Computational and Structural Biotechnology Journal 19 (2021) 1595–1602

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Dissecting the role of glutamine in seeding peptide aggregation

Exequiel E. Barrera [a], Francesco Zonta [b,*], Sergio Pantano [b,c,*]

[a] Instituto de Histología y Embriología (IHEM) - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CC56, Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina
[b] Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai 201210, China
[c] Biomolecular Simulations Group, Institut Pasteur de Montevideo, Mataojo 2020, CP 11400 Montevideo, Uruguay

ARTICLE INFO

Article history:
Received 17 November 2020
Received in revised form 18 February 2021
Accepted 20 February 2021
Available online 13 March 2021

Keywords:
Polyglutamine diseases
Peptide aggregation
Toxic oligomers
Molecular dynamics
Coarse grained modelling

ABSTRACT

Poly glutamine and glutamine-rich peptides play a central role in a plethora of pathological aggregation events. However, biophysical characterization of soluble oligomers -the most toxic species involved in these processes- remains elusive due to their structural heterogeneity and dynamical nature. Here, we exploit the high spatio-temporal resolution of coarse-grained simulations as a computational microscope to characterize the aggregation propensity and morphology of a series of polyglutamine and glutamine-rich peptides. Comparative analysis of ab-initio aggregation pinpointed a double role for glutamines. In the first phase, glutamines mediate seeding by pairing monomeric peptides, which serve as primers for higher-order nucleation. According to the glutamine content, these low molecular-weight oligomers may then proceed to create larger aggregates. Once within the aggregates, buried glutamines continue to play a role in their maturation by optimizing solvent-protected hydrogen bonds networks.

1. Introduction

Protein misfolding, aggregation and their eponymous diseases have been vastly studied since the pioneering work of Alois Alzheimer in the beginnings of the twentieth century [1]. After Alzheimer's disease, multiple neurodegenerative diseases/protein dyads have been described, including frontotemporal dementia/TDP-43; Creutzfeldt-Jakob's/prion-protein or Parkinson's/α-synuclein. Abnormal protein aggregation is also present in systemic diseases like in the amyloid cardiomyopathy/transthyretin and type-2 diabetes/islet amyloid polypeptide. A special neurodegenerative subcategory is represented by the inherited polyglutamine diseases, where the neurotoxicity correlates with the length of consecutive glutamine tracts in the corresponding proteins. Some of the most studied glutamine related disease/protein dyads include Huntington's/huntingtin; Kennedy's/androgen receptor; spinocerebellar ataxia/ataxin-1, and dentatorubral-pallidoluysian atrophy/atrophin-1.

In the vastly studied Alzheimer's amyloid cascade hypothesis, published by Hardy et al. in 1992, the key event proposed to trigger neurotoxicity is the formation of insoluble protein aggregates [2]. However, Lambert et al. changed this paradigm, in 1998, shifting the investigation focus towards soluble oligomers as the main responsible for inhibition of long-term synaptic plasticity [3]. Since then, all the above-mentioned protein aggregation diseases have been associated with these soluble, small aggregated species [4–12].

Although aggregation is clearly a multifactorial process, it may be determined by the relative abundance of certain amino acids in short protein motifs. In particular, the presence of glutamines is recognized as a potential trigger for the formation of oligomeric species [13].

The experimental structural determination of protein oligomeric species is extremely challenging because of their transient nature and structural heterogeneity [14]. To cope with these difficulties, many computational studies have addressed the interaction modes of numerous aggregating peptides, including polyglutamine peptides, providing useful mechanistic insight [15–24].

Despite significant efforts, a comprehensive picture of the role of glutamine residues in different contexts is still missing. Among other reasons, this void originates in the substantial computational cost of atomically detailed simulations. This problem rapidly upscales, as different peptide sequences, multiple copies, and conditions are necessary to obtain a proper generalization in biologically relevant timescales.

* Corresponding authors at: Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai 201210, China (F. Zonta), and at Institut Pasteur de Montevideo, Montevideo CP11400, Uruguay (S. Pantano).

E-mail addresses: fzonta@shanghaitech.edu.cn (F. Zonta), spantano@pasteur.edu.uy (S. Pantano).

https://doi.org/10.1016/j.csbj.2021.02.014

In 2010, Voth et al. [21] presented a solvent-free coarse-grained (CG) model for polyglutamine peptides using a multiscale coarse-graining method. They derived CG parameters from equilibrium all-atom MD simulations by sampling instantaneous atomic positions and forces. However, bottom-up CG models like this may find difficulties to describe thermodynamic properties, specifically when applied to heterogeneous systems [25]. Nevertheless, they reported an increase in the compaction of the structures when increasing the repeat length; although contrarily to experimental findings, the secondary content obtained were rich α-helices. Using the same CG model Yang et al. [26] studied the influence of concentration and temperature on polyglutamine aggregation. Alternatively, Rickman et al. [19] focused on the effects of supersaturation over the nucleation mechanisms of Q10 peptides using the PLUM implicit solvent CG model [27]. They could simulate large-scale systems formed by 1000 monomers and determine Q10's critical nucleus size. Pappu et al. [18] simulated the modulation exerted by heterogeneous flanking regions in polyglutamine aggregation. They used an ultra-CG model, where each peptide is represented by one spherical particle. While showing interesting insights into how flanking regions in huntingtin modulated its aggregation kinetics and macroscopic shape, no secondary or tertiary structural information was reported because of the minimalist representation of the model.

We circumvent these limitations by employing cutting-edge CG simulations to examine glutamine's role in early aggregation events. For this, we employed the SIRAH force field for CG simulations (see Methods), which has been shown to grant nearly atomistic resolution for folded proteins [28] and a very good description of global and local descriptors of intrinsically disordered proteins and peptides [29]. Worth nothing, our force field has shown a remarkable sensibility to point mutations in proteins and peptides [29,30].

We studied a series of polyglutamine (poly-Q) and glutamine-rich (Q-rich) peptides displaying self-aggregative behavior. A comparative analysis among homogeneous poly-Qs of different lengths and heterogeneous Q-rich peptides shows that glutamines play a double role in seeding aggregation. First, they participate in initial inter-monomer contacts, to then become the dominant residue mediating oligomer association providing solvent-buried electrostatic stabilization.

Taken as a whole, the homogeneous set of simulations presented here offer original insight into the mechanistic role of glutamine in the initiation of pathogenic aggregation. Complementary, we present highly detailed scale and time resolution results that substantiate mechanistic pathways previously proposed by low-resolution experimental techniques [31,32].

2. Results

Aimed to establish a comparative baseline for the aggregation of Q-rich peptides, we first investigated homogeneous poly-Q, progressively increasing the length of the peptides. To this aim we considered three different peptide lengths: Q4, Q11, and Q36. All simulations were performed at the CG level and backmapped to atomistic representation (see Methods) to facilitate the visualization. It is important to note that despite the CG nature of our approach, the mapping maintains the most salient characteristics of each amino acid, in particular those of glutamine (Supplementary Fig. 1). Namely, the presence of partial charges in the side chain and backbone moieties retains the capacity to form electrostatic interactions. These interactions are alike but not completely equivalent to hydrogen bonds because of the obvious absence of explicit hydrogen atoms, which may affect the directionality existent between Donor-Hydrogen-Acceptor.

In the remaining of this section we present the results of our simulations, which are compared against experimental reports in the Discussion and Conclusion section.

2.1. Tetraglutamine peptide (Q4)

Q4 in solution showed a quick transition from their initial, fully extended conformation towards a broad backbone-angle distribution, scattered on the upper-left quadrant of the Ramachandran plot (Supplementary Fig. 2A). Along the time, Q4 peptides showed a continuous interchange between monomeric, and short-lived oligomeric species with a mean cluster size (MCS) of 1.94 ± 0.1 (Fig. 1A). MCS can adopt values ranging from 1 (when all peptides in solution are monomeric species) up to the total number of simulated peptides (27, for a full aggregation in the Q4 case; see Methods). As seen from Fig. 1B, monomeric species are dominant in solution, followed by dimers and trimers. Analysis of the global secondary structure (Fig. 1C) evidenced only a small amount of extended and isolated β-structures, as a result of the fleeting character of the small clusters observed during the simulation (see molecular drawings in Fig. 1C).

2.2. Polyglutamine peptide Q11

Compared with the Q4 system, the longer Q11 peptide explored wider areas of the Ramachandran plot (Supplementary Fig. 2B).

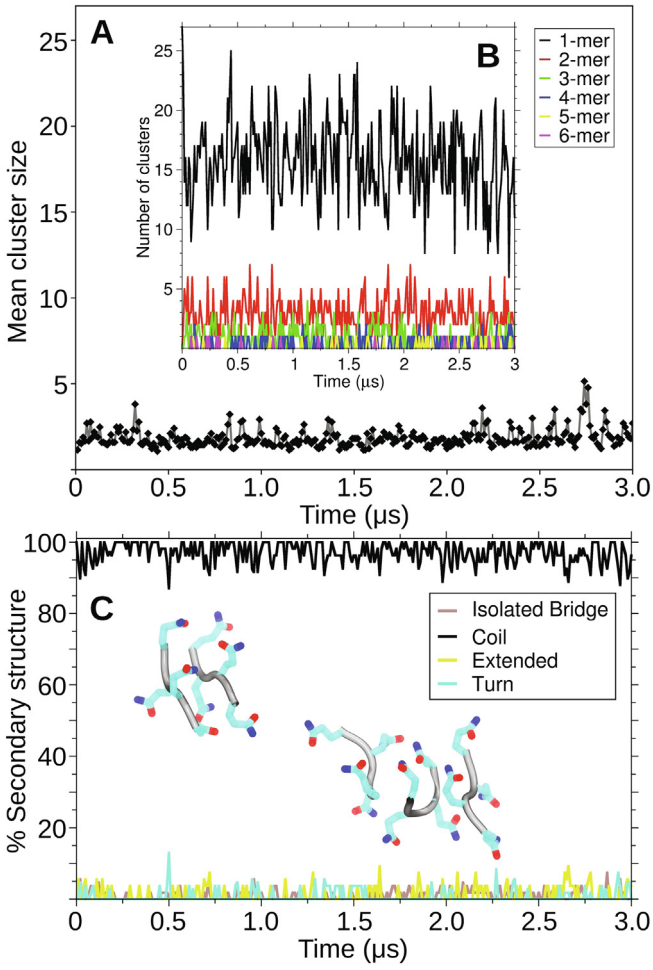

Fig. 1. Aggregation analysis for Q4 peptides. (A) Mean cluster size measured throughout 3 μs MD simulations. (B) Number of clusters classified by size. (C) Global percentage of secondary structure. Representative structures of Q4 dimers and trimers are drawn with a cartoon representation. Side chains are colored by element.

However, its length was insufficient to generate stable β hairpins-like motifs, and Q11 sampled mostly linear configurations. In contrast with the previous case, the MCS calculation indicated that the interaction between Q11 occurs in two well-differentiated fashions (Fig. 2B). Initially, random encounters led to the formation of dimers or trimers (Fig. 2A), which then served as nucleation seeds. Subsequently, all Q11 peptides cooperatively associated within a short time window. The dynamics of this process lead to the MCS's saturation, aggregating all the peptides available in the simulation box. The progressive association of Q11 is accompanied by the passage from a random coil to parallel or antiparallel β-sheet conformations (Fig. 2C). Indeed, the time series of the secondary structure for coil and β-sheets shows a scissors-like graph with a crossing point corresponding to the region where the associations become cooperative. In the final state, almost 80% of the peptides are in β-sheet conformation, i.e., only 1–2 amino acids at the extremities of each peptide remained disordered.

Next, we explored the capacity of smaller peptides to modulate Q11 aggregation. We focused on pentapeptides, as they could bind to the aggregates alone or in couples within the same β-strand, substituting one single Q11. We tested this idea on non-zwitterionic peptides Q5, and QEQQQ. Compared with the homogeneous Q11 system, both pentapetides sustained aggregation although altering the aggregation's kinetics (Fig. 2E, H). Indeed, different dissociation events occurred during the first half of the simulations, indicating that the peptides compete for pairing sites on the aggregates (see insets in Fig. 2E and H). Nevertheless, the content of β -extended conformations converged to indistinguishable values and, in all the cases, all Q11 peptides aggregated into decamers (Fig. 2F, I). However, the pentapeptides' presence, especially QEQQQ, modified the aggregate's final topology (inset Fig. 2H). The procedure presented here may provide a cost-effective computational strategy to test for different peptides' aggregation tendency or explore conditions to modify these propensities.

2.3. Polyglutamine peptide Q36

Increasing the polyglutamine size to 36 amino acids (Q36) conferred the poly-Q the capability to adopt turn conformations and form β-hairpins. The accessibility to a wider conformational variety translates into different interaction possibilities. Despite this, in all aggregation events monitored, monomers always interacted with a pre-formed β-stranded region in a neighboring molecule (Fig. 3). In close analogy with the Q11 case, after these association events, the β-sheet content increased reaching average values of 50.37% ± 5.32. The Q36 aggregates presented intricate structural motifs, showing turn percentages of 9.81% ± 2.5 (Fig. 3 and Supplementary Fig. 3A).

2.4. Q-rich peptides

The results presented in the previous section pointed out that despite the well-known aggregation propensity of poly-Q peptides, there is a threshold below which, they cannot sustain large scale aggregation. Therefore, we sought to characterize how both the glutamine content in Q-rich and Q-poor peptides can affect their aggregative behavior.

2.5. α-gliadin peptide (p31–43)

As the first example of Q-rich peptide, we choose a derivative of the human protein α-gliadin. The peptide named p31–43 carries five glutamines in 13 amino acids (second row, Fig. 4A) and is a proteolytic, gluten derived, peptide generated in the stomach and related to celiac disease [33]. Using biophysical techniques, we previously showed that this peptide undergoes spontaneous aggregation with a concomitant but limited increase in the β-strand content [34].

Here, to focus on the effect of glutamine, we simulated an ensemble of non-zwitterionic p31–43 and monitored the contribution of different amino acids during the aggregation process.

In contrast with poly-Q peptides, heterogeneous peptide aggregation followed different dynamics. According to the aggregation rate, we divide the analysis in three different phases. On phase I, monomers rapidly associated to form low molecular-weight oligomers (top row, Fig. 4A). This was followed by a slower aggregation (phase II), where oligomers progressively interacted with each other until reaching a final amorphous 50-mer aggregate,

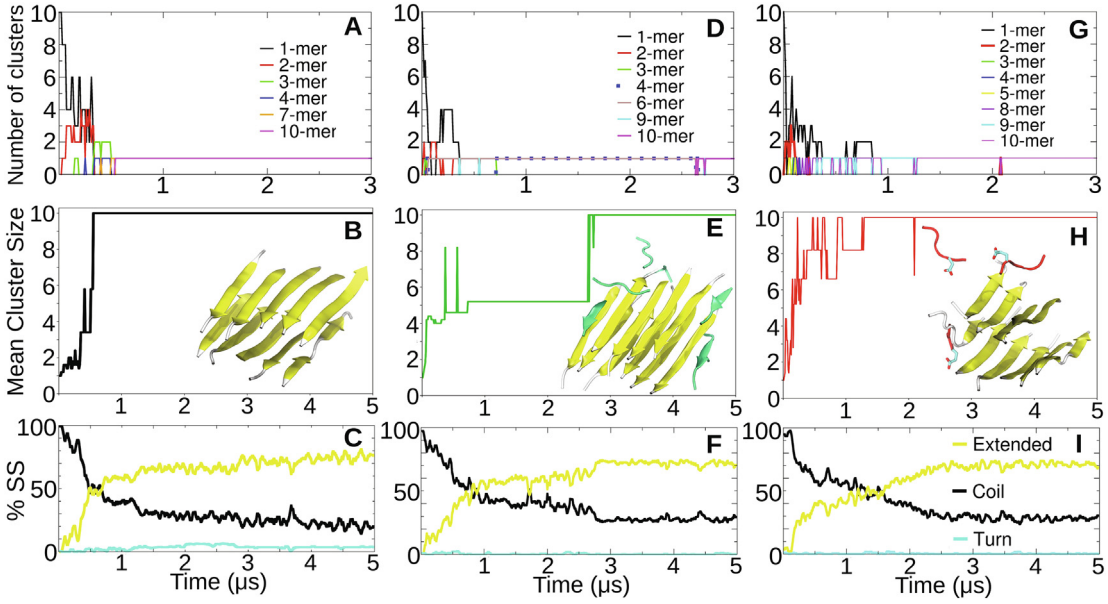

Fig. 2. Aggregation of Q11 peptides. (First row) Number of clusters classified by size. The analysis is restricted to the first 3 μs of the trajectory where all the association-dissociation events occur. (Second row) Mean cluster size measured throughout 5 μs of MD simulations. Final oligomeric structures are represented in cartoons. (Third row) Time evolution of the percentage of secondary structure content. Results correspond to Q11 (first column), Q11 + Q5 (second column) and Q11 + QEQQQ (third column).

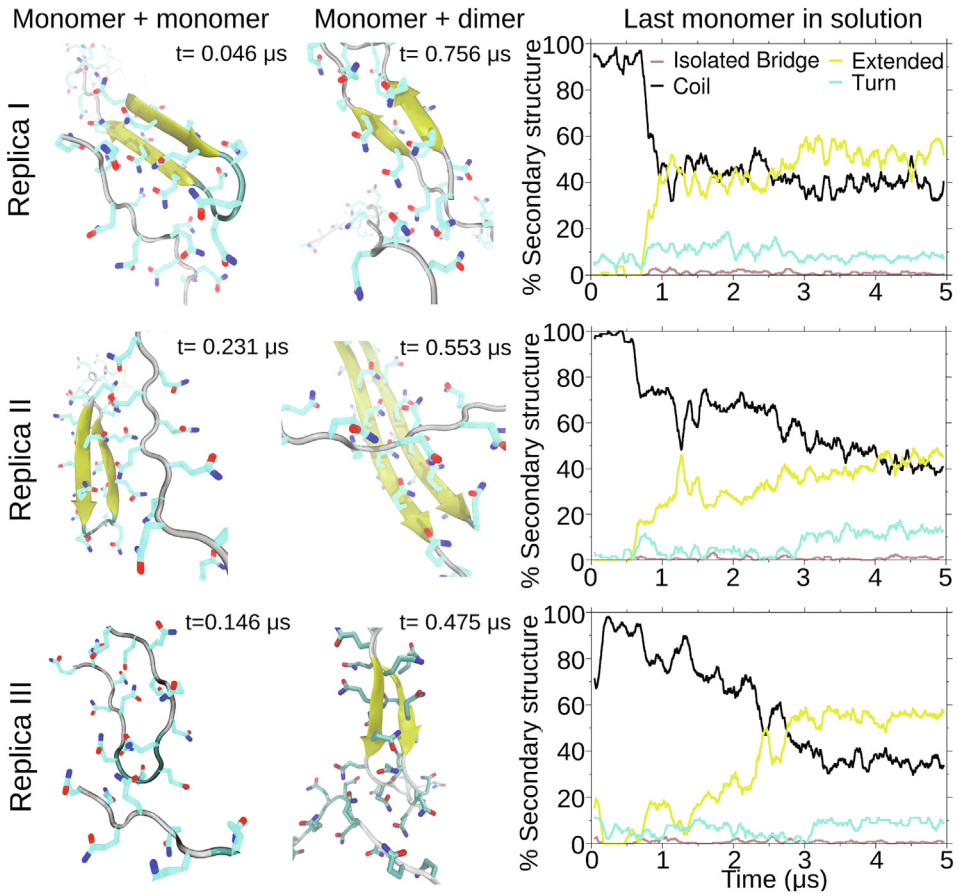

Fig. 3. Aggregation of Q36 peptides. Snapshots of the association events for each simulated replica are shown in the left grid. Q36 monomers first contacts always involved interactions with preformed double stranded regions of their counterparts. Secondary structure plots are shown in the right panels.

presenting a radius of gyration of 2.48 ± 0.01 nm. The global content of β-extended structures rose from zero to 28.08% ± 1.71. The cluster size remained stable until the end of the trajectory (phase III). A per-residue interpeptide contact analysis (third row, Fig. 4A) indicated that glutamines lead the p31–43 aggregation process, followed by hydrophobic residues (Leu, Phe, and Tyr), proline, and glycine, to a much lesser extent. This is in line with the relative amino acid abundances in this peptide. If we divide the contact analysis in the three above mentioned aggregation phases we can observe how glutamine-mediated contacts are increased in phase II (gray shaded time window in the third row of Fig. 4A and Supplementary video 1). In phase III, contacts established by all amino acids reach a plateau, except for glutamine. Therefore, not only glutamines mediate oligomer association but they also continue optimizing internal electrostatic contacts once all peptides are clustered. Indeed, the solvent accessible surface area (SASA) of glutamine residue showed different dynamics (lower row, Fig. 4A). The SASA reaches a plateau after 2.5 μs, decreasing to nearly 70% of its initial value, suggesting a sort of maturation within the final aggregated state.

2.6. Proinsulin's connecting peptide (C-peptide)

Aimed to further explore glutamines' role in an entirely unrelated system, we focused on insulin's C-peptide, which undergoes only limited aggregation. C-peptide is 31 amino acids long with only four—two of them highly conserved—glutamines [35] (second row, Fig. 4B). However, it also contains four conserved glutamic acids, and at pH = 3.2 has been shown to form oligomers and large amyloid-like aggregates with a high β-sheet content [36]. Since at pH = 3.2, all carboxyl groups are expected to be protonated [37], its hydrogen-bonding donor–acceptor capabilities could resemble those present in glutamine (Supplementary Fig. 4). Although this might seem a chemically naïve approximation, quantum calculations yielded similar hydrogen bond energies to protonated carboxyl and amide groups [38]. The same substitution has been proven as a conservative mutation in a glutamate transporter with a deeply buried acidic residue [39]. Hence, at low pH values, we will consider C-peptide as a Q-rich molecule.

Not surprisingly, the aggregation dynamics and kinetics of the C-peptide at pH = 7 differ from those seen in the previous systems (Fig. 4B). Unlike in the other cases, the MCS converged to nearly 3, even though there is a ten-fold excess of free peptides in the solution. On a first phase the monomer population rapidly decreased, forming multiple low molecular weight oligomers (top row, Fig. 4B). After this period the aggregation rate diminished, presenting a monomer-dimer-trimer equilibrium [40], and the presence of two stable tetrameric and pentameric species. A per-residue analysis clearly shows that the limited aggregation is driven by hydrophobic forces and most likely limited by the electrostatic repulsion conferred by the anionic residues (third row, Fig. 4B and Supplementary video 2). This aggregative behavior was reported by Jörnval et al. by performing PAGE and immunoblotting experiments of biotynilated C-peptide which, at pH = 7, aggregated in bands of 6–15 kDa, corresponding to dimers and pentamers [36].

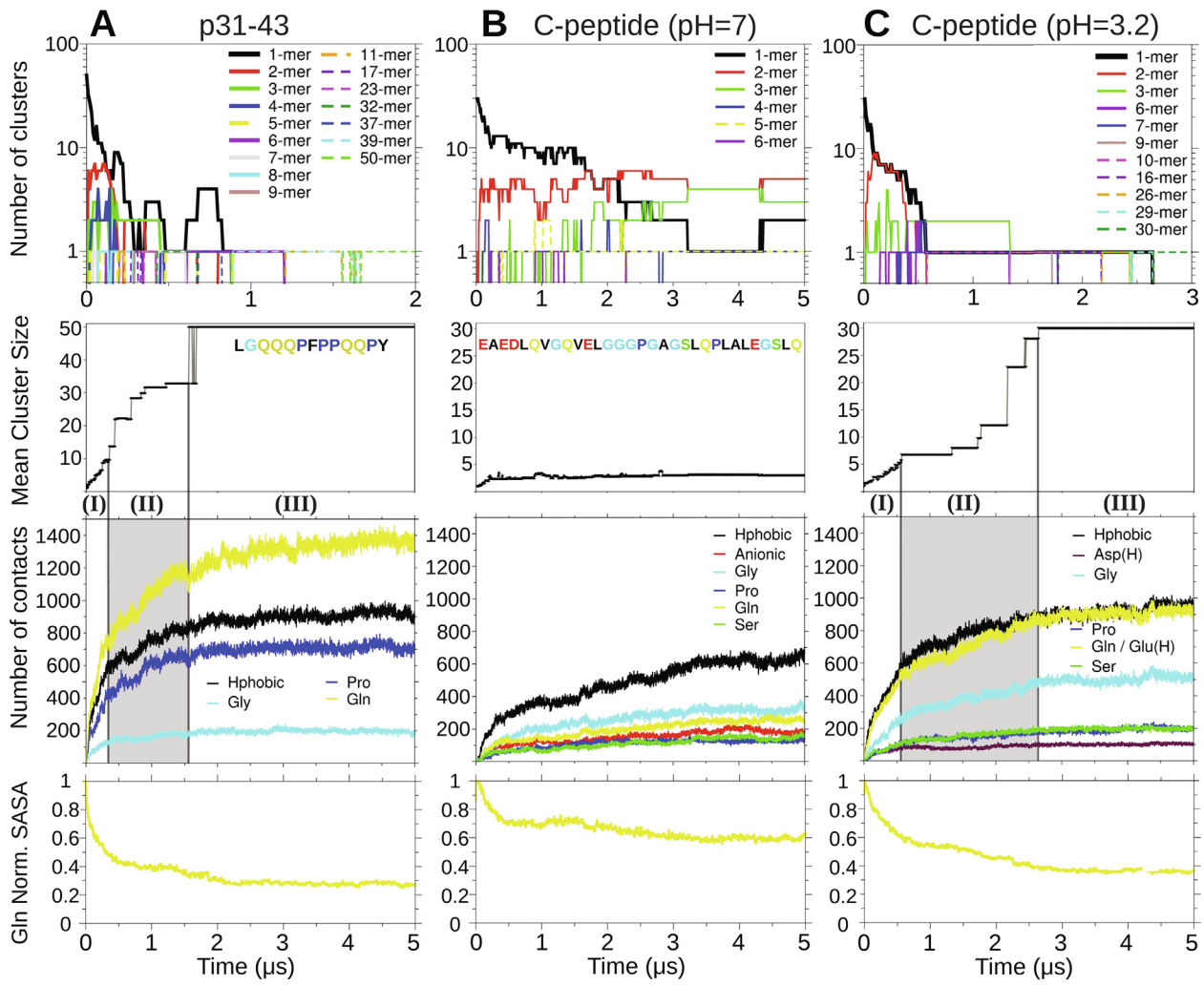

Fig. 4. Aggregation analysis for the p31–43 peptide (A), C-peptide at pH = 7 (B) and C-peptide at pH = 3.2 (C). Top row: Number of clusters classified by size, notice that the y axis is represented in a logarithmic scale. For (A) and (C) the analysis is focused on the first μs of simulation where most of the association events occur. Second row: Mean cluster size is measured throughout 5 μs of MD simulations. For (A) and (C) vertical lines divide the graph in three phases, according to the aggregation rate. The aminoacidic sequence of p31–43 and C-peptide are shown for reference. Third row: Number of interchain contacts defined by residue. Leucine, phenylalanine and tyrosine contacts are grouped as hydrophobic; aspartate and glutamate as anionic. Lower row: Glutamine's solvent accessible surface area; values are normalized by the initial SASA, when peptides are in their monomeric forms.

Converting the C-peptide into Q-rich like peptide by decreasing the solution's pH recovered an aggregation-prone behavior. This time a globular 30-mer aggregate was formed, with a radius of gyration of 3.78 ± 0.06 nm and 15.85% ± 0.84 of β-extended structures. Although with slower kinetics, the evolution of contacts per residue displayed a behavior similar to p31–43, with an initial depletion of monomers forming dimers, trimers and hexamers (top row, Fig. 4C). In close analogy with p31–43, oligomers underwent progressive association during phase II, until one single aggregate was formed. As can be seen in the third row of Fig. 4C, phase I was mainly driven by hydrophobic contacts and contacts between glutamines or protonated glutamic acids in proportions according to their abundances. During aggregation phase II, this contact profile was overturned with glutamine interactions contributing the most to oligomer association as can be seen in the steepest evolution of Q-like contacts. These glutamine-mediated contacts also participate in the compaction of the aggregate (Supplementary video 3 and Supplementary Fig. 4C-D), and might continue over much longer timescales than those accessible to our simulations to produce a mature and more ordered amyloid-like fibril [36]. The SASA per glutamine residue at pH = 7 shows that, in the Q-poor regime, the glutamine burial decreased a 45%. However, in the Q-rich situation (at pH = 3.2) glutamine burial arrived to nearly 70%, which is very similar to the glutamine burial obtained for p31–43 (lower row, Fig. 4B, C).

3. Discussion and conclusions

Our analysis of the role of glutamine within different peptide contexts revealed some general features summarized in Fig. 5.

In the case of poly-Q, the peptide length affects both the extent and the morphology of the aggregates. Q4 peptides showed high solubility, with rapid inter-conversions between monomeric, dimeric, and trimeric species (Fig. 1B). However, it is widely accepted that expanding glutamine tracts increase their aggregation propensity, particularly enriching their β-sheet contents [41]. Therefore, we decided to study two longer poly-Q peptides: Q11 and Q36, representing expansions below and above the toxic threshold in Huntington's disease [42]. Although the Q36 has apparently low disease incidence [43], segments of this length have been reported to acquire high β-sheet content [44]. Q11 peptides initially formed dimers, trimers, and tetramers stabilized by

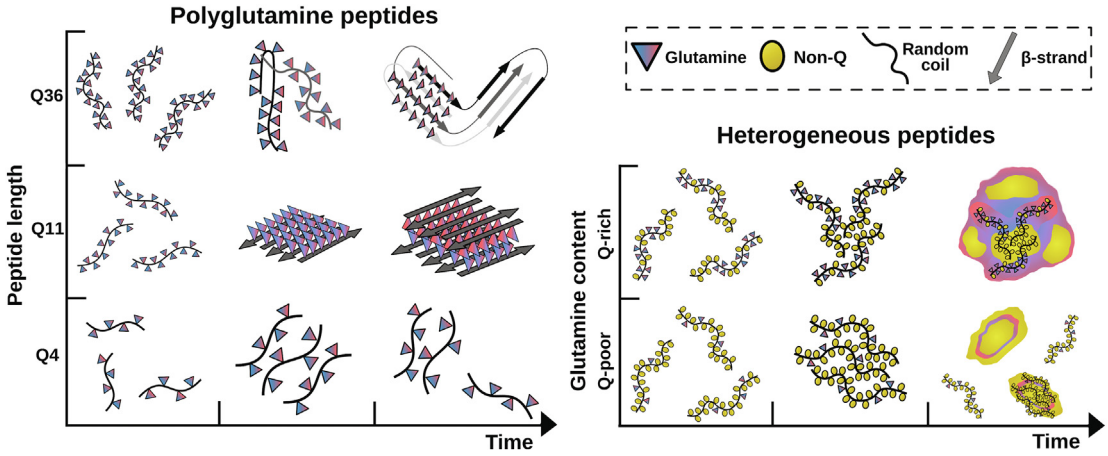

Fig. 5. Schematic description of the different aggregation scenarios studied. The time evolution of the aggregation is classified by peptide length for poly-Q peptides (left grid) and by glutamine content for heterogeneous peptides (right grid). Q4 showed low cluster sizes, presenting a monomer–dimer-trimer equilibrium through the simulation time. Increasing the peptide length to Q11 changed the aggregation behavior, linear peptides initially forming β-sheets then progress to β-sandwiches highly stabilized by side-chain contacts. Q36 presented loops, aggregating into amyloidogenic-like structures. In the case of heterogeneous peptides, Q-rich peptides (e.g. p31–43) formed large amorphous aggregates stabilized by internal glutamine-mediated contacts. Meanwhile Q-poor peptides (e.g. C-peptide, pH = 7) presented a hydrophobic driven aggregation limited by electrostatic repulsion resulting in intermediate cluster sizes.

multiple backbone hydrogen bonds, that ultimately resulted in the recruitment of all peptides in the solution. In line with previous studies, the β-sheets presented extensive surfaces with side chains alternating hydrogen bonds donors and acceptors that recruit new peptides to end in a putative crystalline β-sheet structure [45]. In contrast, Q36 presented higher conformational flexibility, allowing intramolecular β-hairpin formation. These double-stranded motifs acted as aggregation-prone regions. These regions interacted with unstructured soluble monomers and promoted a conformational transition towards β-sheet rich structures, resembling a nucleated-elongation model [31]. This mechanism has been previously proposed by Wetzel et al. studying different length poly-Q peptides using a microtiter plate elongation assay [46]. The distinctive structural characteristics we observed in poly-Q peptides of different lengths are also in remarkable agreement with experimental data obtained by IR spectroscopy, validating the length-dependent aggregation pathways proposed in this low-resolution structural determination [32].

In contrast with homogeneous poly-Qs, for heterogeneous glutamine peptides, the relevant variable is their relative aminoacidic abundances. Analyzing its content in relative terms provides a convenient and quantitative characterization. To this aim, we calculated the ratio between glutamine and non-Q amino acids and compared them with the ratio of new intermolecular contacts established by glutamines vs non-Q amino acids in each aggregation phase. The C-peptide at pH = 7 presents a glutamine ratio of only 0.15 (Table 1). In this Q-poor peptide, glutamine contacts were not dominant and remained insensitive to the aggregation phase. It is somehow predictable that progressively increasing glutamine content –as for the p31–43 and C-peptide at pH = 3.2 cases– could increase their engagement in interpeptide contacts. However, the remarkable observation is how in Q-rich peptides glutamine interactions became dominant during the association of low molecular weight oligomers (phase II), with contact ratios surpassing the residue ratio (gray shadowed row, Table 1). Calculating relative contents between glutamine and hydrophobic amino acids, indicate that initial monomeric association is ruled by the hydrophobic effect and is later switched to a Q-driven regime.

It is worth mentioning that C-peptide has been reported to produce amyloid-like fibrils [36]. However, the timescale required experimentally for this peptide to develop a fully ordered structure ranges from minutes to days [36,47]. Our results are not contradictory with this biophysical evidence, as our simulations can capture only the early phases of aggregation. Indeed, this goes in line with the long-term effect of maturation and optimization of inter-glutamine interactions observed here. Limited time sampling is recognized as one of the limiting factors challenging the comparison between MD simulations and experiments [48]. Despite our CG approach grants over two orders of magnitude in speed up, the sampling problem continues to represent a significant challenge.

In conclusion, the extension of poly-Q peptides is determinant for their aggregation, showing three well-differentiated behaviors. Only above a certain length, poly-Qs may display highly ordered (nearly crystalline) aggregation, after which the possibility to form intramolecular hairpins may lead to large scale but conformationally heterogeneous fibrillar arrays (Fig. 5-left side and Supplementary Fig. 3A). In contrast to poly-Q, heterogeneous peptides may present as many variables as the amino acidic composition can introduce. In this work, we focused on glutamine content and showed how its increase could trigger the low molecular weight

Table 1
Glutamine abundance and contact ratios in heterogeneous systems. Contacts ratios are calculated separately for each aggregation phase. The values are also reported relative to hydrophobic residues (Q/HPhob).

	p31–43		C-peptide (pH = 7)		C-peptide (pH = 3.2)	
	Q/Non-Q	Q/HPhob	Q/Non-Q	Q/HPhob	Q/Non-Q	Q/HPhob
Residue abundance	0.63	1.67	0.15	0.36	0.35	0.73
Contacts phase I	0.66	1.27	0.20	0.42	0.45	0.89
Contacts phase II	0.82	1.79	0.15	0.34	0.53	1.21
Contacts phase III	1.21	2.33	–	–	0.43	0.78

oligomers' association, stabilizing large globular aggregates (Fig. 5-right side and Supplementary Fig. 4C-D).

Several efforts have been devoted to design short peptides with the capability to modulate toxic protein aggregation [49,50]. The topological alteration of the aggregates observed in the simulations of Q11-QEQQQ suggest that the methodology presented here could also be used for cost-effective exploration of small peptides able to modulate the formation of soluble oligomers.

4. Methods

We generated peptides' initial coordinates with Chimera [51] visualization tool, setting the backbone torsional angles as fully extended conformations. C-peptide was studied at pH values of 3.2 and 7, defining their protonation states with PropKa [52]. All simulations were performed at the CG level using the SIRAH force field [28]. SIRAH is a residue-level CG force field in which effective interaction points are mapped from the atomic position, using different numbers of beads for different residues. The readers are referred to Fig. 2 of [28] for a comprehensive view of the force field mapping and interaction parameters.

As the SIRAH force field uses the same Hamiltonian function employed by any standard all-atoms MD simulation, concepts in fully atomistic representations such as hydrogen bonds and hydrophobic interactions are analogous in our CG simulations. As shown in Supplementary Fig. 1, five out of the six CG beads in glutamine carry a partial charge. Therefore, the contacts measured are indicative of the formation of electrostatic interactions or hydrogen bond-like interactions.

Atomic structures were mapped to CG beads using the map files included in SIRAH Tools [53]. Carboxy- and amino- terminals were set as neutral, except for the C-peptide where we focused on the effects of electrostatic interactions over aggregation properties. At pH = 7 systems terminals were set as zwitterionic; while at pH = 3.2 C-terminal was neutral and N-terminal presented a positive charge. MD simulations were performed by triplicate with GROMACS 2018.4 [54]. For the monomeric simulations, all the six peptides were centered in cubic boxes, which sizes were defined setting a distance of 1.5 nm between the solute and the box boundaries. Systems were solvated using a pre-equilibrated box of SIRAH's water model (named WT4) [55]. Forces on the CG beads were balanced applying 5000 iterations of the steepest descent algorithm. The heating step was performed using the V-rescale thermostat [56], keeping the pressure at 1 bar with the Parrinello-Rahman barostat [57]. In order to generate initial conformations for the aggregation studies, 1 μs production runs were simulated, choosing different conformers from the last 0.1 μs of trajectory. Multiple copies of these conformers were placed in simulation boxes setting a distance of 4 nm between their geometric centers. An identical system setup as above described for the monomeric peptides, was applied for the multiple-peptide simulation systems. For a detailed information of systems' setup see Supplementary Table 1. MD trajectories' analysis included secondary structure determinations. To achieve this we first employed the backmapping utility of SIRAH tools and then assign secondary contents to the reconstructed atomic coordinates with STRIDE [58]. To estimate the size of the aggregates along the simulations we calculated the Mean Cluster Size (MCS) as it has been previously defined by Kuroda et al. [59]: $MCS = (\Sigma^{N}_{i=1} CS_{i,t})/N$. "N" corresponds to the total number of peptides in the simulation box and "CS" is the number of peptides forming a cluster to which peptide "*i*" belongs at time "*t*". The generation of a cluster was defined using a distance cutoff of 0.6 nm between beads of different peptides. Interpeptide contacts were measured with the GROMACS utility *gmx mindist* and the MCS calculation was performed with an *in house* Python script. The number of total interpeptide contacts defined by residue was calculated with the Tcl script: *newcontacts.tcl* downloaded from https://www.ks.uiuc.edu/. Here, to better discriminate the residues involved in interpeptide contacts, we used a smaller cutoff value of 0.5 nm. Intrapeptide contacts were not considered. SASA analysis was done with the GROMACS utility gmx sasa.

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.

Acknowledgments

This work was partially funded by FOCEM (MERCOSUR Structural Convergence Fund), COF 03/11 and the National Natural Science Foundation of China (NSFC grant 31770776 to FZ). EEB and SP belong to the SNI program of ANII. Simulations were performed on the National Uruguayan Center for Supercomputing, ClusterUY. We thank Engr. Martin Etchart for his valuable collaboration in Python scripting.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.02.014. Backmapped trajectory files of each system were uploaded to Mendeley Data under the doi: 10.17632/2tmsbchh42.1 (https://data.mendeley.com/datasets/2tmsbchh42/1).

References

[1] Alzheimer, A. Uber einen eigenartigen schweren Erkrankungsprozess der Hirninde. Neurol Cent 1906;25:1134.
[2] Hardy JA, Higgins GA. Alzheimer's disease: The amyloid cascade hypothesis. Science (80-) 1992;256:184–5. 10.1126/science.1566067.
[3] Lambert MP, Barlow AK, Chromy BA, Edwards C, Freed R, Liosatos M, et al. Diffusible, nonfibrillar ligands derived from Aβ1-42 are potent central nervous system neurotoxins. Proc Natl Acad Sci U S A 1998;95:6448–53. https://doi.org/10.1073/pnas.95.11.6448.
[4] Fang YS, Tsai KJ, Chang YJ, Kao P, Woods R, Kuo PH, et al. Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. Nat Commun 2014;5:1–13. https://doi.org/10.1038/ncomms5824.
[5] Bengoa-Vergniory N, Roberts RF, Wade-Martins R, Alegre-Abarrategui J. Alpha-synuclein oligomers: a new hope. Acta Neuropathol 2017;134:819–38. https://doi.org/10.1007/s00401-017-1755-1.
[6] Silveira JR, Raymond GJ, Hughson AG, Race RE, Sim VL, Hayes SF, et al. The most infectious prion protein particles. Nature 2005;437:257–61. https://doi.org/10.1038/nature03989.
[7] Beitel LK, Alvarado C, Mokhtar S, Paliouras M, Trifiro M. Mechanisms mediating spinal and bulbar muscular atrophy: Investigations into polyglutamine-expanded androgen receptor function and dysfunction. Front Neurol 2013;4. https://doi.org/10.3389/fneur.2013.00053.
[8] Lasagna-Reeves CA, Rousseaux MWC, Guerrero-Munoz MJ, Vilanova-Velez L, Park J, See L, et al. Ataxin-1 oligomers induce local spread of pathology and decreasing them by passive immunization slows spinocerebellar ataxia type 1 phenotypes. Elife 2015;4. https://doi.org/10.7554/eLife.10891.
[9] Schonhoft JD, Monteiro C, Plate L, Eisele YS, Kelly JM, Boland D, et al. Peptide probes detect misfolded transthyretin oligomers in plasma of hereditary amyloidosis patients. Sci Transl Med 2017;93. https://doi.org/10.1126/scitranslmed.aam7621.
[10] Haataja L, Gurlo T, Huang CJ, Butler PC. Islet amyloid in type 2 diabetes, and the toxic oligomer hypothesis. Endocr Rev 2008;29:303–16. https://doi.org/10.1210/er.2007-0037.
[11] Leitman J, Ulrich Hartl F, Lederkremer GZ. Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress. Nat Commun 2013;4:1–10. https://doi.org/10.1038/ncomms3753.
[12] Kim YE, Hosp F, Frottin F, Ge H, Mann M, Hayer-Hartl M, et al. Soluble oligomers of PolyQ-expanded huntingtin target a multiplicity of key cellular factors. Mol Cell 2016;63:951–64. https://doi.org/10.1016/j.molcel.2016.07.022.
[13] Lieberman AP, Shakkottai VG, Albin RL. Polyglutamine repeats in neurodegenerative diseases. Annu Rev Pathol Mech Dis 2019;14:1–27. https://doi.org/10.1146/annurev-pathmechdis-012418-012857.

[14] Tycko R. Molecular Structure of Aggregated Amyloid-β: Insights from Solid State Nuclear Magnetic Resonance. Cold Spring Harb Perspect Med 2016;6:. https://doi.org/10.1101/cshperspect.a024083a024083.
[15] Nguyen PH, Sterpone F, Derreumaux P. Aggregation of disease-related peptides. 1st ed. Elsevier Inc.; 2020. 10.1016/bs.pmbts.2019.12.002.
[16] Morriss-Andrews A, Shea JE. Computational studies of protein aggregation: Methods and applications. Annu Rev Phys Chem 2015;66:643–66. https://doi.org/10.1146/annurev-physchem-040513-103738.
[17] Carballo-Pacheco M, Ismail AE, Strodel B. On the applicability of force fields to study the aggregation of amyloidogenic peptides using molecular dynamics simulations. J Chem Theory Comput 2018;14:6063–75. https://doi.org/10.1021/acs.jctc.8b00579.
[18] Ruff KM, Khan SJ, Pappu RV. A coarse-grained model for polyglutamine aggregation modulated by amphipathic flanking sequences. Biophys J 2014;107:1226–35. https://doi.org/10.1016/j.bpj.2014.07.019.
[19] Haaga J, Gunton JD, Buckles CN, Rickman JM. Early stage aggregation of a coarse-grained model of polyglutamine. J Chem Phys 2018;148. https://doi.org/10.1063/1.5010888.
[20] Fluitt AM, De Pablo JJ. An Analysis of Biomolecular Force Fields for Simulations of Polyglutamine in Solution. Biophys J 2015;109:1009–18. https://doi.org/10.1016/j.bpj.2015.07.018.
[21] Wang Y, Voth GA. Molecular dynamics simulations of polyglutamine aggregation using solvent-free multiscale coarse-grained models. J Phys Chem B 2010;114:8735–43. https://doi.org/10.1021/jp1007768.
[22] Flöck D, Rossetti G, Daidone I, Amadei A, Di Nola A. Aggregation of small peptides studied by molecular dynamics simulations. Proteins Struct Funct Genet 2006;65:914–21. https://doi.org/10.1002/prot.21168.
[23] Wen J, Scoles DR, Facelli JC. Molecular dynamics analysis of the aggregation propensity of polyglutamine segments. PLoS ONE 2017;12:10–4. https://doi.org/10.1371/journal.pone.0178333.
[24] Nguyen PH, Ramamoorthy A, Sahoo BR, Zheng J, Faller P, Straub JE, et al. Amyloid Oligomers: A Joint Experimental/Computational Perspective on Alzheimer's Disease, Parkinson's Disease, Type II Diabetes, and Amyotrophic Lateral Sclerosis. Chem Rev 2021:acs.chemrev.0c01122. 10.1021/acs.chemrev.0c01122.
[25] Delyser M, Noid WG. Bottom-up coarse-grained models for external fields and interfaces. J Chem Phys 2020;153:. https://doi.org/10.1063/5.0030103224103.
[26] Deng L, Wang Y, Ou-Yang ZC. Concentration and temperature dependences of polyglutamine aggregation by multiscale coarse-graining molecular dynamics simulations. J Phys Chem B 2012;116:10135–44. https://doi.org/10.1021/jp210683n.
[27] Bereau T, Deserno M. Generic coarse-grained model for protein folding and aggregation. J Chem Phys 2009;130:. https://doi.org/10.1063/1.3152842235106.
[28] Machado MR, Barrera EE, Klein F, Sóñora M, Silva S, Pantano S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. J Chem Theory Comput 2019;15:2719–33. https://doi.org/10.1021/acs.jctc.9b00006.
[29] Klein F, Barrera EE, Pantano S. Assessing SIRAH's capability to simulate intrinsically disordered proteins and peptides. J Chem Theory Comput 2021. https://doi.org/10.1021/acs.jctc.0c00948.
[30] Zonta F, Buratto D, Crispino G, Carrer A, Bruno F, Yang G, et al. Cues to Opening Mechanisms From in Silico Electric Field Excitation of Cx26 Hemichannel and in Vitro Mutagenesis Studies in HeLa Transfectans. Front Mol Neurosci 2018;11:170. https://doi.org/10.3389/fnmol.2018.00170.
[31] Bhattacharyya AM, Thakur AK, Wetzel R. Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction. Proc Natl Acad Sci U S A 2005;102:15400–5. https://doi.org/10.1073/pnas.0501651102.
[32] Yushchenko T, Deuerling E, Hauser K. Insights into the aggregation mechanism of PolyQ proteins with different glutamine repeat lengths. Biophys J 2018;114:1847–57. https://doi.org/10.1016/j.bpj.2018.02.037.
[33] Gómez Castro MF, Miculán E, Herrera MG, Ruera C, Perez F, Prieto ED, et al. P31–43 gliadin peptide forms oligomers and induces NLRP3 inflammasome/caspase 1-dependent mucosal damage in small intestine. Front Immunol 2019;10.. https://doi.org/10.3389/fimmu.2019.00031.
[34] Herrera MG, Gómez Castro MF, Prieto E, Barrera E, Dodero VI, Pantano S, et al. Structural conformation and self-assembly process of p31–43 gliadin peptide in aqueous solution. Implications for celiac disease. FEBS J 2019. https://doi.org/10.1111/febs.15109.
[35] Munte CE, Vilela L, Kalbitzer HR, Garratt RC. Solution structure of human proinsulin C-peptide. FEBS J 2005;272:4284–93. https://doi.org/10.1111/j.1742-4658.2005.04843.x.
[36] Lind J, Lindahl E, Perálvarez-Marín A, Holmlund A, Jörnvall H, Mäler L. Structural features of proinsulin C-peptide oligomeric and amyloid states. FEBS J 2010;277:3759–68. https://doi.org/10.1111/j.1742-4658.2010.07777.x.
[37] Kozlowski LP. IPC – Isoelectric Point Calculator. Biol Direct 2016;11:55–55. 10.1186/s13062-016-0159-9.
[38] Nie B, Stutzman J, Xie A. A vibrational spectral maker for probing the hydrogen-bonding status of protonated Asp and Glu residues. Biophys J 2005;88:2833–47. https://doi.org/10.1529/biophysj.104.047639.
[39] Mwaura J, Tao Z, James H, Albers T, Schwartz A, Grewer C. Protonation state of a conserved acidic amino acid involved in Na + binding to the glutamate transporter EAAC1. ACS Chem Neurosci 2012;3:1073–83. https://doi.org/10.1021/cn300163p.
[40] Betz S, Fairman R, O'Neil K, Lear J, Degrado W. Design of two-stranded and three-stranded coiled-coil peptides. Philos Trans R Soc Lond B Biol Sci 1995;348:81–8. https://doi.org/10.1098/rstb.1995.0048.
[41] Petrakis S, Schaefer MH, Wanker EE, Andrade-Navarro MA. Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. BioEssays 2013;35:503–7. https://doi.org/10.1002/bies.201300001.
[42] Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt BR, et al. Huntington disease. Nat Rev Dis Prim 2015;1:1–21. https://doi.org/10.1038/nrdp.2015.5.
[43] Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana VV, Cassiman JJ, et al. Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. Am J Hum Genet 1996;59:16–22.
[44] Chen S, Ferrone FA, Wetzel R. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. Proc Natl Acad Sci U S A 2002;99:11884–9. https://doi.org/10.1073/pnas.182276099.
[45] Viney C. Natural Protein Fibers. Ref Modul Mater Sci Mater Eng 2016:1–9. 10.1016/b978-0-12-803581-8.02271-2.
[46] Berthelier V, Wetzel R. Screening for modulators of aggregation with a microplate elongation assay. Methods Enzymol 2006;413:313–25. https://doi.org/10.1016/S0076-6879(06)13016-5.
[47] Tsiolaki PL, Louros NN, Zompra AA, Hamodrakas SJ, Iconomidou VA. Unraveling the aggregation propensity of human insulin C-peptide. Pept Sci 2017;108:. https://doi.org/10.1002/bip.22882e22882.
[48] Strodel B. Amyloid aggregation simulations: challenges, advances and perspectives. Curr Opin Struct Biol 2021;67:145–52. https://doi.org/10.1016/j.sbi.2020.10.019.
[49] He RY, Lai XM, Sun CS, Kung TS, Hong JY, Jheng YS, et al. Nanoscopic insights of amphiphilic peptide against the oligomer assembly process to treat Huntington's disease. Adv Sci 2020;7. 10.1002/advs.201901165.
[50] Diao J, Yang K, Lai Y, Li Y, Jiang L, Li D, et al. structure-based peptide inhibitor design of amyloid-β aggregation 2019. 10.3389/fnmol.2019.00054.
[51] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. A visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12. https://doi.org/10.1002/jcc.20084.
[52] Olsson MHM, SØndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions. J Chem Theory Comput 2011;7:525–37. 10.1021/ct100578z.
[53] Machado M, Pantano S. Structural Bioinformatics SIRAH tools: mapping, backmapping and visualization of coarse-grained models. Bioinformatics 2016;32:2–3. https://doi.org/10.1093/bioinformatics/btw020.
[54] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001.
[55] Darre L, Machado MR, Dans PD, Herrera FE, Pantano S. Another coarse grain model for aquous solvation: WAT FOUR?. J Chem Theory Comput 2010;6:3793–807.
[56] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126:14101. https://doi.org/10.1063/1.2408420.
[57] Parrinello MRA. Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 1981;52:7182–90.
[58] Heinig M, Frishman D. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res 2004;32. 10.1093/nar/gkh429.
[59] Kuroda Y, Suenaga A, Sato Y, Kosuda S, Taiji M. All-atom molecular dynamics analysis of multi-peptide systems reproduces peptide solubility in line with experimental observations. Sci Rep 2016;6:1–6. https://doi.org/10.1038/srep19479.