Accepted Manuscript

Establishment of Constraints on Amyloid Formation Imposed by Steric Exclusion of Globular Domains

Rafayel A. Azizyan, Adriana Garro, Zinaida Radkova, Alexey Anikeenko, Anastasia Bakulina, Christian Dumas, Andrey V. Kajava

PII: S0022-2836(18)30533-3
DOI: doi:10.1016/j.jmb.2018.05.038
Reference: YJMBI 65747

To appear in: *Journal of Molecular Biology*

Received date: 19 March 2018
Revised date: 15 May 2018
Accepted date: 27 May 2018

Please cite this article as: Rafayel A. Azizyan, Adriana Garro, Zinaida Radkova, Alexey Anikeenko, Anastasia Bakulina, Christian Dumas, Andrey V. Kajava , Establishment of Constraints on Amyloid Formation Imposed by Steric Exclusion of Globular Domains. Yjmbi (2018), doi:10.1016/j.jmb.2018.05.038

Establishment of constraints on amyloid formation imposed by steric exclusion of globular domains

Rafayel A. Azizyan[a,b], Adriana Garro[c], Zinaida Radkova[d], Alexey Anikeenko[d], Anastasia Bakulina[d], Christian Dumas[e] and Andrey V. Kajava[a,b*]

[a] Centre de Recherche en Biologie cellulaire de Montpellier, UMR 5237 CNRS, Université Montpellier, Montpellier, France,

[b] Institut de Biologie Computationnelle, Université Montpellier, Montpellier, France,

[c] Universidad Nacional de San Luis IMASL-CONICET, San Luis, Argentina,

[d] Novosibirsk State University, Novosibirsk, Russia,

[e] Centre de Biochimie Structurale, CNRS, Montpellier, France

***** To whom correspondence should be addressed: andrey.kajava@crbm.cnrs.fr,

Tel. +33 4 34359538

Keywords

aggregation, amyloids, fibrillogenesis, protein structure, rigid body simulation

Abbreviations

3D, three-dimensional; AR, amyloidogenic region; GFP, Green Fluorescent Protein; MD, molecular dynamics

ABSTRACT

In many disease-related and functional amyloids, the amyloid-forming regions of proteins are flanked by globular domains. When located in close vicinity of the amyloid regions along the chain, the globular domains can prevent the formation of amyloids because of the steric repulsion. Experimental tests of this effect are few in number and non-systematic, and their interpretation is hampered by polymorphism of amyloid structures. In this situation, modelling approaches that use such a clear-cut criterion as the steric tension can give us highly trustworthy results. In this work, we evaluated this steric effect by using molecular modelling and dynamics. As an example, we tested hybrid proteins containing an amyloid-forming fragment of Aβ peptide (17-42) linked to one or two globular domains of GFP. Searching for the shortest possible linker we constructed models with pseudo-helical arrangements of the densely packed GFPs around the Aβ amyloid core. The molecular modelling showed that linkers of 7 and more residues allow fibrillogenesis of the Aβ-peptide flanked by GFP on one side and 18 and more residues when Aβ-peptide is flanked by GFPs on both sides. Furthermore, we were able to establish a more general relationship between the size of the globular domains and the length of the linkers by using analytical expressions and rigid body simulations. Our results will find use in planning and interpretation of experiments, improvement of the prediction of amyloidogenic regions in proteins, and design of new functional amyloids carrying globular domains.

INTRODUCTION

Amyloid fibrils are linked to a broad range of human diseases [1]. In the last decade, a growing body of evidence demonstrated that in some organisms amyloid structures could also play important “beneficial” functional roles [2–7]. Typically, the amyloid fibrils consist of multiple copies of the same protein. In most cases, the amyloidogenic regions (ARs) have two structural states: (i) an unfolded conformation before the aggregation and (ii) cross-β conformation after the fibril-formation. The majority of disease-related and naturally occurring functional amyloids have a structure with β-strands that are stacked one over the other in a parallel and in-register manner [8–12].

Usually, the size of the ARs does not exceed 100 residues [13] while over 90% of proteins are longer than 100 residues (UniProt Release 2017_03). Thus, in the majority of cases, the ARs coexist with the other sequence motifs including intrinsically unstructured regions and structured domains (Fig. 1). Among the examples of proteins containing both ARs and globular structures are Sup35, Ure2p, RIP1/RIP3, Prp, huntingtin, Het-s proteins, Pmel-17, TAR DNA-binding protein [6,14–21]. In some cases, for example, in α-synuclein and Aβ peptide, the ARs are flanked only by unstructured regions [22,23]. It is apparent that flanking regions of ARs can affect the formation and the stability of cross-β amyloid. For example, the effect of domains flanking the polyglutamine tract on structure and toxicity of huntingtin was described in several papers [21,24]. However, despite considerable interest and importance of this issue, the interplay between these regions remains poorly understood.

In general, we can consider five major situations in AR-containing proteins (Fig. 1). (1) The amyloidogenic region (AR) completely covers the sequence of a protein (Fig. 1a). (2) If an AR has very long unfolded flanking regions (Fig. 1b), their motion should slow down the amyloidogenesis due to the increase of the entropic barrier [25]. (3) The similar decelerating

effect is expected for proteins with ARs that are linked to globular structures by long unfolded region (Fig. 1c). (4) If the structures flanking AR tend to interact with each other, this can facilitate the amyloidogenesis (Fig. 1d). (5) Finally, in proteins with a very short linker between AR and globular domain, the steric repulsion of the globular structures can prevent formation the amyloid fibrils (Fig. 1e). It is especially true for the cross-β structures of amyloid fibrils with the parallel and in-register arrangements of β-strands. It is important to keep in mind that ARs can be located not only in the intrinsically unstructured regions but also can be hidden within the natively folded structures. In the latter case, the ARs can be exposed to the solvent and form cross-β amyloids because of the destabilized effect of amino acid mutations or external conditions [26,27]. The remaining part of the native structure can be either unfolded, corresponding to the situation in Fig. 1b, or continue to form the stable structure fitting one of the cases shown in Fig. 1c, d, e.

In this work, we focused on the analysis of the steric repulsion of the globular structures which can prevent the formation of the parallel and in-register cross-β amyloid fibrils (Fig. 1e). The experimental examination of these effects is hampered by high polymorphism of amyloid structures formed by the AR with the same amino acid sequence depending on the conditions [4,28–31]. A majority of the studied disease-related and functional amyloid fibrils have a recurring "β-arcade" structure produced by stacking of β-strand-β-arc-β-strand elements called 'β-arches' (reviewed in [8]). In a β-arch two β-strands interact via their side chains, not via H-bonds of the polypeptide backbone as in a conventional β-hairpin. The parallel and in-register cross-β amyloid fibrils can have either a single β-arcade or an arrangement with several adjacent β-arcades [8,11–13,23]. The polymorphism brings uncertainty in our knowledge about the beginning and end of the AR and linker in a given experiment. In this situation, theoretical modelling approaches are becoming essential to analyze the steric effects, because in the models one can stipulate and fix the boundaries of the AR, linker and globular domain. Correct

theoretical estimation of the attraction between globular domains or the entropic effect of the flanking regions is a difficult task, but the steric tension is so clear-cut that the modelling approaches can give us highly trustworthy results. Earlier, we brought attention to the steric effect of the globular domains on amyloidogenesis [32]. Though this phenomenon would intuitively seem to be apparent, the exact relationship between the size of the globular domain, the length of the linker between this globular structure and ARs, which would allow fibrillogenesis, are unknown. As a result of this ignorance, some experimental results may be misinterpreted. Furthermore, the knowledge of constraints on amyloid formation imposed by the steric exclusion of globular domains must be used in combination with the methods for the prediction of ARs based on their amino acid sequences. Indeed, if the predicted AR is located in close vicinity of the globular domain, it is not clear whether this region will assemble or not in reality. Thus, the goal of this work was to evaluate the steric effect of the neighboring globular structures on amyloid formation by using molecular modelling, dynamics, and mesoscopic rigid body models.

RESULTS AND DISCUSSION

Choice of a model

To demonstrate the constraints on amyloid formation imposed by steric exclusion of globular domains we used a hybrid protein containing an amyloid-forming Aβ(17-42) peptide and *Aequorea* Green Fluorescent Protein (GFP) as a globular domain (Fig. 2). The choice of Aβ peptide is explained by the fact that this is one of the most studied amyloid-forming peptides with well-established 3D structures of the amyloid fibrils. Depending on the conditions, this peptide forms different structures consisting of stacks of β-arches [23] or the other β-arc-containing blocks [12,33–35]. For our structural model, we selected one of these structures

consisting of Aβ(17-42) β-arches [23]. As concerns the globular structure, the GFP domain is the most frequently fused to different amyloid-forming or aggregating fragments. For example, Waldo and colleagues showed that in several hybrid proteins folding of GFP is sensitive to the aggregation of the fused proteins [36]. Wurth and colleagues have studied aggregation of mutated variants of Aβ42-peptide fused with GFP [37]. Fox and colleagues have used the similar approach to study the amyloid formation of amylin (also known as islet amyloid polypeptide) [38]. Recently, GFP molecule has also been used as a reporting protein in fusion constructs with Aβ peptides for visualizing Aβ oligomers *in vivo* [39]. However, in these and other studies [40–43], the impact of linker length on oligomerization or fibril formation was not evaluated. Although some GFPs have a weak dimerization tendency, if necessary, monomeric GFPs can be easily obtained by the dimer interface breaking mutations [44]. In addition, GFP represents a good choice to extend the conclusions on the other types of globular domains, because GFP with its 238 residues is close to the average size of the globular domain (190 residues in MODBASE) [45].

In our model, the linker between these domains contains repetitive sequence of small amino-acids $(Ser\text{-}Ser\text{-}Pro)_n$ which has a potential of flexible and intrinsically unfolded conformation. Ser residue is hydrophilic and flexible, and Pro residue lacks amide hydrogen preventing the formation of the secondary structures, leading to unfolded conformations. The analysis of the 3D structure of GFP showed that the first residue (Ser) does not have any internal structural constraints and, therefore, this residue of GFP was assigned to the flexible linker (Fig. 2a).

In general, the cross-β amyloid fibrils can vary in the number of protofibrils [8,46,47]. Here, the term “protofibril” is used for a structurally independent fibrillar entity that usually consists of one peptide (or protein) per increment. It is important to mention that the known amyloid domains fused with globular domains form single-protofibrils because the subsequent association of the protofibrils is inhibited by these globular domains surrounding the protofibril [29]. Therefore, in

our model, the Aβ(17-42) fragments fused with GFP form a structure with a single-protofibril. In this amyloid structure, each polypeptide chain is folded into β-arches and stacked in a parallel and in-register manner to form a β-arcade (see Methods).

Establishment of the minimal length of the linker that allows the formation of Aβ(17-42)-linker-GFP amyloid fibrils

To find the shortest possible linker for the infinite amyloid fibrils of Aβ(17-42)-linker-GFP molecule we used the following consideration. The shorter is the linker the closer is the packing of GFPs around the core of the fibril. In this case, the densest packing of GFPs can be achieved by a regular pseudo-helical arrangement of the GFP domains. Furthermore, the size of the GFP protein and the location of its N-termini allowed us to determine the optimal orientation and position of the GFP (Fig. 2) relative to the cross-β structure of the amyloid core. This fact, in turn, made possible to determine the number of GFPs in the repetitive unit, which, by translation along the fibril axis can generate infinite fibril. For a diameter of about 30 Å for the β-barrel of GFP, this repetitive unit should contain ~30 Å /4.8 Å = 6 monomers, where 4.8 Å corresponds to the monomer axial rise (Fig. 2b).

In the axial projection, these six GFPs and the core of the fibril should wind around an axis crossing the last residue of the amyloidogenic region (residue 42 of the Aβ-fragment) (Fig. 2c). The globular structures form a helical arrangement with the monomer axial rise of 4.8 Å and the angle of rotation equal to about 51° (360°/(6 GFPs + Aβ core). Following this arrangement, the repetitive unit of 6 monomers was constructed by molecular modelling and manual adjustment to avoid strong overlaps of the monomers. Then, this hexamer unit was multiplied along the fibril axis with a translation step of 28.8 Å, i.e. 6 x 4.8 Å, to generate an infinite fiber. Having obtained

this structure, we can generate the corresponding atomic model for the shortest possible linker connecting residues 42 of Aβ to the first residue of the corresponding GFP.

It is known that the amyloid/prion protofibrils usually have a slight left-handed twist [29,48]. This axial twist per 4.8-Å step of the stack is the same for a given fibril and range from about 0.5° to 4° in different fibrils. Therefore, we also constructed these hybrid amyloid fibrils with the twist angle at the mean value of 2° (Fig. 2d) using a procedure described in Methods section. The analysis of both flat and twisted fibrils led to the conclusion that the twist does not affect the results and the shortest possible linker of infinite Aβ-GFP fibril has 7 residues. Obviously, in reality, the number of residues in the shortest possible linker may be slightly bigger due to the imperfection of the GFP packing within the cylinder and larger dimensions of the GFP domain due to the ordered shell of water around the protein structures.

Our conclusions about the GFP-containing fibrils can be extended to a big number of the other globular structures as the size and shape of GFP structures represents the most typical case [45]. However, it is still desirable to be able to establish the relationship between the length of the linkers and the size/shape of any globular domains. The molecular modelling is not appropriate in this general case as it requires an insurmountable amount of work. In the next sections, we will first overview situations with possible different sizes of the globular structures then we will describe mesoscopic rigid body simulation procedure that can provide a rapid and reliable estimation of constraints on amyloid formation imposed by any type of the globular structures.

Extension of the conclusions to a general case by using analytical representation

Simplification of the model may allow extending our conclusions to molecules having globular structures of any sizes. For example, the globular structures can be approximated by spheres,

spheroids or ellipsoids. In the fibril formed by an AR with a linker and globular structure on one side, the spheres having minimal linkers form densely packed pseudo-helical arrangement. The number of the spheres (n) in one coil of this pseudo-helical arrangement can be calculated as $n=d/4.8$, where d is a diameter of the sphere in Å and 4.8 Å corresponding to the monomer axial rise (Fig. 3). The amyloid core can be represented by axially stacked rectangular parallelepipeds. Considering the Aβ(17-42) amyloid core with dimensions 22 Å x 40 Å x 4.8 Å depending on the diameter of the spheres we can distinguish 3 cases. First, when the diameter of the spheres is less than 40 Å, the core of the amyloid fibril will be a part of the cylinder formed by the circular arrangement of congruent spheres (Fig. 3). In this case, the length of the minimal liker (L) can be expressed by the equation: $L=(d\times n+W)/2\pi-d/2$, where $d\times n+W$ is a circumference of the circle passing through the centers of the congruent spheres, and W is the width of the amyloid core (Fig. 3). It is, for example, the case for the GFP with about 30 Å in diameter. The graph in Figure 3 has W = 22 Å, as in the amyloid core of the Aβ(17-42)-linker-GFP molecule. To obtain the number of amino acid residues in the linker, we divide L to 3.15 Å that corresponds to the average length per residue of the unfolded polypeptide chain. It was interesting to compare the linker lengths obtained from our molecular modelling for the Aβ(17-42)-linker-GFP molecule with the ones derived from the previous analytical expression. The modelling proposes 7 residues in the minimal linker while the calculated number is 6 residues. The difference can be explained by the fact that the modelled linkers cannot adopt any extended conformation due to the covalent and steric limitations in the polypeptide chain. The difference in the shape of the real GFP domain and its spherical approximation may also be an origin of some discrepancy.

Second, if the diameter of the spheres is over 40 Å, the number of the densely-packed spheres in the cylinder coil increases to 8 or more and, at the minimal linkers, the cross-β core of the amyloid can completely enter the space inside of the cylinder of the densely-packed spheres

(Fig. 3). In this situation, the width of the amyloid core (*W*) disappears from the equation and this leads to a formula $L=d\times n/2\pi-d/2$. Finally, when the diameter of the spheres is more than 60 Å, some spheres can enter in the linker space inside of the cylinder, and this makes the situation with the analytical estimations of the packing more complicated (Fig. 3). In the next section, we will describe an approach to mesoscopic modelling that can be used in this case.

Rigid body simulation by using Low Poly 3D models

The atomic resolution modelling can provide the most reliable answer on the relationship between the size of the globular domains and the length of the minimal linkers in the amyloid fibrils (see previous sections). However, this approach is time and resource consuming. The analytical expression derived in the previous section can give a general estimate about the situation with globules of different sizes. However, this estimation is approximate and needs to be verified by the other approaches. In this section, we analyzed several examples of amyloid fibrils with the globules by rigid body simulation using Low Poly 3D models implemented in Blender package (www.blender.org).

The amyloid core Aβ(17-42) was approximated by axially stacked and left-hand twisted (2°) rectangular parallelepipeds with dimensions of 22 x 40 x 4.8 units (Fig. 4). In the hybrid proteins containing Aβ(17-42)-linker-globular domain, the linker was built of the chain of small spheres, and the globular structure was represented as a bigger sphere. In different tested models the diameter of the globular domain was either 30, 40, 50 or 60 units (for details see Methods and https://ab-linker-globula.github.io/). The modelling showed that the minimal linker for the globule-sphere of 30 units is 8 residues, for 40 units is 12, for 50 units is 16 and for 60 units is 21 residues. The established minimal linkers for the analyzed models were mapped on the graph with the analytical dependencies (Fig. 3). The relationship between the diameters of the

spheres and the linker lengths is almost linear. In general, the comparison demonstrated a good fit with the results obtained by the other approaches. At the same time, the linkers for the smaller globular structures of 30 and 40 Å that were obtained by the rigid body simulations is slightly longer (8 and 12 residues) than the ones deduced from molecular modelling of GFPs (7 residues) and analytic expressions (6 and 11 residues). The visual analysis of the fibrils generated by Blender showed that the globular structures are not packed in a perfectly regular manner (Fig. 4c). This observation suggests that with the short linkers allowing only one layer of globular domains around the protofibril, we do not reach the optimal packing due to the limited space, probably leading to the "kinetic traps". Importantly, for the larger globular domains (50 and 60 Å), the one layer analytical estimations are not appropriate anymore. In these cases, the packed globular structures are located at different distances from the protofibril (Fig. 3). For these molecules, the mesoscopic modelling is becoming the only suitable approach to evaluate the packing and the minimal linkers. Thus, the mesoscopic modelling demonstrates its efficiency to rapidly get a reasonable estimation of the structure of amyloid fibrils with globular domains of different size and shapes.

Establishment of the minimal linkers that allows the formation of GFP-linker-Aβ(17-42)-linker-GFP amyloid fibrils

In the previous sections, we analyzed molecules with the AR flanked by a single C-terminal globular domain. A similar relationship between the linkers and the size of the globules is expected for the globular domains connected N-terminally to the amyloid region. At the same time, the ARs can be located between two globular domains of a protein. By using a molecular modelling approach, we analyzed the Aβ(17-42) fragment containing two flanking GFP domains at both sides. There are two distinct cases for this type of hybrid molecules: (1) one linker is

significantly longer than another one, (2) two linkers are of a similar length.

Our modelling showed that in the first case the length of the minimal shorter linker did not change much in comparison with the molecule having GFP on one side (see the previous section). The GFPs with the long linker are located far apart from the fibril and its closely packed GFPs with the minimal linker. The only difference is that the longer linkers go throughout the pores within these packed GFPs and slightly increase the density of the packing (Fig. 5a). Modelling showed that the more dense packing might lead to a one-residue increase of the minimal short linker (from 7 to 8 residues). The minimal length of the long linker (for example, at the N-terminus of Aβ) is determined by the steric clash between the N- and C-terminal GFPs, not between the N-terminal GFPs. The minimal long linker can be established by reducing it until the N-terminal GFPs start to clash with the layer of the C-terminal GFPs (Fig. 5a). The modelling showed that the minimal long linker is 24 residues, taking into account residues between the Cα-atom of the C-terminal GFP residue and the Cα-atom of Aβ residue 17 in the amyloid core.

In the second case, when two linkers are of a similar length, the GFPs from both the C-termini and the N-termini will be located at more or less the same distance from the center of the fibril (Fig. 5b) and as a result, they may interact with each other side by side. Therefore, at the minimal lengths of the linkers, all GFPs together will be densely packed around the amyloid core. The modelling showed that the free space between the core and the GFP-wall is not large enough to accommodate at least one GFP. It means that all GFPs must be located within the closely packed layer (Fig. 5b). The minimal length of the linkers for this GFP-linker-Aβ(17-42)-linker-GFP molecule is 18 residues.

Our modelling also revealed that there are a number of the intermediate GFP packings between two described cases when the shorter linker is more than 8 residues and less than 18 residues.

These cases require additional investigations. Our rough estimation proposes that the sum of two linkers in a given GFP-linker-Aβ(17-42)-linker-GFP molecule should be more than 34 residues. This conclusion can be further verified by the rigid-body simulation implemented in Blender.

CONCLUSIONS

In this work, we demonstrated the importance of the steric effects imposed by the globular structures located near ARs on the fibrillogenesis. In particular, we tested hybrid proteins containing an amyloid-forming Aβ(17-42) fragment linked to one or two globular domains of GFP and found that linkers of 7 and more residues allow fibrillogenesis of the Aβ-peptide flanked by GFP on one side and 18 and more residues when Aβ-peptide is flanked by GFPs on both sides. Furthermore, we were able to establish a more general relationship between the size of the globular domains and the length of the linkers by using analytical expressions and mesoscopic protein modelling. In reality, the established values can be slightly bigger due to, for example, the presence of the ordered layers of water around the proteins that increase the effective volumes of the structures. Our modelling also showed that in the fibrils with the minimal linkers the globules are packed so densely that can protect the linkers from the intermolecular interaction including the interaction with proteases.

These results have a number of applications. In particular, they can improve the understanding of the factors that enable a protein to form amyloids. When an AR, which is predicted based on its amino acid sequence, is located in close vicinity from the globular domain, it is not clear whether this AR will trigger assembly of the fibrils or not. Therefore, the constraints on amyloid formation imposed by the steric exclusion of globular domains must be used in combination with these prediction methods. Our study also points out on a possible mechanism of the control of

the amyloid formation by the programmed cleavage of the globular domains located in close vicinity of the ARs. We expect that this mechanism can be used to regulate the formation of functional amyloids. The other application of the established constraints is in the design of new functional amyloids not yet produced by natural evolution with applications in nanobiotechnology, material science, and synthetic biology. Finally, this knowledge will be used in the structural interpretation of the experimental results and planning of the experiments. Due to the high polymorphism of the amyloid fibrils, the experimental test of this effect is hampered, however, with the structural insight provided here, it is possible, now, to design future experimental studies refining our theoretical estimations.

The other result of this work is development and comparison of different theoretical approaches allowing the determination of the relationship between the structure of the amyloid core, the length of the linkers and the size of the globular domain. Our analysis showed that the amyloid fibrils with large (over 50 Å) globular domains (Fig. 3) could be exclusively tested by the mesoscopic modelling. The rigid body approximation can also be used to examine the steric exclusion of globular domains in the fibrils formed by proteins with two globular domains on both sides of the amyloidogenic region. Previously, Blender software package has been successfully used in BioBlender project [49]. In future, we will improve this approach by taking into consideration more exact shapes of the protofibrils and globular domains using the atomic coordinates of the structures in a module “Atomic Blender”. It will also allow us to analyze all known fibril polymorphs of a given AR. In addition to the shapes, the amyloid fibrils can have different structural arrangements. Here we considered the most common parallel and in-register arrangement of naturally occurring amyloids, but in some amyloids, the monomers can be also packed in an anti-parallel manner [50], or via β-solenoidal structures [51,52]. Some amyloid fibrils are formed by the co-aggregation of two different proteins [6]. The structures of the amyloid core can consist of stacks of β-arches, β-serpentines or the other β-arc-containing

structures [8,11–13,23]. All these arrangements can be analyzed and compared by the mesoscopic modelling and will be the subject of the next work. Finally, with this approach, we envisage the creation of the web-based tool that could analyze the user-defined structure of the amyloid, connected globular domain(s) and the length of the linker.

MATERIALS AND METHODS

Molecular modelling of fibril with a pseudo-helical packing of GFPs

We used the 3D structure of Aβ(17-42) amyloid fibril determined by solid-state NMR spectroscopy (model 1 of PDB entry 2BEG) [23] as a template for the amyloid core fibril. Two types of the core fibril were built: untwisted and left-hand twisted. The untwisted fibril was created by positioning stacks of three successive chains B, C and D of 2BEG model with a mean translation of 14.4 Å (3x4.8 Å) along the Z-axis. The axial 4.8 Å translation of monomers was chosen based on the analysis of several known crystal structures of β-solenoids. The Z-axis corresponds to main inertia axis of the fibril. The left-hand twisted fibril was built with 2° angle between the monomers. The twisting operator was imposed around a Z-axis located at the center of mass of the main-chain atoms from the two β-strands of the Aβ(17-42) fragment (residues 17 to 25 and 32 to 40). A 180-mer fibril that corresponds to one complete helix turn (360°) was then built. The infinite fibril can be obtained by the translation along the Z-axis by the helical pitch of 864 Å.

To build models of hybrid protein Aβ(17-42)-linker-GFP we used the crystal structure of GFP from *Aequorea victoria* protein (pdb entry 1GFL) [53] and a linker with SPSSPSS sequence (Fig. 2a). To position the GFP structures around the amyloid core and evaluate their packing, we built a model of two successive hexameric layers of the subunits. The first GFP subunit was

manually docked on one side of the fibril stack while keeping the axis of the GFP β-barrel perpendicular to the Z-axis. The five other subunits were positioned by applying the successive rotation of ~51° around an axis that intersects the X-Y plane at the Cα atom of residue 42 of the first monomer and their successive translations of 4.8 Å along the Z-axis (Fig. 2b). The rotation angle 51° and the distance between the Cα-atom of Aβ residue 42 and the Cα-atom of the GFP N-terminus were chosen so that they provide the most optimal close packing of the GFP molecules within the hexamer (Fig. 2c). To evaluate the packing of GFP domains between the hexameric units, an adjacent hexameric layer was generated by a 28.8 Å translation along the Z-axis. The 28.8 Å distance agrees well with the corresponding distance between the GFPs from *Aequorea victoria* in the crystal packing (pdb entry 1GFL) [53].

The first hybrid and GFP-linker- Aβ(17-42)-linker-GFP model with a short linker of the C-terminal GFP and a long linker of the N-terminal GFP was built using as a template the previously described Aβ-linker-GFP fibril. The additional N-terminal GFP domains were symmetrically positioned around an axis crossing X, Y coordinates of the Cα-atom of residue 17 of Aβ-fragment. Then, the subunits were translated in the X-Y plane towards the fibril until they contacted the C-terminal GFPs using a similar docking protocol. The second hybrid GFP-linker-Aβ-linker-GFP model containing two linkers of equivalent length was built using the same strategy. In this case, the successive translation of GFP monomers along the Z-axis was 4.8/2 Å.

MOLEMAN [54] and CHIMERA [55] programs were used for the symmetrical positioning of GFPs and molecular modelling procedures. The linkers connecting Aβ-peptides to GFPs were generated using MODLOOP [56,57].

Molecular dynamics simulations of fibrils

The initial models were further fine-tuned using short all-atom molecular dynamics (MD) simulations. The large size of these systems was challenging for MD. Thus, we used an implicit solvent model combined with periodic boundary conditions. For the untwisted fibril, our periodic box contains a layer constituted of the hexamer occupying a rectangular cell of c=28.8 Å and a=b=250 Å. The *c* axis is parallel to the pseudo-helical (or segmented helical) symmetry axis. For the left-hand twisted model, the box dimension along c was 864 Å corresponding to the twist pitch. The MD simulations of the implicitly solvated proteins were performed using NAMD2 program [58] with the CHARMM27 force field [59]. First, an energy minimization step was used to relax strained conformations of linkers and remove side-chain steric clashes created during GFP docking and linker building procedures. The temperature was maintained at 310°K using the Langevin thermostat with a damping coefficient 5ps^{-1}. The van der Waals interactions were modeled with a Lennard-Jones potential using a smooth cutoff (switching distance of 14 Å and cutoff radius of 16 Å). The generalized Born implicit solvent (GBIS) method [60] was used with dielectric constants 80 for solvent and 1 for protein, and ion concentration at 0.3 M.l^{-1}. The solvent molecular electrostatics was calculated by the Poisson-Boltzmann equation, which models water as a dielectric continuum. All the production runs were performed with 2 fs time step in conjunction with the SHAKE algorithm to constrain covalent bond lengths between heavy and hydrogen atoms [61]. Positional restraints (force constants of 1000 kcal/mol/Å^2) were applied to the main chain atoms of the Aβ(17-42) core domain during all simulations to guarantee that conformation of the Aβ arches and the interlayers backbone hydrogen bonds were maintained along the fibril axis. The overall conformation of the GFP β-barrel domain was preserved using harmonic restraints for dihedral angles of amino acid residues in β-sheets or α-helices, as well as restraints for hydrogen bonds involving backbone atoms from the same residues (force constants of 50 kcal/mol/rad^2 for backbone angles and 20 kcal/mol/Å^2 for H-bonds distances). Two independent short 2 ns simulations were performed on an Intel Xeon

cluster including a Tesla M2090 GPU card for the untwisted models. For the twisted models, the simulation ran for only 0.2 ns due to the large size of the system. Trajectory data from the simulations were recorded every 2 ps and analyzed in VMD [62]. The final snapshots from the simulations were subjected to an idealization step using REFMAC5 program [63] for fine refinement. The quality and consistency of the generated models were checked with the PROCHECK program [64]. The analysis of buried surface areas was done using the AREAIMOL [65] program in CCP4 package [66]. To analyze the structural quality of the twisted 180-subunit models, a system of eighteen adjacent subunits was chosen.

Rigid Body Simulation

The Rigid Body Simulation by using the Low Poly 3d models of the amyloid fibrils formed by hybrid proteins Aβ(17-42)-linker-globula was performed by using Blender software package [67]. The Aβ(17-42) amyloid protofibril was represented as a stack of parallelepipeds having the size of 40x22x4.8 units (Fig. 4). In our case, 1 unit means 1 Å and the dimension of 4.8 corresponds to the axial stacking of the parallelepipeds. The neighboring parallelepipeds were twisted 2° relative to each other around the fibril axis that crosses the centers of the parallelepipeds. The linker was modeled as a chain of connected spheres with the diameter of 3.5 units that corresponds to the dimension of an amino acid residue along the polypeptide chain. The links between the spheres of linkers had constraints “Limit Distance” and “Connect Rigid Bodies” (type point). The globular protein structures were represented as spheres of bigger sizes. In different tested models their diameters were 30, 40, 50 or 60 units. All these objects were placed in Rigid Body World of Blender as “rigid bodies”. The following properties of collision between the objects were set up: Shape “Mesh”, Source “Final”, Friction 0.0, Bounciness 0.5, Collision margin 0.001.

The positions of the parallelepipeds were fixed, and all the other elements could move preserving their connections. We analyzed a large fragment of the fibril consisting of 100 molecules. The estimations of the minimal number of spheres-residues in the linker with given size of the globular part were made as follows: at the beginning of the procedure all linkers were straight and protruded perpendicularly to the fibril axis in an equivalent manner (Fig. 4b). The globular parts had the diameter of 0.7 units. During the 5-second simulation, the diameters of the globular parts gradually grew up to the given size. As a result, the globular parts moved apart from each other to avoid the overlaps (Fig. 4c). The final structures were checked for the absence of intersections between the elements of models with Mesh Analysis tool of Blender.

ACKNOWLEDGMENT

The authors acknowledge Armenian Communities Department of Calouste Gulbenkian Foundation for providing “Global Excellence Scholarship” to R.A., Erasmus Mundus program for providing travel grant to A.G. This work was also supported by Russian Ministry of Science and Education under 5-100 Excellence Program, by EU COST Action BM1405 NGP-net, a grant PRC1524 PRC from PRC CNRS, France.

REFERENCES

[1] M.B. Pepys, Amyloidosis, Annu. Rev. Med. 57 (2006) 223–241.

[2] S.K. Maji, M.H. Perrin, M.R. Sawaya, S. Jessberger, K. Vadodaria, R.A. Rissman, P.S. Singru, K.P.R. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, P. Sawchenko, W. Vale, R. Riek, Functional amyloids as natural storage of peptide hormones in pituitary secretory granules, Science. 325 (2009) 328–332.

[3] C.P.J. Maury, The emerging concept of functional amyloid, J. Intern. Med. 265 (2009)

329–334.

[4] L.P. Blanco, M.L. Evans, D.R. Smith, M.P. Badtke, M.R. Chapman, Diversity, biogenesis and function of microbial amyloids, Trends Microbiol. 20 (2012) 66–73.

[5] S. Sugiyama, M. Tanaka, Self-propagating amyloid as a critical regulator for diverse cellular functions, J. Biochem. 155 (2014) 345–351.

[6] J. Li, T. McQuade, A.B. Siemer, J. Napetschnig, K. Moriwaki, Y.S. Hsiao, E. Damko, D. Moquin, T. Walz, A. McDermott, F.K.M. Chan, H. Wu, The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis, Cell. 150 (2012) 339–350.

[7] D. Otzen, P.H. Nielsen, We find them here, we find them there: Functional bacterial amyloid, Cell. Mol. Life Sci. 65 (2008) 910–927.

[8] A. V. Kajava, U. Baxa, A.C. Steven, arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils, FASEB J. 24 (2010) 1311–1319.

[9] R. Riek, D.S. Eisenberg, The activities of amyloids from a structural perspective, Nature. 539 (2016) 227–235.

[10] F. Weirich, L. Gremer, E.A. Mirecka, S. Schiefer, W. Hoyer, H. Heise, Structural characterization of fibrils from recombinant human islet amyloid polypeptide by solid-state NMR: The central FGAILS segment is part of the β-sheet core, PLoS One. 11 (2016).

[11] A.W.P. Fitzpatrick, B. Falcon, S. He, A.G. Murzin, G. Murshudov, H.J. Garringer, R.A. Crowther, B. Ghetti, M. Goedert, S.H.W. Scheres, Cryo-EM structures of tau filaments from Alzheimer's disease, Nature. 547 (2017) 185–190.

[12] Y. Xiao, B. Ma, D. McElheny, S. Parthasarathy, F. Long, M. Hoshi, R. Nussinov, Y. Ishii, Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease, Nat. Struct. & Mol. Biol. 22 (2015) 499–505.

[13] S.A. Bondarev, O. V Bondareva, G.A. Zhouravleva, A. V Kajava, A. Valencia, BetaSerpentine: a bioinformatics tool for reconstruction of amyloid structures,

Bioinformatics. 34 (2018) 599–608.

[14] C. Wasmer, A. Schütz, A. Loquet, C. Buhtz, J. Greenwald, R. Riek, A. Böckmann, B.H. Meier, The Molecular Organization of the Fungal Prion HET-s in Its Amyloid Form, J. Mol. Biol. 394 (2009) 119–127.

[15] C.M. Pfefferkorn, R.P. McGlinchey, J.C. Lee, Effects of pH on aggregation kinetics of the repeat domain of a functional amyloid, Pmel17, Proc. Natl. Acad. Sci. 107 (2010) 21447–21452.

[16] A.K.H. Chen, R.Y.Y. Lin, E.Z.J. Hsieh, P.H. Tu, R.P.Y. Chen, T.Y. Liao, W. Chen, C.H. Wang, J.J.T. Huang, Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis, J. Am. Chem. Soc. 132 (2010) 1186–1187.

[17] U. Baxa, R.B. Wickner, A.C. Steven, D.E. Anderson, L.N. Marekov, W.M. Yau, R. Tycko, Characterization of beta-sheet structure in Ure2p1-89 yeast prion fibrils by solid-state nuclear magnetic resonance, Biochemistry. 46 (2007) 13149–13162.

[18] F. Shewmaker, D. Kryndushkin, B. Chen, R. Tycko, R.B. Wickner, Two prion variants of Sup35p have in-register parallel β-sheet structures, independent of hydration, Biochemistry. 48 (2009) 5074–5082.

[19] J.J. Helmus, K. Surewicz, M.I. Apostol, W.K. Surewicz, C.P. Jaroniec, Intermolecular alignment in Y145Stop human prion protein amyloid fibrils probed by solid-state NMR spectroscopy, J. Am. Chem. Soc. 133 (2011) 13934–13937.

[20] B.R. Groveman, M.A. Dolan, L.M. Taubner, A. Kraus, R.B. Wickner, B. Caughey, Parallel in-register intermolecular β-sheet architectures for prion-seeded prion protein (PrP) amyloids, J. Biol. Chem. 289 (2014) 24129–24142.

[21] K. Shen, B. Calamini, J.A. Fauerbach, B. Ma, S.H. Shahmoradian, I.L. Serrano Lachapel, W. Chiu, D.C. Lo, J. Frydman, Control of the structural landscape and neuronal proteotoxicity of mutant Huntingtin by domains flanking the polyQ tract, Elife. 5 (2016).

[22] A. Der-Sarkissian, C.C. Jao, J. Chen, R. Langen, Structural organization of alpha-

synuclein fibrils studied by site-directed spin labeling, J. Biol. Chem. 278 (2003) 37530–37535.

[23] T. Luhrs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert, R. Riek, 3D structure of Alzheimer's amyloid- (1-42) fibrils, Proc. Natl. Acad. Sci. 102 (2005) 17342–17347.

[24] M.L. Duennwald, S. Jagadish, P.J. Muchowski, S. Lindquist, Flanking sequences profoundly alter polyglutamine toxicity in yeast., Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 11045–11050.

[25] Rubinshtein and Colby, Polymer Physics, Oxford University Press, 2003.

[26] A. Diehl, Y. Roske, L. Ball, A. Chowdhury, M. Hiller, N. Molière, R. Kramer, D. Stöppler, C.L. Worth, B. Schlegel, M. Leidert, N. Cremer, N. Erdmann, D. Lopez, H. Stephanowitz, E. Krause, B.-J. van Rossum, P. Schmieder, U. Heinemann, K. Turgay, Ü. Akbey, H. Oschkinat, Structural changes of TasA in biofilm formation of Bacillus subtilis, Proc. Natl. Acad. Sci. (2018) 201718102.

[27] J. Lu, Y. Yu, I. Zhu, Y. Cheng, P.D. Sun, Structural mechanism of serum amyloid A-mediated inflammatory amyloidosis, Proc. Natl. Acad. Sci. 111 (2014) 5189–5194.

[28] R. Diaz-Avalos, C.-Y. King, J. Wall, M. Simon, D.L.D. Caspar, Strain-specific morphologies of yeast prion amyloid fibrils., Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 10165–10170.

[29] U. Baxa, T. Cassese, A. V. Kajava, A.C. Steven, Structure, Function, and Amyloidogenesis of Fungal Prions: Filament Polymorphism and Prion Variants, Adv. Protein Chem. 73 (2006) 125–180.

[30] M. Anderson, O. V. Bocharova, N. Makarava, L. Breydo, V. V. Salnikov, I. V. Baskakov, Polymorphism and Ultrastructural Organization of Prion Protein Amyloid Fibrils: An Insight from High Resolution Atomic Force Microscopy, J. Mol. Biol. 358 (2006) 580–596.

[31] N.J. Cobb, M.I. Apostol, S. Chen, V. Smirnovas, W.K. Surewicz, Conformational stability

of mammalian prion protein amyloid fibrils is dictated by a packing polymorphism within the core region, J. Biol. Chem. 289 (2014) 2643–2650.

[32] A. V Kajava, U. Baxa, R.B. Wickner, A.C. Steven, A model for Ure2p prion filaments and other amyloids: the parallel superpleated beta-structure, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 7885–7890.

[33] M.A. Wälti, F. Ravotti, H. Arai, C.G. Glabe, J.S. Wall, A. Böckmann, P. Güntert, B.H. Meier, R. Riek, Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril, Proc. Natl. Acad. Sci. 113 (2016) E4976–E4984.

[34] M.T. Colvin, R. Silvers, Q.Z. Ni, T. V. Can, I. Sergeyev, M. Rosay, K.J. Donovan, B. Michael, J.S. Wall, S. Linse, R.G. Griffin, Atomic Resolution Structure of Monomorphic Aβ(42) Amyloid Fibrils, J. Am. Chem. Soc. 138 (2016) 9663–9674.

[35] J. Lu, W. Qiang, W. Yau, C.D. Schwieters, S.C. Meredith, R. Tycko, Molecular Structure of b -Amyloid Fibrils in Alzheimer ' s Disease Brain Tissue, Cell. 154 (2013) 1257–1268.

[36] G.S. Waldo, B.M. Standish, J. Berendzen, T.C. Terwilliger, Rapid protein-folding assay using green fluorescent protein., Nat. Biotechnol. 17 (1999) 691–695.

[37] C. Wurth, N.K. Guimard, M.H. Hecht, Mutations that reduce aggregation of the alzheimer's Aβ42 peptide: An unbiased search for the sequence determinants of Aβ amyloidogenesis, J. Mol. Biol. 319 (2002) 1279–1290.

[38] A. Fox, T. Snollaerts, C. Errecart Casanova, A. Calciano, L.A. Nogaj, D.A. Moffet, Selection for Nonamyloidogenic Mutants of Islet Amyloid Polypeptide (IAPP) Identifies an Extended Region for Amyloidogenicity, Biochemistry. 49 (2010) 7783–7789.

[39] T. Ochiishi, M. Doi, K. Yamasaki, K. Hirose, A. Kitamura, T. Urabe, N. Hattori, M. Kinjo, T. Ebihara, H. Shimura, Development of new fusion proteins for visualizing amyloid-β oligomers in vivo, Sci. Rep. 6 (2016).

[40] R.M. Hussein, R.M. Hashem, L.A. Rashed, Evaluation of the amyloid beta-GFP fusion protein as a model of amyloid beta peptides-mediated aggregation: a study of DNAJB6

chaperone, Front. Mol. Neurosci. 8 (2015) 40.

[41] W. Kim, M.H. Hecht, Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Abeta42 peptide, Proc. Natl. Acad. Sci. 103 (2006) 15824–15829.

[42] C.D. Link, V. Fonte, C.M. Roberts, B. Hiester, M.A. Silverman, G.H. Stein, The beta amyloid peptide can act as a modular aggregation domain, Neurobiol. Dis. 32 (2008) 420–425.

[43] P.E. Arslan, A. Chakrabartty, Probing Alzheimer amyloid peptide aggregation using a cell-free fluorescent protein refolding method., Biochem. Cell Biol. 87 (2009) 631–639.

[44] D.A. Zacharias, J.D. Violin, A.C. Newton, R.Y. Tsien, Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells, Science. 296 (2002) 913–916.

[45] U. Pieper, B.M. Webb, G.Q. Dong, D. Schneidman-Duhovny, H. Fan, S.J. Kim, N. Khuri, Y.G. Spill, P. Weinkam, M. Hammel, J.A. Tainer, M. Nilges, A. Sali, ModBase, a database of annotated comparative protein structure models and associated resources, Nucleic Acids Res. 42 (2014) 336–346.

[46] M.R. Smaoui, F. Poitevin, M. Delarue, P. Koehl, H. Orland, J. Waldispühl, Computational assembly of polymorphic amyloid fibrils reveals stable aggregates, Biophys. J. 104 (2013) 683–693.

[47] A. Kahler, H. Sticht, A.H.C. Horn, Conformational Stability of Fibrillar Amyloid-Beta Oligomers via Protofilament Pair Formation - A Systematic Computational Study, PLoS One. 8 (2013).

[48] T.P.J. Knowles, A. De Simone, A.W. Fitzpatrick, A. Baldwin, S. Meehan, L. Rajah, M. Vendruscolo, M.E. Welland, C.M. Dobson, E.M. Terentjev, Twisting transition between crystalline and fibrillar phases of aggregated peptides, Phys. Rev. Lett. 109 (2012) 158101–158105.

[49] R.M. Andrei, M. Callieri, M.F. Zini, T. Loni, G. Maraziti, M.C. Pan, M. Zoppè, Intuitive

representation of surface properties of biomolecules using BioBlender, BMC Bioinformatics. 13 (2012).

[50] W. Qiang, W.-M. Yau, Y. Luo, M.P. Mattson, R. Tycko, Antiparallel β-sheet architecture in Iowa-mutant -amyloid fibrils, Proc. Natl. Acad. Sci. 109 (2012) 4443–4448.

[51] A. Sen, U. Baxa, M.N. Simon, J.S. Wall, R. Sabate, S.J. Saupe, A.C. Steven, Mass analysis by scanning transmission electron microscopy and electron diffraction validate predictions of stacked beta-solenoid model of HET-s prion fibrils, J Biol Chem. 282 (2007) 5545–5550.

[52] C. Wasmer, A. Lange, H. Van Melckebeke, A.B. Siemer, R. Riek, B.H. Meier, Amyloid fibrils of the HET-s(218-289) prion form a β solenoid with a triangular hydrophobic core, Science. 319 (2008) 1523–1526.

[53] F. Yang, L.G. Moss, G.N. Phillips, The Molecular Structure of Green Fluorescent Protein, Nat. Biotechnol. 14 (1996) 1246–1251.

[54] G.J. Kleywegt, Experimental assessment of differences between related protein crystal structures, Acta Crystallogr. Sect. D Biol. Crystallogr. 55 (1999) 1878–1884.

[55] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A Visualization System for Exploratory Research and Analysis, J Comput Chem. 25 (2004) 1605–1612.

[56] A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, Protein Sci. 9 (2000) 1753–1773.

[57] A. Fiser, A. Sali, ModLoop: Automated modeling of loops in protein structures, Bioinformatics. 19 (2003) 2500–2501.

[58] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.

[59] A.D. MacKerell, D. Bashford, J. Dunbrack Roland, L., J.D. Evanseck, M.J. Field, S.

Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R.H. Stote, J.. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, J. Phys. Chem. B. 102 (1998) 3586–3616.

[60] D.E. Tanner, J.C. Phillips, K. Schulten, GPU/CPU algorithm for generalized born/solvent-accessible surface area implicit solvent calculations, J. Chem. Theory Comput. 8 (2012) 2521–2530.

[61] V. Kräutler, W.F. Van Gunsteren, P.H. Hünenberger, A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations, J. Comput. Chem. 22 (2001) 501–508.

[62] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.

[63] A.A. Vagin, R.A. Steiner, A.A. Lebedev, L. Potterton, S. McNicholas, F. Long, G.N. Murshudov, REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use, Acta Crystallogr. Sect. D Biol. Crystallogr. 60 (2004) 2184–2195.

[64] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.

[65] B. Lee, F.M. Richards, The interpretation of protein structures: Estimation of static accessibility, J. Mol. Biol. 55 (1971) 379–400.

[66] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, Acta Crystallogr. Sect. D Biol. Crystallogr. 67 (2011) 235–242.

[67] blender.org - Home of the Blender project - Free and Open 3D Creation Software, Blender.org. (2015). https://www.blender.org/.

[68] W. DeLano, Pymol: An open-source molecular graphics tool, CCP4 Newsl. Protein Crystallogr. 700 (2002).

FIGURE LEGENDS

Fig. 1. General scheme of amyloid fibrils with different types of the flanking regions. The fiber axis is indicated by vertical thin arrow. (a) Stack of ARs (as boxes of green color) without flunking regions. (b) Stack of ARs with intrinsically disordered regions (as black lines). (c) Stack of ARs with globular domains that are displayed as spheres of brown color. (d) Stack of ARs with globular domains that interact with each other. (e) Situation when the stacking of ARs is hampered by the steric repulsion of globular domains.

Fig. 2. Model of amyloid fibril formed by hybrid molecule Aβ(17-42)-linker-GFP with minimal allowed linker and pseudo-helical packing of GFPs. (a) Sequence and 3D structure of Aβ(17-42)-linker-GFP monomer with Aβ(17-42) peptide (*PDB entry* 2BEG) (in red), a linker (underlined) composed of SerProSerSerProSer sequence (in green) and one additional residue from the N-terminus of GFP, and GFP domain from *Aequorea victoria* (PDB entry 2Y0G) (in blue). (b) Lateral projection of the repetitive structural element of the fibril containing six Aβ(17-42)-linker-GFP molecules. For the sake of clarity, only two GFP domains (in green) are shown (1^{st} and 7^{th}). The fibril can be generated by the 28.8 Å translation of this hexameric element along the fibril axis. (c) Axial projection of the repetitive element. (d) Model of the fibril with 2° twist (Aβ(17-42) is in blue and GFPs are in green). Image is generated by VMD software [62].

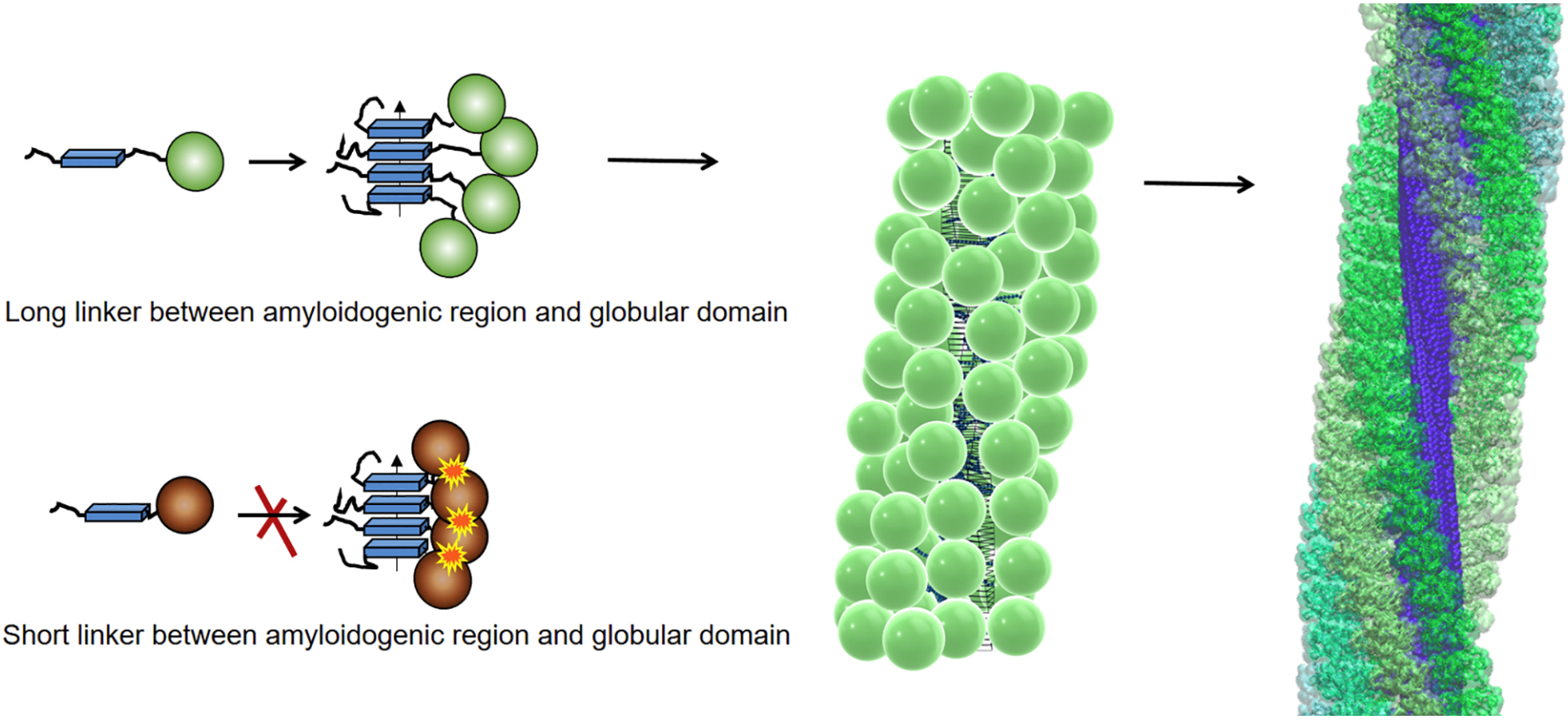
Fig. 3. Dependences between the minimal length of the linker and the size of the globular domain obtained by analytical estimation (blue and red stepped lines), by molecular modelling of Aβ(17-42)-linker-GFP fibrils (green circle) and by rigid body simulation (black rectangles).

Fig. 4. Models used in the rigid body simulation. An example with globular domain diameter of 50 Å is shown. (a) Monomer. (b) Fibril before the simulation with extended linkers and small spheres of globular domains. (c) Fibril after the simulation.

Fig. 5. Axial projection of molecular models of amyloid fibrils formed by GFP-linker-Aβ(17-42)-linker-GFP molecules. (a) Molecules with one linker significantly longer than another one. The minimal short linker has 8 residues, and the minimal length of the long linker (at the N-terminus of Aβ) is 24 residues. (b) Two linkers are of a similar length. The minimal length of the linkers for this GFP-linker-Aβ(17-42)-linker-GFP molecule is 18 residues. Images were generated by using PyMol software [68].

Highlights

We show the importance of steric effect of globular structures on amyloidogenesis.
Amyloid-forming region and folded globular structure need to be separated by linker.
We established the relationship between the sizes of globular structures and linkers.
It was demonstrated by using different modelling approaches.
Results can be used for molecular design of new functional amyloids.



Graphics Abstract

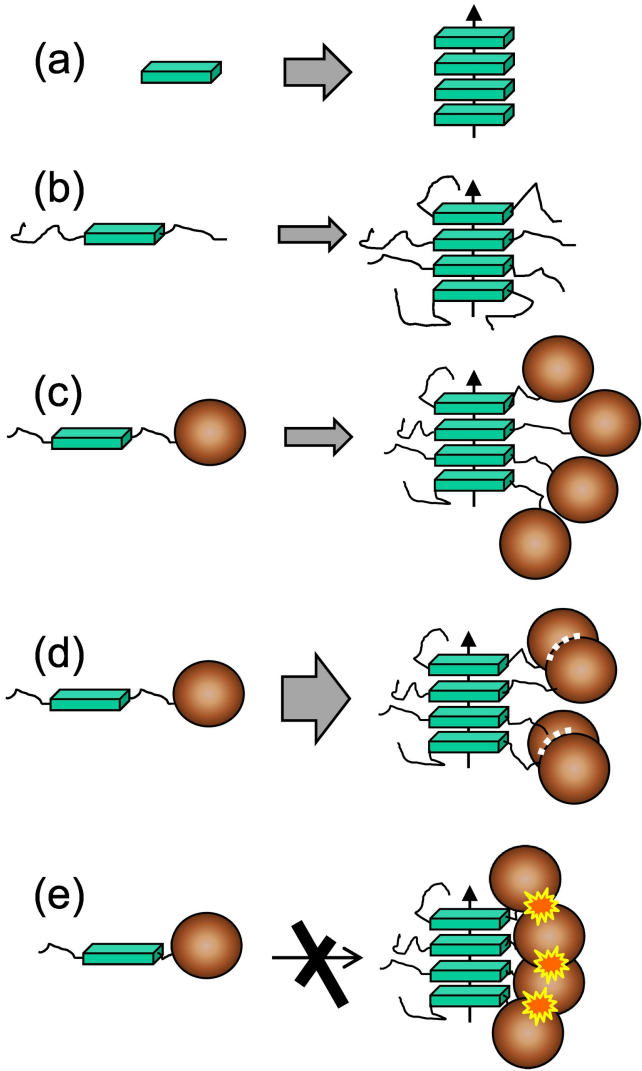


Figure 1

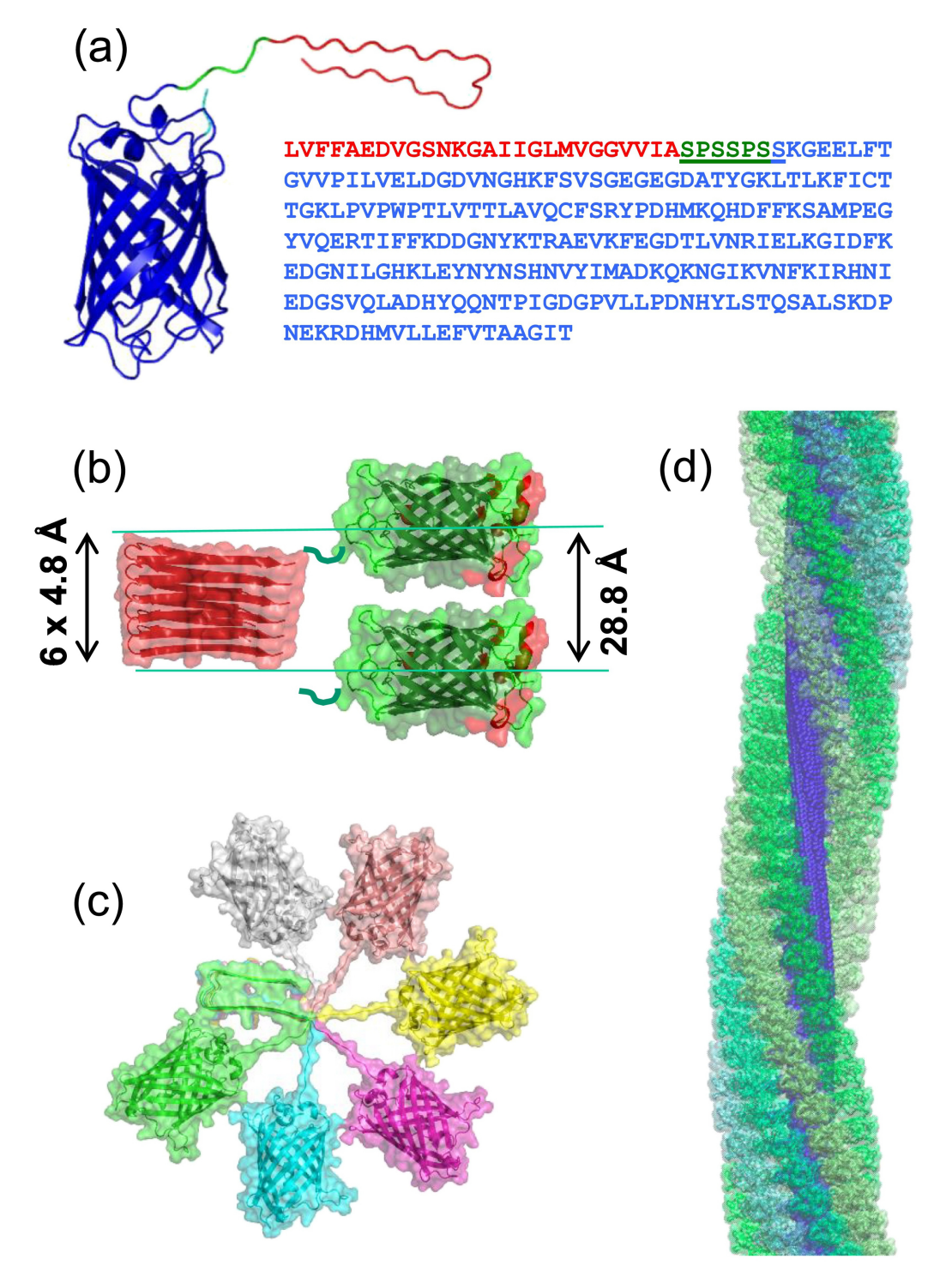

(a)
LVFFAEDVGSNKGAIIGLMVGGVVIASPSSPSSKGEELFT
GVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT
TGKLPVPWPTLVTTLAVQCFSRYPDHMKQHDFFKSAMPEG
YVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFK
EDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNI
EDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDP
NEKRDHMVLLEFVTAAGIT
(b)
6 x 4.8 Å
28.8 Å
(c)
(d)

Figure 2

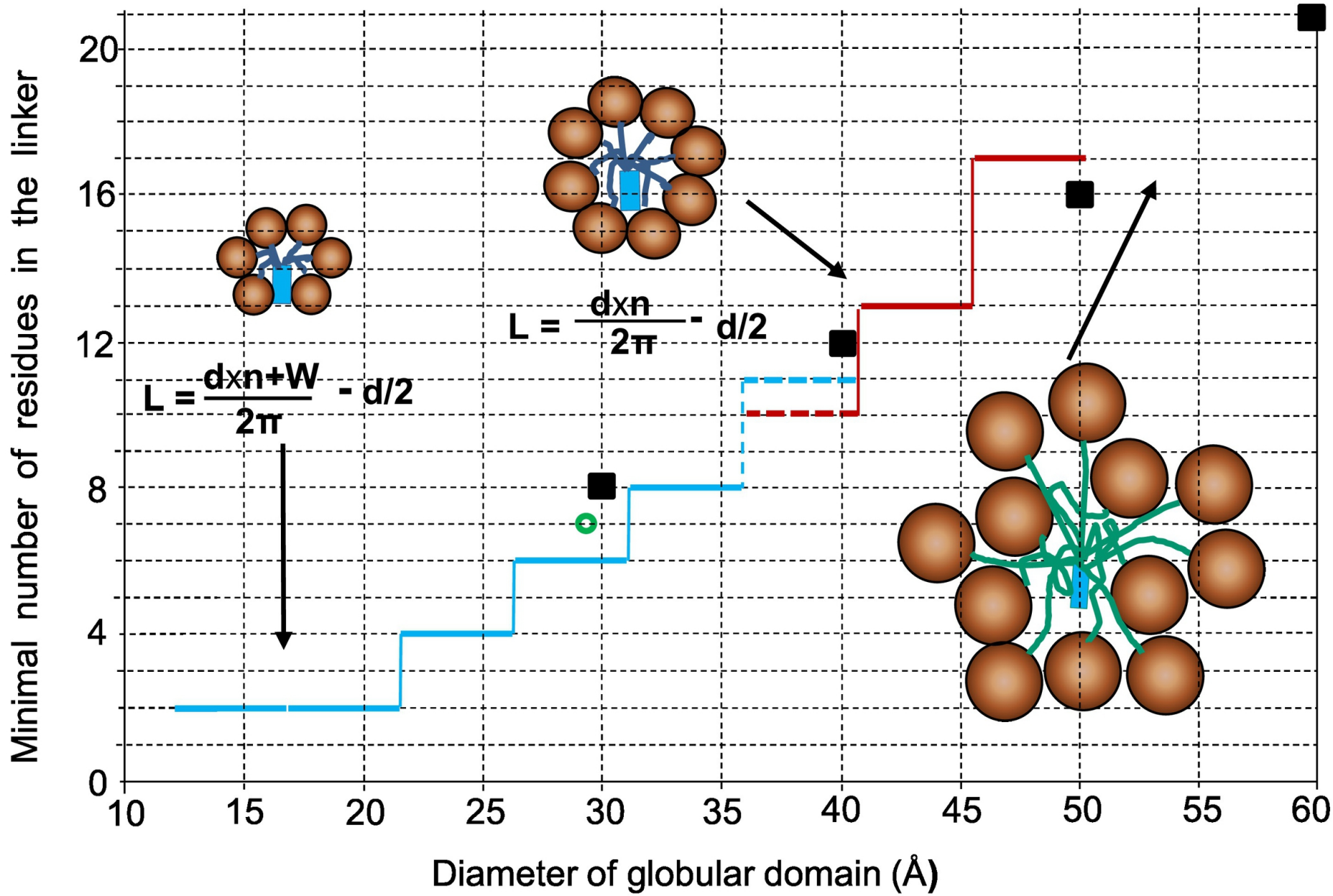

Minimal number of residues in the linker
Diameter of globular domain (Å)
0
4
8
12
16
20
10
15
20
25
30
35
40
45
50
55
60
L = dxn+W / 2π - d/2
L = dxn / 2π - d/2

Figure 3

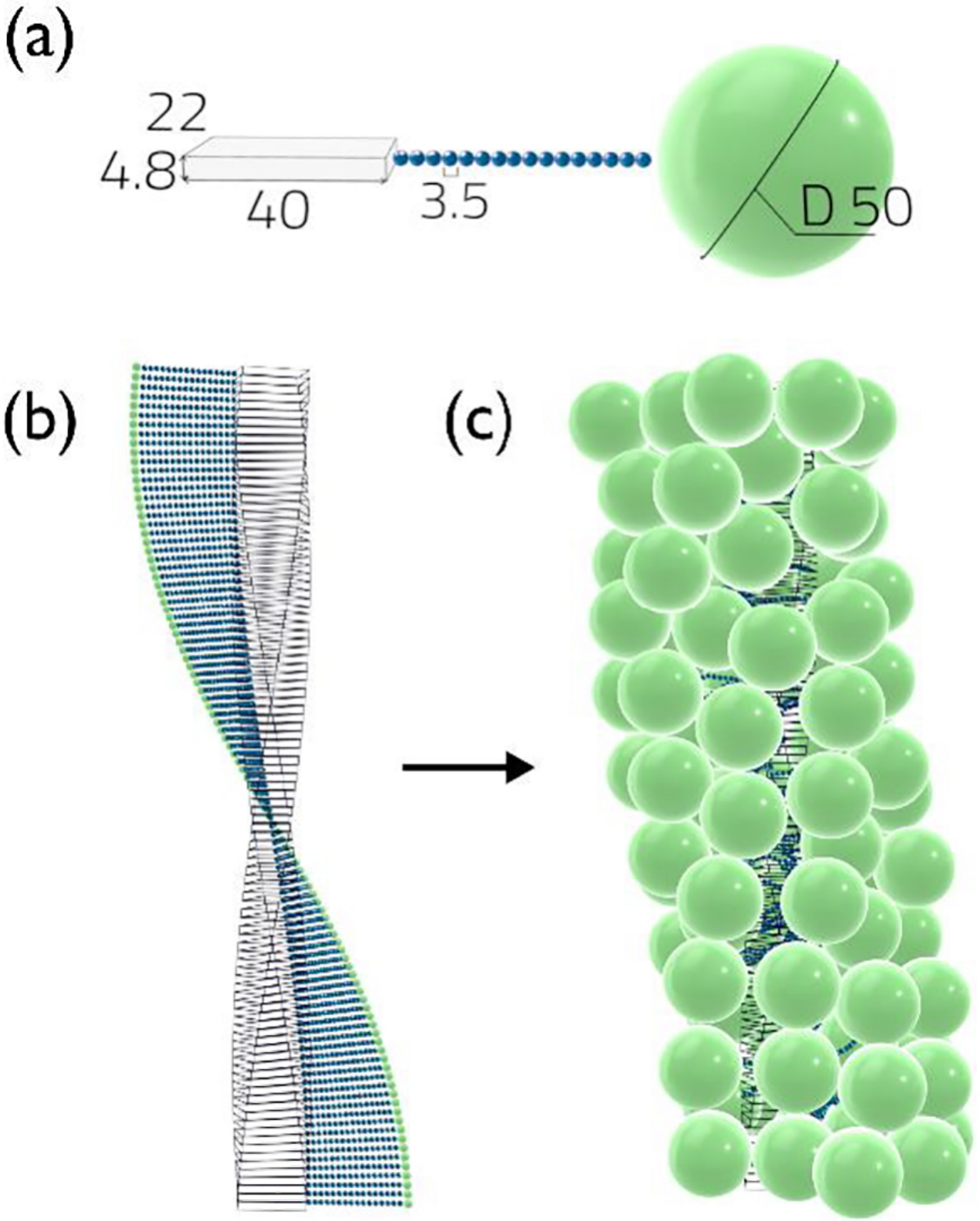


Figure 4

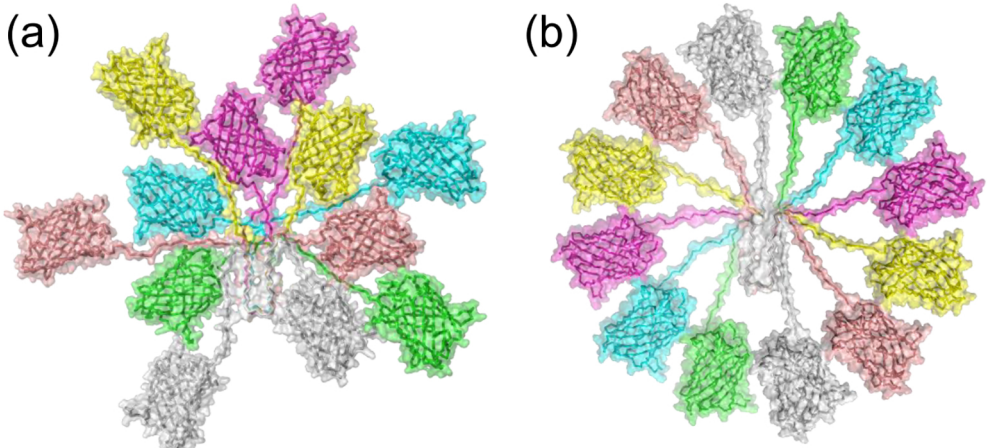


Figure 5