

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/263293700>

Prediction of the structure of a silk-like protein in oligomeric states using explicit and implicit solvent models

ARTICLE *in* SOFT MATTER · JUNE 2014

Impact Factor: 4.03 · DOI: 10.1039/c4sm00384e · Source: PubMed

READS

54

3 AUTHORS:



Jamoliddin Razzokov

University of Antwerp

7 PUBLICATIONS 0 CITATIONS

[SEE PROFILE](#)



Saber Naderi

Technische Universiteit Eindhoven

7 PUBLICATIONS 7 CITATIONS

[SEE PROFILE](#)



Paul van der Schoot

Technische Universiteit Eindhoven

143 PUBLICATIONS 3,174 CITATIONS

[SEE PROFILE](#)

Cite this: DOI: 10.1039/c4sm00384e

Prediction of the structure of a silk-like protein in oligomeric states using explicit and implicit solvent models

Jamoliddin Razzokov,^{*ab} Saber Naderi^{bc} and Paul van der Schoot^{bd}

We perform Replica Exchange Molecular Dynamics (REMD) simulations on a silk-like protein design with amino-acid sequence $[(\text{Gly-Ala})_3\text{-Gly-Glu}]_5$ to investigate the stability of a single protein, a dimer, a trimer and a tetramer made up of these proteins starting from β -roll and β -sheet structures in both explicit (TIP3P) and implicit (GBSA) solvent models. Our simulation results for the implicit solvent model agree with those for the explicit solvent model for simulation times up to the longest tested, being 30 ns per replica. From this we infer that the implicit solvent model that we use is reliable, allowing us to reach much longer time scales (up to 200 ns per replica). We find that the self-assembly of fibers of these proteins in solution must be a nucleated process, involving nuclei made up of at least three monomers. We also find that the conformation of the protein changes upon assembly, *i.e.*, there is a transition from a disordered globular state to an ordered β -sheet structure in the self-assembled state of aggregates containing more than two monomers. This indicates that autosteric effects must be important in the polymerization of this protein, reminiscent of what is observed for β -amyloids. Our findings are consistent with recent experimental results on a protein with an amino acid sequence similar to that of the protein we study.

Received 18th February 2014
Accepted 30th May 2014

DOI: 10.1039/c4sm00384e
www.rsc.org/softmatter

I. Introduction

In designing nano-scale functional systems, two approaches can be considered: a bottom-up one in which one makes use of basic physical principles that govern the nano-scale world to rationally design a system with desired functionalities, and a top-down one, where components of a nano-scale system are controlled and assembled by externally manipulated tools. In the former approach, components of the system are designed in such a way that they self-assemble into larger-scale structures of desired shape and/or functionality for example by inter-molecular interactions.^{1–5} Self-assembly processes of this kind are studied in the context of many fields, including bionanotechnology, drug delivery, nano sensors, self-healing coatings, tissue engineering, and DNA coating.^{6–11} Understanding them is vital for progress in nanophysics and biophysics,^{12,13} as it helps us to find how to take control over macromolecular behavior by tuning physiochemical parameters such as temperature, solvent, ionic strength, acidity and hydrogen bonding.^{14,15}

In biological or biomimetic self-assembly the building blocks are often proteins and oligo peptides,^{16–23} the latter often specifically designed. Towards a rational design of proteinaceous materials it is important to understand the interactions within and between the building blocks that govern their thermodynamic stability.^{20,24–32} For instance, depending on the amino acid sequence, proteins and peptides form different types of secondary structure such as α -helices and β sheets.^{33–35} β -Sheet structures have a tendency to stack on top of each other due to hydrogen bonding and/or hydrophobic interactions, and self-assemble into fibers.³⁶ For some sequences, α -helices morph into β sheets in order to be able to form fibres.³⁷ This could be seen as an example of autosteric binding, *i.e.*, involving conformational switching of a protein facilitating the switching of subsequent proteins that attach to the growing fibre.³⁸ Conformational switching and allosteric are believed to be important factors regulating virus self-assembly.³⁹

Protein fiber formation is the object of intense investigation, *e.g.*, in the context of Alzheimer's disease^{40–42} and of biomimetic synthetic proteins with a structure-forming function.⁴³ In a recent study, Martens and co-workers performed experiments to study the self-assembly of triblock copolymers consisting of two hydrophilic, collagen-like blocks attached to a central silk-like block with the amino acid sequence $[(\text{Gly-Ala})_3\text{-Gly-Glu}]_n$.⁴³ In these experiments, the copolymer self-assembles into fibers in water at low pH, where the glutamic acid residues are largely uncharged. The self-assembly process itself takes about

^aInstitute Ion-Plasma and Laser Technologies, Academy of Sciences of Uzbekistan, Dordon yoli Str. 33, 100125, Tashkent, Uzbekistan. E-mail: jrazzokov@gmail.com

^bFaculteit Technische Natuurkunde, Technische Universiteit Eindhoven, Postbus 513, 5600 MB Eindhoven, The Netherlands

^cDutch Polymer Institute, P.O. Box 902, 5600 AX Eindhoven, The Netherlands

^dInstituut voor Theoretische Fysica, Universiteit Utrecht, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands

20 hours, which indicates that the protein–protein binding must be nucleated and highly cooperative. The self-assembled fibers were found to grow in excess of a micrometer in length, suggesting very strong binding. They turn out to be ribbon-shaped with a height and a width of 2.8 and 13.6 nm, respectively, according to atomic force and electron microscopy (AFM and TEM) and small angle X-ray scattering (SAXS) measurements.⁴³ The core of the fibers contains stacks of the silk-like central blocks but their internal structure and arrangement in the stacks remains unclear.

Here we study the structural details of these proteins by extending a recent simulation work by Schor and collaborators.⁴⁴ In this work, the authors performed all-atom replica exchange molecular dynamics (REMD) simulations to investigate the thermodynamic stability of the silk-like central block in water and in methanol. They concluded that in water the most stable structure is that with two interconnected parallel β -sheets, which they term a β -roll.⁴⁶ These simulations were carried out over a temperature range of 298 K to 500 K starting from a monomer (a single β -roll) or a dimer containing two β -sheets or two β -rolls. For the latter, the simulation time was 17 ns per replica and for the former it was 10 ns per replica.

Free energy landscapes obtained from the simulations starting from a single β -roll showed that this structure is very stable. For the case of a dimer starting from two β -sheets, one of the β -sheets partially turned into β -roll while the other one remained a β -sheet. In this case, the minima in the free energy landscape were of the same depth as for both β -sheet and β -roll states implying a degenerate quaternary structure. In the simulations with a dimer starting from two β -rolls, both monomers most of the time maintain their initial structure although one of them occasionally turned into a β -sheet. This is why the corresponding minimum in the free energy landscape of the β -sheet was not as deep as the one corresponding to the β -roll. From these results, it was concluded that the most stable structure of this protein in water is a β -roll.⁴⁴

It is important to note that the free energy landscapes obtained by Schor and collaborators do not converge, meaning that the one obtained from simulations starting from two β -sheets is different from the one starting from two β -rolls.⁴⁴ This could be an indication that 17 ns simulation time per replica is not sufficient for exploring the phase space of the protein of such a size. One of the great challenges of this type of simulation is that they are extremely computationally time consuming. This is mostly because in explicit solvent simulations most of the simulation box is filled with water molecules and this consumes most of the simulation time. One way to reach longer time scales is to use implicit solvent models.^{47,48} In these models, the effect of solvent molecules is included in the simulations *via* a continuum solvent model. These models usually take into account the effect of electrostatic properties of the solvent and/or the free energy of solvation.

In this work, we invoke both explicit and implicit solvent models in our simulations and extend the study of Schor *et al.* to much longer simulation times. To this end, first, we carry out short (30 ns per replica) simulations with the two solvent models and compare the results to make certain that the

implicit solvent model reproduces the results of the simulations with explicit solvent. Next, we perform replica exchange simulations with our implicit solvent model of choice in order to reach longer timescales, and explore more of the phase space and obtain converged results. Our simulation results indicate that the conclusion of Schor and coworkers might be premature and that the situation is much more complex. Indeed, we find that for the case of the monomers and dimers, the proteins form a globular structure after 50 ns per replica simulation time. So, the most stable structure of a monomer is a globular one, and binding two monomers does not liberate enough free energy to turn them into beta-sheets or beta-rolls.

In addition to the monomers and dimers that were considered by Schor and collaborators, we study the stability of larger aggregates containing three and four proteins at long timescales. For the case of the trimer, the initial β -sheet or β -roll structures do not change significantly after 100 ns (per replica) nor does that of the tetramer after 50 ns. Our findings suggest that our proteins in the trimeric and tetrameric state prefer a β -sheet or β -roll structure, not the globular one. This indicates that there must be a conformational change from the globular to the β -roll or β -sheet structure upon assembly of monomers or dimers into trimers and tetramers.

The remainder of this paper is organized as follows. In Section II we describe our simulation methods and the way we analyze our simulation data. Section III we discuss our simulation results and Section IV is devoted to conclusions.

II. Simulation methods

We perform all-atom molecular dynamics simulations using the GROMACS package (version 4.5.5). In order to create β -sheet and β -roll structures, we used the Xleap program included in AmberTools.⁵⁶ First, we created a long chain with the amino-acid sequence $[(\text{Gly-Ala})_3\text{-Gly-Glu}]_n$ with $n = 5$. In our simulations, all glutamic acid (Glu) residues are protonated, so have a zero charge. Next, we manipulate the dihedral angles to obtain a structure close to a β -sheet and, by pulling the β -strands towards each other in a short simulation run using the Amber simulation package,^{57,58} we create β -sheet structures. To obtain β -roll structures, we use different values for the dihedral angles and pull every second strand towards each other in a short simulation run. In order to create stacks of two to four β -sheets (β -rolls), we copy and translate one β -sheet (β -roll) in the direction perpendicular to the β -sheet plane. We equilibrate all these structures by performing 1 ns MD simulations using a stochastic dynamics (SD) integrator with implicit solvent.⁵⁹ In the following, we describe in more detail our methods for the explicit and implicit solvent simulations, and the way we analyse the obtained simulation data.

A. Explicit solvent

The OPLSAA⁶⁰ force field is employed in combination with the TIP3P explicit water model for all our explicit solvent simulations. Protein and water systems are equilibrated by carrying out 1 ns conventional MD simulations at constant temperature

and pressure of 298 K and 1 bar, using the velocity-rescaling thermostat and the Parinello–Rahman barostat⁶¹ in order to prepare them for our REMD simulations. Rhombic dodecahedra box types are considered for all simulation boxes. All our REMD simulations with explicit solvent are performed in the NVT ensemble. For these simulations there are 48 replicas at 48 temperatures in the range of 298–500 K. To obtain the optimal temperature distribution we make use of a method described by Prakash and coworkers and we choose the temperature range in such a way that the exchange probabilities, p_{ex} , are about 30 percent.⁶² The exchange probabilities are calculated based on the Metropolis criterion where the weight factors are given by the product of Boltzmann factors for each replica. Replica exchange attempts between neighbouring replicas are done every 500 MD steps with each step being 2 fs. For the case of monomers and dimers, the volume of the simulation boxes in explicit solvent simulation are about the same as those in the work of Schor and coworkers.⁴⁴ The particle-mesh Ewald method⁴⁵ is applied to deal with the electrostatic interactions between atoms. The salt concentration in the simulations with explicit solvent is 0.1 M. The total integrated REMD time is 0.96 μ s [48 \times 20 ns] in the explicit solvent simulations with monomers of β -sheet and β -roll initial structure. For the case of dimer and trimer β -sheet and β -roll structures the integrated REMD time is 1.44 μ s [48 \times 30 ns].

B. Implicit solvent

Implicit solvent models have been widely used to decrease the computing time of simulations and to reach long time scales.⁴⁹ In many studies, these models have also been compared to simulations with explicit solvent.^{50–55} For example, Zhou and coworkers compared explicit solvent simulations with the OPLSAA force field and the SPC water model to different combinations of implicit solvent models and force fields, such as OPLS/SGB, AMBER94/SGB, AMBER96/GBSA with GB solvent assessable area (GBSA) and AMBER96/GBSA, AMBER99/GBSA, using a highly parallel replica exchange method. The lowest energy state in the free energy landscapes obtained from simulations with implicit solvent was not the native state, except for the AMBER96/GBSA force field combination.⁵² The results of this type of study show that the implicit solvent model is a useful model for reaching long time scales but it is not always reliable and careful checking against explicit solvent models is required.

For the implicit solvent simulations we used the OPLSAA force field in combination with the generalized Born solvent accessible surface area (GBSA) model. In these simulations the salt concentration is zero. Because the net electric charge of each residue on our protein chain is zero, we expect that the presence or absence of salt is not relevant in the simulations. In order to check this we compare the results of our simulations with explicit solvent and salt concentration of 0.1 M to those with implicit solvent. We find that the free energy landscapes obtained from the two types of simulation are essentially identical indicating that the absence of salt in our implicit solvent simulations is not an issue. In simulations involving a single monomer, there are 9 replicas in the temperature range

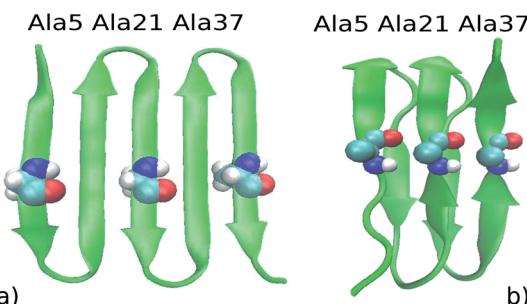


Fig. 1 Schematic representation of (a) a β -sheet and (b) a β -roll structure of our silk-like protein. Alanine residues in the middle of every second strand are highlighted by indicating their atomic structures.

of 289 K to 518 K. For all other implicit solvent simulations involving protein dimers, trimers and tetramers, we use 12 replicas in the temperature range of 298 to 511 K. The temperature distribution is chosen in such a way that exchange probabilities vary between 10 to 25 percent. The simulation time step is set at 2 fs and replica exchange attempts occur every 500 MD steps. The sum of the REMD simulation time for the case of the monomer β -sheet and β -roll is equal to 1.8 μ s [9 \times 200 ns]. The entire REMD simulation time for the dimer β -sheet and β -roll is 2.4 μ s [12 \times 200 ns]. The overall REMD simulation time for the trimer β -sheet and β -roll is 1.2 μ s [12 \times 100 ns]. Integrated REMD simulation time appropriate to the tetramer β -sheet and β -roll case is 0.6 μ s [12 \times 50 ns].

C. Data analysis

To study the most stable structure of our protein in monomeric and aggregated states, we obtain a free energy landscape from each REMD simulation. For this purpose, we first calculate the probability of finding our protein in a state, $p(x, y)$, as a function of two relevant reaction coordinates x and y to be specified below. Next, we obtain the free energy from the Boltzmann distribution, *i.e.*, $F(x, y) \propto -k_B T \ln p(x, y)$ where k_B is the Boltzmann constant and T the absolute temperature. Here, we assume that our simulations are sufficiently long to allow our protein to explore all the possible states. The reaction coordinates that we choose are the distances between the alanine residues in the middle of the β -strands. This enables us to observe conformational changes of our protein, and to distinguish between β -roll and β -sheet structures because the distance between second neighboring strands is different in these structures (see Fig. 1). In Fig. 1 the alanine residues in the middle of strands are shown for three strands of the β -roll and β -sheet structures. For the case of a β -roll the distance between Ala5 and Ala21 is about 0.48 nm. This differs significantly from the distance between the same residues for the case of a β -sheet that measures about 0.8 nm. We checked that our identification is accurate by evaluating snapshots.

III. Results and discussion

In this section we discuss our simulation results for the monomers, dimers, trimers and tetramers separately and

postpone a global overview of our conclusions in the next section.

A. β -Roll and β -sheet monomer in explicit and implicit solvent

We perform REMD simulations to investigate thermodynamic stability of both β -sheet and β -roll structures in explicit and implicit solvent. Free energy landscapes are calculated as a function of the distance between Ala6 and Ala21, and that between Ala21 and Ala37 for single proteins; see Fig. 1. First, we perform short simulations (10 ns per replica) in order to compare explicit and implicit solvent models. We find that in explicit and implicit solvent the β -sheet structure is unstable and melts in our short simulations into a globular structure. However, the β -roll structure is stable in these simulations and the deepest minimum in the free energy landscapes corresponds to this structure (see Fig. 2a and b). This is in agreement

with the simulation results of Schor and collaborators.⁴⁴ By comparing the free energy landscape of a β -roll in explicit solvent (Fig. 2a) to that in implicit solvent (Fig. 2b), we find that more of phase space is explored in the latter case. This is due to the fact that viscosity of water is ignored in the simulations with implicit solvent model and therefore the protein motion faster. This difference in the dynamics of proteins in explicit and implicit solvent is less visible in the simulations with a larger number of monomers in the simulation box (see below). This is presumably because of the inter-protein interactions that restrict the motion of molecules.

In order to check how strongly the results depend on the simulation time, we extended the REMD simulations with explicit solvent to 20 ns per replica. Shown in Fig. 2d is the free energy landscape obtained from our simulation data, corresponding to the last 5 ns of the simulation, *i.e.*, the first 15 ns data are discarded. We do this to avoid as much as possible the

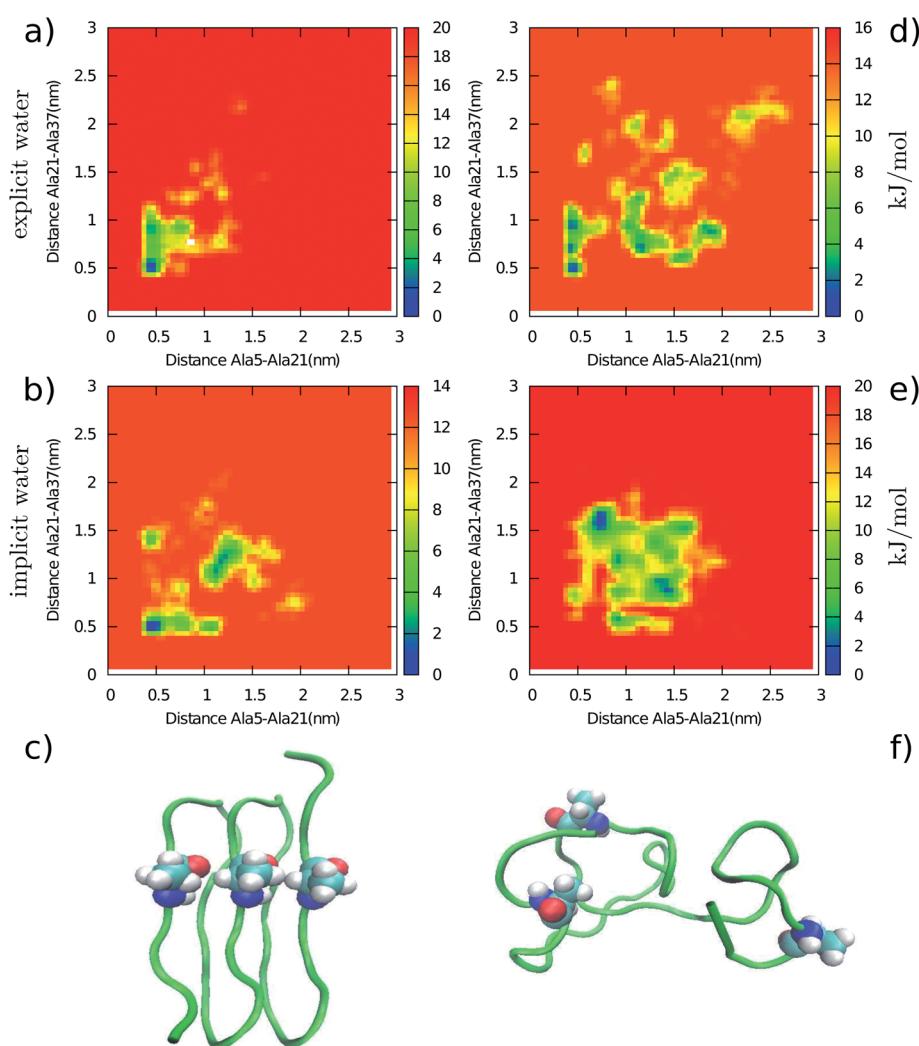


Fig. 2 Free energy landscape of a monomer as a function of the distance between Ala5 and Ala21 and that between Ala21 and Ala37, obtained from simulations starting from a β -roll structure. (a) Free energy obtained from 10 ns per replica simulation with explicit solvent, (b) 10 ns per replica simulation with implicit solvent, (d) 20 ns per replica simulation with explicit solvent, (e) 200 ns per replica simulation with implicit solvent. For the two latter cases the data of the first 15 ns are discarded. (c) Snapshots of simulations with explicit solvent after 10 ns per replica and (f) implicit solvent after 20 ns per replica.

bias towards the initial state of our simulation. This free energy landscape is different from the one that we obtain from the 10 ns simulations, indicating that 10 ns per replica simulation time is not sufficient for determining the most stable structure of the protein.

There are four to six minima in this free energy landscape that have more or less the same depth indicating conformational flexibility (see Fig. 2d). Two of these minima correspond to random globular structures (see, e.g., Fig. 2f). It suggests that the β -roll structure might be a local minimum structure and that the protein might melt on longer time scales. Note that this is in contrast with what Schor and co-workers found at shorter time scales (10 ns per replica) where the β -roll structure is the most stable structure of a monomer.⁴⁴

Extending further our 48 replica REMD simulations to longer time scales is computationally very expensive with explicit solvent. Therefore, to investigate the stability of the β -roll structure on longer time scales, we extend the computationally less demanding implicit solvent simulations to 200 ns per replica. Since the results of our 10 ns per replica simulations with explicit and implicit solvent agree well with each other, we expect to obtain the same results at longer timescales albeit that this is not assumed certain. Shown in Fig. 2e is the free energy landscape obtained from this simulation. It has two major minima corresponding to two random globular structures, showing that the β -roll structure has indeed molten. All in all, the results of longer timescale simulations with explicit and implicit solvent models suggest that not only the β -sheet but also the β -roll structure is not a stable structure in water, at least not in its monomeric state.

In the next section we study the stability of this silk-like protein in an aggregated state that contains two of these proteins on top of each other. As we shall see, even in dimeric form, the proteins melt.

B. β -Roll and β -sheet dimers in explicit and implicit solvent

To evaluate the most stable structure of our protein in a dimeric state, we carry out simulations starting from two β -sheets (denoted “2sheets”) or two β -rolls (“2rolls”) on top of each other. To obtain free energy landscapes, in addition to calculating the distance between Ala5 and Ala21, and between Ala21 and Ala37, which correspond to the first monomer, we in addition calculate the Ala46-to-Ala62 and the Ala62-to-Ala78 distances corresponding to the second monomer. First, we perform 30 ns per replica REMD simulations using explicit and implicit solvent models. For the case of the 2sheets simulations, we obtain consistent results from both solvent models and the deepest minimum in the free energy landscape corresponds to the β -sheet structure for both of them (see Fig. 3). This is in agreement with 17 ns per replica simulations of Schor and collaborators,⁴⁴ except that they find another minimum in the free energy landscape of one of the β -sheets that corresponds to a β -roll structure. In our simulations occasionally β -sheets partially turn into β -rolls, but we do not find a deep minimum corresponding to a β -roll structure. In order to verify if the simulation time affects the structure of our dimer, we extend

the implicit solvent simulations to 200 ns per replica. Interestingly, we observe that in this simulation the two β -sheets melt into a random globular structure, similar to what we observed in the simulations with a single monomer.

To further test if these results depend on the initial structure we perform REMD simulations with a dimer starting from a stack of two β -rolls on top of each other. Shown in Fig. 4a is the free energy landscape of one of the monomers obtained from a 30 ns per replica simulation with explicit solvent. The deepest minimum corresponds to a β -sheet structure, which means that this monomer partially turns into a β -sheet (see Fig. 4c, the monomer is colored yellow). For the case of the other monomer, the β -roll structure appears to be stable (see Fig. 4c, the monomer is colored green). This is similar to what we obtained from the simulation with implicit solvent (see Fig. 4d-f), except that in this simulation the deepest minimum in the free energy landscape of the first monomer does not correspond to a β -sheet structure (see Fig. 4d). This might be due to the short simulation time and poor sampling of phase space; in one of the simulations the protein by chance escapes from a local minimum and finds another one whilst in the other simulation the protein stays in the initial state. The fact that the free energy landscapes obtained from 30 ns per replica simulations starting from stacks of two β -sheets and two β -rolls do not converge, supports this idea. To reach longer timescales, we again extend the implicit solvent simulations to 200 ns per replica. Similar to what we observed for the case of two β -sheets, we find that both monomers form a random globular structure.

Our simulation results show that the silk-like protein forms a globular structure in both monomeric and dimeric states. This must be due to the interaction of hydrophobic residues on the surface of this protein with water in these two states. The number of hydrophobic residues in contact with water decreases in the aggregates as they are buried inside the globular core of the aggregates. To further investigate this, in the next sections, we perform simulations with aggregates containing three and four of these proteins.

C. β -Roll and β -sheet trimers in explicit and implicit solvent

The effect of inter-molecular interactions on the stability of our protein becomes clear by performing 30 ns per replica REMD simulation for the case of a trimer containing three β -rolls in explicit and implicit solvent. We find that this structure is very stable in both solvent models. However, for the case of the implicit solvent we observe that β -roll strands reorient to form inter-molecular β -sheets (see Fig. 5h) similar to those that have been observed in the work of Prieto and coworkers for a different type of protein.⁶³ In this figure β -roll strands from different monomers interact and form β -sheets. This also occurs for the case of simulations with explicit solvent but in these simulations β -strands are less mobile than in the simulations with implicit solvent. This is presumably due to interaction of the β -strands with the water molecules.

We also find that the folded structure monomer in the middle of the trimer is the most stable one (see Fig. 5a). This again indicates that the inter-molecular interactions stabilize

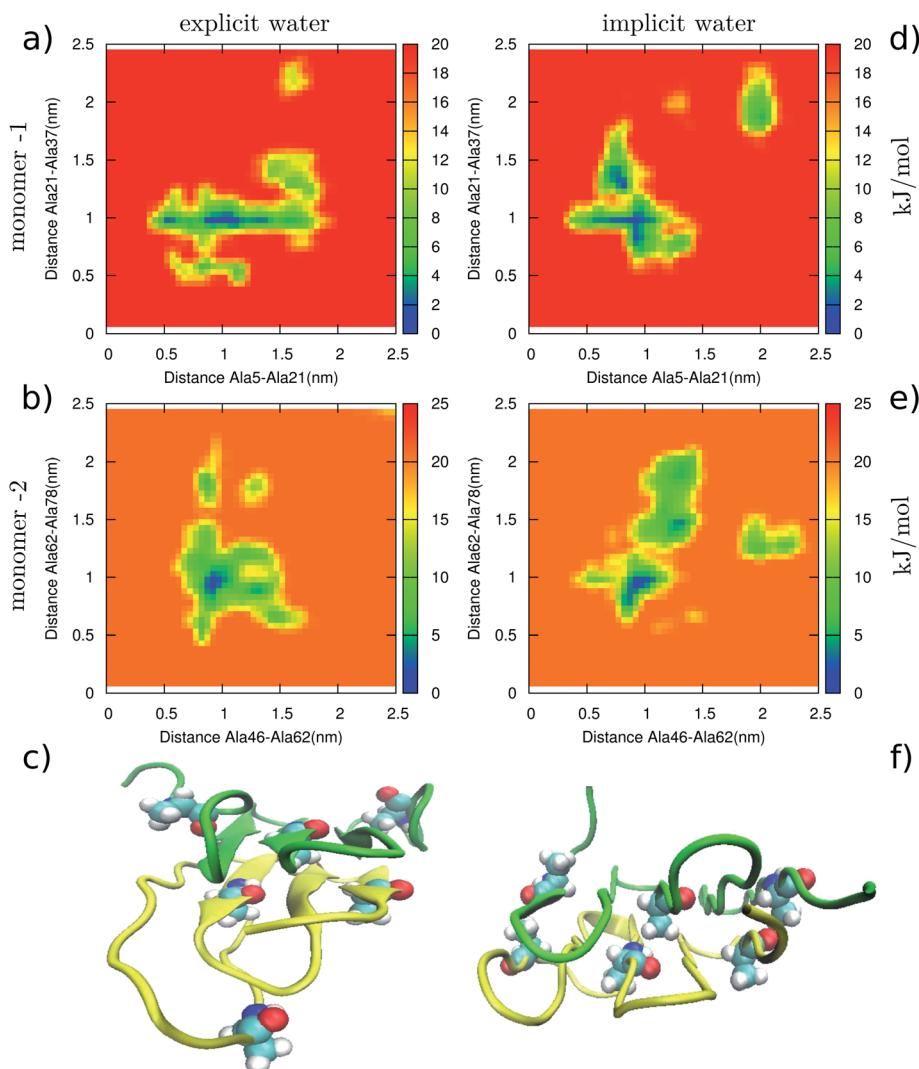


Fig. 3 Free energy landscape of a dimer as a function of the distance between Ala5 and Ala21, and between Ala21 and Ala37 for the first monomer (a and d), and the distance between Ala46 and Ala62 and between Ala62 and Ala78 for the second monomer (b and e), obtained from REMD simulations starting from a β -sheet structure. Free energy landscapes obtained from a 30 ns per replica simulation with explicit solvent (a and b) and 30 ns per replica simulations with implicit solvent (d and e). Snapshots of a dimer after a 30 ns per replica simulation with explicit solvent (c) and with implicit solvent (f).

β -roll (or β -sheet) structures. In order to ascertain that our simulation results do not depend on the simulation time, we extend our implicit solvent simulations to 100 ns per replica. We find that the results do not change significantly and the β -roll trimer with inter-molecular β -sheets remains the most stable structure.

We repeat our simulations starting from another initial state with a trimer containing three β -sheets. First, we perform 30 ns per replica REMD simulations in explicit and implicit solvent to check if we obtain the same results from these simulations. Next, we extend the implicit solvent simulations to 100 ns per replica. Free energy landscapes obtained from these simulations are shown in Fig. 6, which shows that the simulations with implicit and explicit solvent are very similar. Again, we observe that the folded structure of the monomer in the middle of the trimer is more stable than the ones on the top and bottom of the stack. Trimer β -sheets seem to maintain their initial structure

whereas β -rolls slightly change their conformation in order to form inter-molecular β -sheets.

Our results from long timescale simulations show that the silk-like protein that we study forms a globular structure in monomeric and dimeric states. To self-assemble into fibers, which this type of protein does,⁴³ the proteins need to stack on top of each other, but stacking will not be thermodynamically favoured unless the stacks contain sufficiently many proteins of the right conformational structure. This suggests that the process of fiber formation of these proteins observed in the experiments of Martens and collaborators⁴³ on a protein that has a central block that is similar to our protein model must be a nucleated process. Note that one can conclude this because of the fact that fiber formation occurs over the course of 20 hours and seems to have a lag time. In fact, our simulations on tetramers confirm this picture.

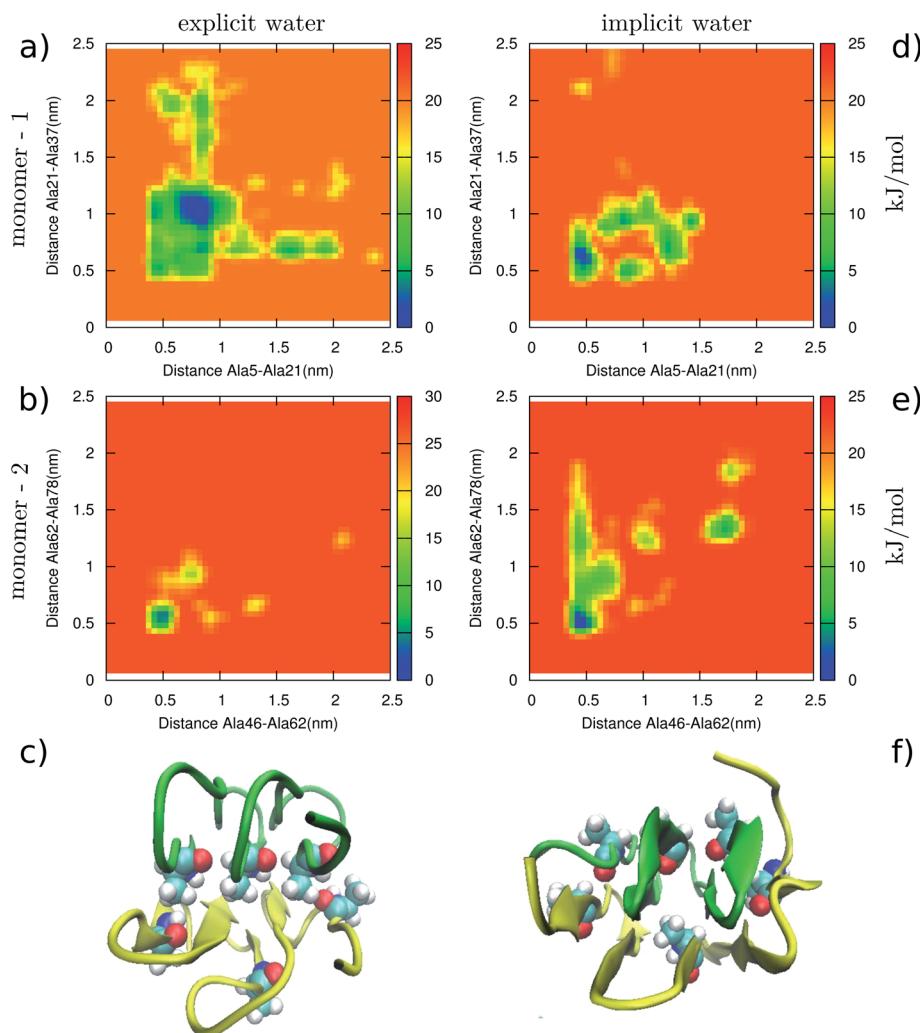


Fig. 4 Free energy landscape of a dimer as a function of the distance between Ala5 and Ala21 and between Ala21 and Ala37 for the first monomer (a and d), and the distance between Ala46 and Ala62 and between Ala62 and Ala78 for the second monomer (b and e) obtained from REMD simulations starting from a β -roll structure. Free energy landscapes obtained from a 30 ns per replica simulation with explicit solvent (a and b) and 30 ns per replica simulations with implicit solvent (d and e). Snapshots of a dimer after a 30 ns per replica simulation with explicit solvent (c) and with implicit solvent (f).

D. Tetramers in implicit solvent

As a final test, we perform 50 ns REMD simulations starting from all β -roll and all β -sheet tetramers in implicit solvent. Our previous simulations have shown that the implicit solvent model that we use is reliable for our purposes. Free energy landscapes obtained from simulations starting from four β rolls are shown in Fig. 7. The initial and final state of this tetramer is shown in Fig. 7c and f. Interestingly, the two β -rolls in the middle of the tetramer turn into an intermolecular β -sheet. This can be seen from the location of the deepest minimum of the free energy landscape corresponding to these two proteins (see Fig. 7b and d).

We also perform 50 ns REMD simulations with a tetramer containing four β -sheets. We find that this structure is very stable in water. Free energy landscapes obtained from these simulations are shown in Fig. 8. The two proteins at the top and bottom of the tetramer are more mobile because each interacts

with one other protein only, so interact more with the implicit solvent (compared to the proteins in the middle), but, interestingly, they seem to be less mobile, *i.e.*, explore a smaller range of conformations compared to those at the top and bottom of the trimer. This is presumably because the core of the tetramer contains two stable structures that suppress fluctuations in the motion of the monomers at the top and bottom through hydrogen bonding, whereas the core of the trimer contains only one. From the simulations with four β rolls, we again observe the formation of inter-molecular β -sheets (see Fig. 7f). These results indicate that the aggregates containing inter- and intra-molecular β -sheets must be degenerate states.

E. Discussion

The simulation results that we presented in previous sections show that for the case of the monomers and dimers, the proteins form a globular structure after 50 ns per replica

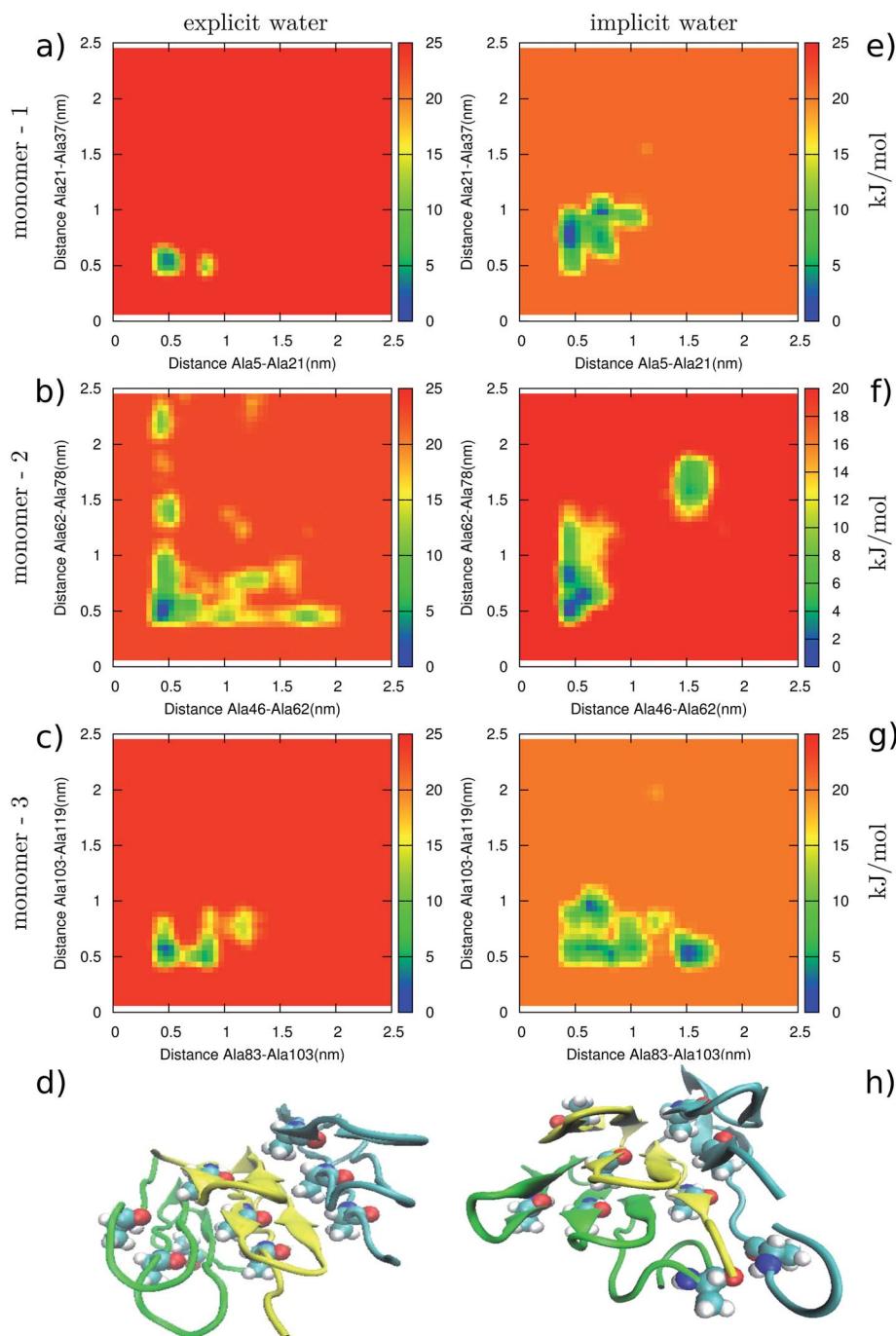


Fig. 5 Free energy landscape of a trimer as a function of the distance between Ala5 and Ala21 and between Ala21 and Ala37 for the first monomer, distance between Ala46 and Ala62 and between Ala62 and Ala78 for the second monomer, between Ala83 and Ala103 and between Ala103 and Ala119 for the third monomer obtained from simulations starting from a β -roll structure. (a)–(c) Free energy obtained from simulation with explicit solvent 30 ns per replica, and (e)–(g) implicit solvent simulations 30 ns per replica simulations. Snapshots of a trimer after 30 ns per replica simulations: (d) explicit solvent simulation and (h) implicit solvent simulation. Notice the inter-protein β -sheet structure.

simulation time. So monomers prefer a globular structure, and binding two monomers does not provide enough energy to turn that into a β -sheet or β -roll. For the case of the trimer, the structure of the proteins does not change significantly after 100 ns (per replica) nor does that of the tetramer after 50 ns. This suggests that the most stable structure of our proteins in the trimeric and tetrameric state is a β -sheet or β -roll, not

globular structure. This conformational change from a disordered globular monomer or dimer to an ordered trimer or tetramer must be due to competition between components of free-energy.

To more closely study why the conformational changes occur, we calculate the average number of hydrogen bonds, \overline{HB} , divided by the number of monomers, N , the average number of

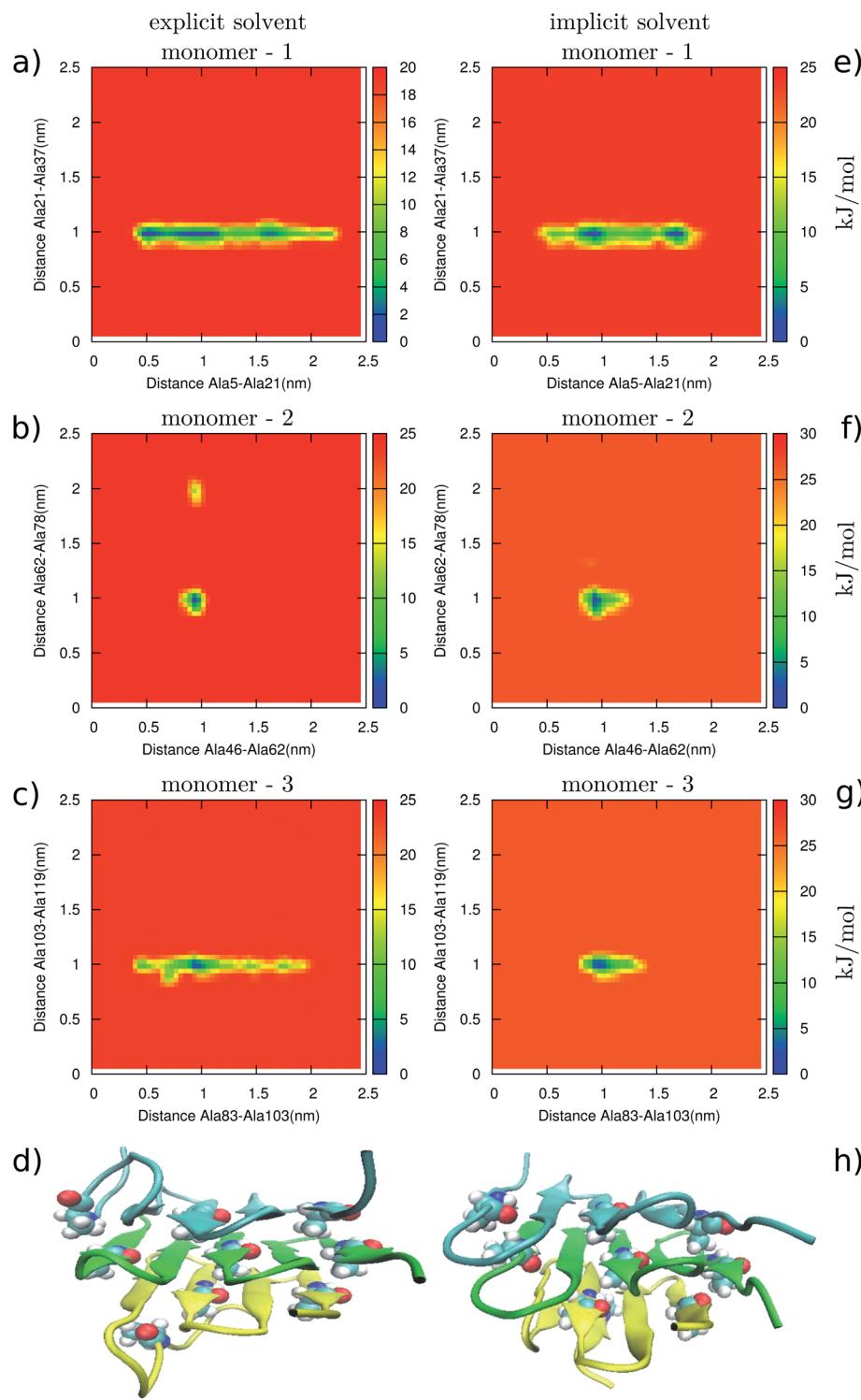


Fig. 6 Free energy landscape of a trimer as a function of the distance between Ala5 and Ala21 and between Ala21 and Ala37 for the first monomer, distance between Ala46 and Ala62 and between Ala62 and Ala78 for the second monomer, between Ala83 and Ala103 and between Ala103 and Ala119 for the third monomer obtained from simulations starting from a β -sheet structure. (a)–(c) Free energy obtained from simulation with explicit solvent 30 ns per replica, and (e)–(g) implicit solvent simulations 30 ns per replica simulations. Snapshots of a trimer after 30 ns per replica simulations: (d) explicit solvent simulation (h) implicit solvent simulation.

hydrogen bonds between neighbouring monomers, $\overline{HB}_{\text{inter}}$, the average root mean square deviation from the initial structure (except for the molten trimer where we compute the RMSD with

respect to the initial structure of trimer β -sheet), $\overline{\text{RMSD}}$ and R_m , which is the replica corresponding to the lowest temperature, T_m , at which the β -sheet or β -roll structures are molten. To

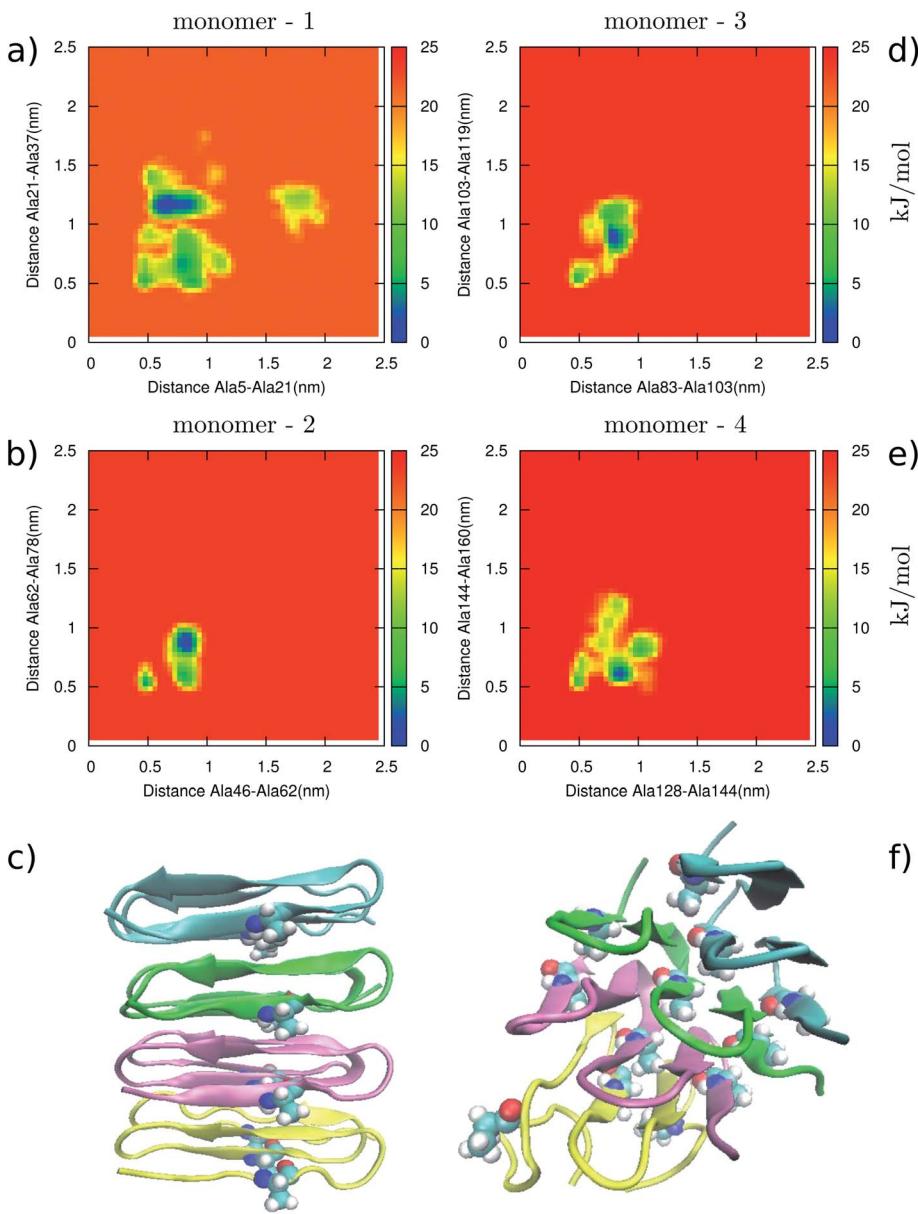


Fig. 7 Free energy landscape of a tetramer as a function of the distance between Ala5 and Ala21 and between Ala21 and Ala37 for the first monomer, distance between Ala46 and Ala62 and between Ala62 and Ala78 for the second monomer, between Ala83 and Ala103 and between Ala103 and Ala119 for the third monomer, between Ala128 and Ala144 and between Ala144 and Ala160 for the fourth monomer obtained from simulations starting from a β -roll structure. (a), (b), (d) and (e) Free energy obtained from simulation with implicit solvent 50 ns per replica simulations. Snapshots of a tetramer: (c) initial structure of a tetramer and (f) a final structure after 50 ns per replica simulations. Notice the interprotein β -sheet structure.

determine whether a structure is molten or not we compute RMSD at each temperature. The data is shown in Table 1.

The average number of hydrogen bonds per monomer, \overline{HB}/N , increases as a function of number of monomers. This is because for assemblies that contain more than one monomer hydrogen bonds form also between monomers. For a given number of monomers, \overline{HB}/N has approximately the same value for β -sheets and β -rolls. In contrast, the average number of hydrogen bonds between two neighbouring monomers of trimers and tetramers, $\overline{HB}_{\text{inter}}$, is larger for those that contain β -rolls compared to β -sheets. This is because β -rolls form inter-

molecular β -sheets. Because the total number of hydrogen bonds is the same for these two structures, this also means that the number of hydrogen bonds within β -roll monomers in the trimeric and tetrameric states is less than that within the β -sheet monomers in these states. For the case of dimers, $\overline{HB}_{\text{inter}}$ is the same for both the simulations starting from β -sheets or β -rolls because they both form globular structures in which the two monomers are entangled.

The relative stability of β -rolls and β -sheets can also be compared in terms of the strength of hydrophobic interactions. To study this, we calculated the non-polar contribution of the

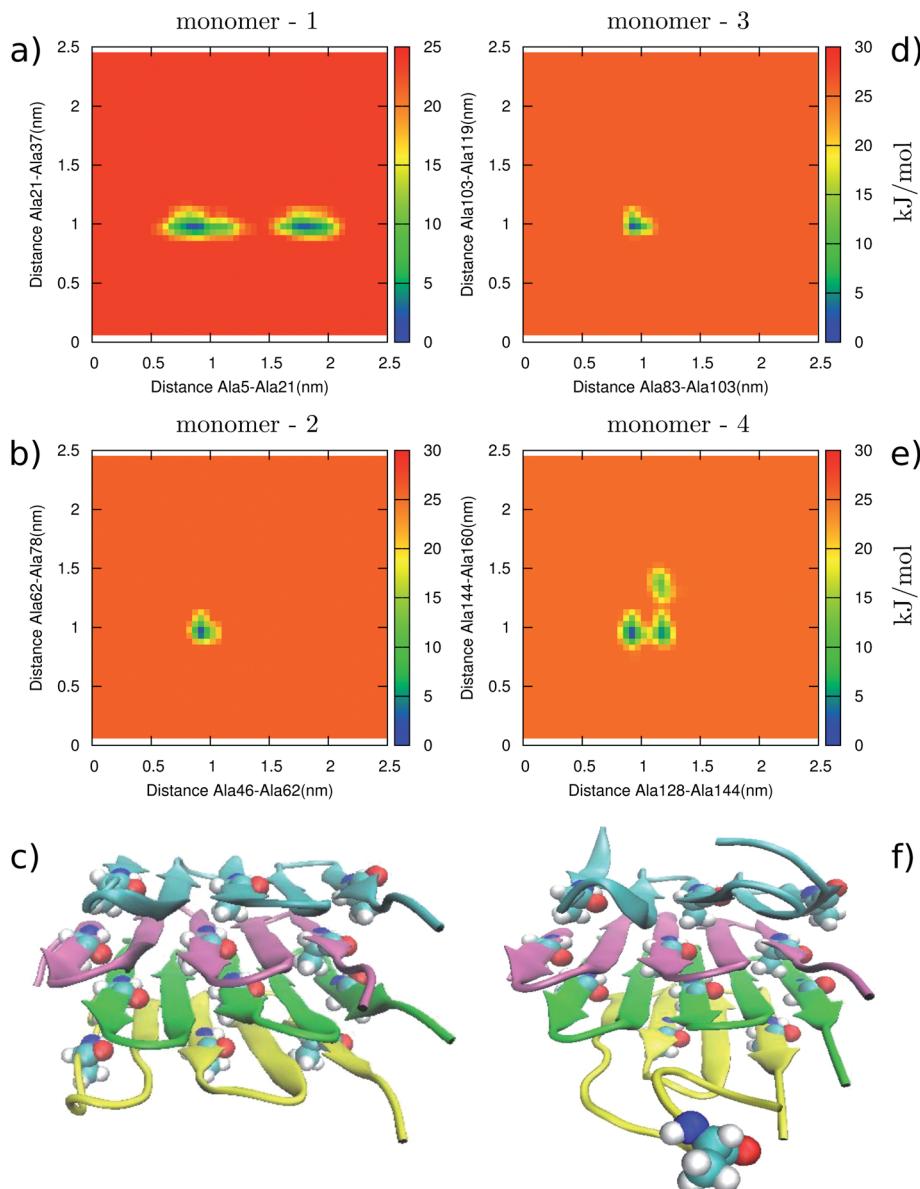


Fig. 8 Free energy landscape of a tetramer as a function of the distance between Ala5 and Ala21 and between Ala21 and Ala37 for the first monomer, distance between Ala46 and Ala62 and between Ala62 and Ala78 for the second monomer and between Ala83 and Ala103 and between Ala103 and Ala119 for the third monomer, between Ala128 and Ala144 and between Ala144 and Ala160 for the fourth monomer obtained from simulations starting from a β -sheet structure. (a), (b), (d), (e) Free energy obtained from simulation with implicit solvent 50 ns per replica simulations. Snapshots of a tetramer: (c) initial structure of a tetramer and (f) a final structure after 50 ns per replica simulations. Notice the inter-protein β -sheet structure.

solvation free energy divided by the number of monomers averaged over simulations time, G_{np}/N . As the number of particles increases, G_{np}/N decreases because by increasing the number of monomers the volume of the aggregates grows faster than their surface and therefore for larger assemblies more hydrophobic residues are buried inside the assembly. These hydrophobic interactions compete with conformational entropy gain of aggregates forming a globular random structure. For the case of trimers and tetramers, the hydrophobic interactions win the competition whereas for the case of dimers and monomers conformational entropy wins.

The average root mean square deviation from the initial structure, $\overline{\text{RMSD}}$, for the case of trimers and tetramers is much smaller than that of dimers and monomers. This is because trimer and tetramers do not melt during the simulations. For both trimers and tetramers, $\overline{\text{RMSD}}$ for β -sheets is smaller than that of β -rolls because β -sheets mostly maintain their initial conformation during the simulations whereas β -rolls slightly change their conformations to form inter-molecular β -sheets.

The melting temperature, T_m , of the protein structures can be used as a measure of their relative stability. In our REMD simulations the melting temperature of tetramers is higher than that of trimers. This indicates that as these aggregates

Table 1 The properties of different protein structures in implicit solvent. \overline{HB}/N is the average number of hydrogen bonds per monomer. $\overline{HB}_{\text{inter}}$ is the average number of hydrogen bonds between neighbouring monomers. RMSD is the average root mean square deviation of a monomer from its initial structure (for the trimers and tetramers we calculate RMSD for the monomer that is in the middle of the aggregates). R_m is the replica corresponding to the lowest temperature, T_m , at which the β -sheet or β -roll structures are molten. G_{np}/N is the non-polar contribution of the solvation free energy

Initial structure	\overline{HB}/N	$\overline{HB}_{\text{inter}}$	RMSD (nm)	R_m	T_m (K)	G_{np}/N (kJ mol ⁻¹)
Monomer sheet	17.0	0	1.00	—	—	43.98
Monomer roll	17.0	0	0.93	—	—	42.24
Dimer sheet	18.5	10	0.90	—	—	33.58
Dimer roll	18.5	10	0.72	—	—	33.23
Trimer sheet	20.0	7	0.26	8	443	29.16
Trimer roll	20.0	13	0.34	5	383	28.17
Molten trimer	20.0	5	0.28	—	—	28.44
Tetramer sheet	21.8	6	0.16	10	488	24.40
Tetramer roll	20.3	11	0.34	9	465	24.67

grow in size they become more stable in solution. Interestingly, for both trimers and tetramers, T_m is higher for the case of aggregates that contain β -sheets compared to those that contain β -rolls although the difference between the melting temperature of β -rolls and β -sheets is smaller for the case of the tetramers.

We also perform 50 ns per replica REMD simulations starting with three unfolded monomers that we put in close proximity of each other in the implicit solvent. The proteins were positioned such as to be able to form a linear trimer structure. We found the monomers to move closer together to form hydrogen bonds among them and to reduce their overall solvent accessible surface area. We did not see any beta-sheet formation during our simulations. Also, the final product was no longer linear but resembled a triangle. This is in our view exactly because the folded structure is stabilized by inter-molecular interactions that are not there unless all proteins have switched conformation. The \overline{HB}/N , $\overline{HB}_{\text{inter}}$, RMSD and G_{np}/N calculated for this system is shown in Table 1.

We can also compare the relative stability of trimers with different starting configurations. The total number of hydrogen bonds, \overline{HB} , is the same for these structures. However, for the case of β -rolls the number of hydrogen bonds between two monomers is larger than that of β -sheets and molten structure. It seems that the number of inter and intra monomer hydrogen bonds are exchanged for the β -sheets and β -rolls.

IV. Conclusion

We performed extensive replica exchange molecular dynamics simulations in order to investigate the stability of a protein aggregate that contains one, two, three or four copies of a silk-like protein. For each of these protein aggregates we carried out simulations starting from two initial states: aggregates made up of β sheets or β rolls. We first performed short simulations with

explicit and implicit water models to test if we obtain the same results from these two models. Next, we extended our simulations with implicit solvent model to much longer time scales to obtain consistent results from simulations with different initial states. We found that monomers and dimers melt and form a globular structure. Interestingly, the structure of trimers and tetramers does not change significantly during our simulations and we observed that inter- and intra-molecular β -sheets form. One of the differences between the inter- and intra-molecular β -sheets is that the width of the former is smaller and therefore thickness of the fibers formed by these type of β -sheets must be smaller. This could explain the thickness of aggregates in the experiments of Martens and co-workers where the fiber dimensions do not match with the β -sheet structure.⁴³ From our simulations we also find that because the conformation of this silk-like protein changes from a random globular state to a β -sheet structure upon binding in an aggregate of three or more monomers, the self-assembly process of this protein must be a nucleated process. This is in line with the recent experimental observations of Martens and co-workers.⁴³ Our simulation approach can be used in future studies for predicting relative stability of different fiber forming polypeptides and can help the design of new materials.

Acknowledgements

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 Wetenschappelijk Onderzoek (Netherlands Organization for Scientific Research, NWO).

References

- 1 A. Lakshmanan, S. Zhang and C. A. Hauser, *Trends Biotechnol.*, 2012, **30**, 155–165.
- 2 V. F. Segers and R. T. Lee, *Drug Discovery. Today*, 2007, **12**, 561–568.
- 3 P. Ringler and G. E. Schulz, *Science*, 2003, **302**, 106–109.
- 4 K. Rajagopal and J. P. Schneider, *Curr. Opin. Struct. Biol.*, 2004, **14**, 480–486.
- 5 U. B. Sleyter, B. Schuster and D. Pum, *IEEE Eng. Med. Biol. Mag.*, 2003, **22**, 140–150.
- 6 S. Zhang, *Nat. Biotechnol.*, 2003, **21**, 1171–1178.
- 7 J. C. M. van Hest and D. A. Tirrell, *Chem. Commun.*, 2001, **19**, 1897–1904.
- 8 N. C. Seeman and A. M. Belcher, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 6451–6455.

- 9 N. C. Seeman, *Nature*, 2003, **421**(6921), 427–431.
- 10 R. P. Goodman, I. A. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt and A. J. Turberfield, *Science*, 2005, **310**, 1661–1665.
- 11 J. Kisiday, M. Jin, B. Kurz, H. Hung, C. Semino, S. Zhang and A. J. Grodzinsky, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9996–10001.
- 12 M. Shimomura and T. Sawadaishi, *Curr. Opin. Colloid Interface Sci.*, 2001, **6**, 11–16.
- 13 K. A. Scott, P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid and M. S. Sansom, *Structure*, 2008, **16**, 621–630.
- 14 E. Kokkoli, A. Mardilovich, A. Wedekind, E. L. Rexeisen, A. Garg and J. A. Craig, *Soft Matter*, 2006, **2**, 1015–1024.
- 15 G. W. Vandermeulen and H. A. Klok, *Macromol. Biosci.*, 2004, **4**, 383–398.
- 16 A. Aggeli, I. A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. B. McLeish and N. Boden, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11857–11862.
- 17 S. Zhang, D. M. Marini, W. Hwang and S. Santoso, *Curr. Opin. Chem. Biol.*, 2002, **6**, 865871.
- 18 D. M. Marini, W. Hwang, D. A. Lauffenburger, S. Zhang and R. D. Kamm, *Nano Lett.*, 2002, **2**, 295–299.
- 19 S. Vauthhey, S. Santoso, H. Gong, N. Watson and S. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5355–5360.
- 20 A. P. Nowak, V. Breedveld, L. Pakstis, B. Ozbas, D. J. Pine, D. Pochan and T. J. Deming, *Nature*, 2002, **417**, 424–428.
- 21 S. W. Lee, C. Mao, C. E. Flynn and A. M. Belcher, *Science*, 2002, **296**, 892–895.
- 22 J. D. Hartgerink, E. Beniash and S. I. Stupp, *Science*, 2001, **294**, 1684–1688.
- 23 R. C. Merkle, *Trends Biotechnol.*, 1999, **17**, 271–274.
- 24 S. Zhang, *Nat. Biotechnol.*, 2003, **21**, 1171–1178.
- 25 I. A. Nyrkova, A. N. Semenov, A. Aggeli and N. Boden, *Eur. Phys. J. B*, 2000, **17**, 481–497.
- 26 S. Zhang, T. Holmes, C. Lockshin and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 3334–3338.
- 27 S. Zhang, D. Marini, W. Hwang and S. Santoso, *Curr. Opin. Chem. Biol.*, 2002, **6**, 865871.
- 28 M. R. Caplan, E. M. Schwartzfarb, S. Zhang, R. D. Kamm and D. A. Lauffenburger, *Biomaterials*, 2002, **23**, 219–227.
- 29 W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz and D. A. Tirrell, *Science*, 1998, **281**, 389–392.
- 30 D. M. Marini, W. Hwang, D. A. Lauffenburger, S. Zhang and R. D. Kamm, *Nano Lett.*, 2002, **2**, 295–299.
- 31 J. Kopecek and J. Yang, *Acta Biomater.*, 2009, **5**, 805.
- 32 S. M. Douglas, H. Dietz, T. Liedl, B. Hgberg, F. Graf and W. M. Shih, *Nature*, 2009, **459**, 414–418.
- 33 L. Pauling, R. B. Corey and H. R. Branson, *Proc. Natl. Acad. Sci. U. S. A.*, 1951, **37**, 205–211.
- 34 Y. Loo, S. Zhang and C. A. Hauser, *Biotechnol. Adv.*, 2012, **30**, 593–603.
- 35 R. V. Ulijn and A. M. Smith, *Chem. Soc. Rev.*, 2008, **37**, 664–675.
- 36 J. A. Wallace and J. K. Shen, *Biochemistry*, 2010, **49**, 5290–5298.
- 37 A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz and L. Serrano, *Nat. Biotechnol.*, 2004, **22**, 1302–1306.
- 38 D. L. Caspar, *Biophys. J.*, 1980, **32**, 103–138.
- 39 P. G. Stockley, O. Rolfsson, G. S. Thompson, G. Basnak, S. Francese, N. J. Stonehouse, S. W. Homans and A. E. Ashcroft, *J. Mol. Biol.*, 2007, **369**(2), 541–552.
- 40 G. Bitan, A. Lomakin and D. B. Teplow, *J. Biol. Chem.*, 2001, **276**, 35176–35184.
- 41 G. Bitan, B. Tarus, S. S. Vollers, H. A. Lashuel, M. M. Condron, J. E. Straub and D. B. Teplow, *J. Am. Chem. Soc.*, 2003, **125**, 15359–15365.
- 42 B. Urbanc, L. Cruz, F. Ding, D. Sammond, S. Khare, S. V. Buldyrev and N. V. Dokholyan, *Biophys. J.*, 2004, **87**, 2310–2321.
- 43 A. A. Martens, G. Portale, M. W. Werten, R. J. de Vries, G. Eggink, M. A. Cohen Stuart and F. A. de Wolf, *Macromolecules*, 2009, **42**, 1002–1009.
- 44 M. Schor, A. A. Martens, M. A. C. Stuart and P. G. Bolhuis, *Soft Matter*, 2009, **5**, 2658–2665.
- 45 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 46 M. D. Yoder and F. Jurnak, *FASEB J.*, 1995, **9**, 335–342.
- 47 B. Roux and T. Simonson, *Biophys. Chem.*, 1999, **78**, 1–20.
- 48 T. Wang and R. C. Wade, *Proteins*, 2003, **50**, 158169.
- 49 L. Y. Zhang, E. Gallicchio, R. A. Friesner and R. M. Levy, *J. Comput. Chem.*, 2001, **22**, 591–607.
- 50 R. Zhou and B. J. Berne, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12777–12782.
- 51 R. Zhou, B. J. Berne and R. Germain, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 14931–14936.
- 52 R. Zhou, *Proteins: Struct. Funct. Bioinf.*, 2003, **53**, 148–161.
- 53 Y. M. Rhee, E. J. Sorin, G. Jayachandran, E. Lindahl and V. S. Pande, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 6456–6461.
- 54 M. D. Struthers, R. P. Cheng and B. Imperiali, *Science*, 1996, **271**, 342–345.
- 55 B. D. Bursulaya and C. L. Brooks, *J. Phys. Chem. B*, 2000, **104**, 12378–12383.
- 56 R. SalomonFerrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **3**, 198–210.
- 57 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz and R. J. Woods, *Comput. Chem.*, 2005, **26**, 1668–1688.
- 58 J. W. Ponder and D. A. Case, *Adv. Protein Chem.*, 2003, **66**, 27–85.
- 59 M. Schaefer, C. Bartels and M. Karplus, *J. Mol. Biol.*, 1998, **284**, 835–848.
- 60 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.
- 61 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 62 M. K. Prakash, A. Barducci and M. Parrinello, *J. Chem. Theory Comput.*, 2011, **7**, 2025–2027.
- 63 L. Prieto and A. Rey, *J. Chem. Phys.*, 2009, **130**, 115101.