. Supplementary Table 1 shows predicted pK_a value by H⁺⁺ server. In the present study, we used the

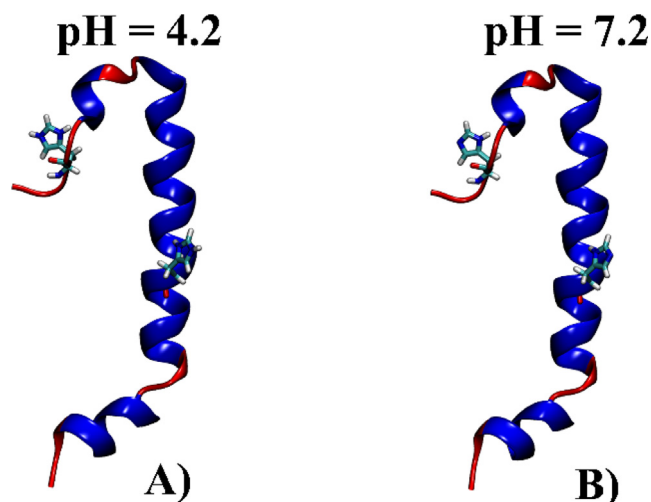


Fig. 1. A) PAP248–286 structure with both histidine residues protonated (pH 4.2), b) PAP248–286 NMR structure (PDB id = 2L77) at pH 7.2, with both histidine residues deprotonated. The helical portion of the peptide is shown in blue color, while the unstructured portions of the peptide are shown in red color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

charmm36m force field [35], which is an updated version of the charmm36 force field [36]. The charmm36m force field has shown improved accuracy in generating conformational ensembles for intrinsically disordered peptides and proteins [35,37]. Our simulation system contains a total of 12,031 water molecules at pH 4.2 and 12,039 water molecules at pH 7.2. To neutralize the overall charge, 8Cl⁻ ions were added in the pH 4.2 system and 6Cl⁻ ions were added in the pH 7.2 system. Both systems were energy minimized using the steepest descent algorithm, followed by two sequential 500 picoseconds (ps) equilibration simulations using the canonical (NVT) and isobaric – isothermic (NPT) ensemble, and production simulations were performed using the NPT ensemble. During energy minimization and the first equilibration step, restraints of 400 and 40 kJ mol⁻¹ nm⁻² were employed on backbone and sidechain atoms of PAP248–286, respectively. In the second equilibration step, restraints were maintained only on backbone atoms. In the production runs, no restraints were employed on protein atoms. The LINCS algorithm was used for constraining the bond lengths from heavy atoms to hydrogen and the SETTLE algorithm was employed for water molecule bond length constraints. The Particle mesh Ewald (PME) algorithm was used for long-range electrostatic interactions, while for short-range van der Waals (vdW) and columbic interactions a cut-off of 12 Å was used. The Nose-Hoover algorithm was used for temperature coupling and the Parrinello – Rahman algorithm was employed for pressure coupling. A 1 ps time constant was used for both temperature and pressure coupling. All production simulations were performed at a pressure of 1 bar and at a temperature of 310.15 K. A leap-frog algorithm was used for integrating Newton's equations of motion. A total of 60 simulations were performed (30 simulations at each pH value) using different initial velocities randomly generated by the GROMACS package. Simulations were performed using GROMACS version 2020.1.

3. Analysis details

Time evolution of backbone RMSD of PAP248–286 was calculated with respect to the NMR structure. Hydrogen (H) bonds analysis was performed using the GROMACS [38] hbond program and employing the default parameters (donor and acceptor atoms dis-

tance cut-off 3.5 Å, and hydrogen-donor-acceptor angle cut-off 30°). The transition of PAP248-286 from an “I-shaped” structure to a “U-Shaped” structure was calculated using the GROMACS [38] angle program and the C α atoms of Val15, Thr28 and Leu36 were used for angle calculation. Residue contact maps, total interaction time, interaction types, and native contacts were calculated with the CONTACT ANalysis (CONAN) tool [39] using backbone atoms and rcut was set at 10 Å, rinter and rhigh-inter were set at 5 Å as used in the reference paper of CONAN tool [39]. Above the rcut value, any pair of residues that does not have at least one atom in this range is not considered. Contact is formed between a pair of residues under rinter and interactions break above rhigh-inter cut-off value [40]. The interaction type lifetime cut-off was set to 0.50 (representing only interactions that are physically relevant and last at least 50 % of the trajectory, i.e. 6 μ s or above). Free energy landscape (FEL) was constructed with the GROMACS [38] sham program using RMSD (backbone atoms) and the fraction of native contacts (Q) as two order parameters. The initial NMR structure of PAP248-286 was used as reference for RMSD and Q calculation. BitClust program [41] was used for the clustering of PAP248-286 structures with a RMSD cut-off of 5 Å. Secondary structure analysis of PAP248-286 was performed using the Dictionary of Secondary Structure of Proteins program (DSSP) [42]. GROMACS rmsf program was used for root mean square fluctuation (RMSF) calculation. For analysis of RMSD boxplot, ROG boxplot H-Bonds length distribution, H-Bonds angle distribution, residue contact maps, total interaction time, interaction types, native contacts, FEL and

cluster analyses all 30 trajectories were combined to make a single trajectory (12 μ s) at both pH values.

4. Results and discussion

4.1. Effect of pH on conformation transition of PAP248-286

RMSD is the most used method to provide a quantitative measure of the similarity between two superimposed atomic coordinates [43]. To investigate the effect of pH on the change in conformation of the PAP248-286 peptide we have plotted the time evolution of its backbone RMSD for all trajectories with respect to its initial structure, whereas to visualize the distribution of RMSD values we have used boxplots (Fig. 2). Time evolution of backbone RMSD shows that large conformational changes take place in PAP248-286 at both pH values. However, at pH 4.2 we observe a larger and a faster conformational change in the structure of the peptide compared to pH 7.2. (Fig. 2A and B).

Fig. 2C provides a further insight into the RMSD distribution, showing that at pH 4.2 the majority of the RMSD values are within the range \sim 8.8 Å to \sim 11.8 Å while at pH 7.2 they are mostly comprised between \sim 7.8 Å and \sim 10.8 Å. Finally, to assess PAP248-286 change in compactness, we calculated the radius of gyration (ROG) of the peptide (Fig. 2D). ROG values at pH 4.2 are mainly distributed between \sim 11 Å to \sim 14.8 Å, while at pH 7.2 they are between \sim 10 Å to \sim 13.8 Å. We calculated average RMSF of each residue at both pH values (Fig. S1): for each residue, RMSF value

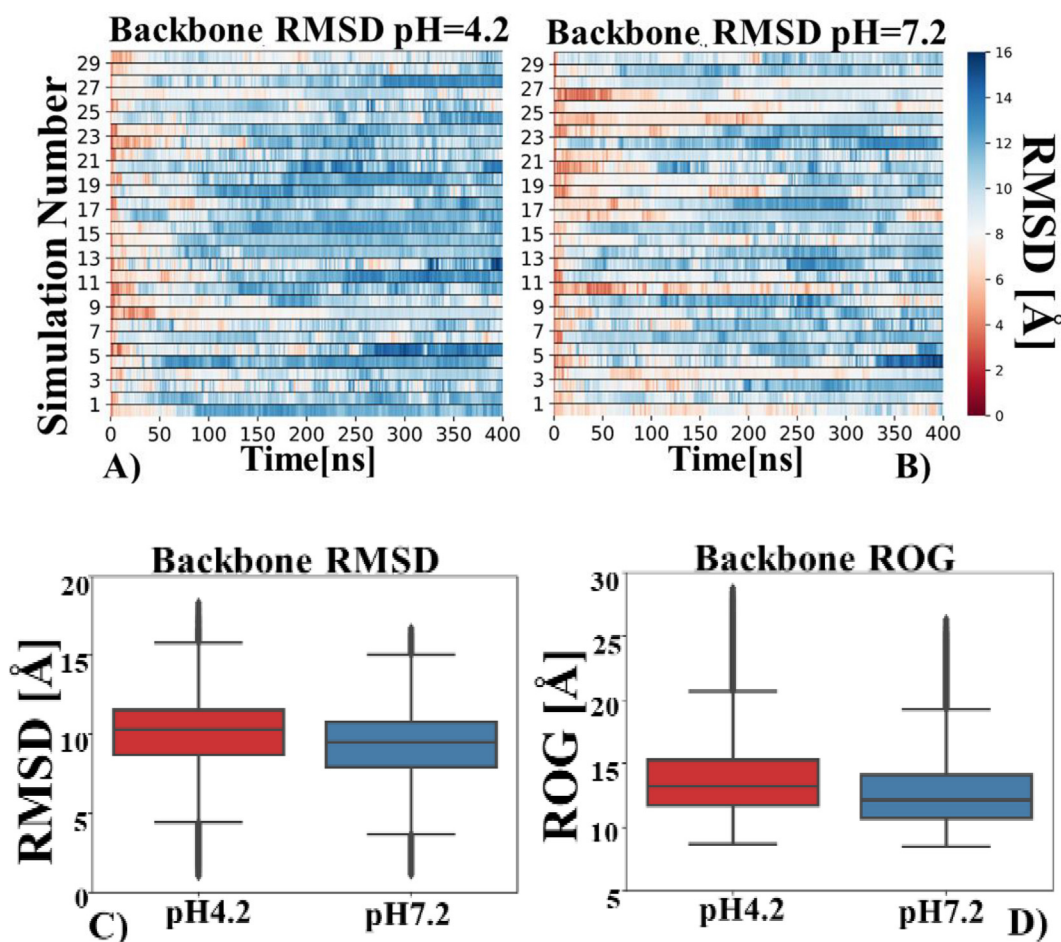


Fig. 2. A) Time evolution of the backbone RMSD of PAP248-286 at pH 4.2. B) Time evolution of the backbone RMSD of PAP248-286 at pH 7.2. More blue color represents high RMSD value. C) Backbone RMSD boxplots at pH 4.2 and pH 7.2. D) Radius of gyration boxplot at pH 4.2 and pH 7.2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

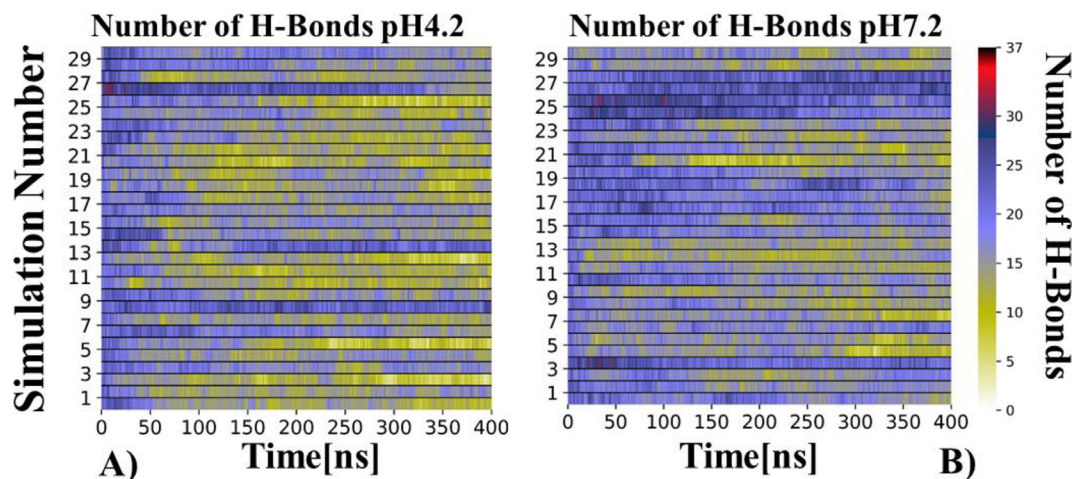


Fig. 3. A) Time evolution of the number of hydrogen bonds at pH 4.2. B) Time evolution of the number of hydrogen bonds at pH 7.2. Blue color indicates a higher number of hydrogen bonds while yellow color indicates a lower number of hydrogen bonds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was slightly higher at pH 4.2 than at pH 7.2. Overall, this quantitative data suggests that, relative to its initial conformation, there are more conformational changes in the PAP248-286 peptide at pH 4.2 than at pH 7.2.

4.2. Intra-peptide hydrogen bonding

Hydrogen bonds play a key role in protein/peptide stability as they provide the majority of directional interactions and help protein/peptide to adopt stable secondary structures such as α -helices and β -sheets [44]. Loss of hydrogen bonds often leads to enhanced flexibility in protein/peptide and may promote transition to unfolded/misfolded state [45]. To evaluate the effect of pH on the change of intra-molecular hydrogen bonds in PAP248-286, we have calculated time evolution H-bonds at the two pH values (Fig. 3). Interestingly, while we see a loss of intra-molecular hydrogen bonds at both pH values, we observe that more H-bonds are lost at pH 4.2 and that this occurs in more trajectories (Fig. 3A). To quantify the total number of hydrogen bonds, we have calculated an average number of hydrogen bonds per timeframe for the last 100 ns for each trajectory and then we have calculated its average value and standard deviation over the whole set of runs. At pH 4.2, there are $\sim 14.01 \pm 3.47$ hydrogen bonds per frame in the last 100 ns, while at pH 7.2 their number changes to $\sim 16.32 \pm 3.32$.

Hence, on average in the last 100 ns there are ~ 2.31 more hydrogen bonds per timeframe of trajectory at pH 7.2 than at pH 4.2. To observe the effect of pH on the lengths and the bond angles of the hydrogen bonds, we have plotted the distribution of these two parameters for all the trajectories combined at both pH values. Fig. 4 shows that there are noticeable differences in the distribution of hydrogen bond lengths and relative frequency of bond angle. At pH 7.2, distribution frequency of bond length at around bond length value ~ 2.95 to 3.0 Å is higher than at pH 4.2. Additionally, relative frequencies of hydrogen bond angles $>10^\circ$ is also higher at pH 7.2. Taken together, this data suggests that at pH 4.2 the peptide is more flexible and unstable than at pH 7.2.

4.3. Transition of PAP248-286 towards a “U-shaped” conformation at pH 4.2 and pH 7.2.

The literature has reported that the first step in the multistage process of amyloid fibril formation is the misfolding of the monomeric peptide and that, in some amyloidogenic species, misfolded monomers may induce conformational changes in other monomeric peptides [46,47]. Previous MD simulations studies have suggested that misfolding in an amyloid monomer starts with the formation of a turn, such as for instance in the α -Synuclein (α S) peptide, where turn formation takes place in the four-residue

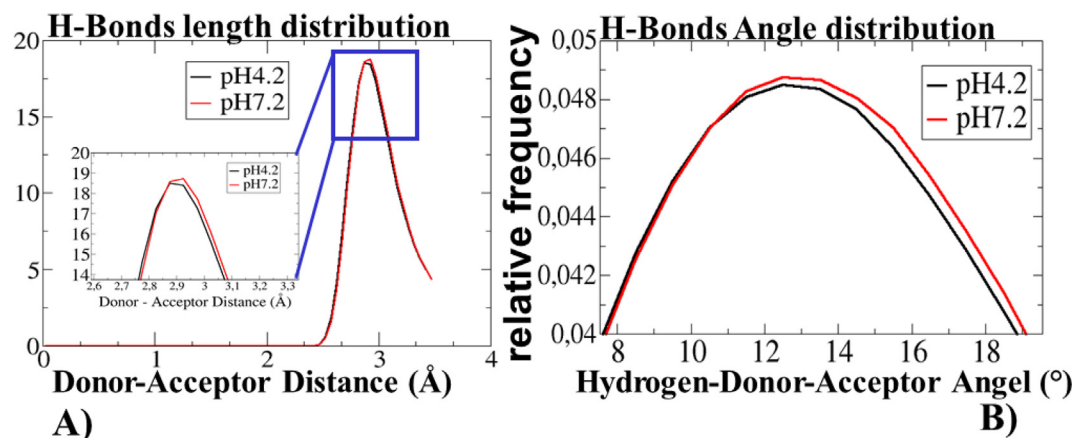


Fig. 4. A) Distribution of hydrogen bond length at pH 4.2 and pH 7.2, B) Relative frequency of hydrogen bond angles at pH 4.2 and pH 7.2.

region Thr44–Gly47 [48], and in the A β peptide, where the most prevalent region for turn formation is between Val24 and Lys28 [29]. In amyloid monomer peptides, turn formation causes a change in conformation from an “I-shaped” structure to a “U-shaped” structure. To investigate the transition of the PAP248–286 peptide structure from “I-Shaped” to “U-shaped” (Fig. 5) and to understand the effect of pH on this transition, we have calculated an angle of bending (see the method section for more details) for all simulation trajectories. Bending angle $\leq 90^\circ$ corresponds to a “U-shaped” topology. We observed that, at pH 4.2, turn formation takes place in all 30 trajectories and that in 27 of them (bending angle $\leq 90^\circ$, green color in Fig. 5A) it remains until the end of the simulation time. At pH 7.2, turn formation also takes place in all 30 trajectories, remaining until the end of the simulation time in 23 of them (Fig. 5B). The model of the PAP248–286 structure at different time points from one of the representative simulations at both pH values is shown in Fig. 6.

To analyze the region involved in turn formation, we performed a visual inspection of all trajectories at the two different pH values. At both pH values, turn formation is most prevalent in the residue region 26RATQ29, albeit with a significantly different frequency. At pH 4.2 and at pH 7.2, turn formation takes place in this region in 21 and in 15 runs, respectively. Turn formation is also observed in the region 29QIP31 at both pH values. Other regions are specific to pH 4.2 (24MK25, 24MKR26, 23HMKR26, 30IPS32, and 23DMK25) or to pH 7.2 (28TQ29, 27ATQ29, 30IPS32, 30IP31, 24MKRA27, 25KRA27, 28TQ130). Overall, this data suggests that there is more variability in turn formation region at pH 7.2 than at pH 4.2. C-terminal residues between 24 and 32 (MKRATQIPS) are likely to be involved in turn formation at both pH values (Supplementary Table 2).

4.4. Changes in secondary structure in PAP248–286

Previous studies have shown that the secondary structure of a protein/peptide is affected by external factors such as temperature, pressure and pH [49,50]. To see the effect of pH on the secondary structure of the PAP248–286 peptide, we have calculated the average secondary structure propensity for each residue at both pH values (Fig. S2). This analysis showed there is higher α -helix propensity at pH 7.2 than at pH 4.2. To see a distribution of α -helix and coil content in all trajectories we have plotted boxplots of the percentage of α -helix and coil content for the last 100 ns (300–400 ns) at both pH values. (Fig. 7). The plot shows that that α -helix content in PAP248–286 is mostly in the range

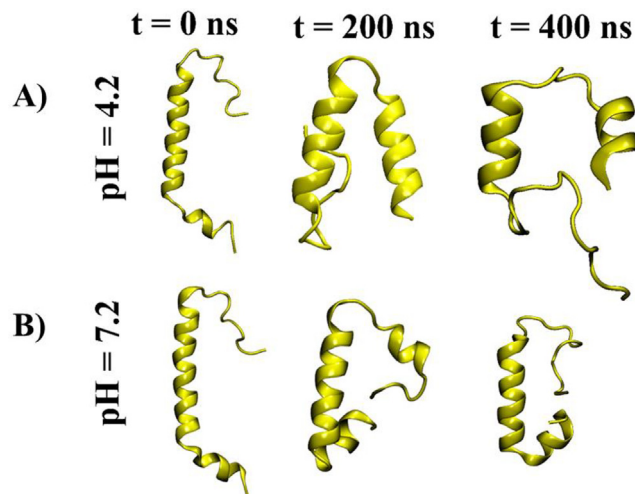


Fig. 6. Model structures of PAP248–286 from one of the representative simulations at three different time points A) at pH = 4.2 and B) at pH = 7.2.

between $\sim 22\%$ and $\sim 28\%$ at pH 4.2 (Fig. 7A, red color), and in the range between $\sim 27\%$ and $\sim 38\%$ at pH 7.2 (Fig. 7A, blue color). On the other hand, coil content is in the range $\sim 42\%$ to $\sim 54\%$ at pH 4.2 (Fig. 7B, red color) and between $\sim 33\%$ and 47% at pH 7.2 (Fig. 7B blue color). Moreover, we also noticed that in two simulations at pH 4.2 the percentage of α -helix is $\sim 2.0\%$ and $\sim 0.05\%$ respectively, and at pH 7.2–3.0 and $\sim 12\%$, respectively. Overall, these boxplots show that PAP248–286 has more coil and less α -helical content at pH 4.2 compared to pH 7.2.

4.5. Residue contact maps

To identify those residues that make favorable contacts, we computed the residue-residue average contact map and interaction time (see method section for more details). Interestingly, we observed a significant difference in intra-peptide contacts at the two pH values (Fig. 8A and B). At pH 4.2, Arg10 and Leu11 in the N-terminal region, Ile20 and Leu21 in the central region, and Met24 in the C-terminal region form contacts with the residues Lys35–Tyr39 for at least 10% of the simulation time (Fig. 8C and D). At pH 7.2, only Ile20 and Met24 form contacts with residues Lys35–Tyr39 for 10% or more of the simulation time. Another sig-

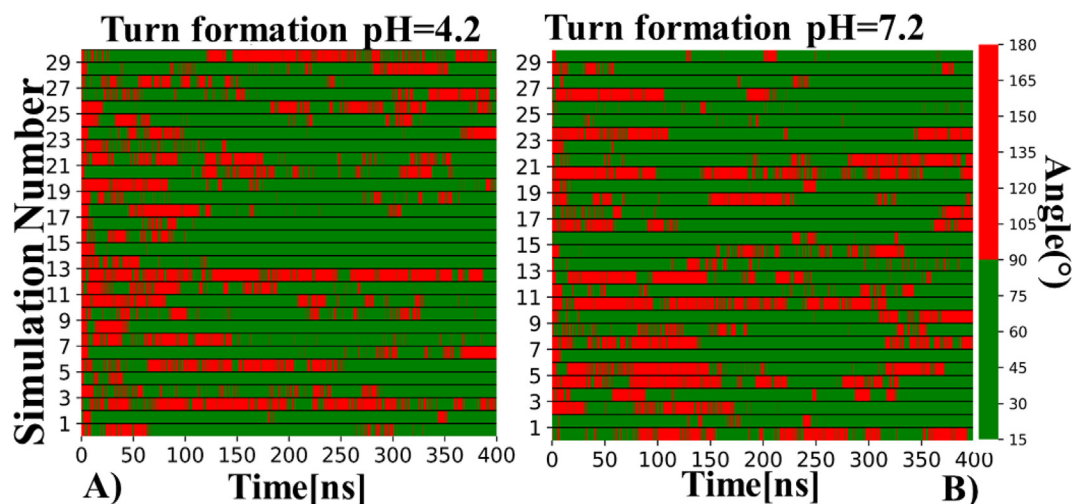


Fig. 5. Time evolution of turn formation in the PAP248–286 peptide at A) pH 4.2 and B) pH 7.2.

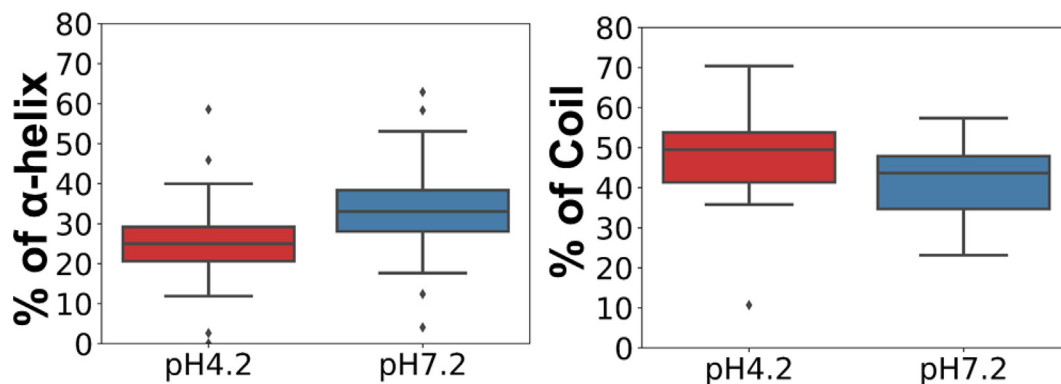


Fig. 7. A) Percentage of α -helix content in PAP248–286 at pH 4.2 and pH 7.2, B) Percentage of coil content in PAP248–286 at pH 4.2 and pH 7.2.

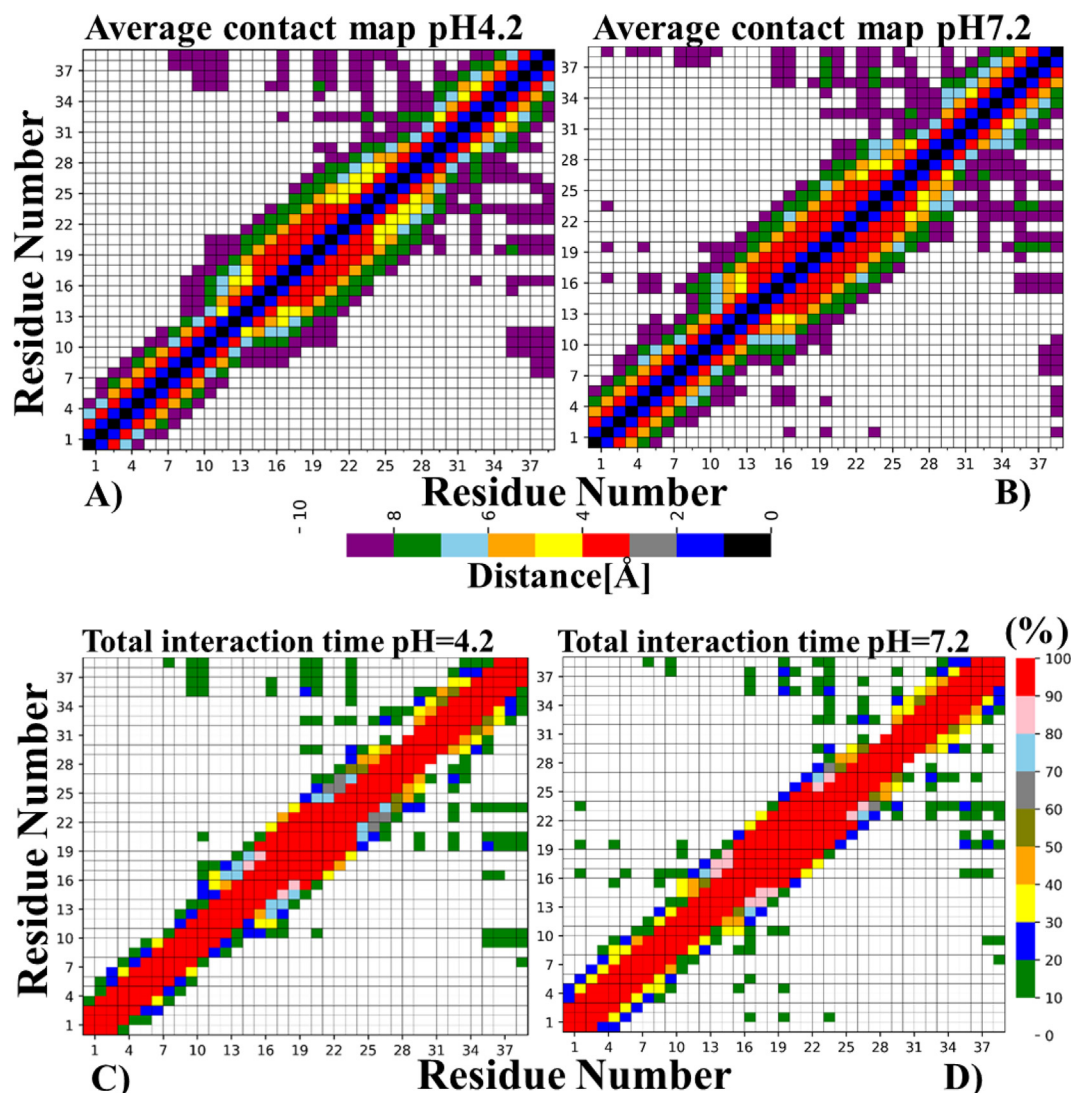


Fig. 8. A) Average contact map of PAP248–286 at pH 4.2. B) Average contact map of PAP248–286 at pH 7.2. C) Total interaction time between residues of PAP248–286 at pH 4.2. D) Total interaction time between residues of PAP248–286 at pH 7.2.

nificant difference in contacts at the two pH values can be observed in the middle region residues Leu16–Ile20 and in the *N*-terminal residues. At pH 7.2 Leu16 forms contacts with Gln5 and Ser9; Val17 forms contacts with Ile2, Lys6, Ser9, and Arg10; Asn18 interacts with Leu11, while Ile20 forms contacts with residues Ile2,

Gln5, Leu11 for 10 % or more of the simulation time. However, these contacts are absent at pH 4.2. Moreover, at pH 7.2 the *N*-terminal residues Gly1 and Gln5 form contacts with each other for ≥ 20 % of the simulation time, while at pH 4.2 their interaction time is below 10 %. Overall, these data suggest that at pH 7.2 more

residues remain within a rcut cut-off range. Additionally, interactions are also formed between the residues for longer periods at pH 7.2 compared to pH 4.2.

4.6. Interaction type

To assess the nature of the interactions that take place between the residues of the PAP248–286 peptide at the two different pH values, we have plotted interaction types between the residues (see methods section for more details) (Fig. 9). In the N-terminal region of the peptide there are hydrogen bonding interactions and two salt bridges at both pH values. The first salt bridge is between Lys6 and Glu7, while the second one is between Lys8 and Glu7. In the middle region and in the C-terminal region of PAP248–286, multiple hydrogen bonding and hydrophobic interactions are formed at both pH values. However, we observe no major difference in the interaction pattern between the two pH values, except the presence of a single hydrophobic interaction between Tyr33 and Leu36 at pH 4.2, which is absent at pH 7.2. On the other hand, at pH 7.2 Ala27 and Ile30 form a hydrophobic interaction that is absent at pH 4.2. Overall, hydrogen bonding and hydrophobic interactions are largely prevalent at both pH values (Fig. 9).

4.7. Free energy landscape of PAP248–286

The energy landscape of a protein is the mapping of the different energy levels corresponding to all its possible conformations [51]. To identify the misfolding pathway and the different conformational states adopted by PAP248–286 at pH 4.2 and 7.2, we constructed a two-dimensional free energy landscape (FEL) contour map using fraction of native contacts (Q) and RMSD (Å) as reaction coordinates (Fig. 10). The free energy landscape of PAP248–286 features three energy basins at pH 4.2 and four energy basins at pH 7.2. At pH 4.2 (Table 1), basin A is in the Q range (~0.869–0.892) and in the RMSD range (~7.16–8.19 Å). In basin A, PAP248–286 contains a long helical region while part of the N- and C-terminal regions of the peptide contain random coil structures. The second basin (B), which is the lowest energy basin (Fig. 10A) of PAP248–286, is in the Q range (~0.7389–0.7946) and in the RMSD range (~11.3 to 11.84 Å). In basin B, the N- and C-terminal regions of PAP248–286 are completely unstructured and the central region of the peptide forms a helix. The third energy basin (C) is in the Q range (~0.6079–0.6257) and in the RMSD range (~12.74 to 14.14 Å). In basin C, the structure of PAP248–286 does not contain any helical motif and most of the peptide forms a random coil,

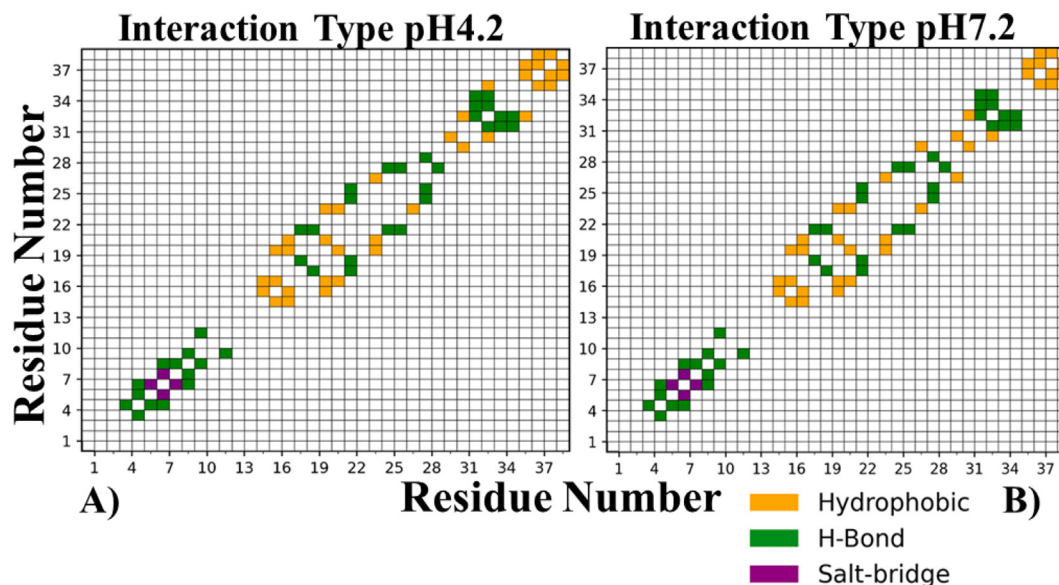


Fig. 9. PAP248–286 interaction types A) at pH = 4.2, and B) at pH = 7.2.

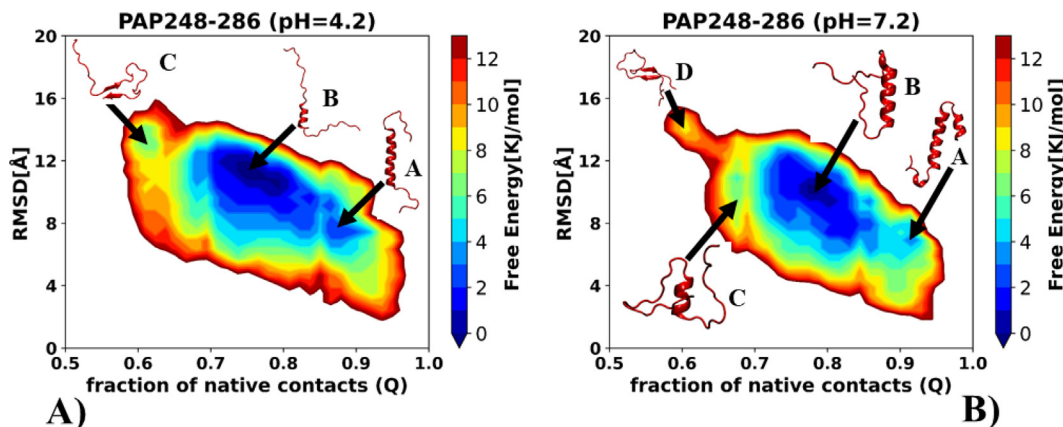

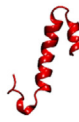


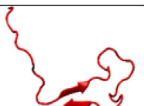
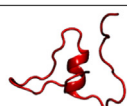



Fig. 10. Free energy landscape of PAP248–286 at A) pH = 4.2 and B) pH = 7.2.

Table 1
Values of native contacts and RMSD and respective structures for each basin at both pH values.

Basins	pH 4.2		pH 7.2	
	Q, RMSD (Å)	Structure	Q, RMSD (Å)	Structure
A	Q = 0.869-0.892 RMSD = 7.16-8.19		Q = 0.9151 RMSD = 6.88	
B	Q = 0.7389-0.7946 RMSD = 11.3-11.84		Q = 0.7669- 0.7885 RMSD = 9.86-10.41	
C	Q = 0.6079-0.6257 RMSD = 12.74-14.14		Q = 0.6756 RMSD = 10.56	
D			Q = 0.6044 RMSD = 14.12	

while the C-terminal residues Leu36, Ile37, Met38 and the central region residues Val15, Leu16, Val17 form a small β -sheet with an antiparallel arrangement. It is worth noting that the energy barrier between basin B and C is significantly higher than the energy barrier between basin A and B. At pH 7.2, basin A is at Q (0.9151) and RMSD (6.88 Å). The basin A structure features a “U-shaped” topology in a predominantly helical form. Basin B, which is the lowest energy basin at this pH value, is in the Q range (~0.7769–0.7885) and in the RMSD range (~9.86–10.41 Å). Hence, it is highly restricted compared to the lowest energy basin (B) at pH 4.2. In

basin B at pH 7.2, the structure of PAP248-286 features an N-terminal random coil region and a long helical portion formed by the central region of the peptide and 6 residues of the C-terminal region. In basin C, which is at Q (0.6756) and RMSD (10.56 Å), the structure of PAP248-286 features N- and C-terminal random coils while part of the central region maintains a helical structure. Basin D is at Q (0.6044) and RMSD (14.12 Å). In this basin, PAP248-286 adopts a random coil and antiparallel β -sheet-rich structure. The antiparallel β -sheet in the structure is formed by the N-terminal residues Arg10, Leu11 and Gln12 and by the C-terminal

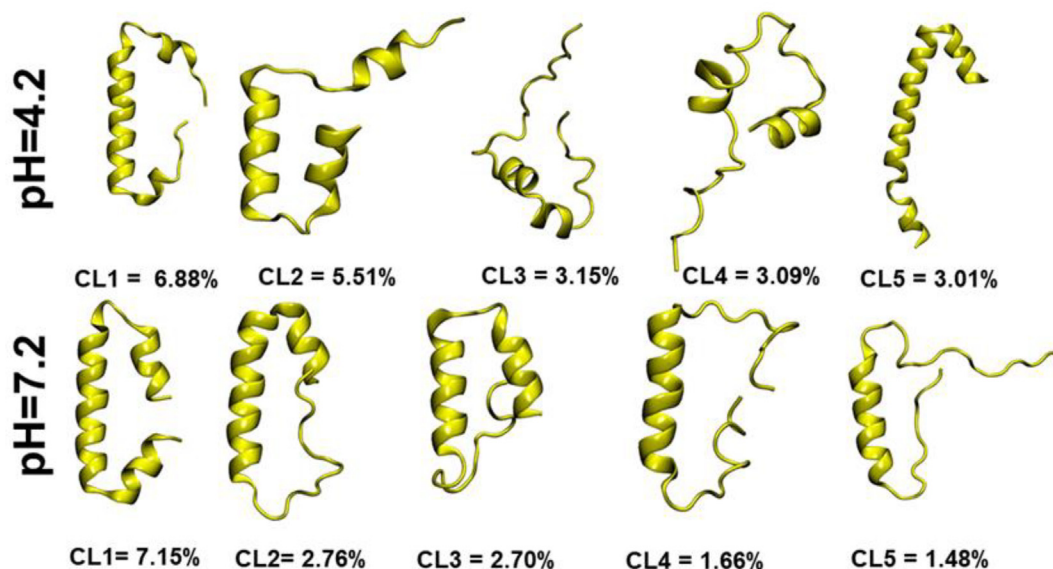


Fig. 11. RMSD-based clusters of PAP248-286 at A) pH = 4.2 and B) pH = 7.2.

residues Ser32, Tyr33 and Lys34. The energy barrier between basin C and D is lower than the energy barrier between basin B and C, indicating that the conversion between the C and the D state could be more rapid than the conversion between the B and the C state. Overall, FEL analysis shows that i) there are more local minima at pH 7.2 than at pH 4.2 and ii) basin C at pH 7.2 and the lowest energy structure at pH 4.2 (basin B) are almost identical to the structure of PAP248-286 determined by NMR in membrane environment (PDB id: 2L3H) [19]. iii) PAP248-286 forms antiparallel β -sheet structures at both pH values.

4.8. Clustering of structures

Clustering of protein conformations from multiple simulation trajectories help to divide protein structures into different groups that share similar topological features [52]. To assess the effect of pH, we have performed RMSD-based clustering using 0.6 million structures at both pH 4.2 and pH 7.2 (Fig. 11). At both pH values, the top five conformational pools show that PAP248-286 adopts different conformations. At pH 4.2, we observe that cluster-3 (CL3) and cluster-4 (CL4) feature the least α -helical content, while cluster-5 has the most. At pH 7.2, cluster-4 (CL4) and cluster-5 (CL5) show the least α -helical content, while cluster-1 (CL1) displays the most. Interestingly, at both pH values most of the structures adopt a “U-shaped” conformation, which suggests that the PAP248-286 peptide adopts prevalently a “U-shaped” topology. Overall, clustering data suggests that the PAP248-286 peptide adopts significantly different conformations at the two pH values.

5. Discussion & conclusions

Our MD simulations of the PAP248-286 peptide at two different pH values have shown that more conformational changes take place at pH 4.2 than at pH 7.2. At the acidic pH, less hydrogen bonds and less intra-molecular contacts are formed than at pH 7.2. In addition, we identified a clear difference in the formation of a hydrophobic interaction between a pair of residues at the two pH values, which could be extremely relevant considering that, typically, native and non-native hydrophobic interactions are a crucial driving force for protein secondary structure folding and misfolding processes [53,54]. Furthermore, we also observed significant differences in the free energy landscape of the peptide at the two pH values, featuring more local minima at pH 7.2 than at pH 4.2. Finally, clustering analysis showed that at pH 7.2 the PAP248-286 peptide has more helical content in different clusters than at pH 4.2. Overall, these data suggest that PAP248-286 adopts significantly different conformations at the two physiological pH values. In previous experimental studies, Ramamoorthy & co-workers [7,16] showed that the PAP248-286 peptide enhances viral infection *in vivo* in the acidic vaginal environment (pH \sim 4.2), where it also promotes interaction with lipid bilayers, but not at neutral pH (\sim 7.2). In general, membrane-active peptides are positively charged, amphiphilic in nature and short in length. Moreover, they are commonly unstructured in solution [55]. Based on previous studies and on the data that is provided herein, we suggest that the larger extent at which secondary structure content, especially α -helical, and hydrogen bonds are lost at pH 4.2 renders the peptide more flexible. Furthermore, the higher net positive charge of PAP248-286 at pH 4.2 than at pH 7.2 could drive the PAP248-286 peptide to form more interactions with lipid membranes and enhance HIV infection at pH 4.2. We also observed that PAP248-286 adopts a “U”-shaped conformation, which is in line with a previous study by Agrawal *et al.* [29] on A β peptide, where early stage misfolding of A β peptide was shown to feature the formation of a “U-shaped” structure.

In summary, this study provides atomic-level insight into structural dynamics of PAP248-286 at two different pH values and helps unravel the physiological factors that may play an important role in enhancing HIV infection at the acidic vaginal pH value.

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

N.A. acknowledges the ERDF postdoc grant (No. 1.1.1.2/VIAA/4/20/757) for funding. E.P. thanks the ERDF project BioDrug (No. 1.1.1.5/19/A/004) and the Latvian Council of Science (grant No. lzp-2020/2-0013) for financial support. We would like to thank Latvian Institute of Organic Synthesis, Riga for supercomputing resources. N.A. would also like to thank CHPC, Cape Town for supercomputing resources.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.08.060>.

References

- [1] Romero RA, Klausner JD, Marsch LA, Young SD. Technology-Delivered Intervention Strategies to Bolster HIV Testing. *Current HIV/AIDS Reports* 2021;1–15.
- [2] Khalid K, Padda J, Fernando RW, Mehta KA, Almanie AH, Al Hennawi H, et al. Stem Cell Therapy and Its Significance in HIV Infection. *Cureus* 2021;13.
- [3] Doncel GF, Joseph T, Thurman AR. Role of Semen in HIV-1 Transmission: Inhibitor or facilitator? *Am J Reprod Immunol* 2011;65:292–301.
- [4] Kariuki SM, Selhorst P, Norman J, Cohen K, Rebe K, Williamson C, et al. Detectable HIV-1 in semen in individuals with very low blood viral loads. *Virology journal* 2020;17:1–5.
- [5] Peebles K, van der Straten A, Palanee-Phillips T, Reddy K, Hillier SL, Hendrix CW, et al. Brief Report: Anal Intercourse, HIV-1 Risk, and Efficacy in a Trial of a Dapivirine Vaginal Ring for HIV-1 Prevention. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2020;83:197–201.
- [6] Münch J, Rucker E, Ständker L, Adermann K, Goffinet C, Schindler M, et al. Semen-derived amyloid fibrils drastically enhance HIV infection. *Cell* 2007;131:1059–71.
- [7] Lee YH, Ramamoorthy A. Semen-derived amyloidogenic peptides—Key players of HIV infection. *Protein Sci* 2018;27:1151–65.
- [8] Arnold F, Schnell J, Zirafi O, Stürzel C, Meier C, Weil T, et al. Naturally occurring fragments from two distinct regions of the prostatic acid phosphatase form amyloidogenic enhancers of HIV infection. *J Virol* 2012;86:1244–9.
- [9] Kim K-A, Yolamanova M, Zirafi O, Roan NR, Staendker L, Forssmann W-G, et al. Semen-mediated enhancement of HIV infection is donor-dependent and correlates with the levels of SEVI. *Retrovirology* 2010;7:1–12.
- [10] E.W. Vane, S. He, L. Maibaum, A. Nath, Rapid formation of peptide/lipid co-aggregates by the amyloidogenic seminal peptide PAP248-286, *bioRxiv*, (2020).
- [11] Roan NR, Münch J, Arhel N, Mothes W, Neideman J, Kobayashi A, et al. The cationic properties of SEVI underlie its ability to enhance human immunodeficiency virus infection. *J Virol* 2009;83:73–80.
- [12] Viswanathan GK, Mohapatra S, Paul A, Arad E, Jelinek R, Gazit E, et al. Inhibitory effect of naphthoquinone-tryptophan hybrid towards aggregation of PAP f39 semen amyloid. *Molecules* 2018;23:3279.
- [13] Mamone S, Glöggler S, Becker S, Rezaei-Ghaleh N. Early Divergence in Misfolding Pathways of Amyloid-Beta Peptides. *ChemPhysChem* 2021;22:2158–63.
- [14] Yu J, Lyubchenko YL. Early stages for Parkinson's development: α -synuclein misfolding and aggregation. *J Neuroimmune Pharmacol* 2009;4:10–6.
- [15] Vane EW, He S, Maibaum L, Nath A. Rapid Formation of Peptide/Lipid Coaggregates by the Amyloidogenic Seminal Peptide PAP248-286. *Biophys J* 2020;119:924–38.
- [16] Brender JR, Hartman K, Gottler LM, Cavitt ME, Youngstrom DW, Ramamoorthy A. Helical conformation of the SEVI precursor peptide PAP248-286, a dramatic enhancer of HIV infectivity, promotes lipid aggregation and fusion. *Biophys J* 2009;97:2474–83.
- [17] Easterhoff D, Ontiveros F, Brooks LR, Kim Y, Ross B, Silva JN, et al. Semen-derived enhancer of viral infection (SEVI) binds bacteria, enhances bacterial phagocytosis by macrophages, and can protect against vaginal infection by a

- sexually transmitted bacterial pathogen. *Antimicrob Agents Chemother* 2013;57:2443–50.
- [18] Brender JR, Nanga RPR, Popovych N, Soong R, Macdonald PM, Ramamoorthy A. The amyloidogenic SEVI precursor, PAP248–286, is highly unfolded in solution despite an underlying helical tendency. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 2011;1808:1161–9.
- [19] Nanga RP, Brender JR, Vivekanandan S, Popovych N, Ramamoorthy A. NMR structure in a membrane environment reveals putative amyloidogenic regions of the SEVI precursor peptide PAP248–286. *J Am Chem Soc* 2009;131:17972–9.
- [20] Blokhin DS, Filippov AV, Antzutkin ON, Karataeva FK, Klochkov VV. Spatial structure of oligopeptide PAP (248–261), the N-terminal fragment of the HIV enhancer prostatic acid phosphatase peptide PAP (248–286), in aqueous and SDS micelle solutions. *J Mol Struct* 2014;1070:38–42.
- [21] Blokhin DS, Aganova OV, Yulmetov AR, Filippov AV, Gizatullin BI, Afonin S, et al. Spatial structure of heptapeptide Glu-Ile-Leu-Asn-His-Met-Lys, a fragment of the HIV enhancer prostatic acid phosphatase, in aqueous and SDS micelle solutions. *J Mol Struct* 2013;1033:59–66.
- [22] Sievers SA, Karanicolas J, Chang HW, Zhao A, Jiang L, Zirafi O, et al. Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* 2011;475:96–100.
- [23] Blokhin DS, Filippov AV, Antzutkin ON, Afonin S, Klochkov VV. Spatial structures of PAP (262–270) and PAP (274–284), two selected fragments of PAP (248–286), an enhancer of HIV infectivity. *Appl Magn Reson* 2015;46:757–69.
- [24] Agrawal N, Skelton AA. Structure and Function of Alzheimer's Amyloid β Proteins from Monomer to Fibrils: A Mini Review. *Protein J* 2019;38:425–34.
- [25] Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. *Science* 1992;256:184–6.
- [26] Agrawal N, Skelton AA. 12-crown-4 ether disrupts the patient brain-derived amyloid- β -fibril trimer: Insight from all-atom molecular dynamics simulations. *ACS Chem Neurosci* 2016;7:1433–41.
- [27] Irwin DJ, Lee V-M-Y, Trojanowski JQ. Parkinson's disease dementia: convergence of α -synuclein, tau and amyloid- β pathologies. *Nat Rev Neurosci* 2013;14:626–36.
- [28] Dobson CM. The amyloid phenomenon and its links with human disease. *Cold Spring Harbor Perspect Biol* 2017;9:a023648.
- [29] Agrawal N, Skelton AA. Binding of 12-crown-4 with Alzheimer's A β 40 and A β 42 monomers and its effect on their conformation: insight from molecular dynamics simulations. *Mol Pharm* 2018;15:289–99.
- [30] Flöck D, Colacino S, Colombo G, Di Nola A. Misfolding of the amyloid β -protein: A molecular dynamics study. *Proteins: Structure, Function, and Bioinformatics* 2006;62:183–92.
- [31] Lump E, Castellano LM, Meier C, Seeliger J, Erwin N, Sperlich B, et al. A molecular tweezer antagonizes seminal amyloids and HIV infection. *Elife* 2015;4:e05397.
- [32] Li J, Yang Z, Liu H, Qiu M, Zhang T, Li W, et al. ADS-1J disaggregates semen-derived amyloid fibrils. *Biochem J* 2019;476:1021–35.
- [33] Jakob CG, Lewinski K, Kuciel R, Ostrowski W, Lebioda L. Crystal structure of human prostatic acid phosphatase. *Prostate* 2000;42:211–8.
- [34] J.C. Gordon, J.B. Myers, T. Folta, V. Shoja, L.S. Heath, A. Onufriev, H++: a server for estimating pK as and adding missing hydrogens to macromolecules, *Nucleic acids research*, 33 (2005) W368–W371.
- [35] Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 2017;14:71–3.
- [36] Huang J, MacKerell Jr AD. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem* 2013;34:2135–45.
- [37] Krupa P, Quoc Huy PD, Li MS. Properties of monomeric A β 42 probed by different sampling methods and force fields: Role of energy components. *J Chem Phys* 2019;151:055101.
- [38] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* 2005;26:1701–18.
- [39] Mercadante D, Gräter F, Daday C. CONAN: a tool to decode dynamical information from molecular interaction maps. *Biophys J* 2018;114:1267–73.
- [40] Fas BA, Maiani E, Sora V, Kumar M, Mashkoo M, Lambrughi M, et al. The conformational and mutational landscape of the ubiquitin-like marker for autophagosome formation in cancer. *Autophagy* 2021;17:2818–41.
- [41] González-Alemán R, Hernández-Castillo D, Rodríguez-Serradet A, Caballero J, Hernández-Rodríguez EW, Montero-Cabrera L. BitClust: fast geometrical clustering of long molecular dynamics simulations. *J Chem Inf Model* 2019;60:444–8.
- [42] Frishman D, Argos P. Seventy-five percent accuracy in protein secondary structure prediction. *Proteins: Structure, Function, and Bioinformatics* 1997;27:329–35.
- [43] Kufareva I, Abagyan R. Methods of protein structure comparison, in: *Homology Modeling*, Springer 2011:231–57.
- [44] R.E. Hubbard, M.K. Haider, Hydrogen bonds in proteins: role and strength, *eLS*, (2010).
- [45] Rader A, Hespeneheide BM, Kuhn LA, Thorpe MF. Protein unfolding: rigidity lost. *Proc Natl Acad Sci* 2002;99:3540–5.
- [46] Uddin MS, Al Mamun A, Rahman M, Behl T, Perveen A, Hafeez A, et al. Emerging proof of protein misfolding and interactions in multifactorial Alzheimer's disease. *Curr Top Med Chem* 2020;20:2380–90.
- [47] Maleki R, Khedri M, Rezvantabab S, Afsharchi F, Musaei K, Shafiee S, et al. β -amyloid targeting with two-dimensional covalent organic frameworks: Multi-scale in-silico dissection of nano-biointerface. *ChemBioChem* 2021;22:2306.
- [48] Agerschou ED, Schützmann MP, Reppert N, Wördehoff MM, Shaykhalishahi H, Buell AK, et al. β -Turn exchanges in the α -synuclein segment 44-TKEG-47 reveal high sequence fidelity requirements of amyloid fibril elongation. *Biophys Chem* 2021;269:106519.
- [49] Makarov A, LoBrutto R, Karpinski P. Effect of pressure on secondary structure of proteins under ultra high pressure liquid chromatographic conditions. *J Chromatogr A* 2013;1318:112–21.
- [50] Francisco C-C, Luis C-L-J, Marina E-B-J, Javier C-M-F, Alexis L-Z-A, Del Carmen S-O-H, et al. Effect of Temperature and pH on the Secondary Structure and Denaturation Process of Jumbo Squid Hepatopancreas Cathepsin D. *Protein Pept Lett* 2019;26:532–41.
- [51] Tsai C-J, Nussinov R. The free energy landscape in translational science: how can somatic mutations result in constitutive oncogenic activation? *PCCP* 2014;16:6332–41.
- [52] Phillips JL, Colvin ME, Newsam S. Validating clustering of molecular dynamics simulations using polymer models. *BMC Bioinf* 2011;12:1–23.
- [53] Dyson HJ, Wright PE, Scheraga HA. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc Natl Acad Sci* 2006;103:13057–61.
- [54] Dias CL, Karttunen M, Chan HS. Hydrophobic interactions in the formation of secondary structures in small peptides. *Phys Rev E* 2011;84:041931.
- [55] Kabelka I, Vácha R. Advances in Molecular Understanding of α -Helical Membrane-Active Peptides. *Acc Chem Res* 2021;54:2196–204.