

Early oligomerization stages for the non-amyloid component of α -synuclein amyloid

Cindie Eugene,¹ Rozita Laghaei,^{1,2} and Normand Mousseau²

¹*Département de Physique and Groupe de recherche sur les protéines membranaires (GEPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec H3C 3J7, Canada*

²*Department of Chemistry, University of Pittsburgh, 319 Eberly Hall, Pittsburgh, Pennsylvania 15260, USA*

(Received 4 February 2014; accepted 12 September 2014; published online 7 October 2014)

In recent years, much effort has focused on the early stages of aggregation and the formation of amyloid oligomers. Aggregation processes for these proteins are complex and their non-equilibrium nature makes any experimental study very difficult. Under these conditions, simulations provide a useful alternative for understanding the dynamics of the early stages of oligomerization. Here, we focus on the non- $A\beta$ amyloid component (NAC) of the monomer, dimer, and trimer of α -synuclein, an important 35-residue sequence involved in the aggregation and fibrillation of this protein associated with Parkinson's disease. Using Hamiltonian and temperature replica exchange molecular dynamics simulations combined with the coarse grained Optimized Potential for Efficient peptide structure Prediction potential, we identify the role of the various regions and the secondary structures for the onset of oligomerization. For this sequence, we clearly observe the passage from α -helix to β -sheet, a characteristic transition of amyloid proteins. More precisely, we find that the NAC monomer is highly structured with two α -helical regions, between residues 2-13 and 19-25. As the dimer and trimer form, β -sheet structures between residues 2-14 and 26-34 appear and rapidly structure the system. The resulting conformations are much more structured than similar dimers and trimers of β -amyloid and amylin proteins and yet display a strong polymorphism at these early stages of aggregation. In addition to its inherent experimental interest, comparison with other sequences shows that NAC could be a very useful numerical model for understanding the onset of aggregation.

© 2014 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4896381>]

I. INTRODUCTION

Amyloid aggregates and insoluble fibrils are associated with many neurodegenerative diseases such as Alzheimer's and Parkinson's diseases.^{1,2} The latter is characterized by the presence of Lewy bodies and neurites composed, in large parts, of the α -synuclein protein.³ The presence of α -synuclein has also been confirmed for a number of other important neurodegenerative disorders such as Down's syndrome and familial Alzheimer's disease, making this protein a major focus in the study of amyloid diseases.⁴⁻⁷ α -syn is a 140-residue protein characterized by three major regions: an N-terminal region with KTEKEGV imperfect repeats, a hydrophobic center region named non- $A\beta$ amyloid component (NAC), and a highly negatively charged C-terminal region.⁸ The NAC is also the second major component found, after β -amyloid, in the amyloid components associated with Alzheimer's disease (AD).⁹

The polymerization of α -synuclein takes place through conformational changes that bring the protein from disordered conformations to β -sheet structures as it assembles into toxic oligomers.^{10,11} Yet, little direct information is known about the first steps of oligomerization as its dynamical nature makes it difficult to characterize experimentally.^{12,13} To overcome this limitation, it is useful to turn to computer simulations that can provide detailed information about these small assemblies.

For α -synuclein, most simulations focused on the full-length sequence. A combination of paramagnetic relaxation enhancement, nuclear magnetic resonance (NMR) spectroscopy, and ensemble molecular dynamics (MD) simulations were used to probe the topology of native α -synuclein.¹⁴ More recently, molecular simulations combined with NMR were used to characterize the structure of full-length monomeric α -synuclein at low and neutral pH. It was shown that, while the protein is more structured at low pH,¹⁵ it is far from random coil, even though it is disordered, at neutral pH, in agreement with a Bayesian ensemble analysis.¹⁶ Molecular dynamic was also used to vary the stability and structural properties of several α -synuclein oligomers and mutated sequences.¹⁷⁻²¹

In this article, we are interested in characterizing the first steps of α -synuclein oligomerization. More precisely, we focus on the 35-residue NAC fragment. As mentioned above, this fragment represents the second major component of amyloid deposits associated with Alzheimer's diseases, with a ratio of up to 10%.²² The NAC forms α -synuclein's most hydrophobic region. It was early discovered that it self-aggregates into insoluble amyloid β -pleated structures.²³ The NAC also plays a central role in α -synuclein aggregation as is demonstrated by the behavior of β -synuclein. This highly homologous protein to α -synuclein lacks an 11-residue segment located in the NAC domain at position 73-83,²⁴ which overlaps with the minimal toxic aggregate-prone segment,

residues 68-78.²⁵ It does not aggregate²⁶ and even inhibits α -synuclein aggregation.²⁷ NAC's *in vivo* function is still unknown. The neurotoxicity of extracellular NAC is well established, however.²⁸⁻³⁰ A recent study suggests that NAC could activate p53, Cdk5, and Bas-dependent apoptotic signaling pathways, causing cell death.³¹

Previous studies tried to identify a signature motif and a particular part of NAC responsible for the fibrillation propensity^{26,32} and the interaction and the conformational state when NAC is linked with a membrane.³³ Others have focused on the characteristic mutations present in this region.^{19,20,34} Yet, there is still very little structural information regarding the onset of oligomerization for the NAC peptide with only rare simulations done to serve as a reference model for experiments.¹⁷ The simulations presented here form therefore a first atomistic description of the onset of oligomerization for the NAC peptide.

Here, we use REMD coupled with the Optimized Potential for Efficient peptide structure Prediction (OPEP) coarse-grained potential,³⁵⁻³⁷ to characterize the fragment's structural changes as the system goes from monomer to dimer and trimer. These unbiased simulations allow us to follow aggregation and generate a rich set of structures that provide a first picture as to the assembly mechanism for this sequence. Comparison with dimers of other amyloid peptides simulated in the same conditions also offers an indication of generic and specific properties for early aggregates.

II. METHODS

In this study, we investigate the structure and the thermodynamics of monomeric, dimeric and trimeric NAC of α -synuclein using Hamiltonian-temperature replica exchange molecular dynamics (HT-REMD) coupled with the coarse-grained OPEP potential.^{35,38,39} The 35 amino acid sequence of the non-A β amyloid component of α -synuclein is given by EQVTN-VGGAVVTGVTAVAQKTVEGAGSIAAATGFV. The GXXX motif is associated with amyloidogenic propensity in many amyloidogenic proteins where X might be the uncharged residues, GLY, ALA, VAL, ILE, LEU, PHE, TYR, TRP, THR, SER, or MET. For the NAC, we find

three GXXX motifs ⁸GAVV¹¹, ¹³GVTA¹⁶, and ²⁴GAGS²⁷ corresponding to positions 68-71, 73-76, 84-87, respectively, in the α -synuclein protein.^{26,40}

To describe intra and interpeptide interactions, we use the coarse-grained force field OPEP version 3.2. OPEP reduces most amino acids to six beads, focusing on the heavy backbone atoms: C α , N, H, C, and O and one bead for the side chain except for the proline amino acid which is represented by all heavy atoms.⁴¹ It takes into account the propensities of each residue to adopt α and β conformations, a crucial aspect for prediction of the aggregation rate of amyloid-forming proteins like synuclein. This potential has been applied with success to study the aggregation of many amyloid peptides such polyglutamine,⁴² amylin,^{35,38} and different segments of A β .^{39,43-48}

The OPEP force field is combined with Hamiltonian-temperature replica exchange molecular dynamics, HT-REMD, a hybrid of temperature replica exchange (T-REMD), and Hamiltonian replica exchange (H-REMD), which was shown to accelerate sampling compared to T-REMD.³⁸ In the α -synuclein simulation, a set of 35 replicas are launched at logarithmically-distributed temperatures ranging from 240 to 555 K. Five other replicas are run at the highest temperature with progressively reduced non-bonded attractive forces scaled by 0.8, 0.7, 0.6, 0.4, and 0.2 of the initial interaction, respectively. Following the standard REMD procedure, at regular intervals, exchanges are attempted between adjacent replicas with a Metropolis probability.^{35,38}

The simulated temperature is controlled using the Berendsen thermostat.⁴⁹ The integration timestep is set at 1.5 fs with an external coupling constant of 100 fs. We use the RATTLE algorithm for covalent constraints⁵⁰ and replica exchanges are attempted every 7.5 ps. The monomer is simulated in a 40 Å-radius sphere and the dimer and trimer are placed in a 140 Å-radius sphere, all with reflecting boundary conditions. The monomer is simulated for 500 ns per replica (total 20 μ s), the dimer for 700 ns (total 28 μ s), and the trimer for 1200 ns (total 48 μ s). The three systems are started from an extended configuration with chains, for the dimer and the trimer, placed at random in the box. Initial configurations, after a first zero-temperature energy minimization, are shown in Fig. 1.

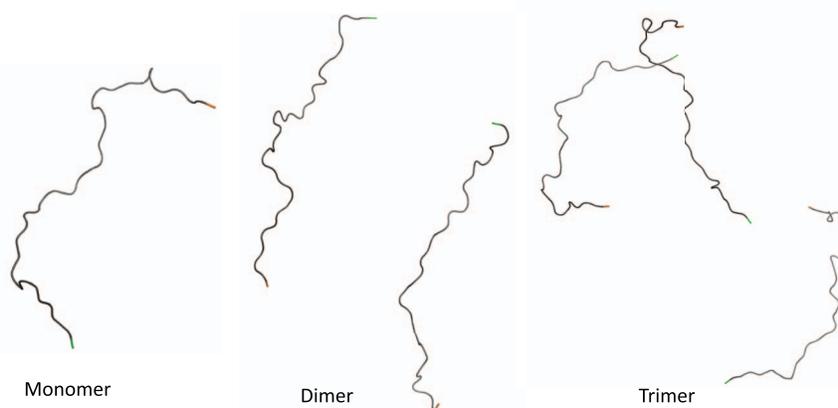


FIG. 1. Initial configurations from monomer to trimer. The N-terminal and the C-terminal are shown in green and orange.

The reconstruction of the thermodynamical properties of the NAC monomer, dimer, and trimer is done using the replica-exchanged adapted weighted histogram method adapted to Hamiltonian, PTWHAM.⁵¹ Secondary and tertiary structures are predicted using STRIDE, a protein secondary structure assignment based on the combined use of hydrogen bond energy and statistically derived backbone torsional angle information.⁵² Clustering is performed using the $C\alpha$ root-mean square deviations (rmsd) and following Daura's procedure:⁵³ the largest cluster with a rmsd of 2.5 Å is first identified and its member configurations removed; this procedure is repeated as long as there are remaining configurations.

We verify the convergence of the simulations by ensuring that the entropy as a function of temperature is constant. Entropy is obtained from the first law of thermodynamics, $F = E - TS$ and the free energy calculated by WHAM as implemented by Chodera *et al.*⁵¹ We compute the entropy over non-overlapping time intervals and identify the minimum simulation time over which $S(T)$ is time independent. As shown in Figure 2, the monomer reaches equilibrium after 100 ns/replica, the dimer 200 ns/replica, and the trimer 1000 ns/replica. Error bars are estimated with the bootstrapping algorithm.⁵⁴

To compute the chemical shifts, we first reconstruct the full atomic configuration of each side-chain using SCWRL4.⁵⁵ The resulting structure is then analysed with SPARTA+, a packaged based on artificial neural network that was trained to establish quantitative relation between chemi-

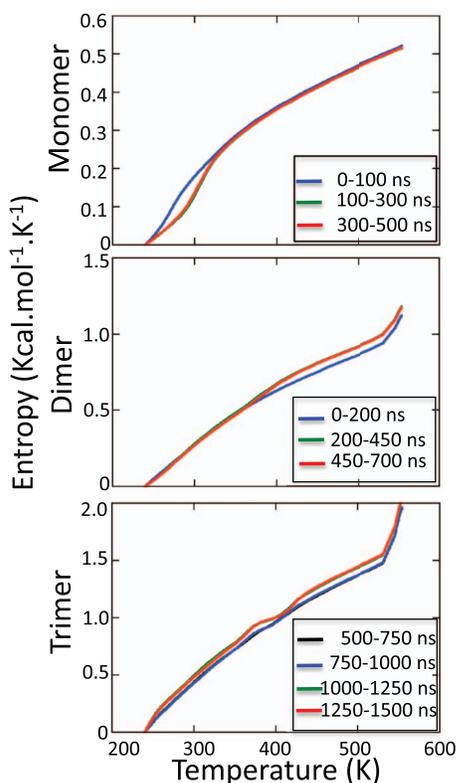


FIG. 2. Entropy as a function of temperature taken over various time intervals for the NAC monomer, dimer, and trimer. From top to bottom, we see the monomer converged after 100 ns/replica, the dimer after 200 ns/replica, and the trimer after 1000 ns/replica. Statistical error (not shown for legibility) by bootstrapping is ± 0.01 kcal/(mol*K).

TABLE I. Percentage of secondary structure taken at 300 K over the full NAC peptide. Results for the monomer, dimer, and trimer are averaged over their respective interval of convergence (see Sec. II). Statistical error is estimated at $\pm 1\%$.

Secondary structure (%)	Monomer	Dimer	Trimer
α -helix	40	8	7
β -strand	<1	24	23
Turn	39	48	49
Random coil	21	20	21

cal shifts and protein structures: it predicts chemical shifts for ^{15}N , $^1\text{H}^{\text{N}}$, $^1\text{H}^{\alpha}$, $^{13}\text{C}^{\alpha}$, $^{13}\text{C}^{\beta}$, and $^{13}\text{C}'$. Here, we focus on the chemical shifts for N, C_{α} , and C_{β} .^{56,57}

III. RESULTS

We are interested in the characterization of the structural and thermodynamical properties in the early step of aggregation the α -synuclein NAC. For this, we follow the evolution of its monomer, dimer, and trimer at 300 K. We first present results for each system separately and compare those in Sec. IV.

A. Monomer

We simulated the monomer for 500 ns/replica, collecting statistics over the last 400 ns. Table I gives the average secondary structure propensity over all residues. α -helices dominate, with a propensity of 40%, while β -strands are present at less than 1%. The secondary structure per residue (Fig. 3) reveals that α -helices are dominant in two regions: between residues 2-13 and residues 19-25 with a respective probability of $\sim 75\%$ and $\sim 90\%$.

As expected with the high proportion of secondary structure, clustering analysis for the monomer (Fig. 4) finds relatively well-defined conformations. The first cluster (M1), with a weight of 43%, shows the two dominant α -helical structures at residues 2-13 and residues 19-25 (Table II) positioned in a perpendicular conformation, with a disordered C-terminal. The same helices (at positions 3-13 and 19-24), in an antiparallel organization, form the second largest clusters, M2, at 16% probability with the loop at residues 14-18 being present in both clusters. No secondary structure is observed in the third cluster, M3, already much smaller at 8.5%. Yet, the contact map for M3 overlaps significantly with that of M2 (Figs. 5 and 6). This suggests that the M3 cluster is simply a slightly destabilized version of M2 and that the NAC monomer, while unstable against any specific native state, samples a relatively small number of well-defined conformations.

In the absence of direct experimental measurement on the monomeric NAC fragment, it is useful to provide the predicted chemical shifts for the equilibrium ensemble. Figure 9 shows the predicted NMR chemical shifts per residue for the non-amyloid component of α -synuclein amyloid at 300 K. These shifts are a very sensitive measure of local structure and can serve as a relatively straightforward tool for establishing the plausibility of numerically derived structures.^{56,58}

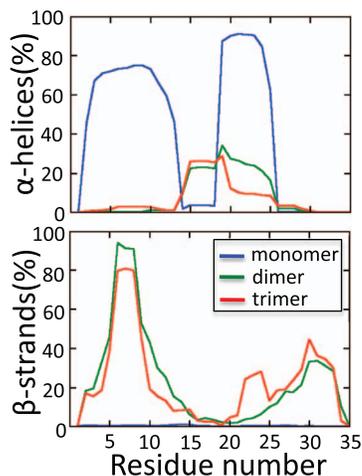


FIG. 3. Evolution of α -helix and β -strands as a function of residue number at 300 K from monomer to trimer. Statistical error (not shown for legibility) obtained by bootstrapping is estimated at $\pm 1.0\%$.

B. Dimer

We simulated the NAC dimer for 700 ns/replica, collecting statistics over the last 500 ns. Table I shows that the presence of a second peptide affects strongly the secondary structure. While the monomer is α -helical at 40% with almost no β -sheet, the dimer is dominated by β -strands, with a propensity of 24%, and only 8% for α -helices. The first β -strand, at residues 2 to 14, overlaps completely with the corresponding α -helix found for the monomer while the second β -strand is found at residues 26 to 34 and does not overlap with the

monomer's second α -helix. As shown in Fig. 3, maximum stability is observed in the 6-9 ($\sim 90\%$) and 30-33 ($\sim 30\%$) regions while α -helices form mostly between residues 15 and 25, a wider region than for the monomer, with a relatively small probability of 20% to 30%.

The dimer displays a much more diverse set of structures than the monomer. In the second row of Fig. 4, we show the four clusters present with a probability of more than 5% (D1 to D4) defined with the 2.5 Å-rmsd cut-off as for the monomer. Together, they represent only 30% of all sampled conformations. Another 30% of the conformations are member of clusters present with a probability between 1% and 5%. While presenting some structure, they typically show less β -sheet secondary structure than clusters D1 to D4.

As can be seen in Fig. 4 and Table II, clusters D1 and D4 display purely β -strand conformations. In the first case, the strands are fully interdigitated while D4 shows rather the two NAC peptides side by side. For their part, clusters D2 and D4 show a mixture of α -helical and β -strand structures. Here again, one of the clusters is intertwined (D2), leading to quaternary structure, while the other offers rather the image of two tertiary structures stabilized once against the other (D3).

Two α -helical regions are identified in clusters D2 and D4: residues 14 (or 15) to 19 is present on both chains in D2 and on one chain in D3. The second helix, in the latter cluster, is shifted to residues 19 to 25. Antiparallel β -sheets are also well defined: residues 6 to 8 form a sheet on both chains in all four clusters with, in some case, extensions (4-8 and 6-10 in D4 and 6-11 in D2). The region 30-34 also forms β -strands in D1 and D3 while 25-29 and 27-29 are seen in D2 and D4,

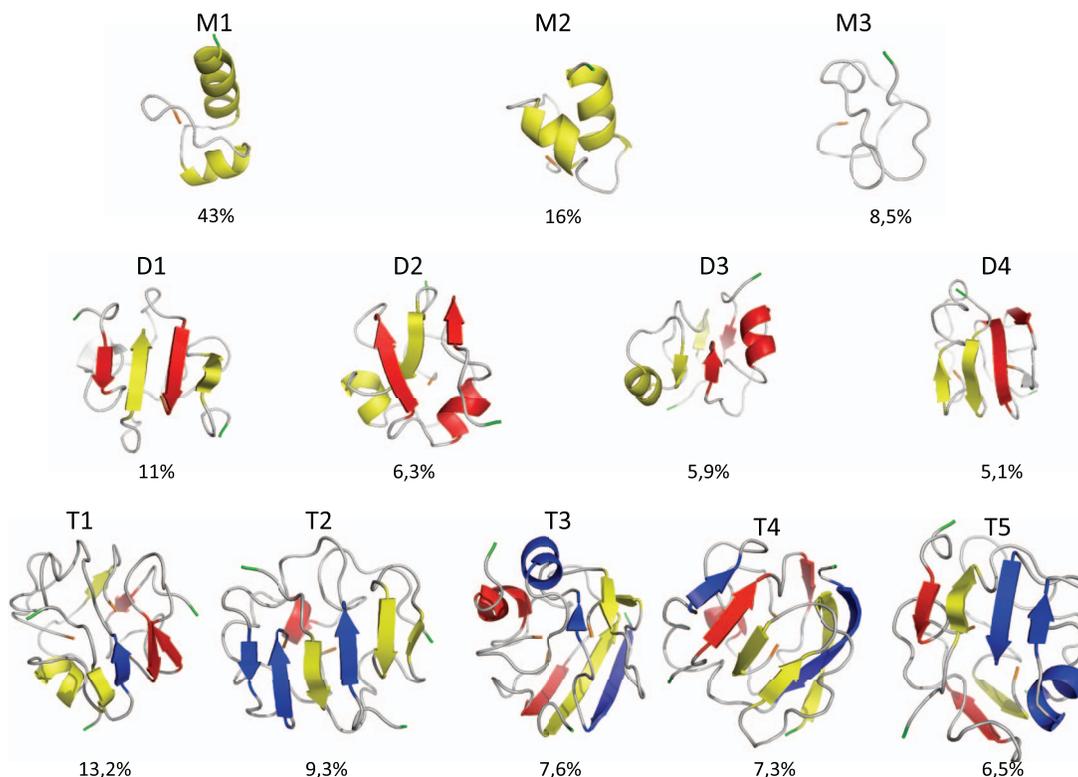


FIG. 4. The clusters' center are shown for NAC of monomer (first row), dimer (second row), and trimer (third row). The N-terminal and the C-terminal are shown in green and orange. The first chain is shown in yellow, the second in red, and the third in blue.

TABLE II. This table gives the region implicated in secondary structure for monomer, dimer, trimer, see in Fig. 4, and identified them to the 7 regions described before (see Sec. II). The letters A, B, and C allow to distinguish the different chains and extract some information about 3D organization in the secondary structure.

	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\beta 1$	$\beta 2$	$\beta 3$	$\beta 4$	Other
Monomer								
M1	2-13	19-25						
M2	3-13	19-24						
Dimer								
D1				6-8(B)/31-33(A) 31-33(B)/6-8(A)	30-34(A)/30-34(B)			
D2			15-19(A) 14-19(B)			6-8(A)/6-8(B)		25-29(B)/6-11(A) β
D3		19-25(A)	14-19(B)		31-33(A)/31-33(B)	6-8(A)/6-8(B)		
D4						4-8(A)/6-10(B)		13-15(A)/5-7(A) β 12-14(B)/27-29(B) β
Trimer								
T1			15-19(A)	31-33(B)/6-8(C)		6-8(A)/6-8(C)		28-30(A)/6-8(B) β 20-22(B)/31-33(B) β
T2			15-18(B)	6-9(A)/30-33(C) 31-33(A)/6-8(C)		6-8(A)/6-8(C)	22-24(A)/31-33(A) 22-24(C)/31-33(C)	
T3	5-10(B)		14-19			2-8(C)/6-12(A)		6-10(A)/23-27(B) β 20-21(A)/32-33(A) β 21-22(A)/14-15(C) β
T4				30-32(A)/5-7(B)		2-10(A)/4-12(C)	22-25(A)/30-33(A)	25-28(B) α 7-9(A)/13-15(B) β 6-8(B)/26-28(C) β
T5			14-19(C)	6-9(A)/30-33(C)		6-9(A)/6-9(B)	22-24(C)/31-33(C)	30-33(A)/26-29(B) β 32-33(A)/2-3(C) β

respectively. D4 also displays a β -strand at 13-15 and 12-14, each on one chain.

Table I gives some additional information about the region involved in the secondary structure and its spatial organization. The highest contact density region is associated with interchain hydrophobic interactions between the region 6-8 with itself (D2 and D3). This region can undergo additions of residue at the beginning or the end as in D4 where regions 4-8 and 6-10 are linked across chains. To a lesser extent, as for D1, 6-8 region may also be linked to the 31-33 region although this region (31-33 or 30-34) prefers to bind with its copy on the second monomer in D1 and D3. Intrachain β -sheets are only observed in D4, where the region 13-15 (12-14) binds to both 5-7 in chain A and 27-29 in chain B.

While interchain and intrachain contacts vary from cluster to cluster, we observe a striking similarity in total map contacts (Fig. 7), particularly between D1, D2, and D4. The variations between these intrachain and interchain contact maps reflect a richness in the organizations at the tertiary and quaternary structure level observed in Table II. Yet, these are constrained by the stability of the secondary structures that enforce a specific set of contacts that can be satisfied either through intra or interchain contacts. This explains the similar total contact map for very different clusters, when projected on the monomeric sequence. In terms of salt-bridges, while the intramolecular bridge at LYS20-GLU23 is present on both chains for all dominant clusters, we also observe a contact between GLU1 and LYS20 on either chain 1 or 2 for all clusters. While these few charged residues play an important role for

determining structure of this sequence, their behavior is almost the same for all clusters, however.

C. Trimer

The NAC trimer is run for 1500 ns/replica, with equilibrium being reached in the last 500 ns. As seen in Table I, the trimer not more structured than the dimer: β -strands are found with an average propensity of 23% and α -helices with a propensity of 7%. These values are, within the error margin, equivalent to those found for the dimer. However, β -sheets are found mostly in region 2-10, with a peak at 6-8 reaching a $\sim 80\%$ probability, and region 22-33, with two peaks at 22-24 ($\sim 25\%$) and 28-33 ($\sim 45\%$) as shown in Figure 3. Interestingly, the first peak in this region is absent in the dimer suggesting that more regions can be stabilized into β -strands as monomers are being added. α -helices, for their part, display a weak but non-zero formation probability for almost the whole sequence with a peak reaching 30% between residues 15 and 19.

The third row of Fig. 4 shows the center of the five clusters, T1 to T5, accounting for at least 5% of the total population. Together, they represent 44% of all sampled structures at 300 K. Like for the dimer, clusters counting at least 1% of all configurations represent 30% of the total set and add diversity mostly in the tertiary structures determined by the β -sheet organization. Globally, as seen in Table II, α -helices spanning residues 14-19 are present in T1, T2, T3, and T5. α -helices spanning residues 2-12, also observed in the monomer, appear only in T3.

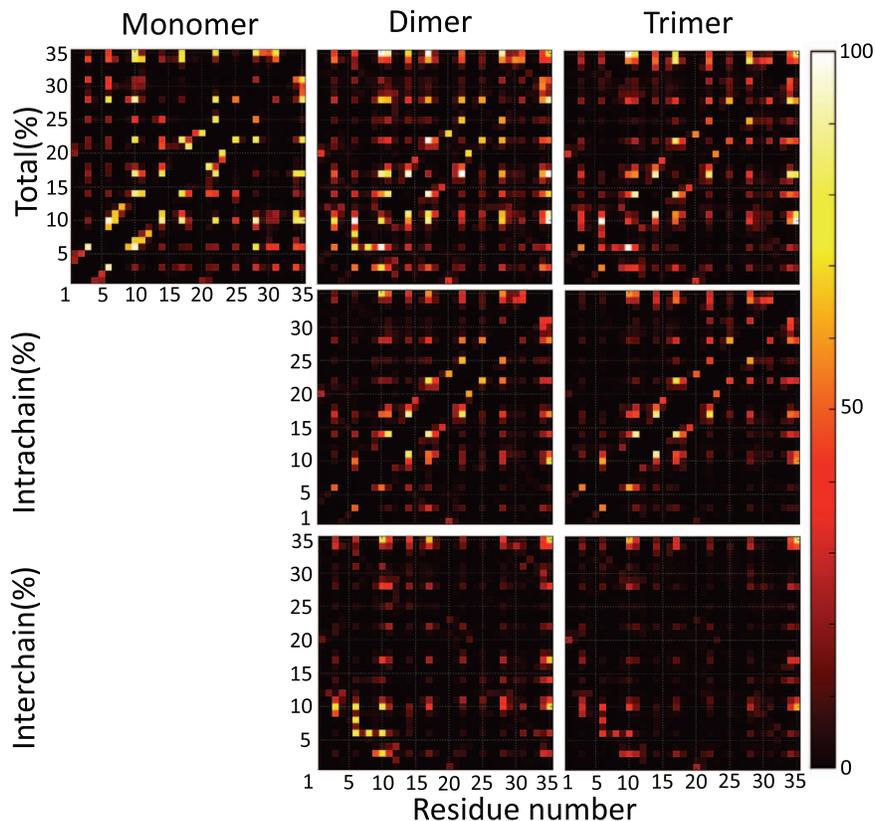


FIG. 5. The contact maps display the contact propensity of total (first row), intrachain (second row), and interchain (third row) contacts for NAC of monomer (first column), dimer (second column), and trimer (third column). The total contact propensity is calculated from the intrachain and interchain contacts.

As in the dimer, we observe that a few β -strand prone regions adopt a rich diversity of contacts for forming exclusively anti-parallel β -sheets. For example, region 6-8 is the region with the highest contact density. In all five dominant clusters (T1 to T5), we observe at least one antiparallel β -sheet formed by the 6-8 region of two chains. The sheets extend to four residues (6-9/6-9) in T5, to seven residues (2-8/6-12) in T3, and to nine residues (2-10/4-12) in T4. This region also forms β -sheet with other NAC segments. In one case (T2), we observe a six-strand β -sheet formed by the 6-8, 30-33, and 22-24 regions of two chains. As with the dimer, the region 6-8 also interacts with the 31-33 region (T1, T2, T4, and T5) and the 26-29 region (T5), always in anti-parallel orientation. We also note that a region of high contact density not observed for the dimer appears at residues 22-24 and

binds with the 31-33 region (T2, T4, and T5) into intrachain antiparallel β -sheets. As for the dimer, interchain β -sheets are found in all clusters, showing the importance of these interactions for breaking α -helices.

Figure 8 shows the contact maps for the intrachain, interchain, and total contacts for all five clusters. While the contact density is smaller than for the dimer, explaining the reduced structural stability observed in the clustering, for example, we can still observe the same interplay between intrachain and interchain contacts that lead to very similar total contact maps for the six clusters. The role of salt-bridges for the trimer is similar to the dimer: they occur with a high probability both intrachain and interchain, but with no specificity attached to particular clusters, beyond what is described above.

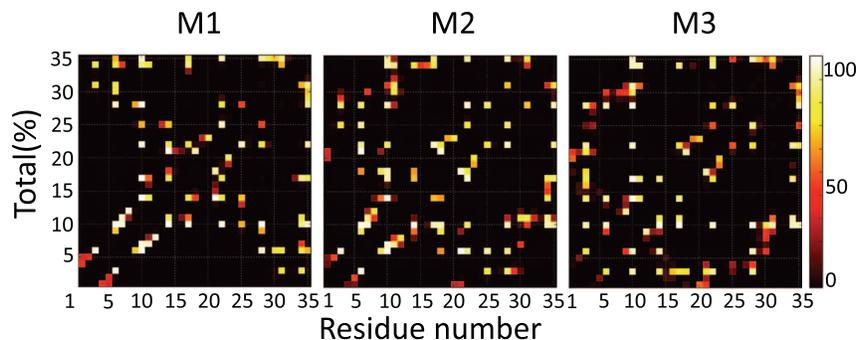


FIG. 6. The contact maps display the contact propensity of total (first row) contacts for the dominant morphology of monomer, see the first column of Fig. 5.

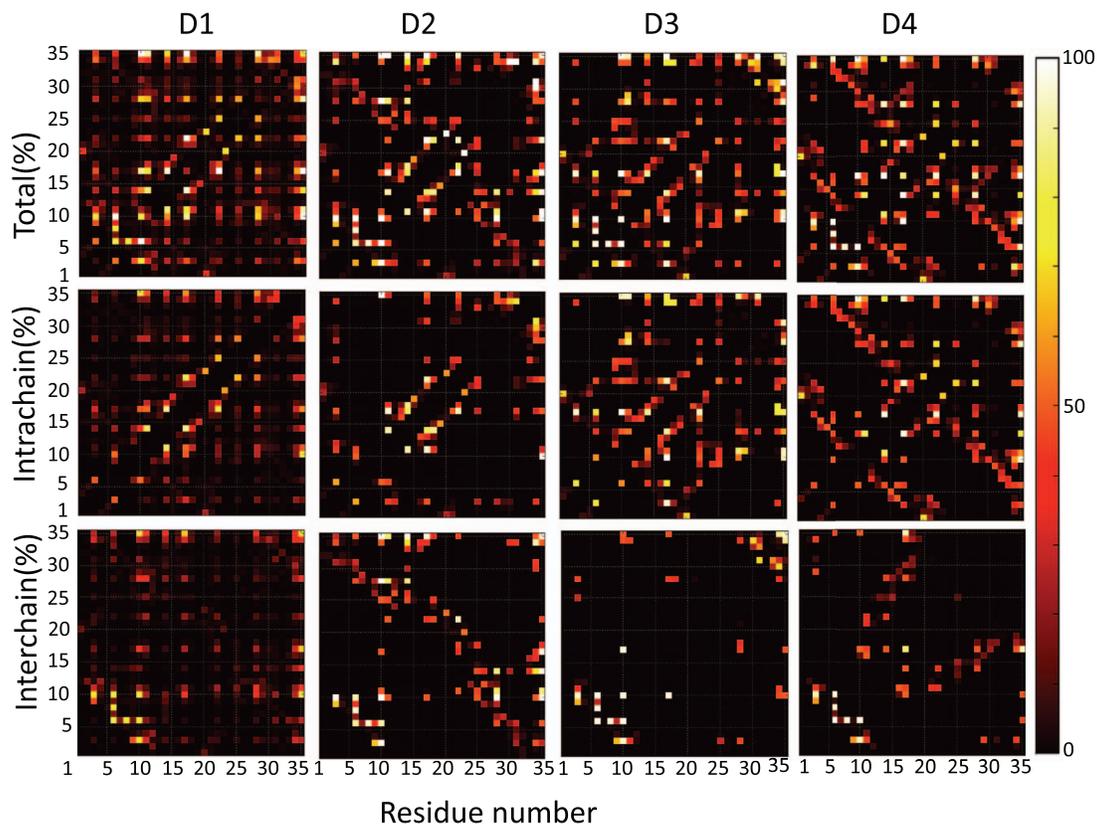


FIG. 7. The contact maps display the contact propensity of total (first row), intrachain (second row), and interchain (third row) contacts for the dominant morphology of dimer, see the second column of Fig. 5. The total contact propensity is calculated from the intrachain and interchain contacts.

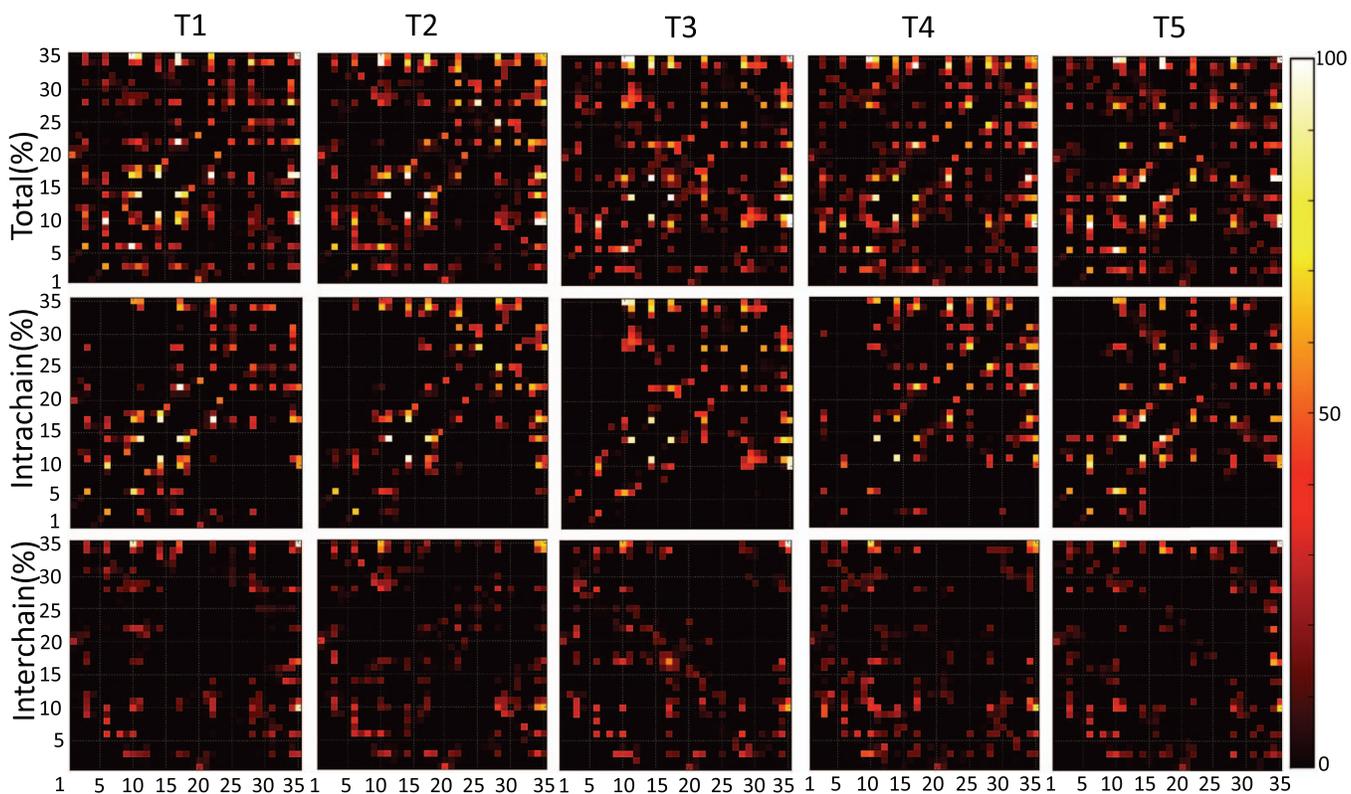


FIG. 8. The contact maps display the contact propensity of total (first row), intrachain (second row), and interchain (third row) contacts for the dominant morphology of trimer, see the third column of Fig. 5. The total contact propensity is calculated from the intrachain and interchain contacts.

IV. DISCUSSION

A. Evolution of the secondary structure

We first look at the structural changes that occur as NAC goes from monomer to dimer to trimer. Previous studies have shown that the formation of β -sheet was an important factor promoting the initiation of nucleation causing fibrillation.⁵⁹⁻⁶¹ Following these studies, we expect that the ability to form β -sheets increases with the number of available chains.

In the case of the NAC fragment, the monomer shows no β -strand at all but two predominantly α -helical regions at residues 2-13 and 19-25 (positions 62-73 and 79-85 with respect to the full α -synuclein).

Full-length α -synuclein simulations and experiments indicate that the monomer displays a β -sheet propensity of 11% and an α -helical propensity of only 2%–3%, with little helices in the NAC region.¹⁶ Yet, an NMR study of an extended NAC fragment (57-102 as compared to 61-95), with both the N-terminal and the C-terminal dominated by charged residues, finds no β -strand and suggests three-helix structures at 58-63, 70-80, and at 88-92, with the latter one, rather unstable.³³ The first helix overlaps with a relatively high α -helical propensity region for the full length α -synuclein, and the third one is in a region that is found to form helices with a non-zero but low probability (around 5%).¹⁶ Results on the extended NAC suggest therefore that, for the monomer, β -strands require contacts outside of the NAC region to be stabilized. This is consistent with one of the β -rich conformation shown in Figure 6 of Ullman *et al.*¹⁶

The first helix observed in our simulation of the NAC monomer is shifted compared to both experiments. Although heteronuclear single quantum coherence measurements would indicate the NAC structure is not affected by those,^{33,62} the shift on this first helix is to be expected since the α -helices observed in both experiments are centered on residue 60 and extend only to residues 63 or 64 (residue 3 or 4 of the NAC region). Interestingly, the second helix, while shifted with respect to Bisaglia *et al.* overlaps with a region found to display a low but non-zero helical propensity in full-length α -synuclein.¹⁶ While our structure for the monomer is compatible with available experiments, direct measurement on the structure of the NAC monomer, an important fragment, is needed to confirm these conclusions and provide more detailed information. In particular, chemical shifts (Fig. 9), which are predicted using the highly reproducible SPARTA+ package,^{56,57} should offer a direct comparison with experiment.

We observe a clear evolution from α -helical to β -strand secondary going from monomer to trimer (see Table I). In total, the α -helix propensity goes from 40% to 8%, between the monomer and dimer, and remains almost constant, at 7%, for the trimer, while the β -strand propensity moves from less than 1% in the monomer to plateau around 24% and 23% for the dimer and trimer, respectively, a transition characteristic of amyloid aggregation.⁶³⁻⁶⁵ As we see in Fig. 2, however, the entropy for the trimer still shows some small but visible fluctuations when computed over 250 ns time intervals. This suggests that the relevant phase space for this system is significantly larger than for the dimer. Longer sim-

ulations, which are beyond our current computational capabilities, could therefore lead to further increase in secondary structure at the trimer level. Already, however, the addition of a third monomer increases the regions where β -sheet is significant.

The first α -helix at residues 2-13, present at 75% in the monomer, disappears and is replaced, instead, by a β -strand most stable at residues 6-8 that form exclusively interchain sheets. The second helix, at residues 19-25, more stable in the monomer (90% propensity) is shifted mostly to residues 14-19, and is present in most dominant clusters except D1, D4, and T4, with an overall 25% propensity.

The β -strands are mostly found between residues 6-9 (90%) and 30-33 (30%) in both dimers and trimers although a third important region, between 22-24 is also found in the trimer, extending the β -strand prone regions with respect to the dimer (25%). Previous studies define five important regions of β -strands appear in oligomerization of α -synuclein, with three in the NAC region corresponding to residues 2-6, 8-17, and 30-35.^{59,60,66} The last region is in agreement with our findings. The first two are shifted by a few residues. This could be due to the fact that the experimental results are for the full protein and allow for interactions with other segments.²⁶

A number of α -synuclein studies emphasize the role of the GXXX motif, generally associated with amyloidogenic propensity in many amyloidogenic proteins.^{26,40} As discussed in Sec. II, the NAC region contains three GXXX motifs at positions 8, 13, and 24. These motifs do not overlap significantly with any of the dominant β -strands observed for the dimer and trimer, in agreement with various experimental observations that find that this motif is not involved in the fibrillation process of α -synuclein.^{26,32}

B. The polymorphism

Polymorphism appears to be a hallmark for a majority of amyloid proteins and applies to the existence of distinct morphologies with well-defined structures.^{67,68} It can be observed during the formation of amyloid and affect the final form of fibrils.⁶⁹ Studies have been conducted to link polymorphism and multiple misfolding pathway⁷⁰ and the neurodegenerative diseases in which they appear.⁷¹ Experimentally, polymorphism means that the final structure observed experimentally does not only depend on the sequence but also on the concentration and the preparation details. It reflects the possibility that there are several mechanisms of oligomerization leading to aggregation.⁷²

Polymorphism should not be confused with the diversity of partially disordered conformations that are generally observed for the small oligomers of amyloid peptides such as those studied here. Beyond this rich structural diversity, however, polymorphism is clearly observable in the NAC dimer and trimer. Indeed, already at the dimer level, the NAC peptide presents a very defined secondary structure, with a few dominant β -strand and α -helical regions. These regions, however, interact to generate a rich set of tertiary and quaternary structures while maintaining a distribution of interchain

and intrachain contacts that project an almost constant overall contact map and salt-bridges (Figs. 7 and 8). For example, the 30-34 region has great flexibility and can bind to very different regions: 6-8 (D1, T1, T2, T4, and T5), 30-34 (D1 and D3), and 22-24 (T2, T4, and T6). For the monomer, we observe that the relative position of the helices then change it to be composed of the same residues. In the dimer, the antiparallel β -strands can alternate between chains (D1 and D2) or show a strong symmetry (D3 and D4). For the trimer, T2 and T5 have the same regions involved in their secondary structure. However, T5 structure uses three chains in interchain contact, while T2 uses only two chains for its β -sheets and the third adopts an α -helix conformation.

These results show that polymorphism can appear at the very early stages of aggregation. Even when structures are still partially disordered, we observe that some regions show very stable secondary structure elements that can assemble into a rich set of well-defined tertiary structures. This suggests that the onset of oligomeric growth can play a defining role in determining the final amyloid structure in agreement with experiments.^{67,68}

C. Comparison with other amyloid proteins

It is interesting to compare the structural changes taking place from the monomeric to the dimeric state for various amyloid sequences. Oligomers share a number of structural signatures⁷³ but very little is known, at the moment, about the microscopic similarities and differences in the growth of these sequences. Beyond the fundamental interest, there are also two more pragmatic advantages: (1) comparing the behavior of various peptides simulated with the same conditions, it is possible to separate, at least partially, the technical limitations, associated with the forcefield and the simulation approach, from the sequence-related behavior, and to extract, from their global and sequence-specific properties; (2) the comparison also allows to identify which sequence is more accessible from a numerical point of view, given the computational timescale and size limitations that we are still facing. This is particular useful when using a coarse-grained potential, such as OPEP, which allows better equilibrium on large systems such as those presented here, but at the costs of more uncertainty with respect to the final predictions. In spite of the advantage of comparing results over many sequences, most comparisons have been limited to very close sequences such as $A\beta_{1-40}$ and $A\beta_{1-42}$ ⁷⁴⁻⁷⁶ or mutations.³⁷ Here, we compare our results on the NAC region of α -synuclein to results obtained to two other amyloid proteins studied with the same potential and computational techniques: $A\beta$ and human islet amyloid polypeptide (hIAPP), giving us a first insight on these questions.^{35,37,39,42}

$A\beta$ is a 40 or 42 residue protein and the major component of the fibrils involved in Alzheimer disease. It contains many charged residues, but the structure is globally neutral and has the same number of hydrophobic residues as the NAC region. Pancreatic-amyloid deposits in humans consist mainly of β -sheet fibrillar aggregates of the 37-residue polypeptide, hIAPP. This protein is positively charged and counts fewer hydrophobic residues than the NAC (Fig. 10).

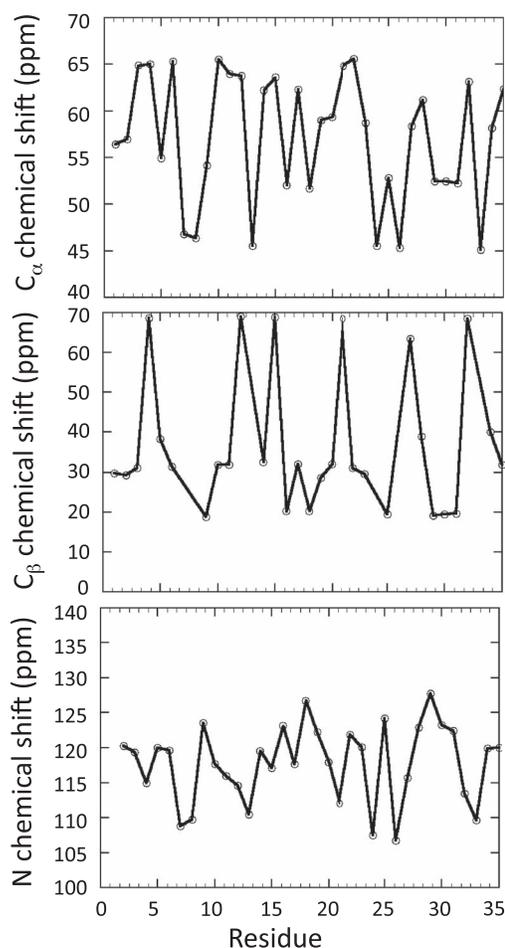


FIG. 9. From top to bottom Chemical shift of the C_{α} , C_{β} , and N of the 35 residues of NAC monomer at 300 K computing using the SPARTA+ chemical shift analysis package.^{56,57}

REMD-OPEP computational studies of $A\beta_{1-40}$ and $A\beta_{1-42}$ show, in agreement with most other simulations of these systems, that the monomer is mostly disordered with very low α -helical propensity (between 5% and 6%) and 11%–12% β -strand propensity due, in large part, to a relatively stable β -sheet between residues 2-5 and 10-13. With little secondary structure, $A\beta$ monomer samples a much more diverse set of structures than NAC, with its top 3 clusters representing between 28% and 42% of all visiting conformations, compared with 68% for NAC.³⁹ This changes as a second chain is added: The top four clusters for $A\beta$ represent 38% ($A\beta_{1-40}$) to 53% ($A\beta_{1-42}$) of all conformations while they account only for 28% of all visited conformations for NAC.

Looking at β -strand propensity, we find that it increases only slightly for $A\beta_{1-40}$, from 11.8% to 12.6%, and triples for



FIG. 10. We compare the sequence of NAC of α -synuclein with hIAPP and $A\beta_{1-40}$. We colored the hydrophobic residue in red and added the charge of charged residue which are implicated in the hydrophobic interaction and ionic bond, respectively.

$A\beta_{1-42}$, reaching 30.8% while, for NAC's dimer, it goes from less than 1% to 24%, a much higher increase. $A\beta$ dimerization for both sequences is characterized by a global increase in intermolecular hydrophobic contacts, mostly in the hydrophobic core (17-21) and the C-terminal (30-40) regions. Nevertheless, as observed also, for example, by Barz and Urbanc, in MD simulations with OPLS-AA forcefield and explicit solvent, intrachain contacts are more important than interchain contacts and the dimerization can be in large part described as a docking process.^{37,75} We observe a similar behavior with hIAPP. The presence of a disulfide bridge at Cys2-Cys7 stabilizes an α -helix at residues 5-19 in the N-terminal, a structure that is maintained during the dimerization (with propensity above 75%). This helix is combined with a β -strand (17-26 and 30-35) that sees its propensity going from 20% to 40% as the monomer combines into a dimer, for an overall β propensity going from about 5% to around 12%.^{35,42} In spite of this co-stabilization due to a higher interchain contact probability for the β -strand regions, the long N-terminal α -helix seems to limit the interdigitation between the two chains. The final dimeric structure is therefore akin to stacked monomers with structural elements stabilized by the presence of a nearby chain. At 300 K, the four dominant clusters account, here, for only 18% of all structures, in large part because of the high flexibility of α -helical N-terminal that shows very limited interchain interactions but remains very stable.³⁵

This comparison with other sequences simulated with the same potential and method helps identify the specificities of the NAC sequence. While all three proteins do form more β -sheets as they go from monomer to dimer, the transformation is most important for the NAC with the clearest α to β transition. This transformation is accompanied by a much more significant interchain contact map than is observed for $A\beta$ and hIAPP. If, for the latter two sequences, interchain contacts occur at a few points to stabilize secondary structure, NAC shows significant interdigitation, associated with quaternary structures, that is not compatible with a simple docking of one peptide over another one and an oligomerization process based on well-defined secondary structures then can form a richer set of tertiary structures than the two other sequences as can be seen while maintaining a very stable overall contact map. This suggests that, from a numerical point of view, the onset of aggregation for NAC might be much easier to simulate and characterize than for other sequences, allowing for much better complement between simulation and experiments.

V. CONCLUSION

In this study, simulated the onset of aggregation of the non-amyloid component of the α -synuclein protein (NAC). Our simulations show a clear α to β transition from the NAC monomer to the dimer as well as the presence of well-defined secondary structures, in agreement with the general consensus. Long simulations (up to 1,2 μ s per temperature for the trimer) allow us to reach equilibrium and to identify a rich polymorphism characterized by stable secondary structures that assemble to form a wide set of tertiary organizations with very similar contact maps. Overall, the NAC presents

a much clearer aggregation pathway, at the dimer and trimer level, than other amyloid sequences, such as $A\beta$ and hIAPP, that have been much more studied recently. This suggests that more attention, numerically, but also experimentally, should be given to this sequence to develop a first route to aggregation.

ACKNOWLEDGMENTS

This work is supported in part by the Natural Science and Engineering Research Council of Canada and the Canada Research Chair Foundation. This work was made possible through generous allocation of computer time by Calcul Québec/Compute Canada.

- ¹C. A. Ross and M. A. Poirier, *Nat. Med.* **10**(Suppl.), S10–S17 (2004).
- ²C. A. Ross and M. A. Poirier, *Nat. Rev. Mol. Cell Biol.* **6**, 891–898 (2005).
- ³G. M. Spillantini, M. L. Schmidt, V. M.-Y. Lee, J. Q. Trojanowski, R. Jakes, and M. Goedert, *Nature* **388**, 839–840 (1997).
- ⁴C. F. Lippa, H. Fujiwara, D. M. A. Mann, B. Giasson, M. Baba, M. L. Schmidt, L. E. Nee, B. O'Connell, D. A. Pollen, P. St. George-Hyslop, B. Ghetti, D. Nochlin, T. D. Bird, N. J. Cairns, V. M.-Y. Lee, T. Iwatsubo, and J. Q. Trojanowski, *Am. J. Pathol.* **153**, 1365–1370 (1998).
- ⁵C. F. Lippa, M. L. Schmidt, V. M. Lee, and J. Q. Trojanowski, *Ann. Neurol.* **45**, 353–357 (1999).
- ⁶R. Resende, S. C. F. Marques, E. Ferreira, I. Simões, C. R. Oliveira, and C. M. F. Pereira, *Neurochem. Res.* **38**, 797–806 (2013).
- ⁷N. Ramakrishna, H. C. Meeker, S. Patel, T. W. Brown, and A. El Idrissi, *J. Neurosci. Res.* **90**, 1589–1596 (2012).
- ⁸E. A. Waxman and B. I. Giasson, *Biochim. Biophys. Acta* **1792**, 616–624 (2009).
- ⁹J. E. Duda, V. M. Lee, and J. Q. Trojanowski, *J. Neurosci. Res.* **127**, 121–127 (2000).
- ¹⁰A. M. Bodles, D. J. S. Guthrie, B. Greer, and G. B. Irvine, *J. Neurochem.* **78**, 384–395 (2001).
- ¹¹J. P. Taylor, J. Hardy, and K. H. Fischbeck, *Science (N.Y.)* **296**, 1991–1995 (2002).
- ¹²R. P. Seung, J.-H. Lee, D.-H. Kim, C.-S. Chang, and Y.-S. Kim, *FEBS Lett.* **421**, 73–76 (1998).
- ¹³D.-P. Hong, A. L. Fink, and V. N. Uversky, *J. Mol. Biol.* **383**, 214–223 (2008).
- ¹⁴M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, *J. Am. Chem. Soc.* **127**, 476–477 (2005).
- ¹⁵K.-P. Wu, D. S. Weinstein, C. Narayanan, R. M. Levy, and J. Baum, *J. Mol. Biol.* **391**, 784–796 (2009).
- ¹⁶O. Ullman, C. K. Fisher, and C. M. Stultz, *J. Am. Chem. Soc.* **133**, 19536–19546 (2011).
- ¹⁷J. Yoon, S. Jang, K. Lee, and S. Shin, *J. Biomol. Struct. Dyn.* **27**, 259–270 (2009).
- ¹⁸I. F. Tsigelny, P. Bar-On, Y. Sharikov, L. Crews, M. Hashimoto, M. A. Miller, S. H. Keller, O. Platoshyn, J. X.-J. Yuan, and E. Masliah, *FEBS J.* **274**, 1862–1877 (2007).
- ¹⁹J. D. Perlmutter, A. R. Braun, and J. N. Sachs, *J. Biol. Chem.* **284**, 7177–7789 (2009).
- ²⁰D. Balesh, J. Ramjan, and W. B. Floriano, *J. Biophys. Chem.* **2**, 124–134 (2011).
- ²¹T. Gurry, O. Ullman, C. K. Fisher, I. Perovic, T. Pochapsky, and C. M. Stultz, *J. Am. Chem. Soc.* **135**, 3865–3872 (2013).
- ²²K. Ueda, H. Fukushima, E. Masliah, Y. Xia, A. Iwai, M. Yoshimoto, D. A. Otero, J. Kondo, Y. Ihara, and T. Saitoh, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11282–11286 (1993).
- ²³A. Iwai, M. Yoshimoto, E. Masliah, and T. Saitoh, *Biochemistry* **34**, 10139–10145 (1995).
- ²⁴A. M. Bodles, O. M. A. El-Agnaf, B. Greer, D. J. S. Guthrie, and G. B. Irvine, *Neurosci. Lett.* **359**, 89–93 (2004).
- ²⁵O. M. A. El-Agnaf and G. B. Irvine, *Biochem. Soc. Trans.* **30**, 559–565 (2002).
- ²⁶V. N. Uversky and A. L. Fink, *FEBS Lett.* **522**, 9–13 (2002).
- ²⁷R. Shaltiel-Karyo, M. Frenkel-Pinter, N. Egoz-Matia, A. Frydman-Marom, D. E. Shalev, D. Segal, and E. Gazit, *PLoS One* **5**, e13863 (2010).

- ²⁸Y. Liu and D. Schubert, *J. Neurochem.* **71**, 2322–2329 (1998).
- ²⁹O. M. A. El-Agnaf, R. Jakes, M. D. Curran, D. Middleton, R. Ingenito, E. Bianchi, A. Pessi, D. Neill, and A. Wallace, *FEBS Lett.* **440**, 71–75 (1998).
- ³⁰A. Kazmierczak, J. B. Strosznajder, and A. Adamczyk, *Neurochem. Int.* **53**, 263–269 (2008).
- ³¹A. Kazmierczak, G. A. Czapski, A. Adamczyk, B. Gajkowska, and J. B. Strosznajder, *Neurochem. Int.* **58**, 206–214 (2011).
- ³²O. M. A. El-Agnaf and G. B. Irvine, *J. Struct. Biol.* **130**, 300–309 (2000).
- ³³M. Bisaglia, A. Trolio, M. Bellanda, E. Bergantino, L. Bubacco, and S. Mammi, *Protein Sci.* **15**, 1408–1416 (2006).
- ³⁴D. P. Karpinar, M. B. Gajula Balija, S. Kügler, F. Opazo, N. Rezaei-Ghaleh, N. Wender, H.-Y. Kim, G. Taschenberger, B. H. Falkenburger, H. Heise, A. Kumar, D. Riedel, L. Fichtner, A. Voigt, G. H. Braus, K. Giller, S. Becker, A. Herzig, M. Baldus, H. Jäckle, S. Eimer, J. B. Schulz, C. Griesinger, and M. Zweckstetter, *EMBO J.* **28**, 3256–3268 (2009).
- ³⁵R. Laghaei, N. Mousseau, and G. Wei, *J. Phys. Chem. B* **115**, 3146–3154 (2011).
- ³⁶G. Liang, J. Zhao, X. Yu, and J. Zheng, *Biochemistry* **52**, 1089–1100 (2013).
- ³⁷S. Côté, R. Laghaei, P. Derreumaux, and N. Mousseau, *J. Phys. Chem. B* **116**, 4043–4055 (2012).
- ³⁸R. Laghaei and N. Mousseau, *J. Chem. Phys.* **132**, 165102 (2010).
- ³⁹S. Côté, P. Derreumaux, and N. Mousseau, *J. Chem. Theory Comput.* **7**, 2584–2592 (2011).
- ⁴⁰J. T. Jarrett and T. L. Peter, Jr., *Biochemistry* **31**, 12345–12352 (1992).
- ⁴¹J. Maupetit, P. Tuffery, and P. Derreumaux, *Proteins: Struct., Funct., Bioinf.* **69**, 394–408 (2007).
- ⁴²R. Laghaei, N. Mousseau, and G. Wei, *J. Phys. Chem. B* **114**, 7071–7077 (2010).
- ⁴³W. Chen, N. Mousseau, and P. Derreumaux, *J. Chem. Phys.* **125**, 084911 (2006).
- ⁴⁴G. Wei, N. Mousseau, and P. Derreumaux, *Prion* **1**, 3–8 (2007).
- ⁴⁵X. Dong, W. Chen, N. Mousseau, and P. Derreumaux, *J. Chem. Phys.* **128**, 125108 (2008).
- ⁴⁶Y. Chebaro, N. Mousseau, and P. Derreumaux, *J. Phys. Chem. B* **113**, 7668–7675 (2009).
- ⁴⁷A. Melquiond, X. Dong, N. Mousseau, and P. Derreumaux, *Curr. Alzheimer Res.* **5**, 244–250 (2008).
- ⁴⁸Y. Lu, G. Wei, and P. Derreumaux, *J. Phys. Chem. B* **115**, 1282–1288 (2011).
- ⁴⁹H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684–3690 (1984).
- ⁵⁰H. C. Andersen, *J. Comput. Phys.* **52**, 24–34 (1983).
- ⁵¹J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, *J. Chem. Theory Comput.* **3**, 26–41 (2007).
- ⁵²D. Frishman and P. Argos, *Proteins: Struct., Funct., Genet.* **23**, 566–579 (1995).
- ⁵³X. Daura, R. Suter, and W. F. van Gunsteren, *J. Chem. Phys.* **110**, 3049–3055 (1999).
- ⁵⁴M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).
- ⁵⁵G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack, Jr., *Proteins: Struct., Funct., Bioinf.* **77**, 778–795 (2009).
- ⁵⁶Y. Shen and A. Bax, *J. Biomol. NMR* **38**, 289–302 (2007).
- ⁵⁷Y. Shen and A. Bax, *J. Biomol. NMR* **48**, 13–22 (2010).
- ⁵⁸D. S. Wishart and A. M. Nip, *Biochem. Cell Biol.* **76**, 153–163 (1998).
- ⁵⁹J.-E. Suk, S. B. Lokappa, and T. S. Ulmer, *Biochemistry* **49**, 1533–1540 (2010).
- ⁶⁰M. Vilar, H.-T. Chou, T. Lühns, S. K. Maji, D. Riek-Loher, R. Verel, G. Manning, H. Stahlberg, and R. Riek, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8637–8642 (2008).
- ⁶¹M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucalè, L. Bubacco, and B. Samorì, *PLoS Biol.* **6**, 99–108 (2008).
- ⁶²M. Bisaglia, I. Tessari, L. Pinato, M. Bellanda, S. Giraud, M. Fasano, E. Bergantino, L. Bubacco, and S. Mammi, *Biochemistry* **44**, 329–339 (2005).
- ⁶³K.-M. Pan, M. Baldwin, J. Nguyen, M. Gasset, A. N. A. Serban, D. Groth, I. Mehlhorn, Z. Huang, R. J. Fletterick, F. E. Cohen, and S. B. Prusiner, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10962–10966 (1993).
- ⁶⁴D. M. Walsh, *J. Biol. Chem.* **274**, 25945–25952 (1999).
- ⁶⁵M. D. Kirkitadze, M. M. Condrón, and D. B. Teplow, *J. Mol. Biol.* **312**, 1103–1119 (2001).
- ⁶⁶S. M. Celej, R. Sarroukh, E. Goormaghtigh, G. Fidelio, J.-M. Ruyschaert, and V. Raussens, *Biochem. J.* **443**, 719–726 (2012).
- ⁶⁷M. R. Smaoui, F. Poitevin, M. Delarue, P. Koehl, H. Orland, and J. Waldispühl, *Biophys. J.* **104**, 683–693 (2013).
- ⁶⁸R. Tycko and R. B. Wickner, *Acc. Chem. Res.* **46**, 1487–1496 (2013).
- ⁶⁹R. Kodali and R. Wetzel, *Curr. Opin. Struct. Biol.* **17**, 48–57 (2007).
- ⁷⁰K. Hun Lim, *J. Phys. Chem. Biophys.* **3**, 2–3 (2013).
- ⁷¹Y. Zhang, B. Wang, H. Wan, Q. Zhou, and T. Li, *Neurosci. Lett.* **541**, 132–137 (2013).
- ⁷²J. S. Jeong, A. Ansaloni, R. Mezzenga, H. A. Lashuel, and G. Dietler, *J. Mol. Biol.* **425**, 1765–1781 (2013).
- ⁷³R. Kaye, E. Head, J. L. Thompson, T. M. McIntire, S. C. Milton, C. W. Cotman, and C. G. Glabe, *Science (N.Y.)* **300**, 486–489 (2003).
- ⁷⁴B. Urbanc, M. Betnel, L. Cruz, G. Bitan, and D. B. Teplow, *J. Am. Chem. Soc.* **132**, 4266–4280 (2010).
- ⁷⁵B. Barz and B. Urbanc, *PLoS One* **7**, e34345 (2012).
- ⁷⁶Y. Atsushi, O. Akisumi, N. Kazuya, H. Shin'ichi, and K. Noriyuki, *Chem. Phys. Lett.* **595–596**, 242–249 (2014).