scientific reports

Check for updates

OPEN

Unveiling the phenylalanine coaggregation mechanism for a deep understanding of phenylketonuria disease

Haruna L. Barazorda-Ccahuana[1], Francesc Mas[2] & Sergio Madurga[2]✉

The abnormal accumulation of phenylalanine is a defining feature of phenylketonuria (PKU) and is Linked to the formation of toxic, amyloid-like fibrils. To investigate the molecular mechanisms underlying this aggregation, we performed all-atom molecular dynamics simulations of zwitterionic phenylalanine at physiological temperature. Systems with varying phenylalanine concentrations were simulated over 500 ns to assess aggregation dynamics, structural stability, and non-covalent interactions. Our results show that phenylalanine rapidly self-assembles into fibrillar structures stabilized by hydrogen bonding and π–π stacking. Higher concentrations led to more compact aggregates, as indicated by radial distribution functions and solvent-accessible surface area analyses. We further examined the coaggregation of alanine with phenylalanine fibrils and found that alanine preferentially binds to zwitterionic terminal regions via hydrogen bonds. This interaction may contribute to the enhanced toxicity of phenylalanine aggregates. These findings provide molecular-level insights into phenylalanine aggregation in PKU and support the development of strategies to mitigate its pathological effects.

The aggregation of aromatic amino acids, particularly phenylalanine, tyrosine, and tryptophan, has been linked to the pathogenesis of various diseases. These residues are aromatic and hydrophobic, properties that contribute to their tendency to self-assemble through π-π stacking and hydrophobic interactions[1–3]. In many cases, protein misfolding and abnormal aggregation lead to the formation of toxic amyloid-like structures, which contribute to cellular dysfunction. In particular, protein aggregation plays a central role in neurodegenerative diseases such as Alzheimer's and Parkinson's diseases, where misfolded proteins form β-sheet-rich amyloid fibrils and intracellular inclusions[4–6]. Similarly, in type 2 diabetes (T2D), the misfolding and aggregation of islet amyloid polypeptide (IAPP) are implicated in cell dysfunction and disease progression[7–9].

While many amyloid-related disorders stem from protein misfolding, phenylketonuria (PKU) represents a different case, in which toxicity is linked to the self-assembly of free phenylalanine (Phe) rather than protein aggregation[10–16]. Experimental studies have demonstrated that Phe can spontaneously form amyloid-like structures under physiological conditions[12–15,17,18], suggesting that its intrinsic self-assembly properties contribute to PKU-associated toxicity[14,15]. The process is influenced by external factors, including solvent composition, which modulates the interactions driving molecular aggregation. Research on aromatic peptide coassembly of diphenylalanine and triphenylalanine peptides at various mass ratios has revealed that interactions between hydrophobic residues and solvent molecules play a key role in determining nanostructure morphology[15]. In biological systems, the analysis of non-covalent interactions is crucial for rationalizing the structure, dynamics, and function of macromolecules, as well as for elucidating the mechanisms underlying molecular aggregation and self-organization[19]. In these systems, water is crucial for stabilizing chiral supramolecular structures through hydrogen bonding and π-π interactions[20–24], which could be significant in Phe aggregation. In particular, the self-assembly of amino acids often results in the formation of well-ordered nanoscale architectures, which exhibit distinct structural and functional properties[25,26]. These findings underscore the importance of understanding the molecular mechanisms underlying phenylalanine aggregation and its role in PKU pathology.

Anand et al.[14] demonstrated that Phe fibrils can cross-seed the aggregation of native proteins under physiological conditions, promoting their conversion into β-sheet-rich assemblies. This phenomenon is

[1]Computational Biology and Chemistry Research Group, Universidad Católica de Santa María, Arequipa 04000, Perú. [2]Materials Science and Physical Chemistry Department & Research Institute of Theoretical and Computational Chemistry (IQTCUB), University of Barcelona, Barcelona 08028, Spain. ✉email: s.madurga@ub.edu

particularly relevant to PKU, where elevated Phe levels may facilitate the formation of toxic, amyloid-like fibrils *in vivo*. The study further showed that these Phe aggregates induce pronounced hemolytic activity, deforming erythrocyte membranes and causing severe lysis. In addition to interacting with proteins, Phe fibrils were found to coaggregate with other amino acids, acting as molecular scaffolds or "seeds" for their incorporation into fibrillar structures. These observations underscore two critical pathological features of phenylalanine aggregation in PKU: its ability to trap proteins and other biomolecules into irreversible aggregation pathways, and the cytotoxic nature of both primary Phe fibrils and the protein fibrils they induce. Similar findings have been reported by Tomar et al.[13], Singh et al.[10], and Nigdelioglu et al.?, who observed the membrane-disruptive potential and aggregation-seeding capacity of aromatic amino acid assemblies under physiological and pathological conditions.

Computer simulations are increasingly providing valuable insights into the behavior of biologically important molecules. In line with the work of Sahin Uyaver et al., who explored the self-assembly of aromatic amino acids such as tyrosine, tryptophan, and phenylalanine at different temperatures using molecular dynamics simulations to study fibril-like aggregate formation[27–29], our research takes a different approach. We investigate the coaggregation mechanism of Phe residues. First, we use conditions that promote the formation of large Phe aggregates, studying systems containing 100 to 500 phenylalanine monomers–more than in previous computational studies–which allows us to identify different fibril structures. We then explore the coaggregation process using preformed Phe fibrils and alanine (Ala) residues, which typically do not self-aggregate but can aggregate in the presence of Phe fibrils, as experimentally observed.

Given the critical role of phenylalanine fibrils in diseases like PKU and the limited understanding of their toxicity mechanisms, our study aims to investigate the behavior of phenylalanine fibrils in solution and their ability to promote the aggregation of other amino acids. A central focus is understanding how non-aromatic amino acids, such as alanine, interact with phenylalanine and contribute to the aggregation process despite lacking the intrinsic self-assembly properties of aromatic residues. To achieve this, we use all-atom molecular dynamics simulations, to study high concentrations of phenylalanine in its zwitterionic state alongside alanine molecules at physiological temperature. By elucidating the molecular interactions between phenylalanine and alanine, our work provides critical insights into the coaggregation mechanisms that may underlie the toxicity of phenylalanine fibrils in biological systems.

Results and discussion

This investigation aimed to analyze the coaggregation of alanine within phenylalanine fibres. Using all-atom molecular dynamics (MD) simulations, we first examined the self-assembly of zwitterionic phenylalanine monomers at varying concentrations (100 to 500 monomers) to understand their intrinsic aggregation behavior. Subsequently, we investigated the ability of phenylalanine fibrils to promote the incorporation of alanine into these structures, providing insights into how non-aromatic amino acids participate in the aggregation process. Simulations were performed using the OPLS-AA force field, which offers accurate treatment of torsional angles and non-covalent interactions such as hydrogen bonding and π–π stacking–features particularly beneficial for modeling aromatic residue aggregation. To better capture the solvation environment, we employed the TIP5P water model, which provides an improved description of the tetrahedral structure and hydrogen-bonding network of water. This combination allowed us to simulate the aggregation dynamics and solute–solvent interactions with greater physical realism under near-physiological conditions.

Formation of phenylalanine fibres

The results of molecular dynamics simulations of the zwitterionic form of phenylalanine over a period of 500 ns showed that the monomers rapidly aggregated within 100 ns. Furthermore, during the production phase of the simulation, these aggregates progressively improved their stability.

Root-mean-square deviation (RMSD) analysis was performed for all simulations to evaluate the mobility of Phe residues in each system (See Fig. 1). Table 1 presents the RMSD values for Phe residues during the last 100 ns of the simulations. The RMSD calculation was carried out using all recorded frames from the last 100 ns of each independent replicate for all systems (Phe100, Phe200, Phe300, Phe400, and Phe500). For each system, the average RMSD was first computed separately for each trajectory. Standard deviations were then calculated across these trajectory-level averages, treating each trajectory as an independent sample. The results reveal that systems with higher phenylalanine concentrations exhibit lower RMSD values, indicating reduced mobility and greater structural stability. In contrast, more diluted systems show a progressive increase in RMSD, reflecting higher mobility of Phe residues. For instance, the Phe100 system exhibited the highest RMSD value (8.3±1.4 nm), while the Phe500 system, with five times the concentration of phenylalanine, displayed the lowest RMSD value (1.2±0.1 nm). These variations in RMSD are related to the phenylalanine concentration, residue mobility, and the stability of the resulting aggregates.

To evaluate the spatial distribution of Phe and assess the degree of aggregation, we measured the solvent-accessible surface area (SASA) *per residue* for all simulations. SASA represents the surface area of a biomolecule that is accessible to a solvent and serves as a key metric for understanding molecular aggregation. The average SASA values *per residue* are presented in Table 1 and Fig. 2. The results show that the Phe100 system exhibited the highest SASA value, while systems with higher phenylalanine concentrations (Phe200, Phe300, Phe400, and Phe500) displayed progressively lower SASA values. Figure 2 also illustrates the SASA *per residue* for each system during the last 100 ns of simulation, revealing that lower phenylalanine concentrations correspond to larger solvent-accessible areas. This trend indicates that as the concentration of Phe increases, the residues become more aggregated, reducing their exposure to the solvent. Specifically, the solvent-accessible surface area (SASA) *per residue* showed a marked decrease–approximately threefold–as the number of phenylalanine monomers increased from 100 to 500. According to Table 1, the SASA values decreased from 0.9 nm^2 at Phe100 to 0.5 nm^2 at Phe500, despite the fivefold increase in concentration. This inverse correlation underscores a strong

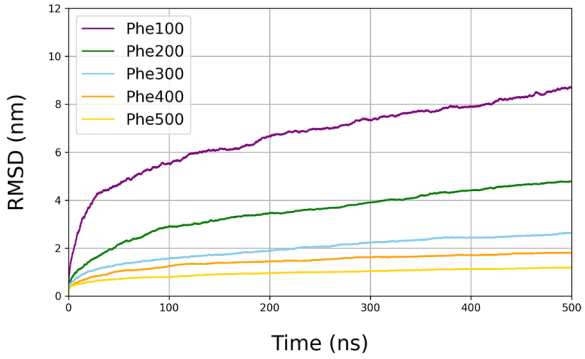

Fig. 1. Root-mean-square deviation (RMSD) as a function of time of five systems. The RMSD values were analyzed based on the average of each independent replicate for all systems (Phe100, Phe200, Phe300, Phe400, and Phe500).

Phenylalanine (molec.)	RMSD (nm)	SASA *per residue* (nm^2)
100	8.3±1.4	0.9±0.9
200	4.6±0.9	0.8±0.7
300	2.5±0.1	0.7±0.6
400	1.8±0.1	0.6±0.5
500	1.2±0.1	0.5±0.4

Table 1. Average values of RMSD and SASA *per residue* for phenylalanine systems during the last 100 ns of MD simulations, where the RMSD values (in nm) reflect the structural stability of the aggregates, while SASA values (in nm^2) indicate the degree of solvent accessibility. The reported error ranges correspond to the standard deviation: for RMSD, it was calculated based on the independent replicates of each system (Phe100, Phe200, Phe300, Phe400, and Phe500), while for SASA, it was computed considering the number of phenylalanine residues in each system.

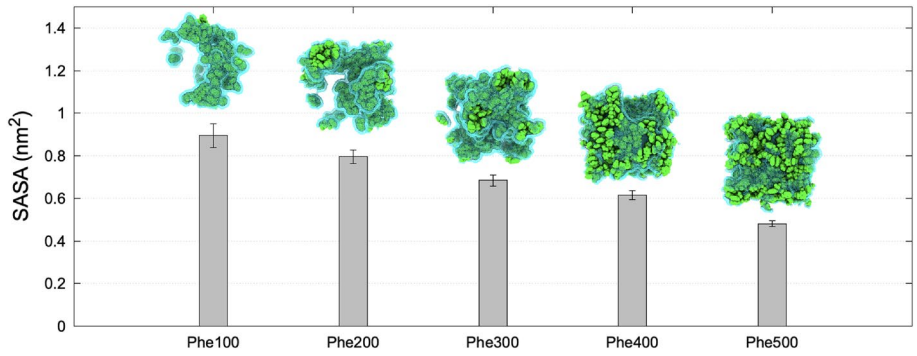

Fig. 2. SASA *per residue* for phenylalanine systems during the last 100 ns of five different systems. Representative snapshots of each system are included, visually demonstrating the increased degree of aggregation in more concentrated systems. Error bars represent the standard deviation.

relationship between monomer concentration, aggregation, and solvent accessibility. As more monomers are introduced, the aggregation becomes more compact and ordered, resulting in reduced exposure of individual residues to the solvent. This trend also aligns with the observed decrease in RMSD values, which reflect greater structural stability of the aggregates at higher concentrations.

As the concentration of Phe increases, a progressive transition in the aggregation pattern is observed, ranging from linear assemblies at low concentrations to densely packed, layered structures at higher concentrations (Fig. 3). In the Phe100 system, Phe monomers primarily form elongated, linear-like aggregates. These structures exhibit a parallel arrangement, where the NH_3^+ termini (blue) and COO^- termini (red) interact electrostatically, forming aligned chains. The hydrophobic side chains of Phe residues are positioned between these chains, minimizing their exposure to the solvent and contributing to the overall stability of the aggregates.

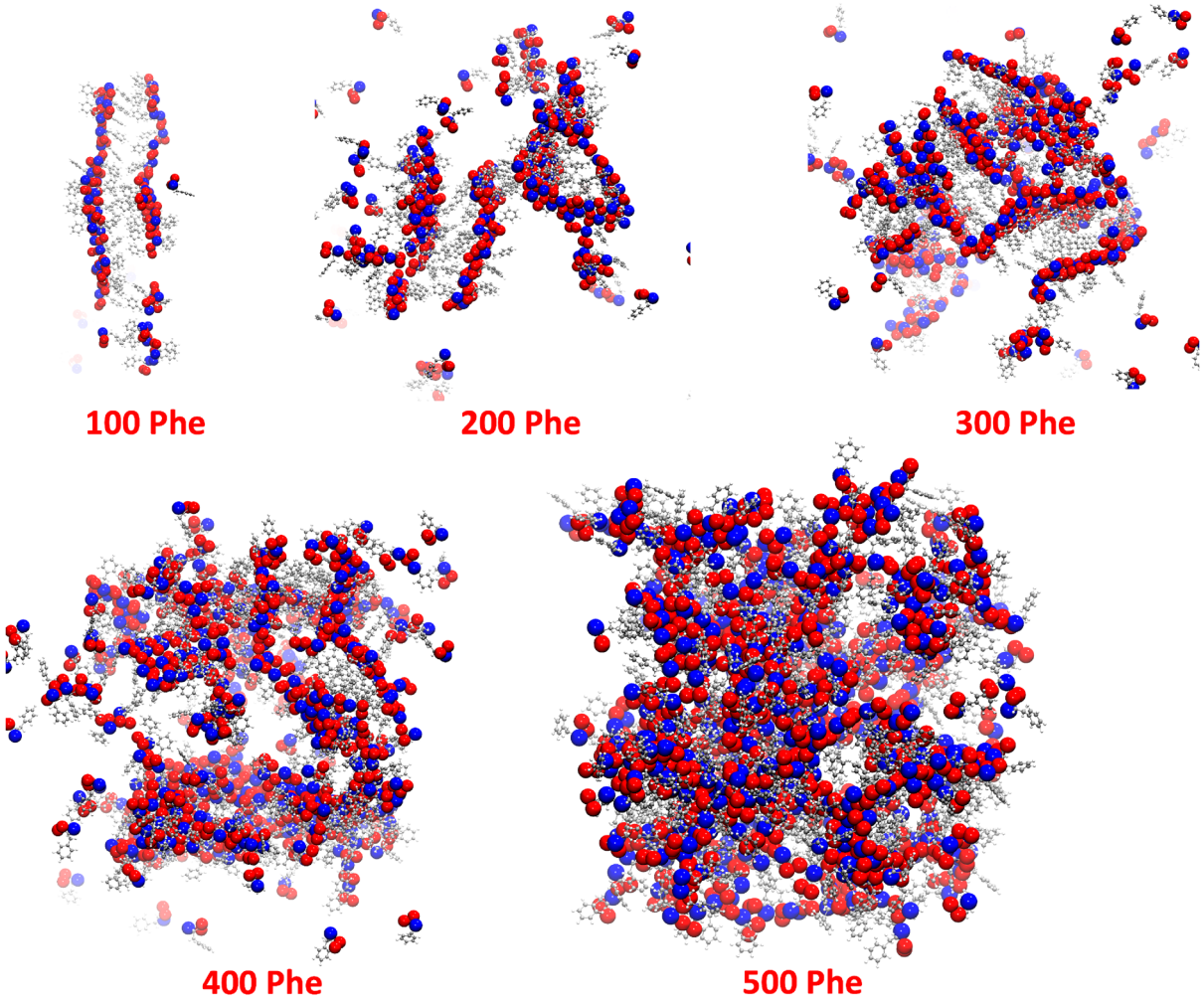

Fig. 3. Last frame of the molecular dynamics simulation for phenylalanine systems at varying concentrations (100 to 500 monomers). Blue spheres represent the nitrogen atoms of the NH_3^+ termini, while red spheres indicate the oxygen atoms of the COO^- termini of Phe residues.

In the Phe200 system, a more interconnected network of chains emerges due to the increased number of electrostatic interactions between NH_3^+ and COO^- termini. This interconnection results in a more stabilized and compact aggregation, though individual chain-like structures are still discernible. The interactions between hydrophobic side chains also become more evident, suggesting an early-stage organization of hydrophobic and electrostatic domains.

With a further increase in concentration, the Phe300 system exhibits more complex aggregation patterns. The individual chains observed at lower concentrations become increasingly entangled, forming multilayered structures. At this stage, the organization of Phe residues begins to resemble a stacked arrangement, where alternating electrostatic and hydrophobic domains become apparent. The NH_3^+ and COO^- termini facilitate the stabilization of these layers through strong electrostatic interactions, while the hydrophobic Phe side chains cluster together, further stabilizing the overall structure.

At Phe400, the aggregation reaches a highly interconnected state, where distinct layers are clearly visible. These layers are composed of alternating electrostatic interactions between NH_3^+ and COO^- termini and hydrophobic interactions between Phe side chains. The system becomes more compact, significantly reducing solvent exposure, indicating a strong self-assembly tendency. This organization suggests a shift from simple chain-like assemblies to a more defined fibril-like structure as the concentration increases.

Finally, in the Phe500 system, a highly compact and organized aggregate is observed. The structure is dominated by well-defined layers, where hydrophobic Phe side chains cluster to form a hydrophobic core, while NH_3^+ and COO^- termini interact extensively to stabilize the assembly. This results in a highly ordered self-assembled structure with alternating electrostatic and hydrophobic layers. The drastic reduction in solvent exposure further confirms the increased stability and compactness of the aggregates at high concentrations.

The formation of hydrogen bonds (H-bonds) between nitrogen (N) and oxygen (O) atoms in phenylalanine (Phe) monomers across different systems is illustrated in Fig. 4. In the Phe100 system (Fig. 4a), the monomers approach each other and maintain stable non-covalent intermolecular interactions via their amino (NH_3^+) and

carboxyl (COO^-) groups. These interactions result in the formation of linear assemblies, where the H-bonds are represented as dashed lines, with nitrogen atoms in blue and oxygen atoms in red.

In the Phe200 system (Fig. 4b), a zigzag pattern of hydrogen bonds is observed between the monomers. This pattern indicates a more interconnected network of chains due to the increased number of electrostatic interactions between NH_3^+ and COO^- termini.

As the concentration increases further, the Phe300 system (Fig. 4c) exhibits more complex aggregation patterns. The intermolecular interactions via amino and carboxyl groups occur more rapidly due to the higher concentration, leading to the formation of multilayered structures. The H-bonds form a more complex and stabilized network, contributing to the overall stability and compactness of the aggregates.

These findings highlight the concentration-dependent mechanism of Phe aggregation, where electrostatic interactions between NH_3^+ and COO^- termini, along with hydrophobic interactions between Phe side chains, drive the formation of ordered aggregates. At low concentrations, linear structures predominate, while at higher concentrations, hierarchical, layered organizations emerge. This structural transformation suggests that Phe self-assembly is governed by a balance of electrostatic and hydrophobic interactions, leading to the formation of fibril-like aggregates.

In addition to other non-covalent interactions, π-π interactions between the aromatic rings of phenylalanines are observed. These π-π interactions, though weak between two π systems, play a significant role in the association between aromatic rings. Two types of π interactions can occur: face-to-face and edge-to-face. In this work, we primarily observed the face-to-face interaction, where the aromatic rings are located parallel to each other.

Figure 5 illustrates this type of interaction between aromatic rings, represented by green surfaces. Face-to-face π-π stacking interactions contribute to the stability and organization of the phenylalanine aggregates, highlighting the importance of these interactions in overall structural formation.

In order to verify the interactions formed between the phenylalanine monomers, radial distribution function (RDF) analysis was employed. The g(r) between the alpha carbons of phenylalanine (CA) was investigated to clarify the arrangement or stacking of aromatic rings. Figure 6a shows the g(r) Analysis of CA-CA atomic pairs, revealing two prominent symmetrical peaks. The first peak occurs at a distance of 4.9 Å, And the second at

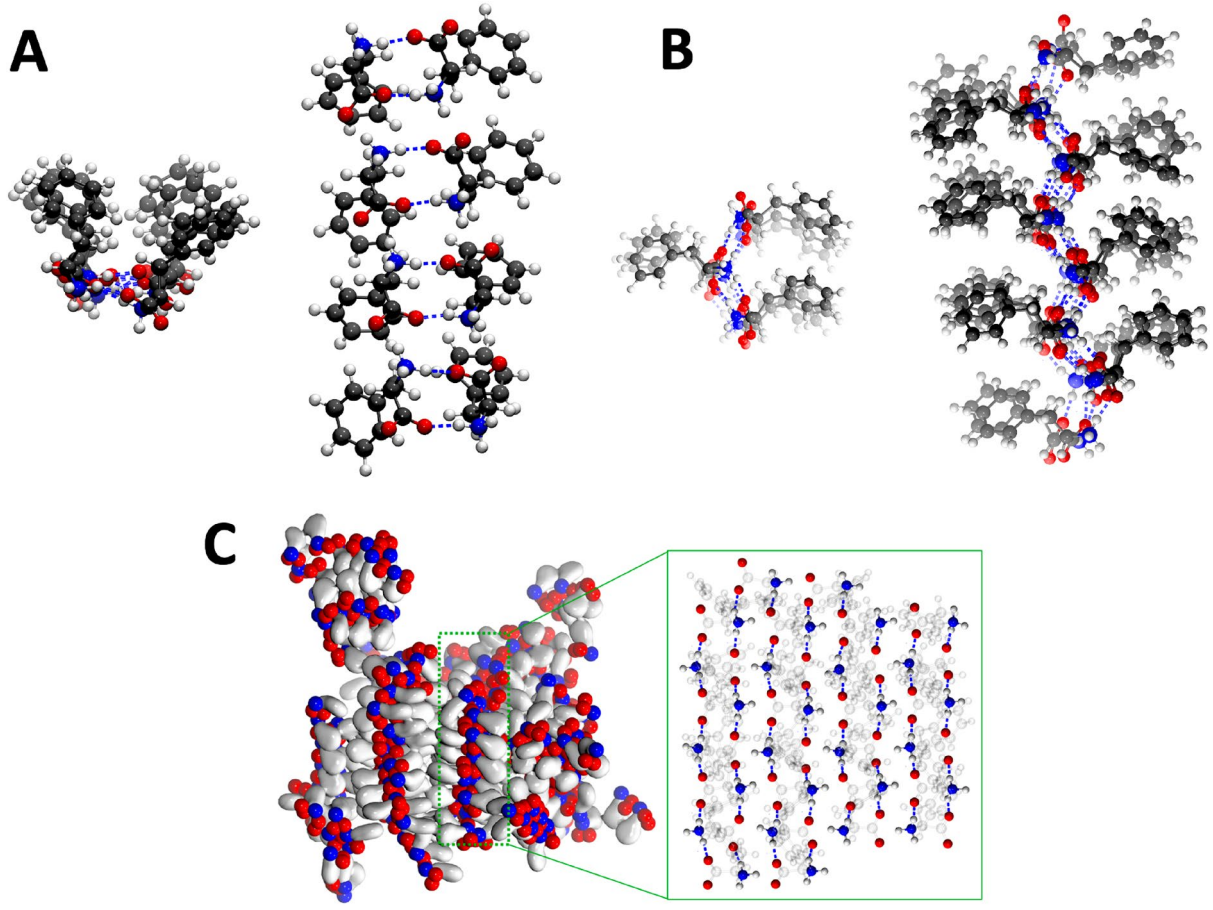

Fig. 4. Hydrogen bond representation between N–O. (**a**) and (**b**) Top and side views of the hydrogen bonds formed by phenylalanine monomers in the Phe100 and Phe200 systems, respectively, taken from the last frame. (**c**) Conformation of the phenylalanine fiber in the Phe300 system. Nitrogens are represented in blue, oxygens in red, and hydrogen bonds as dashed lines.

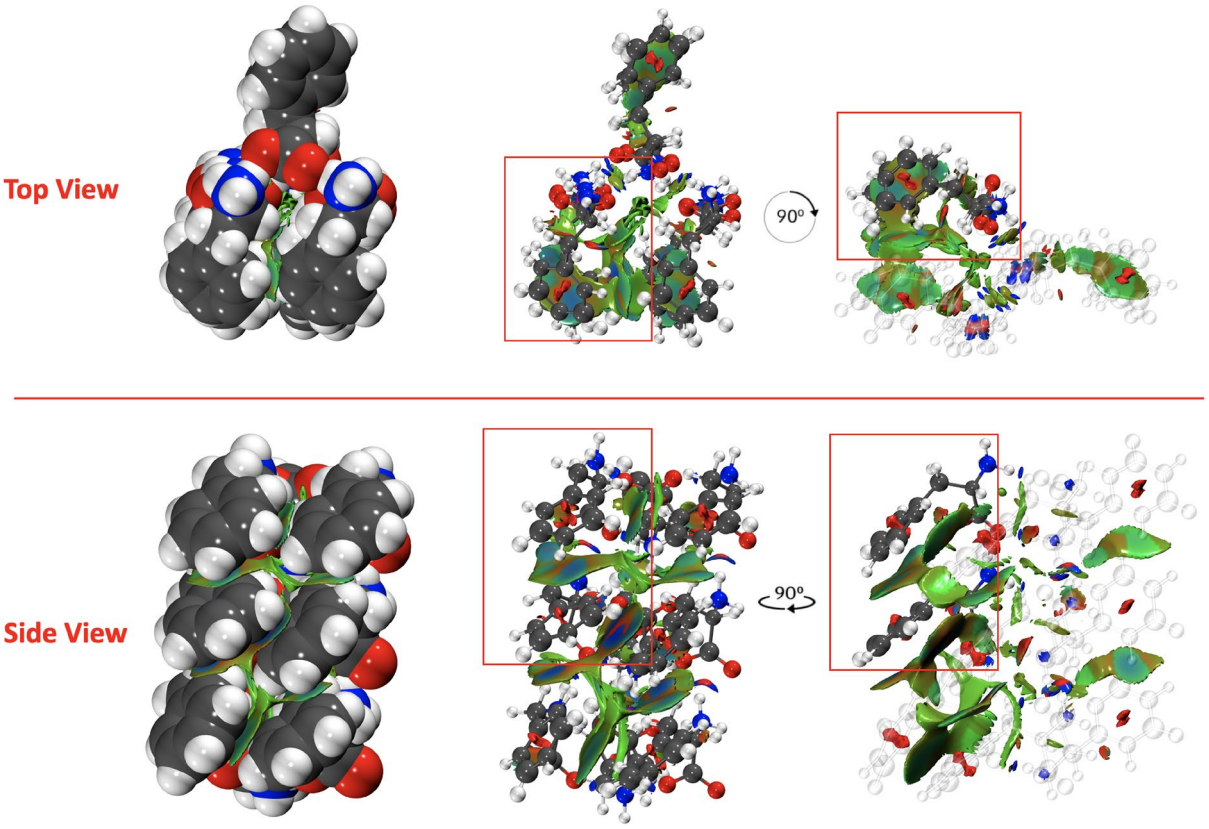

Fig. 5. π-π stacking interaction between neighbouring aromatic rings represented by green surfaces of Phe200 system.

5.4 Å. The corresponding cumulative curve of CA-CA atomic pairs (Fig. 6c) indicates that, on average, 1 to 2 phenylalanine molecules are coordinated.

The accumulative number for the two peaks is relatively independent of concentration, suggesting that the pattern of ring interactions does not change significantly with concentration. The cumulative number represents the average number of neighboring atoms or molecules located within a given distance from a reference atom. The higher g(r) values for systems with lower concentration indicate that once these interactions are formed, they remain in solution, resulting in a higher density in comparison to the average density of a uniform distribution of Phe.

Furthermore, the RDF of nitrogen and oxygen atoms in phenylalanine suggests the presence of structural ordering (Fig. 6b). This is evidenced by two to three well-defined peaks in the RDFs for atomic pairs of N-O, with a prominent peak at a distance of 2.7 Å. This first peak corresponds to the intermolecular interaction between N-O atomic pairs. The corresponding cumulative number for N-O interactions (Fig. 6d) shows that, on average, 1 cumulative number is observed in all situations. Thus, independently of the concentration, the electrostatic interaction of the positive terminus is always counterbalanced by the negative charge of another residue in close contact.

Overall, this result suggests that once aggregation begins to form, it becomes highly stable. Therefore, any therapeutic inhibition strategy should act before aggregates develop to prevent the accumulation of Phe monomers in biological systems.

Additionally, the phenylalanine has a para-substituted phenyl ring, with a hydrogen atom at the CZ carbon and the CG carbon opposite to CZ. The g(r) analysis for the CZ-CZ pair atoms of the phenylalanine rings (Fig. 7a) shows a prominent peak at a distance of 4.5 Å. This result indicates that the phenylalanine rings are preferentially positioned one above and one below each phenylalanine monomer, suggesting a stacked arrangement. The CZ-CZ pattern is similar to that of CA-CA; however, CA-CA shows more order with two better-resolved peaks. This increased order can be attributed to the reduced mobility of CA in comparison to CZ, as CA is closer to the electrostatic termini of the Phe.

Figure 7b shows the g(r) of CZ-CG, where the highest peak occurs at a distance of 6.4 Å. This average value represents the distance between the phenylalanine rings, indicating the spatial arrangement and interaction distance between the aromatic rings of phenylalanine monomers. The cumulative number pattern for CZ-CG is similar to that of CZ-CZ, indicating the mobility of this arrangement during simulation. These distances are compatible with ring-ring stacking, although the rings are not fixed during the simulation and exhibit some shifting.

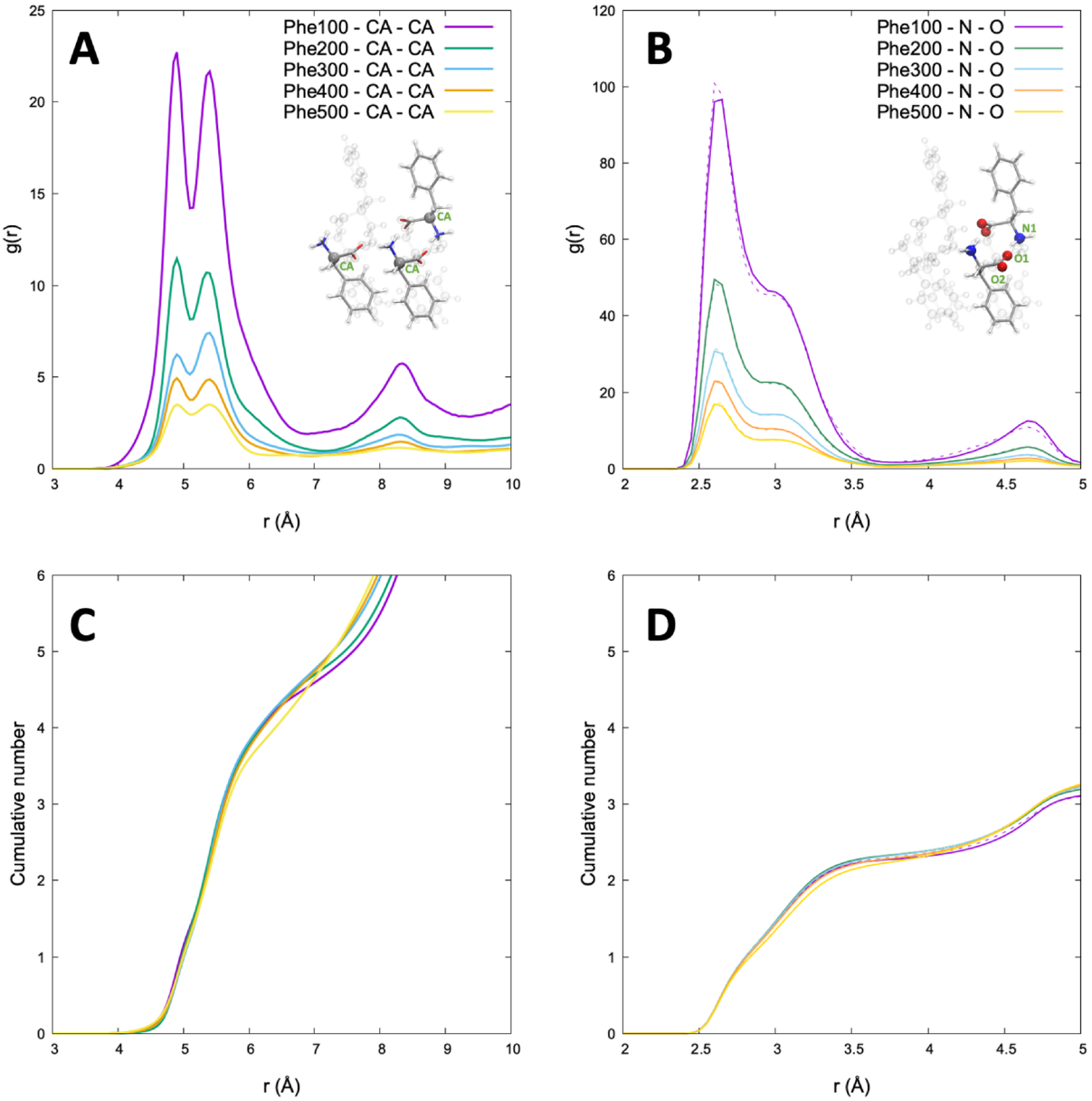

Fig. 6. Radial distribution function between the atoms of CA-CA and N-O for zwitterionic phenylalanine. (**a**) g(r) of CA-CA, with the inset indicating the alpha carbon atoms involved in the corresponding g(r). (**b**) g(r) of N-O, with the inset indicating the nitrogen and oxygen atoms involved in the corresponding g(r). (**c**) Cumulative number of CA-CA. (**d**) Cumulative number of N-O.

In Fig. 7c and d, the cumulative numbers of CZ–CZ and CZ–CG contacts, respectively, are presented. These plots reveal that, at equivalent distances, systems with higher Phe concentrations exhibit greater cumulative numbers. This indicates that elevated concentrations favor an increased number of Phe–Phe contacts, reflecting enhanced intermolecular packing and stronger aromatic interactions among phenylalanine residues.

Coaggregation of alanine with phenylalanine fibers.

Phenylalanine has a well-documented ability to self-assemble into amyloid-like fibrils under physiological conditions, even in the absence of enzymatic or chaperone activity[14,30]. These fibrils, enriched in β-sheet structures, possess a strong tendency to engage in intermolecular interactions such as hydrophobic contacts and hydrogen bonding. Importantly, phenylalanine aggregates can act as nucleating templates that promote the co-aggregation of other amino acids, including those that are not inherently prone to aggregation[31]. Alanine, a small and nonpolar residue, has been shown to coaggregate efficiently with phenylalanine due to its structural compatibility with the hydrophobic microenvironment of the fibrils and its ability to form stable backbone interactions[14]. This templating effect positions phenylalanine fibrils as potential scaffolds for heterogeneous aggregation processes relevant in both pathological and material science contexts.

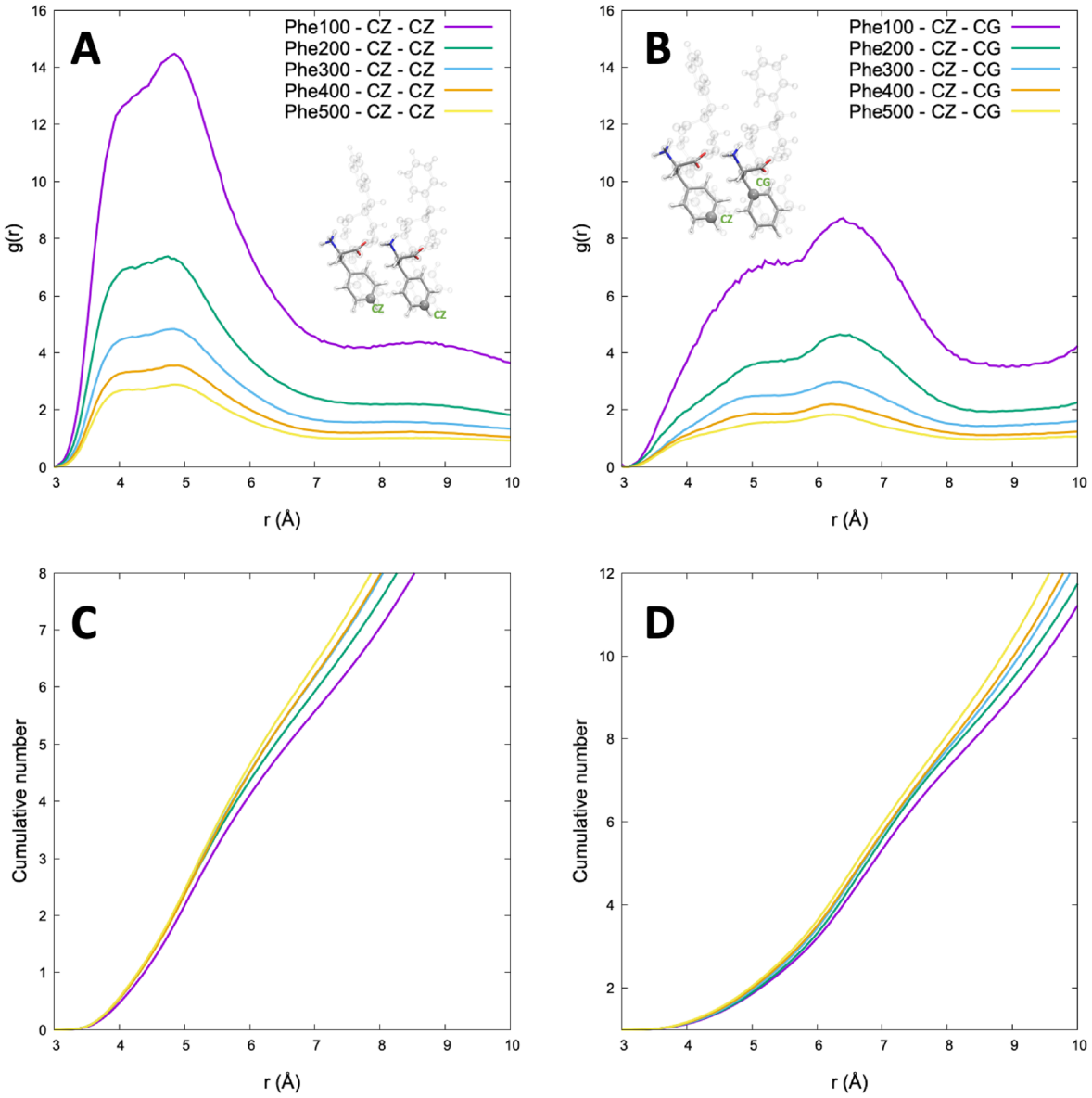

Fig. 7. Radial distribution function between the atoms of CZ-CZ and CZ-CG for zwitterionic phenylalanine. (**a**) g(r) of CZ-CZ, with the inset indicating the carbon atoms involved in the corresponding g(r). (**b**) g(r) of CZ-CG, with the inset indicating the carbon atoms involved in the corresponding g(r). (**c**) Cumulative number of CZ-CZ. (**d**) Cumulative number of CZ-CG.

Phenylalanine fibrils have been shown to promote the rapid aggregation of mixed amino acids under physiological conditions, highlighting their intrinsic ability to drive the co-aggregation of other residues[14]. Based on this property, the present study investigates the co-aggregation behavior of alanine in the presence of preformed phenylalanine fibrils derived from molecular dynamics simulations of the Phe200 system. Ten independent systems were constructed by embedding 100 alanine monomers into the preassembled phenylalanine fibers And subsequently evaluated over 500 ns of molecular dynamics simulations.

Figure 8 depicts the number of hydrogen bonds formed during the final 50 ns of the simulations. Hydrogen bonding was Analyzed across all ten replicas. The individual averages ranged from 108 to 140 hydrogen bonds, with standard deviations between 6 And 10, indicating moderate variability among the configurations. When treating all replicas as equivalent representations of the same system, a global average of approximately 115 hydrogen bonds was observed, with a standard deviation of 15. This result provides a robust estimate of the overall system behavior. The findings suggest that during the late phase of the simulation, the phenylalanine fibers establish stable and consistent interactions with the alanine molecules, with minor fluctuations likely attributable to replica-specific conformational or dynamic differences.

The RDF profile of the N–O distances for phenylalanine And alanine shows a prominent peak at 2.6 Å for both (see Fig. 9a). This peak suggests a substantially similar coordination sphere within the phenylalanine and

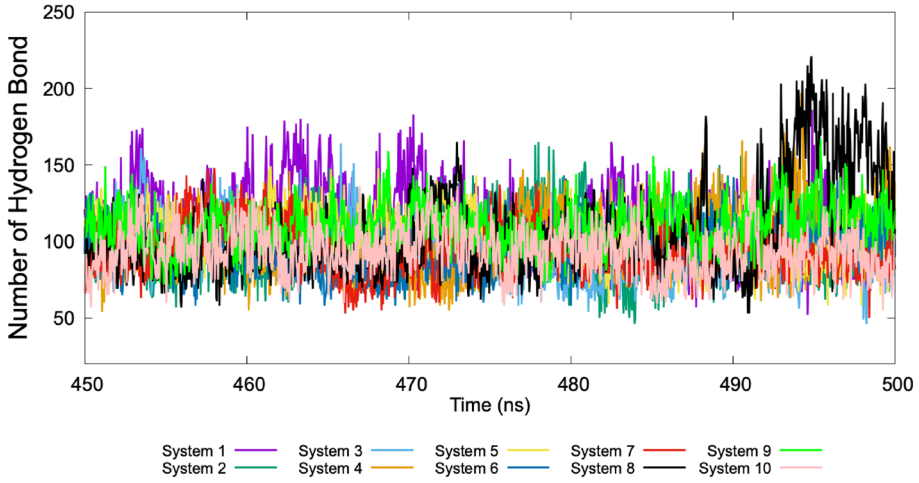

Fig. 8. Number of hydrogen bonds of the phenylalanine fiber with respect to the alanine monomers Analyzed in the last 50 ns.

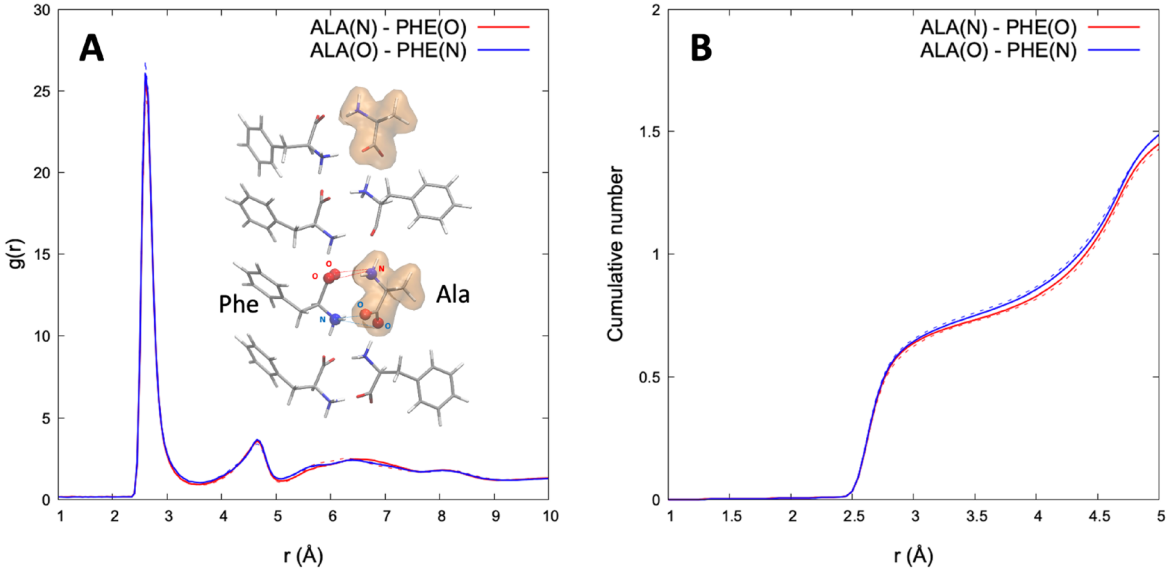

Fig. 9. Radial distribution function and cumulative number for the nitrogen and oxygen atoms of phenylalanine and alanine. (**a**) g(r) of N-O, with the inset indicating the nitrogen and oxygen atoms involved in the corresponding g(r). (**b**) Cumulative number of N-O.

alanine molecular systems. The presence of this peak indicates that the nitrogen and oxygen atoms consistently interact at this distance, reflecting a common structural feature in both amino acids.

Figure 9b presents the cumulative number of N-O interactions, further supporting the RDF findings. The cumulative number graph shows that, on average, the coordination spheres around the nitrogen and oxygen atoms are similar for both phenylalanine and alanine. This indicates that alanine residues interact with phenylalanine residues at either the N terminus or the carboxylic terminus with equal probability. This observation reinforces the idea that the interaction between alanine and phenylalanine is primarily electrostatic.

Figure 10 shows a detailed snapshot illustrating the coaggregation of alanine monomers with phenylalanine fibers. The alanine monomers are depicted as an orange-colored surface, attached to the terminal regions of phenylalanine fibers. The phenylalanine fibers are represented with blue spheres for nitrogen atoms of the amino groups and red spheres for oxygen atoms of the carboxyl groups. The aromatic rings of phenylalanine are visualized as a white van der Waals surface, linked by hydrophobic interactions. This visualization highlights that the phenylalanine fibers facilitate contact between the carboxyl and amino groups of phenylalanine residues situated on the surface of the fiber. The results indicate that the alanine monomers are attracted to the phenylalanine fibers primarily through hydrogen bonds, contributing to the overall stability and organization of the coaggregated structure.

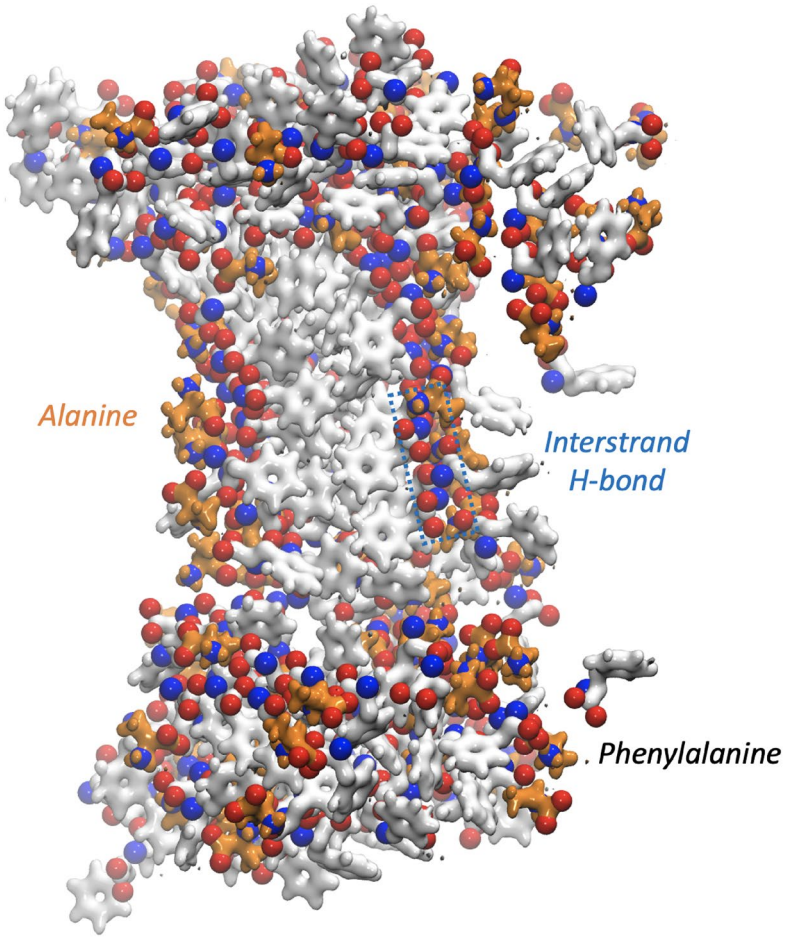

Fig. 10. Final snapshot of the coaggregation of alanine with phenylalanine fibers. Alanine monomers are depicted in orange. In phenylalanine, blue spheres represent nitrogen atoms of the amino groups, and red spheres indicate oxygen atoms of the carboxyl groups. The aromatic rings of phenylalanine are visualized as a white van der Waals surface.

This study provides detailed insights into the intrinsic self-assembly behavior of zwitterionic phenylalanine monomers under physiological temperature using all-atom molecular dynamics simulations. By systematically varying the concentration of Phe monomers and maintaining a constant water box size ($6\times6\times6\ \mathrm{nm}^3$) using the TIP5P water model and the OPLS-AA force field, we observed the progressive formation of ordered aggregates, suggesting that phenylalanine alone, without external cofactors, possesses the capacity to nucleate and stabilize supramolecular assemblies. Notably, coaggregation simulations using preformed Phe fibrils demonstrated the ability of these structures to recruit alanine molecules, further supporting phenylalanine's role as a seeding agent for non-aromatic residues. The use of 10 replicas ensured the robustness of the alanine incorporation analysis, while the dual-replica design for each Phe concentration enhanced the statistical reliability of the self-assembly observations.

Despite these contributions, several limitations must be acknowledged. First, the simulation timescale (500 ns) may not capture slower events such as fibril elongation or long-range rearrangements, which typically occur over micro to millisecond timescales. Second, the simulation environment was intentionally simplified: no ions or crowding agents were included, and pH was represented implicitly by using zwitterionic Phe, omitting dynamic protonation effects that may influence aggregation behavior. Moreover, the finite box size may constrain the full morphological development of extended fibrillar networks. Although the TIP5P model improves solvation realism and OPLS-AA captures relevant non-covalent interactions, additional experimental validation is required to confirm the structural motifs and biological relevance of the observed aggregates. Furthermore, this study does not assess the cytotoxic or functional impact of the aggregates, which is a key consideration in the context of phenylalanine accumulation-related disorders such as PKU.

Conclusion

This study provides a comprehensive analysis of the aggregation behavior of Phe and its interaction with Ala monomers. The RDF analysis revealed significant hydrogen bonding interactions between the amino (NH_3^+) and carboxyl (COO^-) groups of phenylalanine monomers. The presence of well-defined peaks in the RDF profiles indicates stable non-covalent interactions that contribute to the formation of linear and layered

Phenylalanine (molec.)	Solvent TIP5P (molec.)	χ	Molarity (M)
100	6153	0.016	0.77
200	5269	0.037	1.54
300	4387	0.064	2.31
400	3657	0.099	3.08
500	2829	0.150	3.85

Table 2. System composition: number of phenylalanine monomers, TIP5P water molecules, molar fraction (χ), and calculated molarity (M) in a 6 nm cubic simulation box.

aggregates. Additionally, the study identified face-to-face π-π stacking interactions between the aromatic rings of phenylalanine. These interactions were visualized through hydrophobic surfaces and confirmed by the corresponding RDF analysis of the ring moiety atoms. Therefore, identifying the type of interaction between Phe monomers allows the search for molecular compounds that alter the dynamics of the monomers and block the interactions between phenylalanine monomers that lead to aggregation.

The introduction of alanine monomers into phenylalanine fibers demonstrated a strong tendency for coaggregation, primarily mediated by hydrogen bonds. The alanine monomers were attracted to the terminal regions of the phenylalanine fibers, forming stable hydrogen bonds that contributed to the overall structural integrity. The RDF profiles for N-O distances in phenylalanine and alanine showed similar coordination spheres, indicating that alanine interacts with phenylalanine residues at the N terminus or carboxylic terminus with equal probability. This reinforces the idea that the interaction between alanine and phenylalanine is primarily electrostatic. Visualization of the aggregation identified that the coaggregation occurs at the surface of the phenylalanine fibers. Normally, when phenylalanine fibers bind to other dissolved amino acids, a series of interactions can occur that affect the aggregation dynamics, solubility, and toxicity of the aggregates, which have significant implications for health and cell biology. In this case, the interaction between alanine and phenylalanine fibers is stable, which could lead to the formation of pathological structures.

Overall, this study highlights the importance of both electrostatic and hydrophobic interactions in the aggregation behavior of phenylalanine and its coaggregation with alanine. The findings provide valuable insights into the molecular mechanisms driving the formation and stability of these aggregates, which could have implications for understanding similar processes in biological systems.

Computational details

Molecular Dynamics (MD) simulations have been performed with the Gromacs (Groningen Machine for Chemical Simulations) v. 2019.1 software[32] and OPLSAA[33] force field. The periodic boundary conditions (PBC)[34] to minimize edge effects in a finite system were considered. For each system, the molecules were located in the centre of a cubic box of 6 nm^3 and the box system was solvated with TIP5P[35] water model. In all molecular dynamics simulations, a cut-off distance of 0.9 nm was applied for both van der Waals and short-range electrostatic interactions to ensure an optimal balance between computational efficiency and accuracy in capturing relevant non-bonded interactions. Long-range electrostatics were treated using the Particle Mesh Ewald (PME) method with standard GROMACS settings Energy minimization was carried out using the steepest-descent algorithm with 200000 simulation steps. To equilibrate the system, 10 ns MD simulations were carried out in the canonical ensemble NVT with position restraint, And the temperature was regulated with the V-rescale thermostat at 309.65 K. To determine the average properties of the systems, An MD simulation was performed in isothermal-isobaric ensemble with the Parrinello-Rahman barostat with a reference pressure of 1 bar, V-rescale thermostat with 309.65 K And a simulation length of 500ns.

For the analysis of phenylalanine (Phe) self-assembly, five systems of varying concentrations of Phe monomers were prepared, specifically 100, 200, 300, 400, And 500 molecules, which for simplicity the different Phe concentrations will be expressed as the number of monomers required to achieve a specific mole fraction in biological systems, relative to the number of water molecules needed to maintain the appropriate density, depending on the simulation box. See Table 2, which shows the Phe concentration, number of water molecules, mole fraction, and molarity. To ensure reproducibility and assess the robustness of the aggregation dynamics, five independent simulations were performed for the Phe100 and Phe200 systems, whereas two independent simulations were conducted for Phe300, Phe400, and Phe500. Despite the different number of replicates, all systems exhibited consistent aggregation trends, which justified selecting one representative simulation for detailed analysis. Additionally, to explore the coaggregation behavior of alanine (Ala) with Phe fibrils, ten independent 500 ns simulations were carried out. In these simulations, a pre-formed Phe fiber derived from the final configuration of the 200-monomer system was used as the initial template, with 100 Ala residues added to evaluate the potential of Phe fibrils to recruit non-aromatic residues into the aggregate under controlled conditions.

To further corroborate our findings, the RMSD (Root-Mean Squared Deviation), SASA (Solvent Accessible Surface Area), and RDF (Radial Distribution Function) analyses were calculated using Gromacs tools. We also analyze the non-covalent interaction (NCI) using the Multiwfn program[36] to reveal intra And intermolecular interaction between phenylalanine monomers. Finally, the plots were made in Gnuplot v. 5.4[37] and the manuscript images were created and rendered by Visual Molecular Dynamics (VMD)[38] software.

Data availability

The data supporting the findings of this study were generated through all-atom molecular dynamics simulations. The simulation input files, trajectories, and analysis scripts are available from the corresponding author upon reasonable request.

Received: 26 May 2025; Accepted: 10 September 2025
Published online: 29 October 2025

References

1. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890. https://doi.org/10.1038/nature02261 (2003).
2. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. medicine* **10**, S10–S17. https://doi.org/10.1038/nm1066 (2004).
3. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. Amyloid fibrils as building blocks for natural and artificial functional materials. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396. https://doi.org/10.1002/adma.201505961 (2014).
4. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901 (2006).
5. Eisenberg, D. et al. The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Accounts of chemical research* **39**, 568–575. https://doi.org/10.1021/ar0500618 (2006).
6. Selkoe, D. J. *Folding proteins in fatal ways. nature* **426**, 900–904. https://doi.org/10.1038/nature02264 (2003).
7. Mukherjee, A., Morales-Scheihing, D., Butler, P. C. & Soto, C. Type 2 diabetes as a protein misfolding disease. *Trends in molecular medicine* **21**, 439–449. https://doi.org/10.1016/j.molmed.2015.04.005 (2015).
8. Cao, P. et al. Islet amyloid: from fundamental biophysics to mechanisms of cytotoxicity. *FEBS letters* **587**, 1106–1118. https://doi.org/10.1016/j.febslet.2013.01.046 (2013).
9. Westermark, P., Andersson, A. & Westermark, G. T. Amyloid in the islets of langerhans: thoughts and some historical aspects. *Upsala J. Med. Sci.* **116**, 81–89. https://doi.org/10.3109/03009734.2011.573884 (2011).
10. Singh, V., Rai, R. K., Arora, A., Sinha, N. & Thakur, A. K. Therapeutic implication of l-phenylalanine aggregation mechanism and its modulation by d-phenylalanine in phenylketonuria. *Sci. reports* **4**, 1–8. https://doi.org/10.1038/srep03875 (2014).
11. Do, T. D., Kincannon, W. M. & Bowers, M. T. Phenylalanine oligomers and fibrils: the mechanism of assembly and the importance of tetramers and counterions. *J. Am. Chem. Soc.* **137**, 10080–10083. https://doi.org/10.1021/jacs.5b05482 (2015).
12. Adler-Abramovich, L. et al. Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria. *Nat. chemical biology* **8**, 701–706. https://doi.org/10.1038/nchembio.1002 (2012).
13. Tomar, D., Chaudhary, S. & Jena, K. C. Self-assembly of l-phenylalanine amino acid: electrostatic induced hindrance of fibril formation. *RSC advances* **9**, 12596–12605. https://doi.org/10.1039/C9RA00268E (2019).
14. Anand, B. G., Dubey, K., Shekhawat, D. S. & Kar, K. Intrinsic property of phenylalanine to trigger protein aggregation and hemolysis has a direct relevance to phenylketonuria. *Sci. reports* **7**, 1–9. https://doi.org/10.1038/s41598-017-10911-z (2017).
15. Guo, C. et al. Expanding the nanoarchitectural diversity through aromatic di-and tri-peptide coassembly: Nanostructures and molecular mechanisms. *ACS nano* **10**, 8316–8324. https://doi.org/10.1021/acsnano.6b02739 (2016).
16. Bobrova, E. & Kuznetsov, I. Insights into the structural features and toxicity mechanisms of phenylalanine aggregates: A computational perspective. *Int. J. Mol. Sci.* **26**, 512. https://doi.org/10.3390/ijms26020512 (2025).
17. Nigdelioglu, E. et al. Identification of d-and l-phenylalanine enantiomeric mixtures by employing deep neural network models. *J. Mol. Struct.* **1304**, https://doi.org/10.1016/j.molstruc.2024.137628 (2024).
18. Bobrova, N. et al. Experimental animal models of phenylketonuria: Pros and cons. *Int. J. Mol. Sci.* **26**, 5262. https://doi.org/10.3390/ijms261152 (2025).
19. Arias-Olivares, D., Wieduwilt, E. K., Contreras-García, J. & Genoni, A. Nci-elmo: A new method to quickly and accurately detect noncovalent interactions in biosystems. *J. Chem. Theory Comput.* **15**, 6456–6470. https://doi.org/10.1021/acs.jctc.9b00658 (2019).
20. Riley, K. E. & Hobza, P. Noncovalent interactions in biochemistry. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 3–17. https://doi.org/10.1002/wcms.8 (2011).
21. Chen, W. et al. Water-induced formation of a chiral phenylalanine derivative supramolecule. *Phys. Chem. Chem. Phys.* **20**, 4144–4148. https://doi.org/10.1039/C7CP05845D (2018).
22. Al-Hamdani, Y. S. & Tkatchenko, A. Understanding non-covalent interactions in larger molecular complexes from first principles. *The J. chemical physics* **150**, https://doi.org/10.1063/1.5075487 (2019).
23. Juanes, M., Saragi, R. T., Caminati, W. & Lesarri, A. The hydrogen bond and beyond: Perspectives for rotational investigations of non-covalent interactions. *Chem. Eur. J.* **25**, 11402–11411. https://doi.org/10.1002/chem.201901113 (2019).
24. Jena, S. et al. Noncovalent interactions in proteins and nucleic acids: Beyond hydrogen bonding and π-stacking. *Chem. Soc. Rev.* https://doi.org/10.1039/D2CS00133K (2022).
25. Singh, P. et al. Self-assembly of aromatic α-amino acids into amyloid inspired nano/micro scaled architects. *Mater. Sci. Eng. C* **72**, 590–600. https://doi.org/10.1016/j.msec.2016.11.117 (2017).
26. Li, T., Lu, X.-M., Zhang, M.-R., Hu, K. & Li, Z. Peptide-based nanomaterials: Self-assembly, properties and applications. *Bioact. Mater.* **11**, 268–282. https://doi.org/10.1016/j.bioactmat.2021.09.029 (2022).
27. German, H. W., Uyaver, S. & Hansmann, U. H. Self-assembly of phenylalanine-based molecules. *The J. Phys. Chem. A* **119**, 1609–1615. https://doi.org/10.1021/jp5077388 (2014).
28. Uyaver, S., Hernandez, H. W. & Habiboglu, M. G. Self-assembly of aromatic amino acids: a molecular dynamics study. *Phys. Chem. Chem. Phys.* **20**, 30525–30536. https://doi.org/10.1039/C8CP06239K (2018).
29. Uyaver, S. Tyrosine, phenylalanine, and tryptophan undergo self-aggregation in similar and different manners. *Atmosphere* **13**, 1448. https://doi.org/10.3390/atmos13091448 (2022).
30. Wei, G. et al. Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chem. Soc. Rev.* **46**, 4661–4708. https://doi.org/10.1039/C6CS00542J (2017).
31. Singh, P., Pandey, S. K., Grover, A., Sharma, R. K. & Wangoo, N. Understanding the self-ordering of amino acids into supramolecular architectures: co-assembly-based modulation of phenylalanine nanofibrils. *Mater. Chem. Front.* **5**, 1971–1981. https://doi.org/10.1039/D0QM00784F (2021).
32. Van Der Spoel, D. et al. Gromacs: fast, flexible, and free. *J. computational chemistry* **26**, 1701–1718. https://doi.org/10.1002/jcc.20291 (2005).
33. Kaminski, G. & Jorgensen, W. L. Performance of the amber94, mmff94, and opls-aa force fields for modeling organic liquids. *The J. Phys. Chem.* **100**, 18010–18013. https://doi.org/10.1021/jp9624257 (1996).
34. Makov, G. & Payne, M. Periodic boundary conditions in ab initio calculations. *Phys. Rev. B* **51**, 4014. https://doi.org/10.1103/PhysRevB.51.4014 (1995).
35. Mahoney, M. W. & Jorgensen, W. L. Diffusion constant of the tip5p model of liquid water. *The J. Chem. Phys.* **114**, 363–366. https://doi.org/10.1063/1.1329346 (2001).
36. Lu, T. & Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *J. computational chemistry* **33**, 580–592. https://doi.org/10.1002/jcc.22885 (2012).

37. Racine, J. gnuplot 4.0: a portable interactive plotting utility. *J. Appl. Econom.* **21**, 133–141. https://doi.org/10.1002/jae.885 (2006).
38. Humphrey, W. et al. Vmd: visual molecular dynamics. *J. molecular graphics* **14**, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5 (1996).

Acknowledgements
H.L.B.-C. acknowledges the financial support provided by the internal research project of the Universidad Católica de Santa María under the grant 29002-R-2022. S.M. and F.M. acknowledge financial support from the Generalitat de Catalunya (Grant 2021SGR00350) and the Spanish Ministry of Science and Innovation (PID2023-150539OB-I00). S.M., F.M., and H.LB.-C. acknowledge the Spanish Structures of Excellence María de Maeztu program through Grant No. CEX2021-001202-M. The authors gratefully acknowledge the computer resources at Canigo (CSUC, Generalitat de Catalunya) and Xula (CIEMAT) and the technical support provided by Marenostrum at the Barcelona Supercomputing Center (BSC), part of the Spanish Supercomputing Network (RES), through projects QH-2021-2-0007 and QH-2023-2-0007.

Author contributions
H.B.-C. and S.M. wrote the original draft. H.B.-C.: Methodology, Investigation, Conceptualization. F.M.: Methodology and Conceptualization. S.M: Methodology, Conceptualization and Resources. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.