

Relationship between population of the fibril-prone conformation in the monomeric state and oligomer formation times of peptides: Insights from all-atom simulations

Hoang Bao Nam,¹ Maksim Kouza,² Hoang Zung,³ and Mai Suan Li^{2,a)}

¹*Institute for Computational Science and Technology, 6 Quarter, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam*

²*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, Warsaw 02-668, Poland*

³*Computational Physics Laboratory, Vietnam National University, Ho Chi Minh City, 227 Nguyen Van Cu, Dist. 5, Vietnam*

(Received 5 January 2010; accepted 5 April 2010; published online 30 April 2010)

Despite much progress in understanding the aggregation process of biomolecules, the factors that govern its rates have not been fully understood. This problem is of particular importance since many conformational diseases such as Alzheimer, Parkinson, and type-II diabetes are associated with the protein oligomerization. Having performed all-atom simulations with explicit water and various force fields for two short peptides KFFE and NNQQ, we show that their oligomer formation times are strongly correlated with the population of the fibril-prone conformation in the monomeric state. The larger the population the faster the aggregation process. Our result not only suggests that this quantity plays a key role in the self-assembly of polypeptide chains but also opens a new way to understand the fibrillogenesis of biomolecules at the monomeric level. The nature of oligomer ordering of NNQQ is studied in detail. © 2010 American Institute of Physics.
[doi:10.1063/1.3415372]

I. INTRODUCTION

Many structural diseases like Alzheimer, Parkinson, and type-II diabetes are associated with the oligomerization of peptides and proteins.¹ This prompts researchers to study factors that drive the fibril formation process. The ability of a given polypeptide chain to aggregate under specific conditions depends dramatically on its composition and sequence. Common structural characteristics of highly organized aggregates such as fibrils formed from proteins without detectable sequence or structural similarity² suggest that the propensity of proteins to aggregate can be described by general principles.

Recent experiments revealed that the fibril formation times τ_{fib} depend on a number of factors like the hydrophobicity of side chains (SC),³ net charge,⁴ patterns of polar and nonpolar residues,⁵ diverse secondary structure elements,⁶ aromatic interactions,⁷ and the population of the fibril-prone conformation N^* , P_{N^*} , in the monomeric state.⁸ All-atom simulations of short peptides^{9–12} partially support these findings at the qualitative level but not on the quantitative one because due to limitation of computational facility the explicit dependence of oligomerization rates on those factors was not obtained.

Studying amyloid peptide $A\beta_{15-25}$ by all-atom simulations, it was found that P_{N^*} with the lactam bridge D23-K28 is larger than the wild-type case.¹³ Because the fixation of D23 and K26 increases the oligomerization rate by ≈ 1000 times,¹⁴ it was hypothesized that these two effects are related but the fibril formation time was not estimated

theoretically.¹³ Using the simple lattice model Li *et al.*^{15,16} have shown that the self-assembly of polypeptide chains occurs at the temperature where P_{N^*} reaches maximum. Therefore, the enhancement of population of the fibril-prone conformation probably facilitates the aggregation but this conclusion has not been confirmed by all-atom models yet.

In this paper we study the role of population of fibril-prone conformation in the monomeric state in promoting oligomerization using all-atom simulations. To this end we perform all-atom simulations with explicit water for two peptides KFFE and NNQQ with the help of the Gromos96 force field 43a1 (Ref. 17) as well as the OPLS (Ref. 18) and Amber 99 (Ref. 19) force fields. The choice of these short peptides is dictated by the fact that they allow for estimating τ_{fib} for dimers and tetramers with a reasonable amount of CPU time. Therefore, contrary to previous studies, one can obtain the dependence of τ_{fib} on P_{N^*} directly from all-atom simulations. Since the experiments^{20,21} have shown that KFFE and NNQQ are β -strands in the fibril state, we defined N^* as an extended state (see Sec. II for more details).

The self-assembly of peptide KFFE was studied experimentally²⁰ and theoretically,^{9,22,23} but its fibril formation rates have not been estimated. Recent x-ray diffraction analysis by Sawaya *et al.* has shown that NNQQ can form both parallel β -sheet fibrils and closely related structured microcrystals.²¹ However, a theoretical study of this peptide is still missing. So, our goal is not only to find the correlation between τ_{fib} and P_{N^*} , but also to study the nature of self-assembly of NNQQ.

We found that P_{N^*} of KFFE is higher than that of NNQQ. The fibril formation of dimer 2KFFE and tetramer

^{a)}Electronic mail: masli@ifpan.edu.pl.

TABLE I. Durations (in nanoseconds) of trajectories generated in simulations using three different force fields.

Trajectory	Gromos 43a1						OPLS			Amber 99
	KFFE	NNQQ	2KFFE	2NNQQ	4KFFE	4NNQQ	NNQQ	2NNQQ	NNQQ	2NNQQ
1	150	150	150	300	450	500	150	20	150	500
2	150	150	150	300	400	400		60		500
3			100	250	300	400		20		500
4			100	250	300	300		20		500

4KFFE was shown to be faster than 2NNQQ and 4NNQQ. Thus, all-atom models support the fact that the enhancement of P_{N^*} accelerates the oligomerization process. Using the Gromos96 force field 43a1, one can demonstrate that, in accordance with experiments²⁰ and the previous all-atom simulations by the OPLS force field,⁹ the fibril-like structure of KFFE consists of antiparallel β -sheets. In the NNQQ case, within one layer, peptides adopt rather antiparallel than parallel arrangement which has been observed experimentally.²¹ To clarify this departure from experiments, we carried out additional simulations using the OPLS (Ref. 18) and Amber 99 force fields.¹⁹ While the result followed from the simulations by the later force field is not conclusive, the OPLS force field also supports the antiparallel arrangement within one β -sheet. We also estimated the energies of parallel and antiparallel configurations of 2NNQQ and all possible bi-layer arrangements of the octamer (8NNQQ) using six different force fields. It turns out that all of these force fields favor the antiparallel configuration within one sheet.

Although the fibril formation times are different for KFFE and NNQQ, there is a little difference in mechanisms underlying their oligomerization process. For both dimers 2KFFE and 2NNQQ the hydrogen bond (HB) interactions dominate over the SC ones. This result is interesting because since KFFE has opposite charges at termini, the SC interactions are expected to play a more decisive role than hydrogen bonding as in the case of $A\beta_{16-22}$,^{10,24} but this does not happen in our case. For tetramers the contributions of two interactions to the oligomer ordering become compatible for both peptides.

II. MATERIAL AND METHOD

A. Definition of fibril-prone state N^*

It should be noted that there is not a unique microscopic structure that is aggregation prone. In fact there are basins of attraction (usually high free energy structures) many of which can aggregate. Because peptides KFFE and NNQQ adopt the beta-strand shape in the fibril-like state,^{20,21} N^* is defined as an extended state with the end-to-end distance $R \geq 0.9R_{\max}$, where $R_{\max} = 3a$. Here a is a typical distance between two neighboring C_α atoms, $a \approx 3.73$ Å.

B. Simulation details

NNQQ is a fragment derived from Yeast Prion Sup35 (PDB ID: 2OLX) while the initial conformation of KFFE was extracted from the x-ray diffraction structure of KFFEAAAKKFFE peptide (PDB ID: 2BFI). The terminal

residues of the later are oppositely charged (a positive charge on lysine and a negative charge on glutamic acid). The initial conformations of the dimers and tetramers were obtained by replicating the individual monomer structures in random orientations and putting them in space with distances of about 1 nm.

To probe the structural characteristics and fluctuations of monomers and self-assembly of oligomers, the simulation was performed by using mainly Gromos96 force field 43a1 (Ref. 17) for the peptides and the simple point charge water model.²⁵ The system is enclosed in the box with periodic boundary conditions to minimize finite size effects. Typically a monomer was placed in an orthorhombic box with the volume of ≈ 28 nm³ which contains about 900 water molecules. For dimers and tetramers we used 40 nm³- and 78 nm³-boxes which contain approximately 1270 and 2410 water molecules, respectively. The corresponding peptide concentration is ≈ 85 mM which is about three orders of magnitude higher than that used *in vitro* fibril growth conditions (≈ 100 μ M).²⁶ As a result, the interpeptide collision probability is greatly enhanced leading to faster formation of ordered structures. We generated two trajectories for monomer KFFE and NNQQ, four trajectories for 2KFFE 2NNQQ, 4KFFE, and 4NNQQ using Gromos 43a1. To check the robustness of our conclusion about the nature of oligomer ordering of NNQQ, we also made several runs using the OPLS and Amber 99. Durations of these runs are given in Table I, where the longest run is 500 ns.

C. Tools and measures used in analysis of data

Dihedral principal component analysis (dPCA). We use the dPCA that uniquely defines the distance in the space of periodic dihedral angles using the variables^{27,24} $q_{2k-1} = \cos(\alpha_k)$, and $q_{2k} = \sin(\alpha_k)$. Here, $\alpha_k \in \phi_k, \psi_k$ and $k = 1, 2, \dots, N$, with N being the number of backbone and SC dihedral angles. The correlated internal motions are probed using the covariance matrix $\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle$. The free-energy surface along the N -dimensional reaction coordinate $V = (V_1, \dots, V_N)$, obtained by diagonalizing σ , is given by $\Delta G(V) = -k_B T [\ln P(V) - \ln P_{\max}]$, where $P(V)$ is the probability distribution obtained from a histogram of the molecular dynamics (MD) data, and P_{\max} is the maximum of the distribution, which is subtracted to ensure that $\Delta G = 0$ for the lowest free energy minimum. We use dPCA to compute the free energy landscapes (FELs) using mainly the first two eigenvectors V_1 and V_2 .

Contact maps. We monitor the time evolution of the formation of the SC-SC contacts and HB contacts. A SC-SC

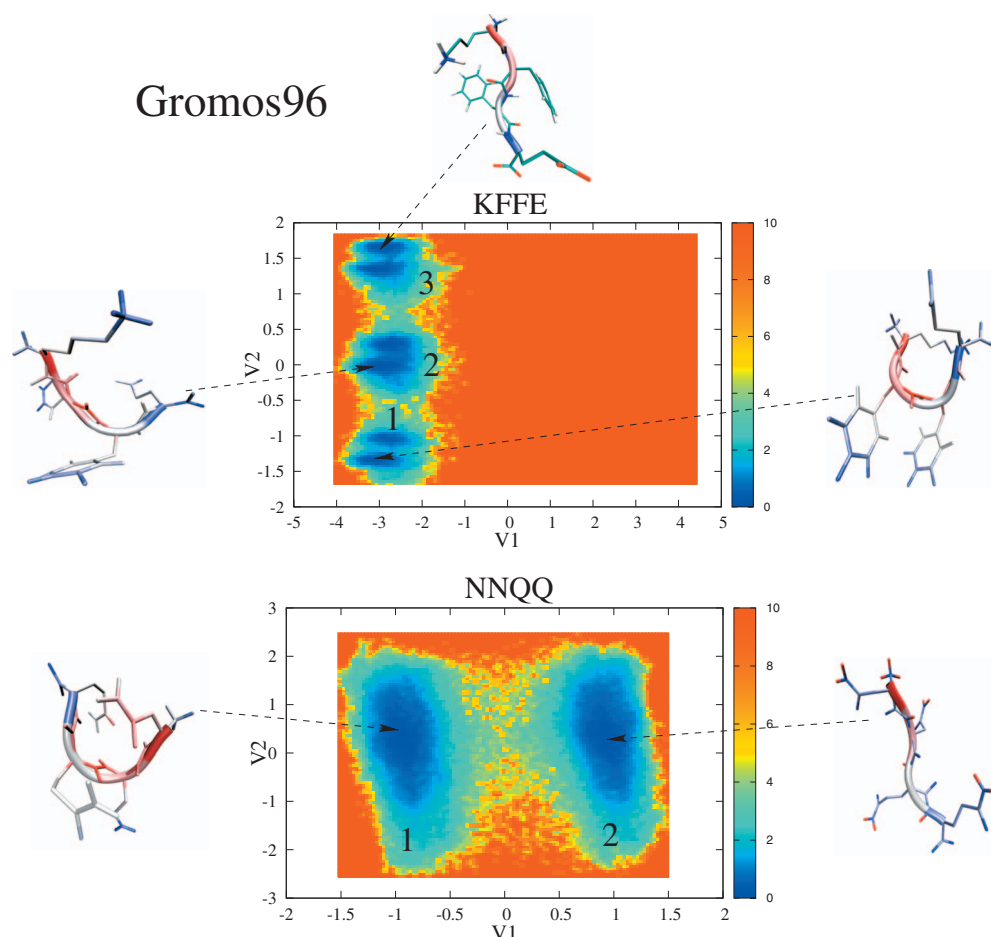


FIG. 1. FEL (in kJ/mol) for monomer KFFE (upper panel) and NNQQ (lower panel) as a function of principal component V_1 and V_2 . The results were obtained using the Gromos96 force field 43a1. Shown are typical snapshots for local minima. For KFFE, the end-to-end distance of snapshots is $R=0.51$, 0.85 , and 1.01 nm for local minima 1, 2, and 3, respectively. For NNQQ, we have $R=0.63$ nm (first minimum) and 1.01 nm (second minimum).

contact is formed if the distance between the centers of mass of two residues is ≤ 6 Å. A HB contact occurs provided the distance between donor D and acceptor A is ≤ 3.5 Å and the angle D-H-A is $\geq 135^\circ$.

Order parameter P_2 . To characterize the fibril state of short peptides we use the “nematic” order parameter P_2 as defined in Ref. 24. If P_2 is bigger than 0.5, then the system has the propensity to be in an ordered state. The fibril formation time, τ_{fib} , is defined as the first passage time to reach $P_2=0.85$.

Probability of the fibril-prone conformation in the monomeric state P_{N^*} . Using the definition of N^* , we define P_{N^*} as a probability for finding conformations with the end-to-end distance R larger than $0.9R_{\text{max}}$. R is computed using equilibrium conformations obtained in simulations of a single monomer.

III. RESULTS

Monomer KFFE is less stable than NNQQ. In this section we present results obtained by the Gromos force field. As evident from Fig. 1, in the monomeric state both peptides are not stable as free energy barriers are of a few $k_B T$. Because the two-dimensional FEL of KFFE has one local minimum more than NNQQ, the former is expected to be less stable. For KFFE, local minimum 1 is the most compact one

having small values of R (Fig. 1). The typical snapshot has the U-shape with two rings almost parallel as observed previously by Bellesia and Shea⁹ using the OPLS force field. Conformations of the second basin have a more extended U-shape compared to the first minimum, while within the basin of the third minimum β -conformations dominate. Interestingly, three similar local minima of FEL of KFFE have been obtained using not only a different force field (OPLS) but also different reaction coordinates. Thus, the FEL of monomer KFFE is robust against different force fields and it may be studied by different reaction coordinates. Folding to the nativelike minimum 1 starting from the unfolded state (minimum 3) proceeds via intermediates presented by the second local minimum. Free energy barriers between β - and U-shape conformations are of 1 kcal/mol.

For NNQQ, the FEL consists of two local basins, 1 and 2. The U-shape conformations largely populate the first basin which corresponds to the compact nativelike states with relatively small end-to-end distances. The second basin is mainly populated by β -extended conformations with larger values of R . As in the KFFE case, they are separated by a low free energy barrier. The folding/unfolding between two basins is not accompanied by intermediates.

Despite the fact that KFFE is bulkier than NNQQ, having more atoms (60 compared to 49) and bigger SCs the

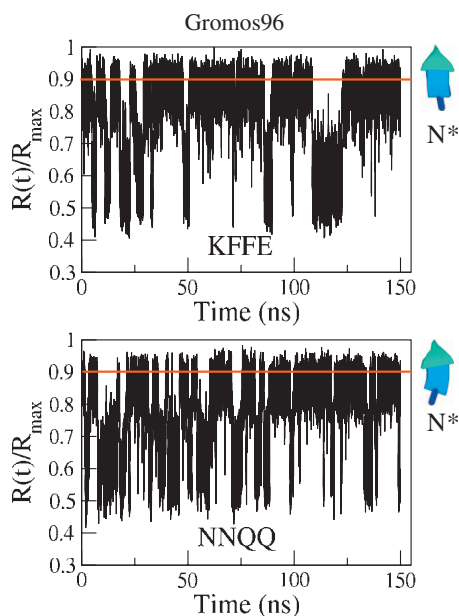


FIG. 2. Time dependence of the end-to-end distance renormalized by R_{\max} for monomer KFFE and NNQQ. The results were obtained using the Gromos96 force field 43a1. Here $R_{\max} \approx 3a$, and $a = 3.73$ Å. Results are shown every 1 ps. The red line refers to $R/R_{\max} = 0.9$. P_{N^*} is defined as the number of snapshots, which have $R/R_{\max} \geq 0.9$, divided by the total number of collected snapshots. A typical snapshot of the fibril-prone conformation N^* is shown in the right.

former is more flexible than the later. The difference in flexibility comes from different sequences. KFFE is composed of two typical kinds of peptides, charged (K and E) and apolar ones (two F residues). The charged residues strongly interact with solvent while the second ones tend to be hydrophobic. This contrast causes the structural instability in water environment. NNQQ, on the other hand, consists of four highly polar residues which interacts more uniformly with water and therefore is more settled.

One of possible principles governing the fibrillogenesis of polypeptide chains is that the instability of the native state of monomer would facilitate the oligomerization.¹ This is because if the monomeric native state is stable then it is hard to get a chain unfolded for aggregation to begin. Therefore, KFFE is expected to have a higher fibril formation rate than NNQQ.

Population of conformation N^ in the monomeric state of KFFE is higher than NNQQ.* In the case of lattice models,^{15,16} P_{N^*} of short enough chains may be obtained by exact enumeration.²⁸ For off-lattice models, the number of all possible conformations becomes infinite and P_{N^*} can be estimated approximately. Since N^* is extended and a chain is short we define it via the end-to-end distance (see Sec. II) and MD sampling. Figure 2 shows the time dependence of $R(t)$ for two peptides. Clearly, the probability of being in the N^* state with high value of R (or high β -content) of KFFE is higher than NNQQ. Averaging over two trajectories, we obtained $P_{N^*} \approx 24.6\%$ and 12.6% for KFFE and NNQQ, respectively. This result is consistent with the fact KFFE is less stable and may be understood as follows. Suppose Δ is a gap between N^* and the native state. Then $P_{N^*} \sim \exp(-\Delta/k_B T)$, the higher value of which would correspond to a smaller gap

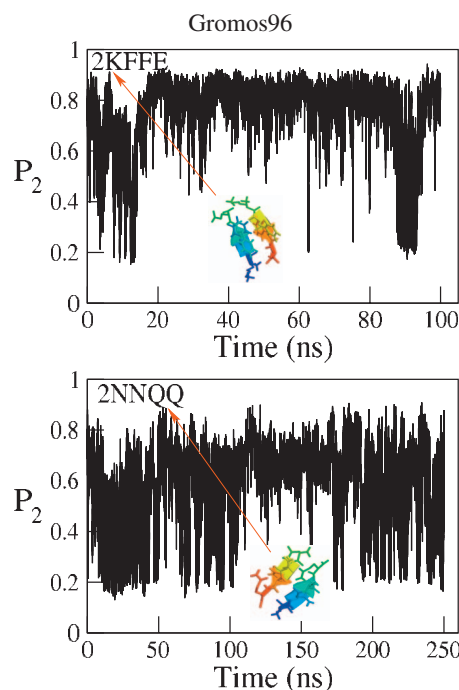


FIG. 3. Time dependence of the order parameter P_2 , obtained by the Gromos96 force field 43a1 for 2KFFE and 2NNQQ. Shown are snapshots of the anti-parallel fibril-like conformations. For these conformations $P_2 \approx 0.9$.

or lower stability of the native state. From this point of view one can use either P_{N^*} or the stability of the monomeric native state to gain insights on propensity to aggregation of biomolecules but the former is easier to obtain numerically. Therefore, we focus on the relationship between P_{N^*} and τ_{fib} .

Dependence of P_{N^} on force fields.* Stability of a monomer and thus P_{N^*} should depend on models we use. To show this we made 150 ns run for NNQQ using the Amber 99 and OPLS force fields within the GROMACS suite. NNQQ is chosen to study by other force fields also because the nature of its oligomeric ordering remains largely ambiguous within the Gromos model (see below). From the time dependence of $R(t)$ [Fig. S1 in the supplementary material (SM)]³⁸ we obtain $P_{N^*} \approx 11.5\%$ and 0.5% for the OPLS and Amber 99 force fields, respectively. The OPLS provides P_{N^*} compatible with the Gromos96 force field 43a1, while the Amber 99 gives considerably lower population of N^* in the monomeric state. This is because the Amber 99 was shown to disfavor the beta content²⁹ (the Gromos96 favors the beta structure while OPLS has intermediate tendency). As evident later, the Gromos and OPLS force fields give compatible short time scales for oligomer formation, but the Amber 99 strongly disfavors self-assembly of NNQQ.

Correlation between the population of N^ in the monomeric state and τ_{fib} .* To characterize the fibril ordering we use the nematic liquid crystal order parameter P_2 .²⁴ Large conformational changes are reflected in its dynamics shown in Fig. 3, where the fibril-like state of 2KFFE occurs earlier than 2NNQQ. The fibril formation time is defined as the first passage time to reach a conformation with $P_2 = 0.9$. Using this definition we obtained $\tau_{\text{fib}} = 6.6 \pm 4.0$ ns and 25.4 ± 9.8 ns for 2KFFE and 2NNQQ, respectively. Here τ_{fib} is the value averaged over four trajectories. 2KFFE shows

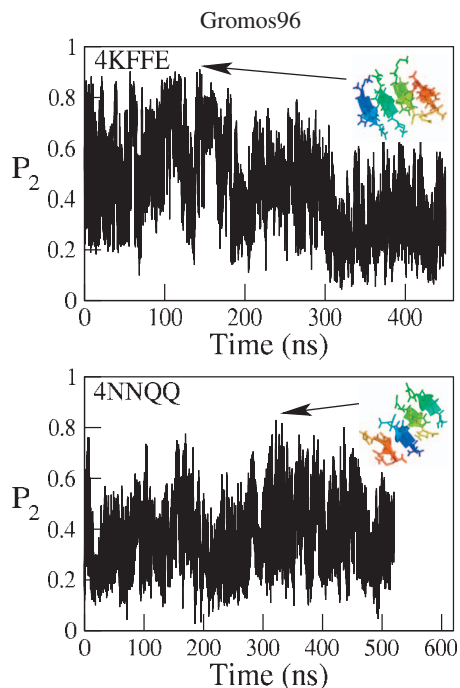


FIG. 4. The same as in Fig. 3, but for tetramers 4KFFE and 4NNQQ. In the fibril-like state antiparallel peptides lie in one layer.

less variation in P_2 compared to 2NNQQ suggesting that the fibril-like state of the former is more stable than the latter. This observation is compatible with the FEL analysis (see below).

In the case of tetramers, P_2 also fluctuates a lot (Fig. 4) presumably because the number of peptides $N=4$ is far below the size of the critical nucleus. One can expect that the critical nucleus size of KFFE and NNQQ is larger than 6 because for longer peptide $A\beta_{16-22}$ it exceeds 6.²⁴ τ_{fib} grows with the oligomer size and we averaged $\tau_{\text{fib}} \approx 74.3 \pm 30.2$ and 288.9 ± 69.1 ns for 4KFFE and 4NNQQ, respectively. Thus, using the Gromos96 force field 43a1, we can demonstrate that the larger population of conformation N^* in the monomeric state, the faster fibril formation. To make this conclusion more convincing, we considered the oligomerization of 2NNQQ using the OPLS (Fig. S2 in SM) and Amber 99 (Fig. S3 in SM) force fields. For OPLS three runs have duration of 20 ns and one run of 60 ns, while for Amber 99 all four trajectories are of 500 ns. Within the OPLS force field the self-assembly occurs at short time scales $\tau_{\text{fib}} \approx 24.3$ ns, which is close to the estimation by the Gromos force field. This is probably because two these force fields provide almost the same value of P_{N^*} . As in the Gromos case, OPLS gives the antiparallel orientation of peptides in the fibril state.

Contrary to the Gromos and OPLS, the Amber 99 strongly disfavors the aggregation of 2NNQQ having very low value of P_{N^*} . Maximum value of P_2 is ≈ 0.7 only in two runs (Fig. S3 in SM) and the fibril ordering, therefore, would appear at $\tau_{\text{fib}} > 500$ ns. Thus from the present MD simulations by Amber 99, it remains unclear if peptides of 2NNQQ are parallel or antiparallel in the fibril-like state. However, using the energetics argument below, we can show that the antiparallel orientation is more favorable.

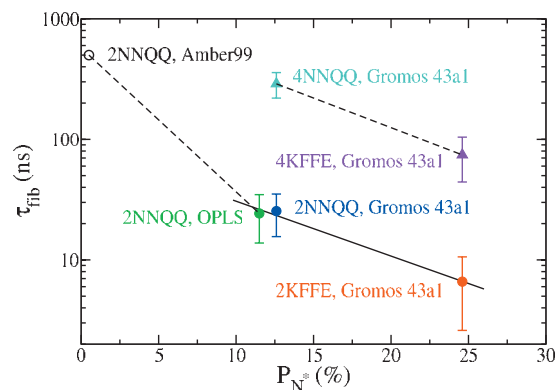


FIG. 5. Dependence of τ_{fib} on P_{N^*} , obtained by different force fields, for dimers (circles) and tetramers (triangles). The open circle refers to 500 ns of four runs for 2NNQQ using Amber 99. The real value of τ_{fib} for this case exceeds 500 ns. The solid straight line is a fit $y=4.472-0.105x$ which was obtained using three points (closed circles, except Amber 99). Dashed lines are for eye guidance. The error bars come from averaging over four trajectories.

The dependence of τ_{fib} on the population of the fibril-prone state in the monomeric state is shown in Fig. 5. Although we have made only four independent runs, relatively small error bars suggest that the sampling is sufficient for studying the relationship between τ_{fib} and P_{N^*} . The fibril formation time for tetramers was not estimated by the OPLS force field but it is probably compatible with that of the Gromos96 as these force fields have almost identical values of P_{N^*} . If P_{N^*} is less than 1% as in the Amber 99 case, the acquisition of fibril state within a reasonable amount of CPU time is almost impossible even for a dimer.

Using the results obtained by OPLS and Gromos force fields for dimers, we obtain $\tau_{\text{fib}} \sim \exp(-cP_{N^*})$, where $c \approx 0.105$ (Fig. 5). This dependence is at least valid for $P_{N^*} > 10\%$. Although our data are not sufficient to obtain the dependence of τ_{fib} on P_{N^*} for the whole region, they suggest that there is a crossover between two regimes at P_{N^*} of a few percents. The exponential dependence presumably always holds but constant c in the large P_{N^*} region is smaller (weaker dependence) than that in the small P_{N^*} region. Clarification of this question is of great interest but beyond our computational facilities.

Because population of N^* is required for oligomerization to begin, the correlation between P_{N^*} and τ_{fib} is not unexpected. Using mutations to change the fibril formation rates of human muscle acylphosphatase (AcP) Chiti *et al.*⁸ showed that τ_{fib} of this protein strongly correlates with the propensity to convert from α -helical to β -sheet structure of a monomer. On the other hand, for those polypeptide chains, fibrils of which consist β -sheets, P_{N^*} is proportional to the beta content in the monomeric state. Therefore, our result is consistent with the mutation experiment on AcP.⁸ The dependence of τ_{fib} on P_{N^*} is also supported by the experiment of Tjernberg *et al.*,²⁰ who reported that the inherent amino acid propensity for β -strand conformation³⁰ promotes amyloid aggregation in small peptides. In the recent experiment it has been shown¹⁴ that the aggregation process in $A\beta_{1-40}$ -lactam [D23-K28], in which residues D23 and K28 are chemically constrained by a lactam bridge, is much faster than in the

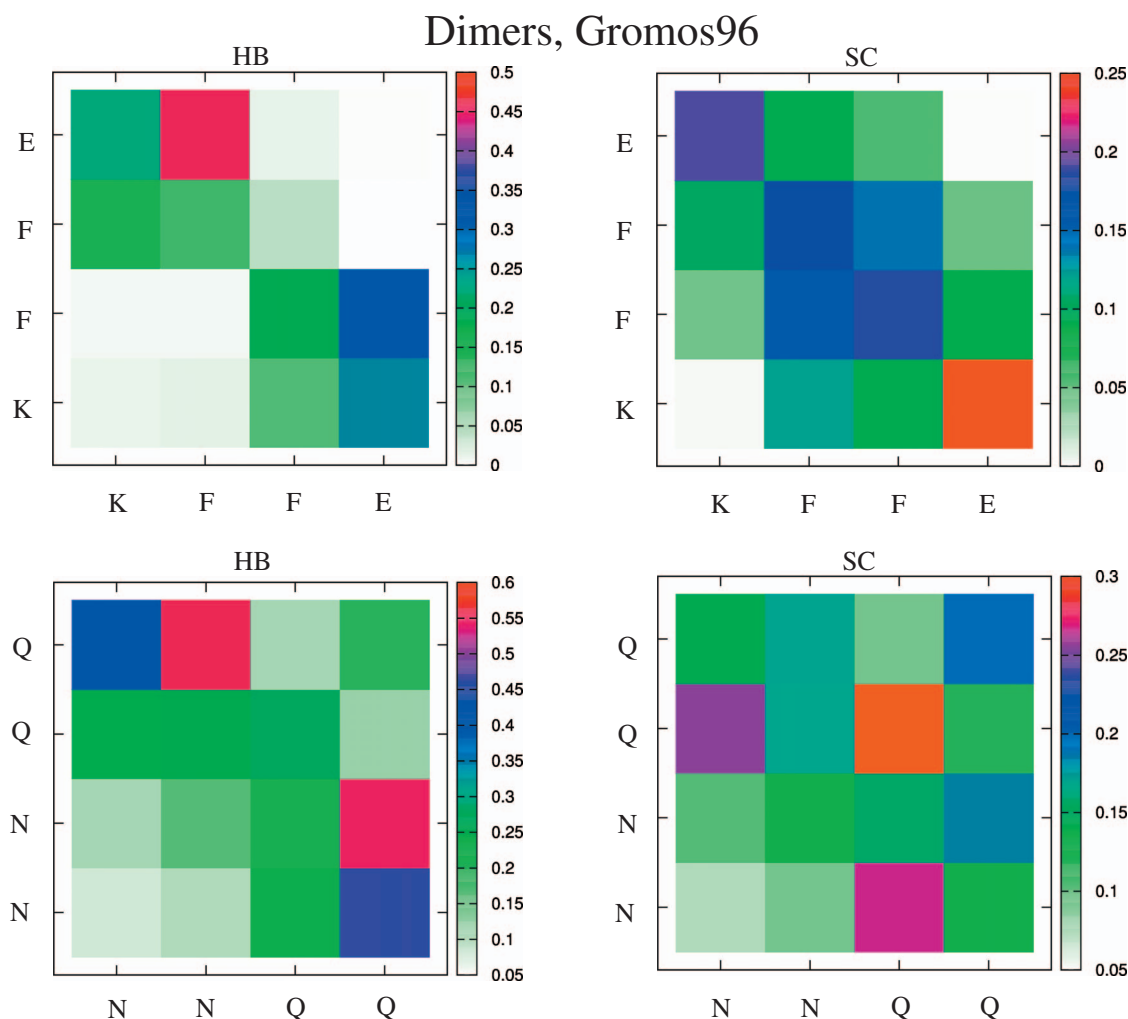


FIG. 6. Shown are HB and SC contact maps for 2KFFE and 2NNQQ. The results were obtained using the Gromos96 force field 43a1 and averaged over four independent trajectories.

wild-type. Since the fixation of the salt bridge may increase the population of the fibril-prone conformation in the monomeric state¹³ our finding is consistent with this experiment. In short, our result implies that one can predict the propensity of polypeptide chains to self-assembly using solely the information about P_N^* obtained in the monomeric state.

Using hydrophobicities of individual amino acids,⁸ we have the hydrophobicity $\text{Hydr}=1.14$ and 6.42 for KFFE and NNQQ, respectively. On the other hand, results followed from simulations using lattice models²⁸ as well as from mutation experiments⁸ suggest that the stronger hydrophobicity, the faster the fibril elongation. From this point of view, the faster aggregation of KFFE compared to NNQQ is also consistent with this trend. The total net charge of both systems is zero and it cannot be used to understand the difference in their oligomerization rates.

A. Nature of ordering of KFFE and NNQQ oligomers

The role of hydrogen bond and side-chain interactions. The question of what interaction drives the self-assembly of biomolecules attracts the attention of many researchers.^{10,11,24,31} The detailed study of the $A\beta_{16-22}$ peptide,^{10,24} e.g., showed that the interpeptide SC interaction

dominates over the HB one. This is associated with the direct and water-mediated charge-charge interaction between oppositely charged termini. For the same reason, the oligomerization of KFFE is expected to be mainly driven by the SC interaction. However, as evident from Fig. 6, the contribution of the HB interaction to the dimerization of this peptide is more important than the SC one. This is probably because the dimer has low stability. As the number of peptides increases the stability of oligomers gets enhanced²⁴ and the role of SC interaction becomes more important. Namely, for 4KFFE, the contributions of the HB and SC interactions become comparable (Fig. 7). The high probability of formation of interpeptide contact K^+-E^- points to the importance to the charge interaction. This is consistent with Bellesia and Shea⁹ who observed that the Coulomb interaction dominates over the aromatic one using the OPLS force field

Similar to 2KFFE, the HB interaction is more relevant in ordering of 2NNQQ than the SC one (Fig. 6). In the 4NNQQ case the hydrogen bonding remains stronger, but the difference in impact of two interactions becomes marginal. For oligomers of larger sizes their contributions are expected to become equivalent. The most important difference between KFFE and NNQQ is that the former has aromatic rings and opposite charges at termini. One can anticipate that the inter-

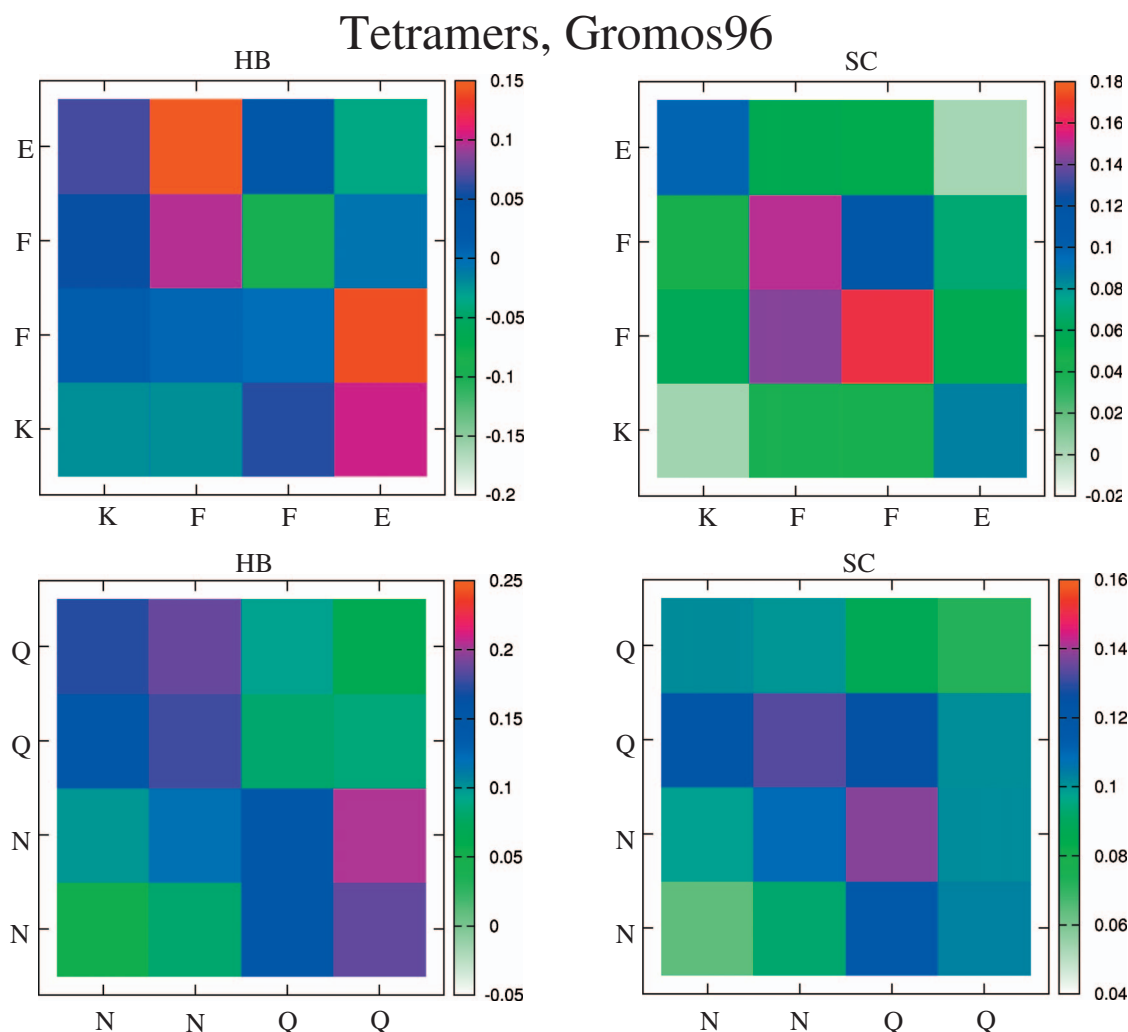


FIG. 7. The same as in Fig. 6, but for tetramers. In this case there are six contact maps formed by six possible pairs of peptides. The result shown here is averaged over six such maps.

play between these two factors washes out differences in their structures leading to a similar nature of ordering of oligomers.

Figure S4 (SM) shows the contact maps obtained by the OPLS force field for dimer 2NNQQ. As in the case of Gromos96 (Fig. 6), the HB interaction is a main driving force in the oligomerization process. Therefore, the nature of ordering is force-field independent.

Low stability of small oligomers. The two-dimensional FEL of 2KFFE is dominated by one wide basin (Fig. 8). This implies that 2KFFE is more stable than the monomer because the FEL of the later has three local minima (Fig. 1, top). However, the stability of 2KFFE remains low as the activation from the shallow minimum requires the energy of ~ 1 kcal/mol. In addition to the fibril-like conformation (β - β shape), within the dominant basin, one can find conformations of U-U and U- β shape which can serve as precursors for the fibril formation. In other words, they are present on pathways to the fibril-like state.

The FEL of 2NNQQ (Fig. 1, bottom) also has one minimum which is sharper than that of 2KFFE. Therefore, as in the monomer case, 2NNQQ is more stable than 2KFFE but the stability of the ground state is low having free energy

barriers of a few $k_B T$. Typical snapshots presented in Fig. 1 show that U-U and U- β conformations occur before the acquisition of the fibril-like state. One can show that 4KFFE and 4NNQQ are more stable than dimers but their stability remains low (results not shown).

B. Energetic argument favoring antiparallel arrangement of peptides NNQQ

Single layer structure. Using snapshots for dimer and tetramer fibril conformations (Figs. 3 and 4 and Fig. S2 in SM), one can show that the typical distance between two neighboring peptides is about 0.47 nm which is close to the experimental value 0.48 nm for peptides within one sheet and clearly smaller than the distance ≈ 0.8 nm between two adjacent sheets.²¹ Thus, results obtained by MD simulations with the Gromos96 43a1 and OPLS force fields support the existence of antiparallel arrangement within one sheet for NNQQ. On the other hand, the experiment of Sawaya *et al.*²¹ showed that peptides belonging to the same sheet are parallel but peptides from adjacent sheets run in opposite directions. From this point of view, our Gromos96 force field 43a1 and OPLS results are in odd with the experiments. The question

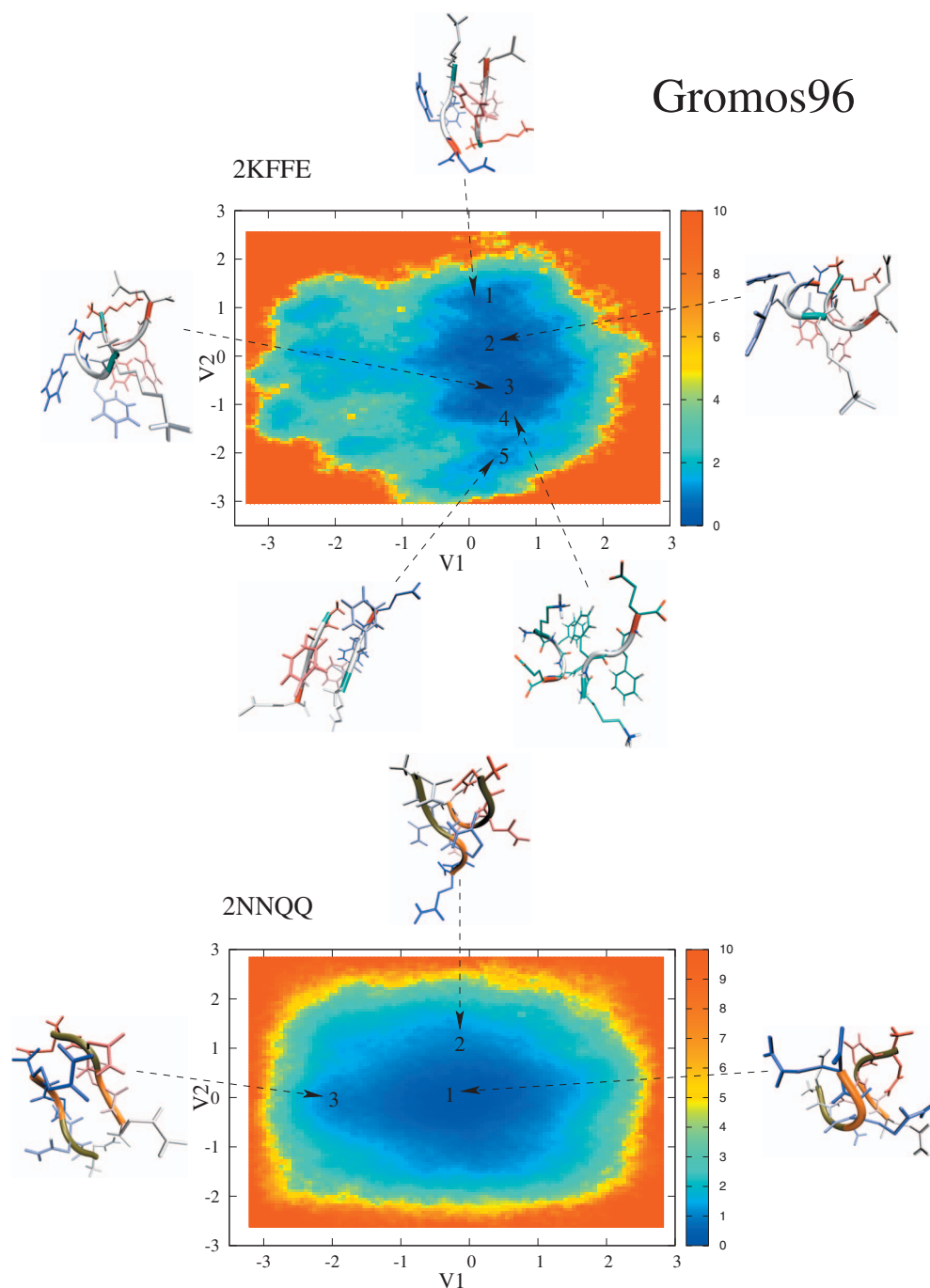


FIG. 8. The FEL obtained by the Gromos96 force field for dimers 2KFFE and 2NNQQ. Typical snapshots have U-U, U- β , and β - β shapes.

arises is whether the in-sheet antiparallel structure is robust against other force fields. To check this we use a simple energetics argument without long MD runs. Our idea is to compute the interaction energy of two antiparallel peptides using different force fields, E_{antipar} and compare it with that for the parallel arrangement. We use the antiparallel configuration obtained from Gromos96 43a1 simulations (Fig. 3) as a starting configuration for finding equilibrium conformations in other force fields. One can show that these conformations may be obtained after short MD runs (≈ 100 ps). The nonlocal interaction energy between two antiparallel peptides NNQQ was computed using the standard GROMACS procedure and different force fields available in this soft-

ware. The results are presented on Table II. Amber 94 and 99 force fields give a comparable value for E_{antipar} . The same is true for two Gromos force fields but with lower energies, while the Charmm27 (Ref. 32) provides the lowest energy for antiparallel configurations. The OPLS is intermediate.

To estimate the interaction energy between two parallel peptides NNQQ, E_{par} , we adopted the following procedure. The parallel configuration was obtained from the antiparallel configuration (Fig. 3) by keeping one peptide fixed, while the second one is rotated and slightly translated along the vector connecting its terminal C_{α} carbons. As in the antiparallel case, using this parallel conformation as a starting structure and different force fields to make short MD runs to find

TABLE II. The interaction energies obtained by different force fields for the dimer and octamer of NNQQ. For the dimer we have the parallel and antiparallel arrangements. P1, P2, A1, and A2 refer to four possible configurations shown in Fig. 9. The numbers in the parentheses correspond to the interlayer interaction energies.

Force field	Interaction energy (kJ/mol)					
	Dimer		Octamer			
	Antiparallel	Parallel	P1	P2	A1	A2
Gromos 43a1	-197.4	-47.5	-723.1(-359.6)	-678.1(-264.4)	-1195.4(-255.8)	-927.7(-149.2)
Gromos 53a6	-192.3	-21.3	-673.0(-251.4)	-644.9(-235.7)	-1212.7(-379.9)	-847.9(-137.6)
OPLS	-241.7	-48.2	-691.3(-286.2)	-691.0(-218.3)	-1332.4(-221.6)	-1358.3(-209.0)
Amber 94	-151.4	-55.4	-882.3(-389.3)	-703.9(-250.6)	-1021.4(-311.5)	-852.0(-292.6)
Amber 99	-152.5	-62.6	-844.1(-312.7)	-609.4(-212.5)	-1160.6(-349.3)	-912.1(-319.9)
Charmm27	-318.0	-63.2	-672.0(-221.3)	-580.7(-181.3)	-1496.0(-298.3)	-1324.5(-148.2)

equilibrium conformations. The interaction energy is calculated and averaged over these conformations. For all of six force fields, E_{par} is higher than E_{antipar} (Table II). Thus, within one sheet the antiparallel configuration of NNQQ is energetically more favorable than the parallel one.

Double layer structure. To see if the interlayer interaction could convert the antiparallel structure within one sheet into the parallel one, we consider four possible double-layer structures for 8NNQQ (Fig. 9). In configuration P1 peptides from the same layer are parallel, while two neighboring layers have opposite orientations. Such a configuration was observed in the experiments of Sawaya *et al.*²¹ In the case of P2 all peptides are parallel. Peptides from the same layer of configuration A1 are antiparallel and two sheets are also antiparallel. Configuration A2 has the same structure as A1 except that two layers have the same orientation (Fig. 9). All configurations were constructed in such a way that the distance between layers is almost the same as in the experiments.²¹

As in the dimer case, the interaction energies of four configurations of the octamer have been estimated using snapshots obtained during short equilibration runs. The results are summarized in Table II. The interlayer interaction is lower than the intralayer one and this is true for all six force

fields. We can rank the total energies in ascending order as $A1 \rightarrow A2 \rightarrow P1 \rightarrow P2$. Thus A1 is the most favorable state but not protofibril P1 which was observed experimentally. One of possible reasons for this discrepancy is that existing force fields are not accurate enough to capture P1 as the ground state. The energy difference between A1 and P1, $\delta E = E(P1) - E(A1)$ obtained by Charmm27 is largest ($\delta E \approx 824$ kJ/mol), while Amber 94 provides the smallest estimate $\delta E \approx 139$ kJ/mol (Table II). This suggests that the improvement of parameters of Amber 94 force field may cure our problem, but this question is left for future study.

IV. CONCLUDING REMARKS

We used all-atom models to elucidate the role of the population of fibril-prone state N^* in the monomeric state in assembly of peptides.

- (1) The measure of population of fibril-prone state N^* in the monomeric state P_{N^*} has been defined using the end-to-end distance. This definition is valid if polypeptide chains adopt shape of β -strand in the fibril state. If they have different shapes then P_{N^*} can be defined using RMSD from the fibril-prone conformation.
- (2) P_{N^*} is found to depend not only on sequences but also on the force fields. We have shown that Gromos96 and OPLS are compatible for studying the oligomerization process where the fibril state contains β -sheets. This result was obtained for short peptides KFFE and NNQQ but it is expected to hold for other systems because these force fields favor beta formation. Amber99 which disfavors beta structures is not recommended to use to study kinetics of formation of fibrils that consist of β -strands, but it may be useful for studying other systems.
- (3) For the first time we demonstrated that P_{N^*} plays a key role in the fibril elongation process using all-atom models. We predict that those molecules that have P_{N^*} less than a few percents have low propensity to oligomerization. From this point of view, our result is useful for elucidating the fibrillogenesis at the single-monomer level. This becomes even more critical taking into account the fact that the fibril formation is an extremely slow process which is difficult for numerical study. Al-

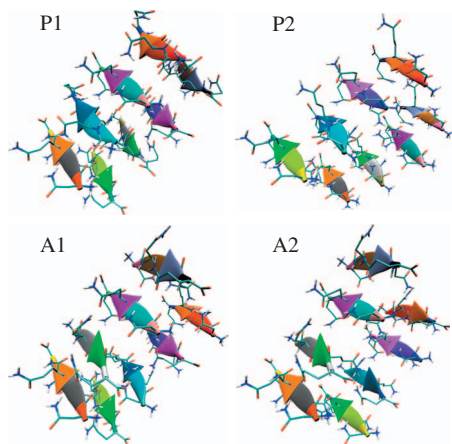


FIG. 9. Four possible two-layer configurations A1, A2, P1, and P2 for the octamer 8NNQQ. Configuration P1 is a fibril-like state observed in the experiments (Ref. 21). A1 with peptides antiparallel within one sheet is the most stable according to our theoretical estimates.

though our main conclusion was obtained for short peptides it should hold for longer peptides and biomolecules due to its universal nature.

- (4) Using MD simulations and the energetics argument with different force fields, we have shown that 2NNQQ, 4NNQQ, and 8NNQQ form the antiparallel fibril within one layer. On the other hand, the x-ray experiments²¹ showed that peptides from one sheet are parallel, while two neighboring layers have opposite orientations (configuration P1 in Fig. 9). As mentioned above, one of possible reasons for this discrepancy is that present force fields are not accurate enough. This problem calls for further investigation. However, the fact that peptides NNQQ in the fibril state are antiparallel in our simulations should not affect our main conclusion that the population of the fibril-prone state in the monomeric state is one of the most important factors governing fibrillogenesis of polypeptide chains. This is because we are interested in the dependence of aggregation rates on P_{N^*} but not in the nature of ordering of fibril states itself.
- (5) It is well known that A β 42 is much more prone to aggregation and much more toxic to neurons than A β 40.^{33,34} Since these peptides are long an estimation of their fibril formation times by all-atom simulations has not been carried out yet. Using the replica exchange molecular dynamics³⁵ and all-atom models it was shown that in the monomeric state the beta content of A β 42 is higher than A β 40.^{36,37} This interesting finding is in line with our main result that the higher P_{N^*} (or higher the beta content in the case of amyloid peptides), the faster is the aggregation process. This example again demonstrates that our theory is useful for predicting the fibrillogenesis of complex systems.

ACKNOWLEDGMENTS

The kind help of Man Hoang Viet in estimation of energies of parallel and antiparallel 2NNQQ conformations is highly appreciated. The work was supported by the Ministry of Science and Informatics in Poland (Grant No. 202-204-234) and Department of Science and Technology at Ho Chi Minh City, Vietnam.

¹F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.* **75**, 333 (2006).

²M. Sunde and C. Blake, *Adv. Protein Chem.* **50**, 123 (1997).

³D. E. Otzen, O. Kristensen, and M. Oliveberg, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9907 (2000).

⁴F. Chiti, M. Calamai, N. Taddei, M. Stefani, G. Ramponi, and C. M. Dobson, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16419 (2002).

⁵M. W. West, W. X. Wang, J. Patterson, J. D. Mancias, J. R. Beasley, and M. H. Hecht, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11211 (1999).

⁶Y. Kallberg, M. Gustafsson, B. Persson, J. Thyberg, and J. Johansson, *J.*

Biol. Chem. **276**, 12945 (2001).

⁷E. Gazit, *FASEB J.* **16**, 77 (2002).

⁸F. Chiti, M. Stefani, N. Taddei, G. Ramponi, and C. M. Dobson, *Nature (London)* **424**, 805 (2003).

⁹G. Bellesia and J. E. Shea, *Biophys. J.* **96**, 875 (2009).

¹⁰K. D. Klimov and D. Thirumalai, *Structure* **11**, 295 (2003).

¹¹M. López de la Paz, G. M. S. de Mori, L. Serrano, and G. Colombo, *J. Mol. Biol.* **349**, 583 (2005).

¹²D. W. Li, S. Mohanty, A. Irback, and S. H. Huo, *PLOS Comput. Biol.* **4**, e1000238 (2008).

¹³G. Reddy, J. E. Straub, and D. Thirumalai, *J. Phys. Chem. B* **113**, 1162 (2009).

¹⁴K. L. Sciarretta, D. J. Gordon, A. T. Petkova, R. Tycko, and S. C. Meredith, *Biochemistry* **44**, 6003 (2005).

¹⁵M. S. Li, D. K. Klimov, J. E. Straub, and D. Thirumalai, *J. Chem. Phys.* **129**, 175101 (2008).

¹⁶M. S. Li, *Physics of Particle and Nuclei Letters* **5**, 328 (2008).

¹⁷W. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. Scott, and I. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag AG an der ETH, Zurich, 1996).

¹⁸W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **118**, 11225 (1996).

¹⁹J. Wang, P. Cieplak, and P. A. Kollman, *J. Comput. Chem.* **21**, 1049 (2000).

²⁰L. Tjernberg, W. Hosia, N. Bark, J. Thyberg, and J. Johansson, *J. Biol. Chem.* **277**, 43243 (2002).

²¹M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Madsen, C. Riekel, and D. Eisenberg, *Nature (London)* **447**, 453 (2007).

²²G. Wei, N. Mousseau, and P. Derreumaux, *Biophys. J.* **87**, 3648 (2004).

²³B. Strodel and D. J. Wales, *J. Chem. Theory Comput.* **4**, 657 (2008).

²⁴P. H. Nguyen, M. S. Li, G. Stock, J. E. Straub, and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 111 (2007).

²⁵H. J. C. Berendsen, J. Postma, W. van Gunsteren, and J. Hermans, *Intermolecular Forces* (Reidel, Dordrecht, 1996).

²⁶A. T. Petkova, Y. Ishii, J. Balbach, O. Antzutkin, R. Leapman, F. Delaglio, and R. Tycko, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16742 (2002).

²⁷Y. Mu, P. Nguyen, and G. Stock, *Proteins* **58**, 45 (2005).

²⁸M. S. Li, N. T. Co, C. K. Hu, G. Reddy, and D. Thirumalai, "Determination of factors governing fibrillogenesis of polypeptide chains using lattice models," *Phys. Rev. Lett.* (submitted).

²⁹T. Yoda, Y. Sugita, and Y. Okamoto, *Chem. Phys. Lett.* **386**, 460 (2004).

³⁰A. G. Street and S. L. Mayo, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9074 (1999).

³¹G. Favrin, A. Irback, and S. Mohanty, *Biophys. J.* **87**, 3657 (2004).

³²B. R. Brooks, C. L. Brooks III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, *J. Comput. Chem.* **30**, 1545 (2009).

³³C. J. Barrow, A. Yasuda, P. T. Kenny, and M. G. Zagorski, *J. Mol. Biol.* **225**, 1075 (1992).

³⁴J. T. Jarrett, E. P. Berger, and P. T. Lansbury, Jr., *Ann. N.Y. Acad. Sci.* **695**, 144 (1993).

³⁵Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).

³⁶N. G. Sgourakis, Y. Yan, S. A. MacCallum, C. Wang, and A. E. Garcia, *J. Mol. Biol.* **368**, 1448 (2007).

³⁷M. Yang and D. B. Teplow, *J. Mol. Biol.* **384**, 450 (2008).

³⁸See supplementary material <http://dx.doi.org/10.1063/1.3415372> for the supplementary figures.