Nanoscale insight into silk-like protein self-assembly : effect of design and number of repeat units

Physical Biology

OFFICIAL JOURNAL OF
Sociedad de Biofísicos Latino Americanos

ACCEPTED MANUSCRIPT

Nanoscale insight into silk-like protein self-assembly: effect of design and number of repeat units

To cite this article before publication: Jamoliddin Razzokov *et al* 2018 *Phys. Biol.* in press https://doi.org/10.1088/1478-3975/aadb5e

Nanoscale insight into silk-like protein self-assembly: effect of design and number of repeat units.

Jamoliddin Razzokov,[1,2] Saber Naderi,[2,3] and Paul van der Schoot[2,4]

[1)] *Department of Chemistry, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium*

[2)] *Faculteit Technische Natuurkunde, Technische Universiteit Eindhoven, Postbus 513, 5600 MB Eindhoven, The Netherlands*[a)]

[3)] *Dutch Polymer Institute, P.O. Box 902, 5600 AX Eindhoven, The Netherlands*

[4)] *Instituut voor Theoretische Fysica, Universiteit Utrecht, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands*

By means of replica exchange molecular dynamics simulations we investigate how the length of a silk-like, alternating diblock oligopeptide influences its secondary and quaternary structure. We carry out simulations for two protein sizes consisting of three and five blocks, and study the stability of a single protein, a dimer, a trimer and a tetramer. Initial configurations of our simulations are β-roll and β-sheet structures. We find that for the triblock the secondary and quaternary structures upto and including the tetramer are unstable: the proteins melt into random coil structures and the aggregates disassemble either completely or partially. We attribute this to the competition between conformational entropy of the proteins and the formation of hydrogen bonds and hydrophobic interactions between proteins. This is confirmed by our simulations on the pentablock proteins, where we find that, as the number of monomers in the aggregate increases, individual monomers form more hydrogen bonds whereas their solvent accessible surface area decreases. For the pentablock β-sheet protein, the monomer and the dimer melt as well, although for the β-roll protein only the monomer melts. For both trimers and tetramers remain stable. Apparently, for these the entropy loss of forming β-rolls and β-sheets is compensated for, the free-energy gain due to the hydrogen-bonding and hydrophobic interactions. We also find that the middle monomers in the trimers and tetramers are conformationally much more stable than the ones on the top and the bottom. Interestingly, the latter are more stable on the tetramer than on the trimer, suggesting that as the number of monomers increases protein-protein interactions cooperatively stabilize the assembly. According to our simulations, the β-roll and β-sheet aggregates must be approximately equally stable.

Keywords: silk-like protein, implicit solvent, replica exchange molecular dynamics, self-assembly, fiber.

[a)]Electronic mail: jamoliddin.razzokov@uantwerpen.be

I INTRODUCTION

A promising route to the rational design of functional nanomaterials is a bottom-up approach that makes use of the self-assembly of molecular building blocks and exploits the physical principles that govern inter- and intra-molecular interactions.[1–3] This is particularly true for biomimetic materials based on proteins and peptides as basic molecular units[4–7] albeit that this does require an understanding of the relation between their molecular structure and self-organization properties. Recent advances in experimental, theoretical and simulation techniques have drawn attention to the role played by factors such as the protein concentration, hydrophobic, hydrophilic and electrostatic interactions, the salinity, temperature, pressure, pH and so on.[8–11] By controlling the primary sequence one can in principle control the way that the designer protein molecule responds to these factors and control its structure and function.[12–16]

An application of interest is gene therapy, relevant in the context of treatments of cancer, HIV (human immunodeficiency virus) and SCID (severe combined immunodeficiency), where a therapeutic gene is delivered into target cells.[17,18] In order to successfully deliver the therapeutic gene into the target cells, a protective shell is required that shields the genetic material from degradation and attack by nucleases.[19,20] One way to do this is to use a virus as a delivery vehicle but this apparently has disadvantages, toxic immunological responses.[21,22] An alternative approach is to design bio-compatible proteins that are able to self-assemble on the genetic material and form an artificial virus-like particle.[23,25]

Very recently a possible candidate for such an artificial DNA coat protein was proposed by de Vries and coworkers.[25] The protein is a triblock copolymer made up of a cationic DNA-binding motif, similar to the nucleic-acid binding domains in viruses, a collagen-like hydrophilic random coil motif that provides colloidal stability in aqueous solution and a silk-like protein block.

The design is based on the hypothesis that the latter self-assembles into β-sheets (at least in the aggregated state), that is, stack on top of each other via hydrogen bonding and hydrophobic interactions, and stabilize the protein aggregate around the DNA.

The work of de Vries et al., that the triblock copolymer is indeed capable of self-organizing into a capsid around the DNA, protect it against the action of nucleases and successfully transfect cells.[25] One of the important design parameters is the length of the silk-like block

that consists of, copies of a repeat unit. Below a critical length the sequence does not seem to form a β-sheet, self-assemble and provide sufficiently dense coverage on the DNA to protect it against attack by the nucleases. However, if the number of repeat units is too large the binding of the protein to the DNA may well become too strong, interfering with the expression of the target DNA in the cell.[25]

To shed light on the influence of the number of repeat units in a silk-like protein similar to that of the core of the protein of de Vries and collaborators, we carry out all-atom replica exchange molecular dynamics simulations on proteins consisting of three and five repeat units that we specify below. We restrict ourselves in this work to aggregates consisting of one to four proteins in free solution and investigate the stability of the folded structure as well as that of the aggregates themselves. In follow-up work we extend this to simulation studies in the presence of DNA, which are computationally much more demanding.

Before presenting our simulations, we note that the influence of the length of β-sheet forming silk-like proteins on their assembly into fibers has also been studied experimentally.[12,24,26]

For instance, Davies and coworkers studied the self-assembly of peptides containing 7 and 9 amino acids with sequences CH_3CO-RLQLQLE-NH_2 and CH_3CO-QRLQLQLEQ-NH_2.[12] They observe that above a critical concentration the 7mers and 9mers, which have a random coil structure in the monomeric state, self-assemble into inter-molecular β-sheet tapes. The critical concentration was lower for the case of the 9mers, indicating that the binding energy for the 9mer must be larger than that of the 7mer. This is plausible because more hydrogen bonds form between the longer peptides compared to the shorter ones.

From these experiments it is not possible to say whether dimers and trimers are also part of the full fiber length distribution that might be dominated by long assemblies. The reason is that the authors focus on circular dichroism spectroscopy that allows one to probe the assembled fraction of proteins and not the size distribution. Whilst informative, measurements like this cannot provide us with an insight at fully atomistic detail of the molecular processes at the root of the stability of the very long assemblies. Computer simulations, however, can provide such atomistic detail. For example, Schor and collaborators studied by means of replica exchange molecular dynamics (REMD) simulations in explicit solvent the most stable structure of a fiber forming silk-like protein, containing six repeat units of an amino acid sequence [$(GA)_3$-GE] in aqueous solution. The authors performed simulations on a single protein starting from β-sheet or β-roll structures as well as a dimer containing two

Number of $(GA)_3G$ blocks	Primary sequence
$n = 3$	$K(GA)_3GQ$-$(GA)_3GK(GA)_3GQ$
$n = 5$	$[K(GA)_3GQ$-$(GA)_3G]_2K(GA)_3GQ$

TABLE I. The primary sequence of our triblock ($n = 3$) and pentablock ($n = 5$) proteins.

β-sheets or β-rolls that are stacked on top of each other to find which of the two structures is more stable. They find that in both monomeric and dimeric states the β-roll structure is the most stable structure.[37]

Recently, we repeated the simulations of Schor and coworkers for a monomer and a dimer and in addition to this we carried out simulations with a trimer and a tetramer of their silk-like protein in both explicit and implicit solvent.[27] For the same amount of simulation time we obtain the same result as Schor et al. but by extending our simulations to longer time scales we find that in monomeric and dimeric states both the β-sheet and β-roll structures melt into random coil structures. This shows that more frequent sampling of the fewer replicas required for simulations with implicit solvent and probing longer time scales that can then be achieved in REMD simulations are crucial for sufficient sampling of the phase space of our protein. Here we present results of replica exchange computer simulations focusing on the self-assembly of a slightly different silk-like protein consisting of two types of repeat unit $K[(GA)_3$-G] and $Q[(GA)_3$-G]. See Table I for the primary sequence of our protein. In our simulations we make use of an implicit solvent model and consider two values of the number of repeat units, $n = 3$ and 5. We carry out simulations starting from one, two, three and four β-sheet or β-roll structures stacked on top of each other.

We find that for the case of the proteins with $n = 5$ repeat units, the monomer and dimer β-sheets melt into a disordered globular state whereas the dimer β-roll and trimers and tetramers of β-sheets and β-rolls are stable. In contrast, the proteins with $n = 3$ repeat units melt into random coil structures and disassemble within the simulation time. This indicates that the protein-protein binding free energy increases with increasing the number of repeat units. This is arguably caused by stronger inter-molecular interactions due to (i) a larger number of hydrogen bonds in the aggregates of the larger proteins and (ii) the fact that the solvent accessible surface area per monomer decreases more rapidly with aggregate size for the proteins with $n = 5$ than those with $n = 3$.

The remainder of this paper is organized as follows. In Section II, we present the methods that we use in our simulations and the way we analyze our data. We discuss obtained results in Section III. Section III A highlights our simulation results for the proteins with five repeat units. These results are compared with our findings on the triblock proteins in Section III B. Finally, Section IV is devoted to conclusions.

II SIMULATION METHOD AND ANALYSIS

We perform all-atom molecular dynamics simulations of a single protein, a dimer, a trimer and a tetramer of a silk-like protein with an alternating amino acid sequence K[$(GA)_3$-GX]$_n$, where X stands for glutamine and lysine, see Table I. We make use of the Amber simulation package[28] and employ the ff99SB force field[29], in combination with the generalized Born (GB) implicit solvent model[30] which represents solvent as a countinuous medium without artifacts from periodicity. In these simulations solvation energies become part of the total energy of the system and the forces driving dynamics contain the derivatives of this term. The GB is widely applied implicit-solvent model in atomic scale simulations, approximates long range electrostatic interactions based on an analytical formula given in literatures[31–33]. The average energetic contribution of solute-solvent hydrogen bond is also included in implicit solvent model.[30,34] This also serves to enhance sampling and rapid convergence due to the absence of solvent frictional forces which is not required re-organization of explicit water molecules in response to protein conformational changes. The effective salt concentration in our simulations equals 0.1 M, which is implemented in the simulations by including a Debye-Hückel-type term[35] in the calculation of the electrostatic interaction energy of the implicit solvent model.

By choosing an implicit solvent model instead of an explicit one we reduce the computational cost of our simulations. This allows us to reach longer timescales compared to simulations with explicit solvent because (i) we need less computing time for each replica and (ii) we need fewer replicas. In a recent study on a similar protein with an amino acid sequence of $[(GA)_3$-GE$]_5$ we have shown that (i) for this protein and for the same amount of simulation time the results of implicit solvent simulations agree well with those of explicit solvent ones and (ii) by extending the implicit solvent simulations to longer timescales we have found that to study the self-assembly of our protein a sufficiently long simulation is required.[27]

We run simulations with $n = 3$ and 5 repeat units starting from β-sheet and β-roll

structures that are created using the Xleap program in AmberTools. The difference between β-sheet and β-roll structures is that in the β-sheet structure each repeat unit forms hydrogen bonds with its nearest neighbor in the protein sequence, whereas in β-rolls the hydrogen bonds are mostly formed between each repeat unit and the next-nearest neighbor. In order to create these structures we make linear protein chains using the Xleap program and pull the nearest (for the β-sheet) or the next-nearest (for the β-rolls) repeat units together during a short simulation run using the Amber simulation package. Cartoon view of monomers of a β-sheet and a β-roll structures is given in Fig. 1. Schematic images of proteins presented

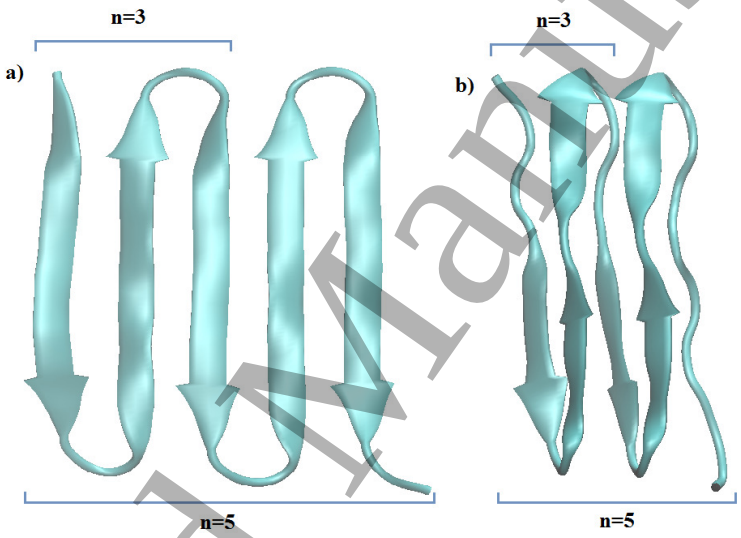

FIG. 1. Schematic representation of (a) a β-sheet and (b) a β-roll structure of our silk-like protein. Here each monomer contains $n = 5$ repeat units. $n = 3$ repeat units of protein is created taking 3 strands of each monomer.

in this paper are created using VMD software.[36]

All simulations are performed using Langevin dynamics in implicit solvent, mimicking conditions where the lysine residues and the proteins as a whole are either charge neutral or the ionic strength is sufficiently large to effectively screen Coulomb interactions between charged groups. Partial positive and negative charges are invoked to model electro-negative and -positive atoms. Multimers of β-sheets and β-rolls are created by copying and translating monomers in the direction perpendicular to the β-sheet or β-roll plane over an arbitrary distance of 0.6 nm. From our simulations we find that for this choice individual proteins are able to find each other and bind within a reasonable simulation time. Note that our

procedure conserves the broken up-down symmetry of the proteins and maximizes the contact area between them. This is necessary because the proteins in cross section are not symmetric. All initial structures are energy minimized prior to our simulations.

For each of the protein aggregates we run replica exchange molecular dynamics (REMD) simulations in the temperature range of 298 to 554 K. This temperature range we distribute over 10 replicas for the monomers and dimers of the aggregates consisting of the protein with $n = 3$ repeat units and for the monomers of proteins with five repeat units ($n = 5$). For the trimers and tetramers of the shorter proteins with $n = 3$ and for the dimers, trimers and tetramers of proteins with $n = 5$ the temperature range is somewhat smaller, 298 to 511 K, and spread over 12 replicas.

The temperature distribution is chosen in such a way that the replica exchange probabilities are between 10 to 30 percent, as is customary.[37] A replica exchange attempt is done every 500 MD steps, with each step representing 2 fs of real time. The exchange probability is determined by a Metropolis algorithm, where the weight factors are given by the product of the Boltzmann factors for each replica. The overall REMD simulation time is different depending on the size of the aggregates, but varies between 20 ns and 40 ns per replica. In our simulation the average time that it takes for a replica at the lowest temperature to diffuse up in the temperature space, reach the highest temperature and diffuse back to the lowest temperature is about 1 to 3 ns. This shows that the overall REMD simulation time is sufficient for our replicas to explore the temperature space. Besides, substantial speedup conformational changes is observed by using implicit solvent models in literatures.[31,38]

To analyze our simulation data we calculate the number of hydrogen bonds within each protein aggregate, the root mean-square deviation (RMSD) of each aggregate from its initial structure and the solvent accessible surface area (SA). The number of hydrogen bonds in each simulation frame is determined by measuring the distance between H-bond donors and acceptors, using a cutoff of 0.3 nm, and the angle between the acceptor, hydrogen and donor atoms for which the cutoff is 135°. For calculating the RMSD, we take the initial configuration of each simulation as a reference structure, which consists of a single or multiple β-sheets or β-rolls depending on the system we study, and compute the RMSD between all atoms of each frame to the atoms in the reference configuration. To probe the solvent accessible surface area (SA) of our protein systems, a "rolling-ball" algorithm is used in which a spherical probe of radius of 0.14 nm moves on the surface of the protein.[39]

In REMD simulations, trajectories are in a way not continuous in time, because they are exchanged between replicas at given intervals with a certain probability. Therefore, instead of calculating the number of hydrogen bonds, the RMSD and the SA as a function of time, we obtain the free energy of our protein systems, ΔG, as a function of each of the quantities mentioned above. To do this, first, from the data of the lowest temperature replica, we count the number of occurrences in which our system is in state X, which corresponds to a value of, e.g., the RMSD. Next, the probability, $\Pi(X)$, of finding our system at a state X is computed from the number of occurrences. Finally, the free-energy is calculated using the Boltzmann relation $\Pi(X) \propto \exp -\Delta G(X)/k_B T$, where k_B is the Boltzmann constant and T is the temperature. The global minimum of the free energy we arbitrarily set to zero as we can only probe free energy differences within a single simulation run.

In the following, we shall first discuss our results for the protein consisting of five repeat units, and after that contrast that to what we find for the triblock protein.

III RESULTS AND DISCUSSION

III A. Cooperative stabilization of folded pentablock proteins

Focusing first on a discussion of our simulation results on the assemblies of our five-repeat unit protein, we remind the reader that for each of the assemblies the initial configurations start off from all β-sheet and all β-roll structures. Shown in Fig. 2 are representative snapshots of a monomer, a dimer, a trimer and a tetramer taken from the simulations with initial β-sheet structures, and similar ones for the initial β-roll structures in Fig. 9. Note that these snapshots are not necessarily the final structures of the simulations, they represent the most stable ones.

In our simulations, the monomers and the dimers of the β-sheets melt into a random coil whereas their trimeric and tetrameric states are much more stable. For the β-rolls only the monomers melt within our simulation time, and dimers, trimers and tetramers are stable albeit one of the outer monomers of the tetramer melts. To quantify these observations we measure the root mean-square deviation (RMSD) of our proteins using snapshots that we take every 500 MD steps starting from their initial structures. Shown in Fig. 3 are results for the β-sheets, we have very similar ones for the β-rolls (not shown).

Fig. 3, shows that the minimum of free-energy for the monomer and one of the monomers in the dimer occur almost at the same value of the RMSD. This is because in our simulations the monomers in the dimer melt into random-coil-like structures, separate and turn into two

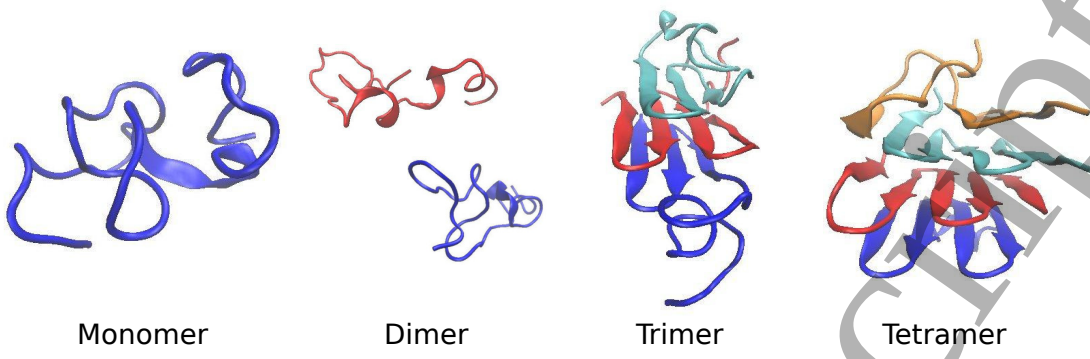

FIG. 2. Representative snapshots of simulations at the temperature $T = 298K$ starting from a monomer, a dimer, a trimer and a tetramer of β-sheets with five repeat units. Here each monomer contains $n = 5$ repeat units.

individual random-coil monomers. The minimum of, free energy moves to lower values of the RMSD for the monomers that reside in the center of the trimer and the tetramer; the RMSD curves corresponding to these monomers are in Fig. 3 referred to as $trimer(middle)$ and $tetramer(middle(1))$.

There are two minima in the free-energy landscape of the $trimer(middle)$ that are indicated in the figure as $5S1$ and $5S2$. One of these minima, is associated with a β-sheet structure and the other one corresponding to the deepest minimum, $5S2$ is a β-sheet with one of the strands molten. Renderings of these structures are given in Fig. 4. For the $tetramer(middle(1))$, there are also two minima approximately on the same values of the RMSD but the deepest one corresponds to the $5S1$ structure. The width of the potential well in this case is smaller compared to that of the $trimer(middle)$ protein indicating that the $tetramer(middle(1))$ protein has less conformational freedom.

Within the trimer and the tetramer structures, the conformations of the monomers in the center fluctuate much less than those at both ends of the aggregates. The fluctuations can be quantified by measuring the RMSD for the individual monomers in the trimer and tetramer. Shown in Fig. 5 is the RMSD for the individual monomers in the tetramer. The monomers in the middle of the tetramer are very stable and their free-energies exhibit a deep minimum at 0.22 nm. The standard deviation of RMSD for these monomers is about 0.04 nm. In contrast, the free-energy landscape of the ones on the top and bottom of the aggregate, which are in contact with water, is much broader. In this case the standard deviation of RMSD is about 0.2 nm. Their free-energy minimum corresponds to the *5S2*

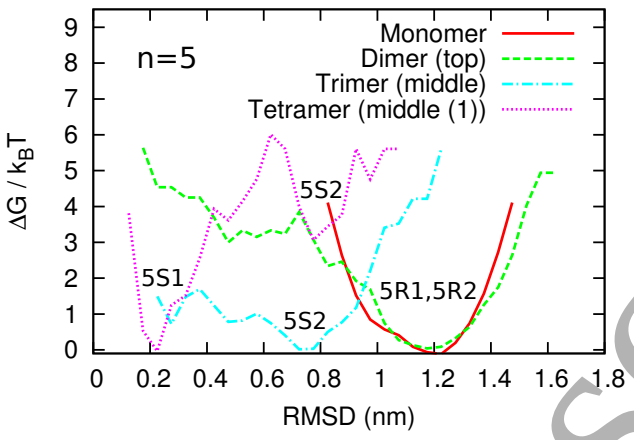

FIG. 3. Free energy, ΔG, obtained from simulations starting from a monomer, a dimer, a trimer and a tetramer of β-sheets consisting of five repeat units, as a function of the root-mean square deviation of snapshots of the monomer, one of the monomers in the dimer and monomers in the middle of the trimer and tetramer from their initial (β-sheet) structures. The structures corresponding to the minima of the free energy are denoted as 5S1, 5S2, 5R1 and 5R2 and shown in Fig. 4.

structure that can turn into the *5S3* structure by overcoming a relatively small free-energy barrier as is clear from Fig. 5.

Our results on structures starting from β-rolls mirror these findings. The only difference between the initial β-roll and β-sheet structures is in the conformational asymmetry of the top and bottom monomers. For the β-sheet structure our results point at a much more symmetric configuration, as evident from Fig. 5. Since both β-rolls and β-sheets are inherently chemically not mirror symmetric, we expect both types of assembly to be symmetry-broken. Obviously we cannot exclude the possibility that in a longer simulation run both structures would exhibit this structural asymmetry.

The monomers in the center of the trimers and tetramers are more stable than the ones on the top and the bottom of the aggregate. However, the free-energy landscapes of the monomers in the center of the trimers are broader than those in the center of the tetramers. This suggests that the trimer is not as stable a structure as a tetramer. Interestingly, the monomers at the top and the bottom of the tetramer are also structurally more stable than the outer monomers of the trimer (data not shown). This indicates that the more stable β-sheets in the center of the aggregates restrict the fluctuations of the outer monomers via inter-molecular interactions and hence stabilize the structure as a whole. Considering that

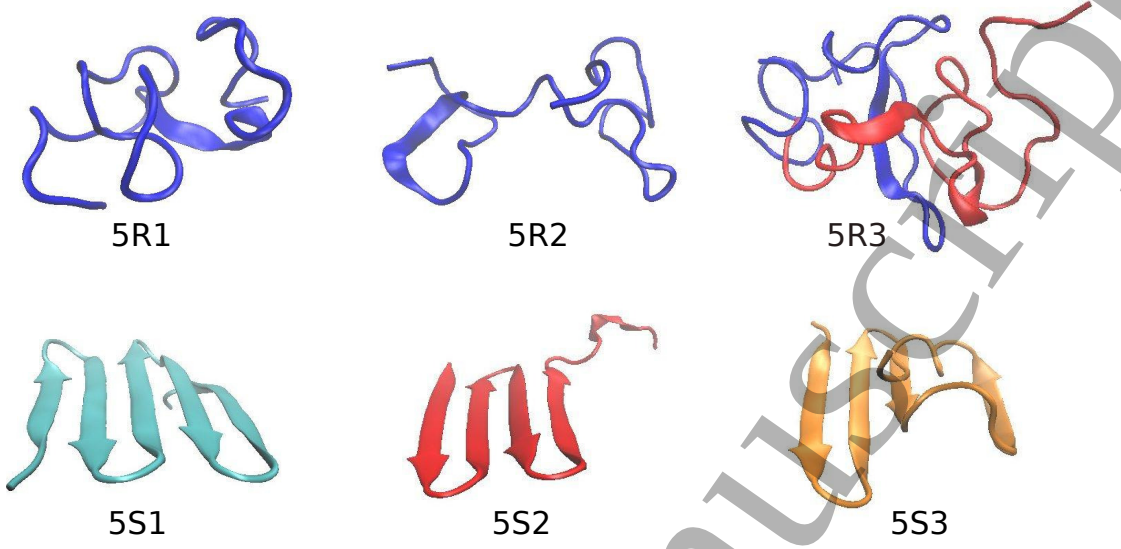

FIG. 4. Snapshots of minimum free energy configurations representing a (5R1) random-coil structure, (5R2) a random coil with more surface exposed to water, (5R3) two random coils bound to each other, (5S1) a β-sheet, (5S2) a β-sheet with a molten strand, (5S3) a β-sheet with two molten strands. See also Fig. 5. These snapshots are taken from the lowest temperature replicas corresponding to a temperature $T = 298\ K$. 5R1, 5R2 are snapshots of simulations started from a monomer β-sheet and 5R3 is taken from a simulation on a dimer starting from β-sheet structures. 5S1, 5S2 and 5S3 structures are taken from simulation on tetramers consisting of β-sheets.

β-sheet dimers are not stable at all, this suggests a gradual increase in stability of the assemblies with increasing degree of polymerization. This is in-line with density functional theory calculations of Filot et al., where the authors find that the absolute value of average interaction energy between self-assembling C_3-symmetrical trialkylbenzene-1,3,5-tricarboxamide-based polymers increases as a function of the number of monomers.[40]

To verify that the stability of our protein increases with increasing degree of polymerization, we calculate the number of hydrogen bonds and the solvent accessible surface area as a function of the number of monomers in the aggregates. Shown in Fig. 6 is the free energy landscape of all of our assemblies as a function of number of hydrogen bonds per monomer, starting off from β-sheet structures. In Fig. 7, only simulation results for the β-roll structures for the monomer and tetramer are shown for clarity. As expected, the location of the minimum of the free energy of the monomer and the dimer configurations is approximately at the same value, corresponding to about four hydrogen bonds per monomer. Again, this is

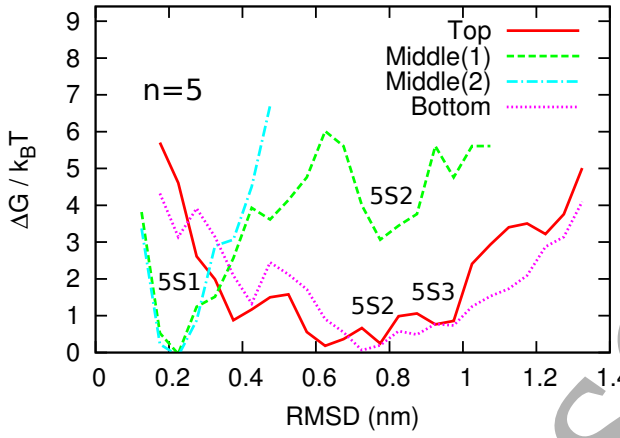

FIG. 5. Free energy, ΔG, obtained from simulations starting from a tetramer of β-sheets with five repeat units as a function of the root-mean square deviation of snapshots of the individual monomers in the tetramer. The structures corresponding to the minima of the free energy that are denoted as 5S1, 5S2, 5R1 and 5R2 shown in Fig. 4. The average of RMSD of the monomers is $RMSD \approx 0.22$ nm with a standard deviation of about 0.04 nm. The average and standard deviation of the top and bottom monomers are about 0.7 nm and 0.2 nm.

because the β-sheet monomers in the dimer melt, separate and form individual monomers. As the number of monomers increases the average number of hydrogen bonds per monomer shifts to higher values and the corresponding standard deviation decreases. This confirms that our protein structures become more stable as the number of monomers in the aggregates increases.

For the trimer and tetramer configurations, we find that the location of the free-energy minimum shifts to larger number of hydrogen bonds as a function of the number of monomers, because of the formation of inter-monomer hydrogen bonds. So, the free energy gain associated with additional hydrogen bonding increases with the number of monomers, arguably offsetting the entropy loss of formation of the β-sheet structures from random coils (molten globules) that also increases with the number of monomers. In fact, hydrogen bonds are also involved in the bonding between the proteins. An even deeper analysis is recurrent.

Assemblies are presumably not only stabilized by inter-molecular hydrogen bonding but also by hydrophobic interactions between the proteins. This we can probe by calculating the solvent accessible surface area (SA) for each of the assemblies because our proteins are

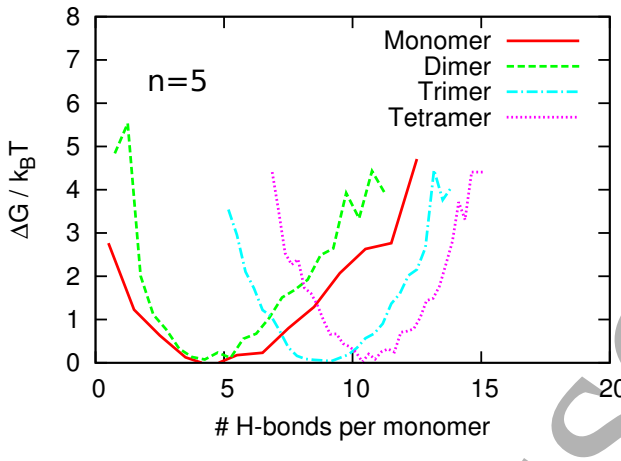

FIG. 6. Free energy, ΔG, obtained from simulations at temperature $T = 298$ K starting from a monomer, a dimer, a trimer and a tetramer of β-sheets of our protein consisting of five repeat units, as a function of the number of hydrogen bonds divided by the number of monomers. The values of the average and standard deviation for a monomer, a dimer, a trimer and a tetramer are (4.6, 2.19), (4.5, 1.7), (8.7, 1.5) and (10.5, 1.4), respectively.

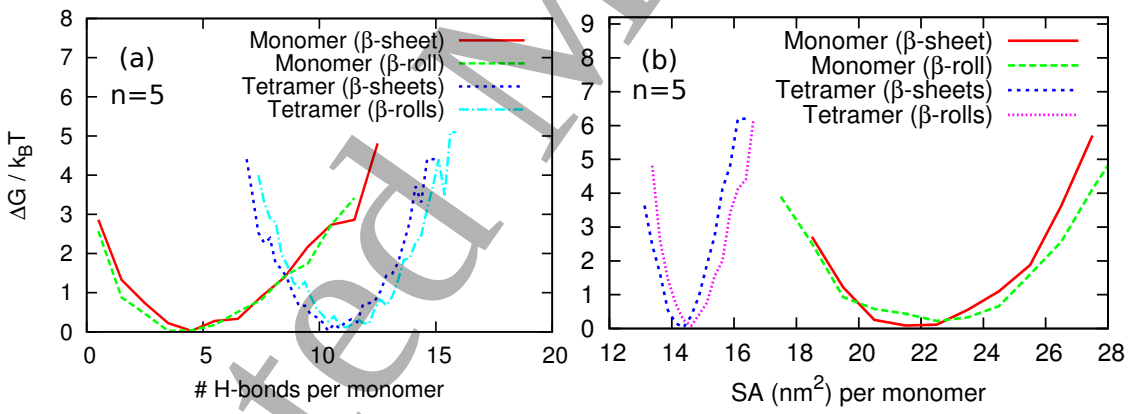

FIG. 7. Free energy, ΔG, as a function of (a) the number of hydrogen bonds per monomer and (b) the solvent accesible surface area, SA, for the monomers and tetramers computed from simulations that start from β-roll or β-sheet structures.

almost entirely made up of hydrophobic amino acid residues (only Glutamine is polar). For all of the structures starting from β-sheets this is shown in Fig. 8, and for the monomer and tetramer of the initial β-roll configurations in Fig. 7. The figures clearly demonstrate that as the assemblies grow in size more of the hydrophobic surface is buried inside the assembly. This means that the overall free-energy cost of the contact of the hydrophobic

residues of the proteins with water decreases as a function of the degree of polymerization of the assemblies. This is herein bound to be utilized to estimate effective binding free energies and a critical polymerization concentration, that, interestingly, are virtually the same for both the β-sheets and the β-rolls.

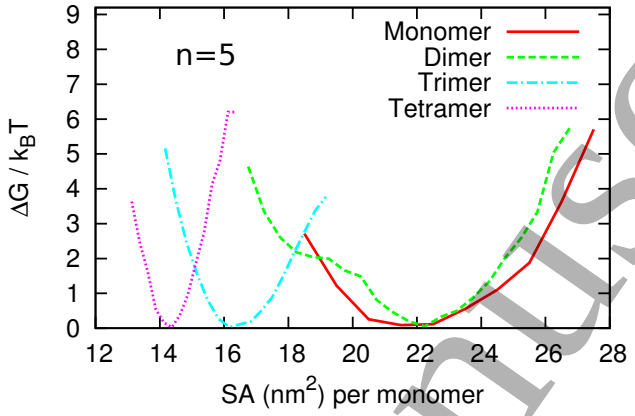

FIG. 8. Free-energy, ΔG, obtained from simulations at temperature T = 298 K starting from a monomer, a dimer, a trimer and a tetramer of β-sheets consisting of five repeat units as a function of the solvent accessible surface area divided by the number of monomers.

All of this indicates that β-sheet and β-roll assemblies must be equally stable, at least within our simulations. This is rather surprising, considering that the large-scale structure of the both types of assembly appear very different. Indeed, as shown in Fig. 9b, the width of a tetramer consisting of the β-rolls is smaller than that of the β-sheets. The former is about 1.3 nm and the latter equals approximately 1.9 nm. On the other hand, if we consider the internal structure of the β-roll tetramer then it becomes evident that the β-rolls form inter-molecular β-sheets, that is, hydrogen bonds are formed between strands of two or more of the β-rolls. See Fig. 9. The β-roll tetramer actually resembles a tetramer consisting of tilted β-sheets. This presumably explains the small difference in stability. An in-depth reconsideration follows for the same issue.

First, however, we discuss stability of the structure of assemblies formed by the triblock protein and compare this with what we found for the pentablock.

III B. Stability of the triblock proteins

The question arises how the results that we obtained in the previous sections depend on the length of the silk-like protein sequence, i.e., on the number of repeat units that make up this protein. As alluded to in the introduction, this is relevant to the experiments of

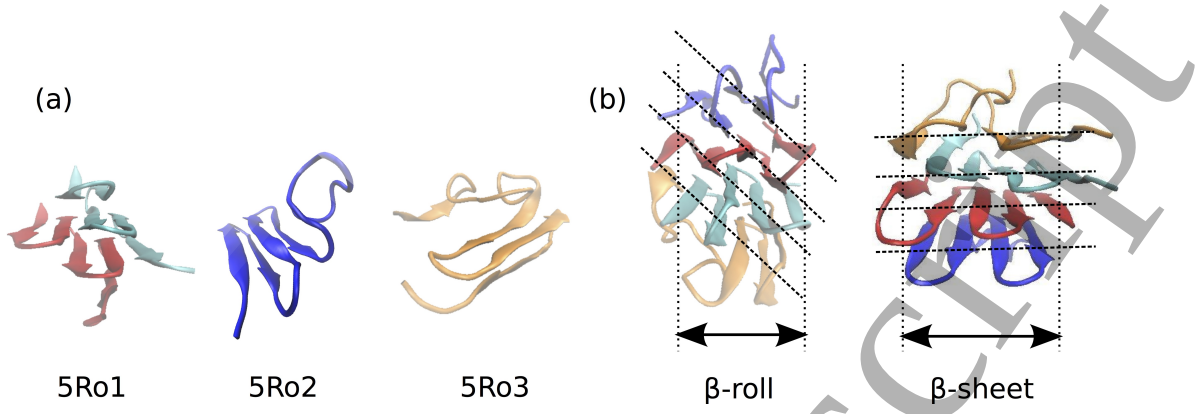

FIG. 9. (a) From the left to the right: snapshots of simulations representing two β-rolls forming inter-molecular β-sheet (5Ro1), a β-roll with a molten strand (5Ro2), a β-roll with a molten strand and a half molten strand (5Ro3). (b) Snapshots of tetramers containing the β-rolls (on the left) and β-sheets (on the right). For the former the approximate width of the aggregate shown by arrows is about 1.3 nm and for the latter it is about 1.9 nm. Orientation of each intra- and inter-molecular β-sheet plane is shown by a dashed line.

de Vries et al., where increasing the number of repeat units of the silk-like block enhances the self-assembly of their proteins on a DNA molecule.[25] To address this question, we again rely on REMD simulations of monomer, dimer, trimer and tetramer of β-sheet and β-roll structures of a smaller sequence consisting of $n = 3$ repeat units. This protein is sufficiently large for formation of a β-sheet and a β-roll structures and is small enough to allow us to study the differences between its self-assembly and that of the protein with $n = 5$ repeat units. For compactness, we only discuss the results we obtained for the β-sheets. Again, as was the case for the pentablock protein, our findings for the β-sheet and β-roll assemblies of the triblock are consistent.

In our simulations, the monomers melt into random coil structures. The monomers in the dimers also melt and subsequently separate. For the trimer, first, the top and bottom monomers of the assembly melt. Next, one of these two monomer detaches itself from the trimer. Following that, the remaining folded monomer of the dimer also melts and eventually the dimer disassembles. This indicates that the protein-protein interactions between the triblock proteins are weaker than those of the pentablocks: trimers of the latter species are indeed stable. The tetramer largely follows the route of the trimer, that is, first the outer two monomers melt and one of them detaches towards the end of the simulation run.

Presumably, the remaining trimer would also completely melt and disassemble in a longer simulation run. However, within our simulation time of 20 ns per replica we are not able to observe the completion of this process. It seems that none of the assemblies upto and including the tetramer are stable under the conditions tested.

To illustrate how we obtained this conclusion, apart from studying snapshots, we analyze how the free energy curves that we obtain depend on the RMSD, the number of hydrogen bonds and the solvent accessible surface area of the protein configurations sampled. Because we sample over the entire simulation time, which includes the actual melting process, our free energy is not a true free energy and contains information on non-equilibrium configurations. This can be seen in Fig. 10 that shows the free-energy of the aggregates as a function of the RMSD of the simulations starting from the β-sheet structures.

The obtained free-energy curves of the monomer and the monomers in the dimer are similar. They all have a minimum at a relatively large value of the RMSD, confirming that the monomers of the dimer separate and form random coil structures. Here, the melting process is so quick that the folded structure does not significantly contribute to the free energy. This is not so for the trimer. The free energy of this structure has two minima, one of which is associated with the original β-sheet and the other one with the molten state. The former is indicated in the figure with the acronym *3S1* and the latter with *3R2* that correspond to structures illustrated in Fig. 11.

The melting and breakup of the tetramer is so slow that the free energy of the proteins in the center retain most of their original configuration. This expresses itself in the free energy landscape of the monomer in the center of the tetramer shown in Fig. 10 in the existence of a minimum at an RMSD between 0.3 and 0.4 nm. This is consistent with an almost perfect β-sheet structure such as the *3S1* structure shown in Fig. 11. A similar picture emerges if we compare the free energy landscapes of all four monomers in the tetramer, shown in Fig. 12. All the monomers in the tetramer retain at least some of their original structures, associated with the β-sheet and β-roll. Conformationally more flexible ones, i.e., ones that have to cross a lower barrier to the actual free energy minimum, develop additional minima associated with different structures (denoted *3S2*, *3R1* and *3R2* in Fig. 11).

More information can be extracted from Fig. 12 allowing us to compare the structures of the triblock and pentablock proteins. For instance, the width of the free energy well for the monomer below the top monomer, which is denoted as *middle(1)* in Fig. 12, is wider than

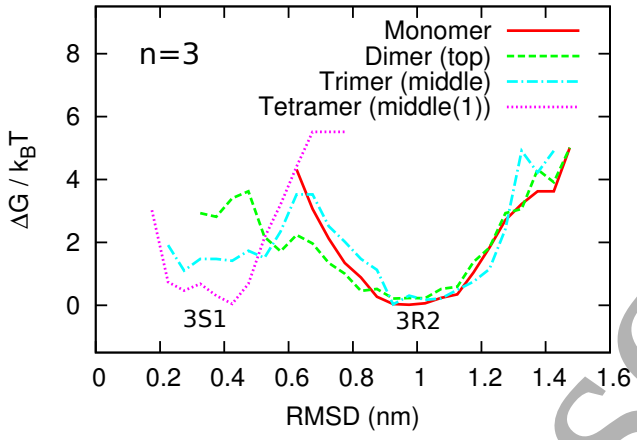

FIG. 10. Free energy, ΔG, obtained from simulations starting from a monomer, a dimer, a trimer and a tetramer of β-sheets with three repeat units as a function of the root-mean square deviation of snapshots of the monomer, one of the monomers in the dimer, the second monomer in the middle of the trimer and tetramer from their initial β-sheet structures. The structures corresponding to the minima of the free energy that are denoted as 3S1 and 3R2 are shown in Fig. 11.

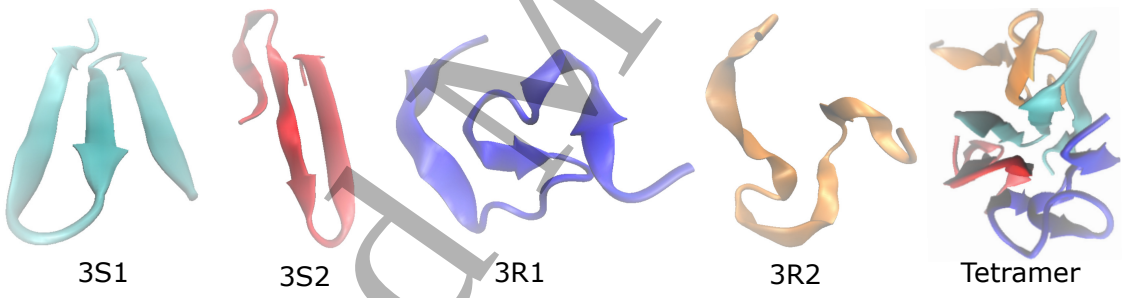

FIG. 11. From the left to the right: snapshots of simulations representing a β-sheet (3S1), a β-sheet with a molten strand (3S2), a half molten β-sheet (3R1), a random coil strucrue (3R2) and a tetramer containing proteins with three repeat units. The snapshots are taken from simulations at the lowest tempereture (T = 298 K) replica.

both monomers in the middle of the tetramer consisting of pentablocks. Hence, the middle triblock protein fluctuates more than the corresponding pentablock ones. Also, the global minimum in the free-energy landscape of the second middle monomer, named *middle(2)* in Fig. 12, is associated with a β-sheet with a molten strand (see *3S2* in Fig. 11).

Moreover, the monomer at the bottom of the aggregate melts into a disordered structure, similar to that of *3R1* in Fig. 11. The free-energy landscape of the monomer on the top has three minima that correspond to three different structures. This includes the initial

3S1 structure that morphs into a partially molten, two-stranded *3S2* structure and the final random coil structure *3R2*. The latter eventually separates from the rest of the aggregate. Apparently, the time evolution of the top and bottom proteins in the tetramer is not the same. Presumably this reflects the up-down asymmetry of the initial β-sheet structures and hence of the entire assembly.

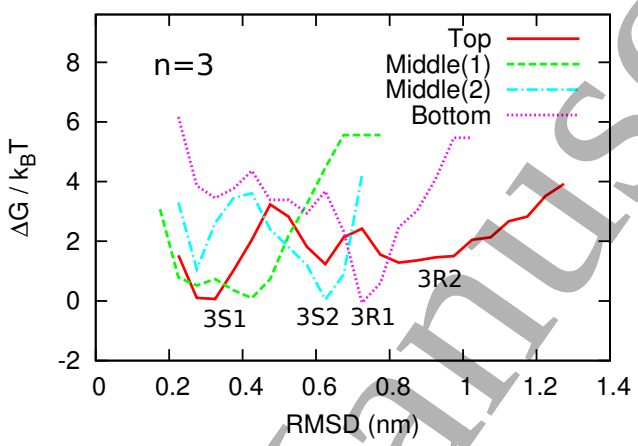

FIG. 12. Free energy, ΔG, obtained from simulations starting from a tetramer of β-sheets with three repeat units as a function of the root-mean square deviation, RMSD, of the individual monomers in the tetramer from their initial β-sheet structure. The structures corresponding to the minima of the free energy that are denoted as 3S1, 3S2, 3R1 and 3R2 are shown in Fig. 11.

To investigate more quantitatively the reason for the structural differences between the aggregates that consist of the two proteins with three and five repeat units, we also compute the free energy of the configurations as a function of the number of hydrogen bonds per monomer. Results are shown in Fig. 13. Again, the free-energy curves of the monomer and the dimer approximately match, because both melt in a short time, but they are different from that of the trimer. This is because the monomers in the trimer melt into random coil structures but they only separate towards the end of the simulation. This causes the shift in the minimum and the asymmetry of the free energy landscape. The shift we observe for the tetramer has the same root cause, amplified by the even slower conformational relaxation processes that take place in this assembly.

For the maximum number of intra- and inter-molecular hydrogen bonds in a fully folded (β-sheet or β-roll) tetramer of the pentablock, we find from our simulations a value of about 15 per monomer. Given the size of the triblock we would expect this to go down to about

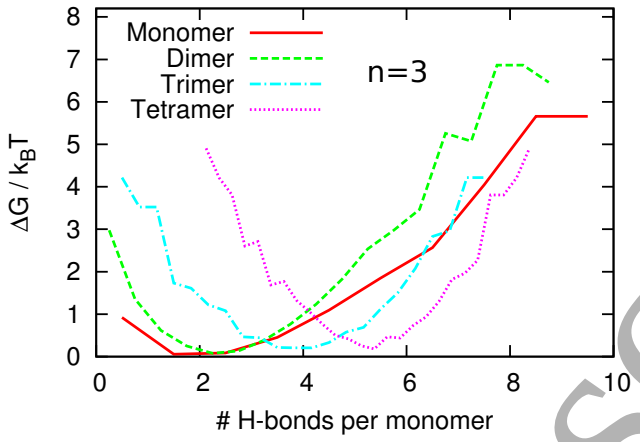

FIG. 13. Free energy, ΔG, obtained from simulations starting from a monomer, a dimer, a trimer and a tetramer of β-sheets with three repeat units as a function of the number of hydrogen bonds divided by the number of monomers.

9. We read off from Fig. 13 that the optimal values are much smaller than that, caused by the melting of the folded structures during the course of the simulation. The melting is partially caused by a reduction in the maximum solvent accessible surface area that can be shielded from water due to the smaller size of the protein.

Fig. 14 confirms this expectation. Comparing with the solvent accessible surface area of the pentablock proteins, we see a fifty per cent reduction in the gain of that quantity upon assembly. Of course, this is not really surprising given the differences in protein size. Also, the ratio of the average solvent accessible surface area, SA, of the molten monomer for the case of triblock to that of the pentablock is about 0.68, which is slightly smaller than the ratio that is expected for collapsed polymers: $(5/3)^{(2/3)} = 0.71$.

For the triblock and pentablock proteins we find that a minimum number of proteins is required to produce a stable aggregate. For the triblock this number is probably larger than four, whilst that for the pentablock it is either two or three depending on whether we are dealing with a β-roll or β-sheet structure. The existence of a difference in configuration of free monomers and the smallest stable folded aggregate implies (i) that this minimum number of disordered free monomers need to come together and (ii) that they have to undergo a conformational change in order to form the smallest thermodynamically stable assembly. Consequently, the linear polymerization of this kind of protein must be nucleated, be it kinetically and/or thermodynamically.[41] Kinetic nucleation occurs through high-free

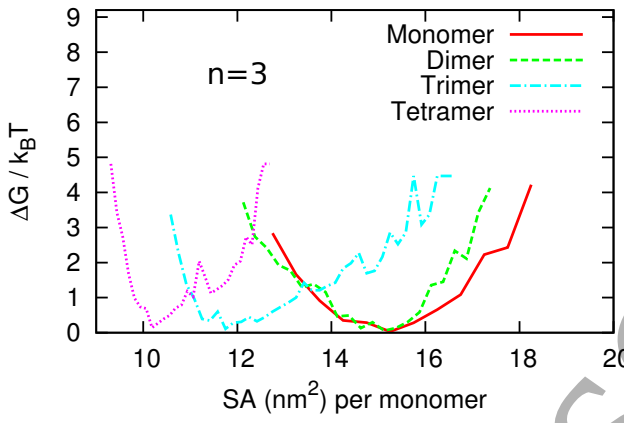

FIG. 14. Free energy, ΔG, obtained from simulations starting from a monomer, a dimer, a trimer and a tetramer of β-sheets with three repeat units as a function of the solvent accessible surface area divided by the number of monomers.

energy intermediate conformational states of monomers between de- and attached states, thermodynamic nucleation through high-free energy conformers in bound state. The latter are stabilized by a binding free energy.

This could explain why the long self-assembled fibers of a triblock construct of two collagen-like disordered sequences sandwiching a silk-like central block take such a long time to form in the experiments of Martens and coworkers.[42] Incidentally, these authors find that the width of the fibers that form in solutions of their protein is almost half of the value that is expected for fibers made out of β-sheets.[42] This suggests that the β-roll structure in the assembled state of the protein is the most stable one albeit that we cannot exclude the possibility that this is due to kinetics rather than thermodynamics.

As already alluded to, from our simulations we have to conclude that β-roll and β-sheet assemblies must be virtually equally stable or very nearly that, at least upto the tetramer level. We can make this statement more quantitative. For this we make use of insights based on a simple coarse-grained model that describes the cooperative binding of monomers into supramolecular polymers.[41] Two free energies describe activation and elongation of these self-assembled polymers. The first is a free energy penalty associated with assuming the polymerization-active state of the molecules and the second a binding free energy between polymerization-active molecules in the assembly. The presumption is that the polymerization-active state represents an excited state or conformation of the molecule that in the assembly is stabilized by the binding to its neighbors. Hence, the model describes

thermodynamic nucleation.

These two free energies we now attempt to estimate from our simulations for the two proteins. Within the nucleated assembly model the total binding free energy, ΔG, depends on the number of monomers in an assembly, N, as,

$$\Delta G(N) = (N-1)\Delta G_b + N\Delta G_c. \quad (1)$$

Here, ΔG_b is the free energy of the formation of a bond between a protein and the assembly, representing elongation of the chain. ΔG_c is the free energy of the conformational change required for the protein to be able to bind. To calculate ΔG_b and ΔG_c, we first compute the $\Delta G(N)$ for our pentablock protein aggregates.

As a first estimate, we presume that binding is driven by hydrophobic interactions that we associate with the reduction of solvent accessible surface area upon binding. We compute the $\Delta G(N)$ by (i) calculating the SA for folded structures upto $N = 4$, (ii) subtract from that the SA of respectively three and four molten monomers, (iii) multiply this by the surface tension of the hydrophobic parts of the protein, γ, (iv) fit Eq. 1 to these two data points. The surface tension, γ, is used in our implicit solvent simulations to calculate the non-polar contribution of the solvation free energy, and has a value of 0.84 k_BT/nm^2 at room temperature (T = 298 K).[43]

Applying this procedure for the $N = 3$ and $N = 4$ β-sheets of the pentablock protein gives values of $\Delta G_c \simeq +8\ k_BT$ and $\Delta G_b \simeq -20\ k_BT$. From this admittedly crude estimate we can indeed conclude that the folded structure represent a high-energy state, stabilized by protein-protein interactions. Interestingly, these values are not all that far off from those obtained by Aggeli and coworkers who analyzed their experimental data on the self-assembly of a β-sheet forming oligopeptide using the same nucleated assembly model: $\Delta G_c \simeq +7\ k_BT$ and $\Delta G_b \simeq -31\ k_BT$.[4] Note that our estimate is based entirely on the contribution from hydrophobic interactions and hence should be considered as a lower estimate. We return to this issue below.

If we repeat this exercise for the $N = 2$, 3 and 4 β-rolls of the pentablock proteins, we find $\Delta G_c \simeq +3 \pm 2\ k_BT$ and $\Delta G_b \simeq -13 \pm 5\ k_BT$. For the β-roll we have error bars because we have three data points. This is not so for the β-sheets for which we only have two data points and hence by definition obtain a perfect fit. See Fig. 16. The estimate of the errors for both protein types we expect to be comparable, hence we do not think there

is a real statistically significance between the different values of the free energies that we obtain. Indeed, the critical assembly concentration, X_c (a mole fraction), we obtain this way is virtually identical for the two conformers[41]: $k_BT \ln X_c = \Delta G_b + \Delta G_c$. For our β-sheets $\ln X_c = -12$ and for the β-rolls $\ln X_c = -10 \pm 6$. That ΔG_c acts as a thermodynamic nucleation free energy is shown in Fig. 15, where we show the mean degree of polymerization, N, as a function of the protein concentration for $\Delta G_b + \Delta G_c = -12\ k_BT$ and two values of $\Delta G_c = +8\ k_BT$ and $0\ k_BT$, obtained from the two constant assembly models.[41] For $\Delta G_c = +8\ k_BT$ the polymerization is sharper than for $\Delta G_c = 0\ k_BT$. For a description of the model the reader is referred to the work of van der Schoot.[41]

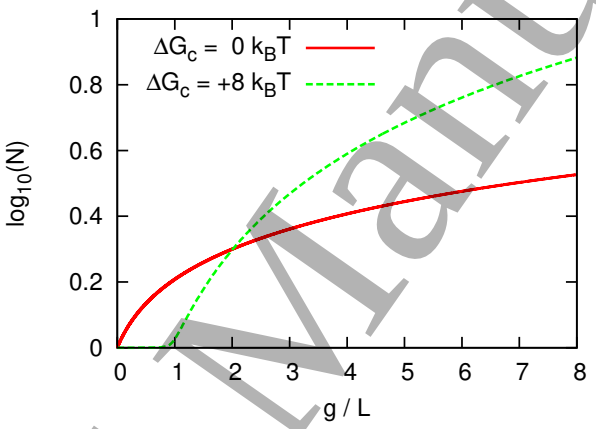

FIG. 15. Mean degree of polymerization, N, as a function of concentration of our pentablock protein for two values of the free energy of conformational change of the protein upon binding, $\Delta G_c = +8\ k_BT$ and $0\ k_BT$. For both curves $\Delta G_b + \Delta G_c = -12$.

Our analysis shows that according to our simulations self-assembled fibrils consisting of β-sheets and those of β-rolls of our protein must be approximately equally stable. From the experiments of Martens et al.[42] on their silk-like protein we obtain a lower bound for $\Delta G_b + \Delta G_c$ of $-18\ k_BT$. This number is much more negative than what we find but this is not surprising because their silk-like protein is about five times larger than ours ($n = 24$). On the other hand, their block is connected to a very large disordered protein blocks that apparently significantly reduce the net binding free energy of that protein. Interestingly, binding energies close to $20\ k_BT$ are typically found in the context of protein assembly of, e.g., viruses.[44,45]

As discussed in Sec. III A, the fibers that Martens and collaborators find must, because

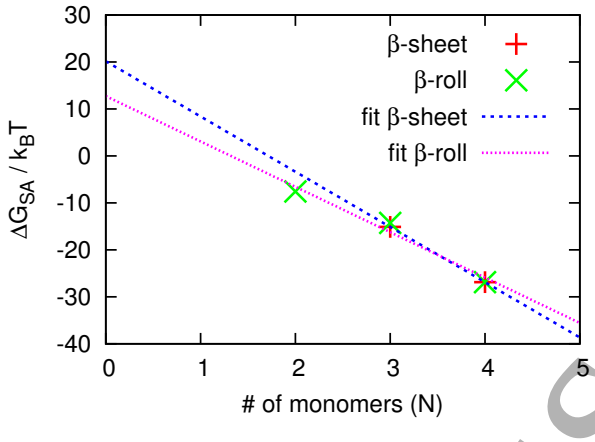

FIG. 16. Free energy contribution of hydrophobic interactions, ΔG_{SA}, obtained from simulations starting from N β-sheets and β-rolls with five repeat units as a function of the number of monomers N. The lines are plotted by fitting our data to eq. 1.

of their dimensions, be of the β-roll variety. This is confirmed by the REMD simulations of Schor and collaborators on a much smaller silk-like protein with $n = 6$.[37] The authors find that in monomeric and the dimeric states the β-roll structures are more stable than the others. However, in a recent study where we repeat the simulations of Schor et al. and we find that the most stable monomer structure is a disordered globule not a β-roll.[27] This is also true for the dimers.

Note that the repeat unit in Schor's work that was inspired by that of Martens et al. is slightly different from the one studied in this work: $(GA)_3$-GE instead of our $(GA)_3$-GX with X alternatingly Q and K. The difference between our previous results and those of Schor on the same sequence we described to differences in simulation time.[27] Our explicit solvent simulations were twice and our implicit solvent ones ten times as long. This highlights the importance of full equilibration of the simulation, in this context meaning that the proteins must visit all the replicas sufficiently often.

In this work, by the same REMD equilibration criterion, our simulations should have completely equilibrated. This is because, as mentioned in Sec. II, the average simulation time that it takes for a replica at the lowest temperature to diffuse up to the highest temperature and again down to the lowest one is about 1 to 3 ns and our simulations are 20 to 40 ns long. Therefore during a simulation run replicas sufficiently explore the temperature space. However, we do observe that one of the monomers of the tetramer consisting of our triblock

proteins melts and detaches to form a trimer. We find that starting from a folded trimer, the trimer melts and dissociates. This implies that the simulation on the tetramer at least as far as disassembly is concerned does not reach its thermal equilibrium. This does not mean that the folded tetramer of the pentablock has not equilibrated but does suggest that it would be useful in the future study to revisit this problem using much longer simulation times analyzing results with a specific tool Wordom[47] that currently are out of our reach.

Finally, we have in our discussion above ignored any contribution from intra- and inter-molecular hydrogen bonding to the stability of the proteins. Their contribution is actually not so trivial to assess because one would need to measure the difference in free energy between the hydrogen bonds between the donors and acceptors on the protein and those between moieties on the protein and water. If we ignore this complication and assign a net binding free energy to every hydrogen bond, G_h, then we find that the contribution to ΔG_c and ΔG_b for the pentablock β-sheet structures amount to approximately -12 G_h and $+24$ G_h, respectively. A reasonable estimate for G_h is between -2.6 and -3.2 k_BT.[46]

This suggests that hydrogen bonding could in principle significantly enhance the stability of the folded aggregate structure but also that hydrogen bonding increases the free energy of the folded state. The latter conclusion is somewhat surprising, because one would naively expect that the folded structure engages in more intra-molecular hydrogen bonds than the molten structure does. Hence, it appears that inter-molecular hydrogen bonds stabilize the β-sheet and β-roll structures of our protein. On the other hand, a naive uncoupling of the effects of burying hydrophobic surface area and the formation of hydrogen bonds ignores the exchange of intra-molecular hydrogen bonds in the globular structure for inter-molecular ones.

IV CONCLUSIONS

We applied replica exchange molecular dynamics simulations to study the influence of the length of the primary sequence of a "blocky" silk-like protein on its quaternary structure. The stability of aggregates containing one, two, three and four of the protein monomers consisting of three or five repeat units was investigated. All simulations were started from two initial configurations: β-sheets and β-rolls. We find that monomers in the assemblies containing the smaller-length proteins melt into random coil structures and separate. Single monomers of both β-sheets and -rolls of the pentablock protein also melt. Dimers of the β-sheet proteins also melt into random coil structures, whereas those of the β-roll seem

stable. Trimers and tetramers of both folds remain folded within our simulation time.

This shows that the inter-molecular interactions stabilizing the protein aggregates must be stronger for the longer proteins. We find that the monomers in the center of the trimers and tetramers are conformationally more stable than the ones on the top and the bottom the assemblies. Interestingly, the ones in the middle of the tetramers are more stable than the ones of the trimers. This results in more conformationally stable folded monomers on the top and bottom of the tetramer due to the inter-molecular interactions with the middle ones and suggest that as the assemblies become longer their stability increases. Presumably this effect levels off for sufficiently long assemblies, as is seen in theoretical studies of assemblies of small molecules.[40]

Our findings allow an in-depth understanding of self-assembly of protein building blocks and provide a new approach for the preparation of bio-inspired nanoscale materials with a unique morphologies.

ACNOWLEDGMENTS

The work of J. Razzokov is supported by Jepa-Limmat Foundation. We thank Sarah Harris (University of Leeds) and Alexey Lyulin (Eindhoven University of Technology), for useful discussions and advice on the simulations. Eindhoven University of Technology is thanked by J. Razzokov for their hospitality. We are grateful for computer time provided by the Dutch National Computing Facilities at the LISA facility at SURFsara. The work of S. Naderi forms part of the research program of the Dutch Polymer Institute (DPI, Project No. 698). This work was supported by NWO Exacte Wetenschappen (Physical Sciences) for the use of supercomputer facilities, with financial support from the Nederlandse Organisatie voor, Onderzoek (Netherlands Organization for Scientific Research, NWO).

References

[1]Cheng, J.Y., Ross, C.A., Smith, H.I., Thomas, E.L. Templated SelfAssembly of Block Copolymers: TopDown Helps BottomUp. *Adv. Mater.*, **2006**, 18, 2505-2521.

[2]Masatsugu Shimomura, Tetsuro Sawadaishi. Bottom-up strategy of materials fabrication: a new trend in nanotechnology of soft materials. *Curr. Opin. Colloid Interface Sci.*, **2001**, 6, 11 - 16.

[3]Bai, F., Wang, D., Huo, Z., Chen, W., Liu, L., Liang, X., Chen, C., Wang, X., Peng, Q., Li, Y. A Versatile Bottomup Assembly Approach to Colloidal Spheres from Nanocrystals. *Angew. Chem. Int. Ed.*, **2007**, 46, 6650-6653

[4]Aggeli, A., Nyrkova, I. A., Bell, M., Harding, R., Carrick, L., McLeish, T. C. B., Boden, N. Hierarchical self-assembly of chiral rod-like molecules as a model for peptide β-sheet tapes, ribbons, fibrils, and fibers. *P. Natl. Acad. Sci.*, **2001**, 98, 11857-11862.

[5]Vauthey, S., Santoso, S., Gong, H., Watson, N., Zhang, S. *P. Natl. Acad. Sci.*, Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles. **2002**, 99, 5355-5360.

[6]Cui, H., Webber, M. J., Stupp, S. I. Selfassembly of peptide amphiphiles: From molecules to nanostructures to biomaterials. *Peptide Science*, **2010**, 94, 1-18.

[7]Ulijn, R. V., Smith, A. M. Designing peptide based nanomaterials. *Chem. Soc. Rev.*, **2008**, 37, 664-675.

[8]Chen, C. L., Bromley, K. M., Moradian-Oldak, J., DeYoreo, J. J. In situ AFM study of amelogenin assembly and disassembly dynamics on charged surfaces provides insights on matrix protein self-assembly. *J. Am. Chem. Soc.*, **2011**, 133, 17406-17413.

[9]Graveland-Bikker, J. F., Schaap, I. A. T., Schmidt, C. F., De Kruif, C. G. Structural and mechanical study of a self-assembling protein nanotube. *Nano Letters*, **2006**, 6, 616-621.

[10]Klein, M. L., Shinoda, W. Large-scale molecular dynamics simulations of self-assembling systems. *Science*, **2008**, 321, 798-800.

[11]Colombo, G., Soto, P., Gazit, E. Peptide self-assembly at the nanoscale: a challenging target for computational and experimental biotechnology *Trends Biotechnol.*, **2007**, 25, 211-218.

[12]Davies, R. P. W., Aggeli, A. Selfassembly of amphiphilic sheet peptide tapes based on aliphatic side chains. *J. Pept. Sci.*, **2011**, 17, 107-114.

[13]Scharnagl, C., Reif, M., Friedrich, J. Stability of proteins: temperature, pressure and the role of the solvent. *BBA-Proteins. Proteom.*, **2005**, 1749, 187-213.

[14]Shukla, D., Schneider, C. P., Trout, B. L. Complex interactions between molecular ions in solution and their effect on protein stability. *J. Am. Chem. Soc.*, **2011**, 133, 18713-18718.

[15]Miklos, A. C., Sarkar, M., Wang, Y., Pielak, G. J. Protein crowding tunes protein stability. *J. Am. Chem. Soc.*, **2011**, 133, 7116-7120.

[16]Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., ... Grimsley, G. R. Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.*, **2011**, 408, 514-528.

[17]Cavazzana-Calvo, M., Fischer, A. Gene therapy for severe combined immunodeficiency: are we there yet?. *J. Clin. Invest.*, **2007**, 6, 1456-1465.

[18]Cavazzana-Calvo, M., Lagresle, C., Hacein-Bey-Abina, S., Fischer, A. Gene therapy for severe combined immunodeficiency. *Annu. Rev. Med.*, **2005**, 56, 585-602

[19]Roos, W. H., Bruinsma, R., Wuite, G. J. L. Physical virology. *Nat. Phys.*, **2010**, 6, 733-743.

[20]Giacca, M., Zacchigna, S. Virus-mediated gene delivery for human gene therapy. *J. Control. Release.*, **2012**, 161, 377-388.

[21]Marshall, E. *Science*, **1999**, 286, 2244-2245.

[22]Check, E. Gene therapy: a tragic setback. *Nature*, **2002**, 420, 116-118.

[23]Thomas, M., Klibanov, A.M. Non-viral gene therapy: polycation-mediated DNA delivery. *Appl. Microbiol. Biotechnol.*, **2003**, 62, 27-34.

[24]Smeenk, J. M., Schn, P., Otten, M. B., Speller, S., Stunnenberg, H. G., van Hest, J. C. Fibril formation by triblock copolymers of silklike β-sheet polypeptides and poly (ethylene glycol). *Macromolecules*, **2006**, 39, 2989-2997.

[25]Hernandez-Garcia, A., Kraft, D. J., Janssen, A. F., Bomans, P. H., Sommerdijk, N. A., Thies-Weesie, D. M., ... & de Vries, R. Design and self-assembly of simple coat proteins for artificial viruses. *Nature nanotechnology*,**2014**, 9, 698-702.

[26]Smeenk, J. M., Otten, M. B., Thies, J., Tirrell, D. A., Stunnenberg, H. G., van Hest, J. Controlled Assembly of Macromolecular Sheet Fibrils. *Angew. Chem. Int. Ed.*, **2005**, 44, 1968-1971.

[27]Razzokov, J., Naderi, S., van der Schoot, P. Prediction of the structure of a silk-like protein in oligomeric states using explicit and implicit solvent models. *Soft matter*, **2014**, 10, 5362-5374

[28]SalomonFerrer, R., Case, D.A. Walker, R.C. An overview of the Amber biomolecular simulation package. *Wiley. Interdiscip. Rev. Comput. Mol. Sci*, **2013**, 3, 198-210.

[29]Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **2006**, 65, 712-725.

[30]Tsui, V., Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers*, **2000**, 56, 275-291.

[31]Anandakrishnan, R., Aleksander, Drozdetski., Ross, Walker., Alexey, Onufriev. Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. *Biophys. J.*, **2015**, 5, 1153-1164.

[32]Feig, M., Im, W., Brooks III, C.L. Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.*, **2004**, 120, 903-911.

[33]Zhang, L.Y., Gallicchio, E., Friesner, R.A., Levy, R.M. Solvent models for proteinligand binding: Comparison of implicit solvent Poisson and surface generalized Born models with explicit solvent simulations. *J. Com. Chem.*, **2001**, 22, 591-607.

[34]Lomize, A.L., Pogozheva, I.D., Mosberg, H.I., Anisotropic solvent model of the lipid bilayer. 1. Parameterization of long-range electrostatics and first solvation shell effects. *J. Chem. Inf. Model.*, **2011**, 51, 918-929.

[35]Srinivasan, J., Trevathan, M.W., Beroza, P. and Case, D.A., 1999. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acco.*, **1999**, 101, 426-434.

[36]Humphrey, W., Dalke, A., Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics.*, **1996**, 14, 33-48.

[37]Schor, M., Martens, A. A., Stuart, M. A. C., Bolhuis, P. G.Prediction of solvent dependent β-roll formation of a self-assembling silk-like protein domain. *Soft Matter*, **2009**, 5, 2658-2665.

[38]Kleinjung, J., Fraternali, F. Design and application of implicit solvent models in biomolecular simulations. *Curr. Opin. Struct. Biol*, **2014**, 25, 126-134.

[39]Shrake, A., Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol*, **1973**, 79, 351 - 371.

[40]Filot, I. A., Palmans, A. R., Hilbers, P. A., van Santen, R. A., Pidko, E. A., de Greef, T. F. Understanding cooperativity in hydrogen-bond-induced supramolecular polymerization: a

density functional theory study. *J. Phys. Chem. B.*, **2010**, 114, 13667-13674.

[41]van der Schoot, P. *Nucleation and Co-Operativity in Supramolecular Polymers. Advances In Chemical Engineering*, **2009**, 35, 45-77.

[42]Martens, A. A., Portale, G., Werten, M. W., de Vries, R. J., Eggink, G., Cohen Stuart, M. A., de Wolf, F. A. Triblock protein copolymers forming supramolecular nanotapes and pH-responsive gels. *Macromolecules*, **2009**, 42, 1002-1009.

[43]Sitkoff, D., Sharp, K. A., Honig, B. Accurate calculation of hydration free-energies using macroscopic solvent models. *J. Phys. Chem.*, **1994**, 98, 1978-1988.

[44]Kegel, W. K., van der Schoot, P. Competing hydrophobic and screened-Coulomb interactions in hepatitis B virus capsid assembly. *Biophys. J.*, **2004**, 86, 3905-3913.

[45]Kegel, W. K., van der Schoot, P. Physical regulation of the self-assembly of tobacco mosaic virus coat protein. *Biophys. J.*, **2006**, 91, 1501 - 1512.

[46]Sheu, S. Y., Yang, D. Y., Selzle, H. L., Schlag, E. W. Energetics of hydrogen bonds in peptides. *Proc. Nat. Acad. Sci.*, **2003**, 100, 12683-12687.

[47]Seeber, M., Cecchini, M., Rao, F., Settanni, G., Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*, **2007**, 23, 2625-2627.

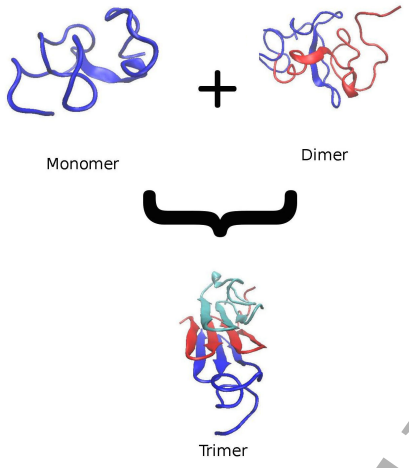

FIG. 17. For Table of Contents Only