

# PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

# Fmoc-RGDS based fibrils: Atomistic details of their hierarchical assembly

David Zanuy,<sup>1,\*</sup> Jordi Poater,<sup>2</sup> Miquel Solà,<sup>3</sup> Ian W. Hamley,<sup>4</sup>

Carlos Alemán<sup>1,5,\*</sup>

<sup>1</sup> *Departament d'Enginyeria Química, ETSEIB, Universitat Politècnica de Catalunya, Av. Diagonal 647, 08028 Barcelona, Spain.*

<sup>2</sup> *Department of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling, Vrije Universiteit Amsterdam, De Boeleaan 1083, NL-1081HV Amsterdam, The Netherlands*

<sup>3</sup> *Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, Campus de Montilivi, E-17071 Girona, Spain*

<sup>4</sup> *School of Chemistry, Pharmacy and Food Biosciences, University of Reading, Reading, RG6 6AD, U.K.*

<sup>5</sup> *Center for Research in Nano-Engineering, Universitat Politècnica de Catalunya, Campus Sud, Edifici C', C/Pasqual i Vila s/n, Barcelona E-08028, Spain*

\* [david.zanuy@upc.edu](mailto:david.zanuy@upc.edu) and [carlos.aleman@upc.edu](mailto:carlos.aleman@upc.edu)

## ABSTRACT

We describe the 3D supramolecular structure of Fmoc-RGDS fibrils, where Fmoc and RGDS refer to the hydrophobic N-(fluorenyl-9-methoxycarbonyl) group and the hydrophilic Arg-Gly-Asp-Ser peptide sequence, respectively. For this purpose, we performed atomistic all-atom molecular dynamics simulations of a wide variety of packing modes derived from both parallel and antiparallel  $\beta$ -sheet configurations. The proposed model, which closely resembles the cross- $\beta$  core structure of amyloids, is stabilized by  $\pi$ - $\pi$  stacking interactions between hydrophobic Fmoc groups. More specifically, in this organization, the Fmoc-groups of  $\beta$ -strands belonging to the same  $\beta$ -sheet form columns of  $\pi$ -stacked aromatic rings arranged in parallel. Eight of such columns pack laterally forming a compact and dense hydrophobic core, in which two central columns are surrounded by three adjacent columns at each side. In addition to such Fmoc $\cdots$ Fmoc interactions, the hierarchically assembly of the constituent  $\beta$ -strands involves a rich variety of intra- and inter-strand interactions. Accordingly, hydrogen bonding, salt bridges and  $\pi$ - $\pi$  stacking interactions coexist in the highly ordered packing network proposed for the Fmoc-RGDS amphiphile. Quantum mechanical calculations, which have been performed to quantify the above referred interactions, confirm the decisive role played by the  $\pi$ - $\pi$  stacking interactions between the rings of the Fmoc groups, even though both inter-strand and intra-strand hydrogen bonds and salt bridges do also play a non-negligible role. Overall, these results provide a solid reference to complement available experimental data, which are not precise enough to determine the fibril structure, and reconcile previous independent observations.

## INTRODUCTION

The comprehension of protein self-assembly and aggregation is relevant in biomedicine and biotechnology to advance in the design of biological materials for a number of applications. A better understanding of the aggregation processes can be achieved by simulation using computational methodologies at atomic resolution. For instance, these methods have been employed to investigate the organization of amphiphilic peptides in aggregates. Thus, atomistic Molecular Dynamics (MD) simulations have been applied to study the dynamics and thermodynamics of fibril formation, nucleation, and polymorphism;<sup>1-4</sup> the inhibition of amyloid aggregation;<sup>5</sup> the assembly of cylindrical nanofibers<sup>6</sup> and fibril bilayers;<sup>7</sup> or nanotube formation from the self-assembly of  $\beta$ -helical peptides.<sup>8,9</sup> Very recently, the enormous capacity of current computational tools has been demonstrated by screening the aqueous self-assembly propensity in 8000 tripeptides and evaluating these by comparison with known examples.<sup>10</sup>

In a recent study we examined the self-assembly of Fmoc-RGDS, where Fmoc and RGDS refer to the hydrophobic N-(fluorenyl-9-methoxycarbonyl) group and the hydrophilic Arg-Gly-Asp-Ser peptide sequence, respectively.<sup>11</sup> While the RGD amino acid sequence is the unit of a cell adhesive activity domain in adherent proteins (*e.g.* fibronectin, fibrin and vitronectin),<sup>12,13</sup> the synthetic peptide RGDS exhibits high cell adhesion activity via binding to the integrin cell receptors.<sup>12,14,15</sup> The binding of RGDS to integrins has previously been shown to trigger chemotactic responses, including cell adhesion and migration *in vitro*.<sup>16</sup> On the other hand, self-assembly properties of Fmoc-protected peptides is a topic of growing interest because of the control exerted by  $\pi$ - $\pi$  stacking interactions induced by aromatic rings on the formed nanostructures.<sup>17-23</sup>

Fmoc-RGDS was found to fibrillise under appropriate conditions.<sup>11</sup> Circular dichroism and FTIR spectroscopy results indicated that the self-assembly at low concentrations being dominated by interactions between Fmoc units.<sup>11</sup> Moreover, fiber X-ray diffraction (XRD) suggested the formation of antiparallel  $\beta$ -sheets at sufficiently high concentrations. This antiparallel organization was in agreement with that proposed for Fmoc-FF<sup>22</sup> (Fmoc-Phe-Phe) and Fmoc- $\beta$ AH (Fmoc- $\beta$ Ala-His).<sup>23</sup>

The assembly of Fmoc-RGDS has also been the subject of preliminary modelling using MD simulations.<sup>11</sup> Modeling of systems formed by 7 or 21 peptide molecules ( $\beta$ -strands) assembled in the same  $\beta$ -sheet indicated that the parallel configuration is significantly less stable than the antiparallel, which was in agreement with XRD information. More recently, two  $\beta$ -sheets, each containing 7 explicit molecules, were packed with the Fmoc-aligned or with the charged side groups oriented face-to-face and, subsequently, simulated.<sup>24</sup> Twelve different packing models were constructed for the MD study, which differ in the parallel or antiparallel configuration of the strands in the sheet, the role played by the Fmoc groups in the assembly of the two sheets, and/or the relative orientation of the charged side groups of Arg and Asp belonging to different sheets.<sup>24</sup> Results suggested that the Fmoc...Fmoc interactions are decisive for the assembly of the two sheets. Among the 12 investigated models, 3 of them showed acceptable stability after 25 ns of MD simulation: two Fmoc-aligned packing models and one laterally packed model in which the sheets with the charged side groups of Arg and Asp are oriented face-to-face. Amazingly, peptide molecules in the  $\beta$ -sheets of such three models were arranged in parallel configuration. Although this was in apparent contradiction with XRD data, it should be remarked that such experimental observations corresponded to high concentration conditions.<sup>11</sup> Thus, the size of the theoretical model

(*i.e.* two sheets containing 7 molecules each one) only represented assemblies stabilized under low concentration conditions.<sup>24</sup>

The current study aims to elucidate and rationalize the 3D supramolecular structure of Fmoc-RGDS fibrils applying theoretical methods and starting from the knowledge acquired in our previous study on a very simple model formed by two interacting  $\beta$ -sheets.<sup>24</sup> In an effort to address the supramolecular association of this peptide, we sought to systematically and comprehensively analyse the organization of Fmoc-RGDS molecules considering realistic models of multiple packing modes for both parallel and antiparallel  $\beta$ -sheet configurations using atomistic MD simulations. Assemblies examined in this work contained from 4 to 8 explicit  $\beta$ -sheets and from 7 to 24 Fmoc-RGDS molecules per sheet. Overall, the MD simulations, which were performed in aqueous solution with explicit solvent molecules, spanned a total simulation time of 1  $\mu$ s. Results have provided a solid reference to complement available experimental data, which are not precise enough to determine the fibril structure,<sup>11</sup> and to reconcile previous independent observations.<sup>11,24</sup> Thus, the structures obtained in this work reveal details with atomic resolution of the packing interactions by which constituent  $\beta$ -sheets assemble hierarchically into fibrils. In addition, quantum mechanical calculations on reduced models of such systems allowed us to characterize and quantify the interactions present between amino acids chains or Fmoc groups of two Fmoc-RGDS units of one  $\beta$ -sheet (stacked  $\beta$ -strand interactions) or between two adjacent  $\beta$ -sheets (lateral  $\beta$ -strand interactions).

## METHODS

*MD simulations.* Intra- and intermolecular interactions were computed using the potential energy functions and the empirical parameters of the AMBER force-field,<sup>25</sup> with the exception of the electrostatic parameters for the Fmoc aromatic group that were obtained from previous work.<sup>11</sup> Bond lengths involving hydrogen atoms were kept at their equilibrium distances using the RATTLE algorithm.<sup>26</sup> Atom pair distance cutoffs were applied at 14.0 Å to compute the van der Waals interactions. To avoid discontinuities in this energy component, the van der Waals energy term was forced to slowly converge to zero by applying a smoothing factor from a distance of 12.0 Å. Electrostatic interactions were extensively computed by means of Ewald summations. The real space term was defined by the van der Waals cutoff (14.0 Å), while the reciprocal space was computed by interpolation of the effective charge into a charge mesh with a grid thickness of 5 points per volume unit (particle mesh Ewald).<sup>27</sup>

Table 1 provides specific information about all systems simulated in this work, including based on explicit and implicit solvent representation. This information refer to the dimensions of the periodic simulation box (minimum image convention), the number of Fmoc-RGDS strands together with the corresponding Na<sup>+</sup> counterions, and the number of water molecules (for simulations with explicit solvent representation). Water molecules were represented using the non-polarizable TIP3 model.<sup>28</sup> Ionic strength was directly represented by the counterions introduced to balance the total charge of each Fmoc-RGDS molecule. Each peptide molecule presented a net charge of -1, which is the outcome of having ionizable groups of both side and main chains at their most favored state at pH  $\approx$  7.0 (*i.e.* Arg side chain protonated with charge +1 and carboxylate groups of both Asp and the C-terminus ionized with charge -1).

Simulations executed with an implicit solvent representation were based on the generalized Born model.<sup>29</sup> The dielectric constant was set at 78.5 (water environment) while the ion concentration kept at 0.07 M that is approximately the concentration of Na<sup>+</sup> ions used in simulations with explicit solvent molecules.

All simulations were performed using the NAMD 2.9 program.<sup>30</sup> Each system was submitted to 5000 steps of energy minimization (Newton–Raphson method) before any MD trajectory was run in order to relax conformational and structural tensions. Both temperature and pressure were controlled by two different strategies depending on which sub-cycle was run. Due to its fast convergence, the weak coupling method<sup>31</sup> (Berendsen thermobarostat) was used to heat the system and to rapidly equilibrate its pressure and temperature around 1 bar and 298 K, respectively. The relaxation times used for the coupling were 1 and 10 ps for temperature and pressure, respectively. For final equilibration and for all production runs, both temperature and pressure were controlled by the Nose–Hoover piston<sup>32</sup> combined with the piston fluctuation control of temperature implemented for Langevin dynamics.<sup>33</sup> Pressure was kept at 1.01325 bars, the oscillation period was set at 1 ps while the piston decay time was set at 0.001 ps. The piston temperature was set at the same value as the thermostat control, 298 K, which used a damping coefficient of 2 ps. The integration step was 2 fs in all simulations.

The temperature, density and pressure of each examined model were equilibrated by three consecutive MD runs. First, a NVT-MD simulation at 298 K was run for 1 ns using Berendsen thermostat, the resulting atom velocities and coordinates being used as the starting point for a 1 ns NPT-MD run using the same procedure (298 K, 1 bar pressure). The end of this simulation was the starting point of a final 1 ns NPT-MD

simulation (298K, 1 bar pressure) using the combination of Langevin dynamics with Nose-Hoover piston oscillator, ensuring their ultimate equilibration through a correct thermodynamics ensemble. Simulations based on implicit solvent were both heated and equilibrated using of Langevin dynamics at 298K. The last step of any last equilibration run (*i.e.* either NVT for implicit solvent and NPT for explicit) was the starting point of the productive trajectories presented in this work (298 K, 1 bar pressure). On the other hand, in order to corroborate the results, MD simulations of all models and systems were run by triplicate. The coordinates of all the production runs were saved every 5 ps for further analysis.

To characterize and quantify the lateral and stacked  $\beta$ -strand interactions, quantum mechanical calculations have been carried out with the Amsterdam Density Functional (ADF) program<sup>34</sup> using dispersion-corrected density functional theory (using Grimme's DFT-D3BJ correction)<sup>35-38</sup> at the BLYP-D/TZ2P level of theory.<sup>39,40</sup> The bonding interactions have been further analysed by means of the energy decomposition analysis.<sup>41-46</sup> The interaction energy  $\Delta E_{\text{int}}$  corresponds to the actual energy change when two separated  $\beta$ -strands at the geometry they have in the assembly are combined either stacked in a  $\beta$ -shell or neighbouring between two different  $\beta$ -shells, and can be decomposed:

$$\Delta E_{\text{int}} = \Delta V_{\text{elstat}} + \Delta E_{\text{Pauli}} + \Delta E_{\text{oi}} + \Delta E_{\text{disp}} \quad (1)$$

Here,  $\Delta V_{\text{elstat}}$  corresponds to the classical electrostatic interaction between the unperturbed charge distributions of the prepared  $\beta$ -strands and is usually attractive. The Pauli-repulsion  $\Delta E_{\text{Pauli}}$  comprises the destabilizing interactions between occupied orbitals and is responsible for the steric repulsions. The orbital interaction  $\Delta E_{\text{oi}}$  accounts

for charge transfer (donor-acceptor interactions between occupied orbitals on one moiety with unoccupied orbitals of the other, including the HOMO–LUMO interactions) and polarization (empty/occupied orbital mixing on one fragment due to the presence of another fragment). Finally, the  $\Delta E_{\text{disp}}$  term (Grimme's DFT-D3BJ correction) accounts for the dispersion interactions.<sup>35-38</sup>

## RESULTS

### Construction of 3D assemblies using 2D building blocks

Amphiphilic peptides present a phenomenon called *segmental polymorphism*,<sup>47</sup> which refers to the identification of a large variety of  $\beta$ -sheet assemblies that serve as the spine (*i.e.* building block) for the fiber formation. Furthermore, remarkably distinct supramolecular assemblies may originate from the same building block, subtle differences in the organization of the same  $\beta$ -sheet assemblies giving place to very different morphologies.<sup>48</sup> The characterization of fibers of peptides with different sequences from simple extrapolation of a common and well-known structure is frustrated by this packing polymorphism, complete exploration of a wide number of  $\beta$ -sheet association models being required for each system. In this section the structure of Fmoc-RGDS fibers has been studied by constructing different 3D packing models using stable pairs of  $\beta$ -sheets as building blocks<sup>24</sup> and, subsequently, examining their stability using MD simulations at atomic resolution.

Specifically, three stable models obtained in our previous investigation on two packed  $\beta$ -sheets (labelled as P-TT2, P-HT2 and P-RD in reference 24 and schematically described in Figure 1) were used as building blocks to construct five new supramolecular models consisting of 6 sheets containing 7 Fmoc-RGDS molecules each

one. The construction of these new models, which are included in Figure 1, was based on the lateral assembly of the building blocks and can be described as follows. The original model P-TT2 (P and T refer to parallel and tail, respectively), which consisted of two parallel  $\beta$ -sheets with their Fmoc groups as far apart as possible, was used as the building block of two basic motifs giving rise to two supramolecular models with six sheets (Figure 1a): *i*) the P-TT2, which preserves the same name since the pair of parallel  $\beta$ -sheets assembled according to a tail-to-tail orientation is replicated laterally by simple translation; and *ii*) the AP-TT2, in which adjacent tail-to-tail pairs of parallel  $\beta$ -sheets are laterally assembled in an antiparallel fashion. The P-HT2 (H refers to head) building block, which consists of two aligned parallel  $\beta$ -sheets with the peptide tails of one sheet facing the peptide heads of the other sheet, was used to construct the P-HT2 and AP-HT2 supramolecular models (Figure 1b) in which pairs of  $\beta$ -sheets are laterally assembled in parallel and antiparallel orientation, respectively. The P-RD building block, which is formed by two parallel  $\beta$ -sheets assembled laterally to favor the formation of inter-sheet salt bridges between Arg and Asp side groups, resulted in a supramolecular model by simple replication of this basic motif (Figure 1c). In addition to these five supramolecular models, an arrangement based on the lateral assembly of two antiparallel  $\beta$ -sheets (labelled AA-RD in reference 24) was used to construct a new supramolecular model with adjacent motifs in antiparallel (Figure 1d), as in AP-TT2 and AP-HT2.

After placing these supramolecular assemblies in a previously equilibrated solvent box with 30464 explicit water molecules, adding the corresponding counterions, and equilibrating the resulting systems using the protocols indicated in the Methods section, production MD simulations were conducted for 10 ns at 298 K (three replicas for each

model, which slightly differ in the inter-sheets distances used to construct the initial structures). After 10 ns of production trajectories, the simulations of models that retain certain organized association (*i.e.* P-TT2, P-HT2 and P-RD) were extended up to 20 ns. The main criterion to evaluate the suitability of each tested model will be centered on the ability of each assembly to retain a coherent association. However, such a criterion must be flexible enough to allow fluctuations in the structure, since the incipient assembly (in biological context mainly referred as complex seed) should be able to sustain the inherent molecular motions of a liquid solvent. Within this context, it should be noted that the only available information on the core organization in the studied system is related to two very distinctive distances,<sup>11</sup> the average inter-strand distance within  $\beta$ -sheets constituted by independent Fmoc-RGDS molecules (between 4.5 and 5.5 Å) and the average lateral spacing of  $\beta$ -sheets (between 8.0 and 12.0 Å). Hence, any model that could explain the final fibre must be able to retain this minimal structural consistency.

Figure 2 represents the structure of the six simulated assemblies at the beginning and at the end of the simulations, while the temporal evolution of the analysed geometric parameters and their average values are provided in the Electronic Supporting Information (ESI). These parameters, which are schematically represented in Figure S1, correspond to the distance between two adjacent  $\beta$ -strands located within the same  $\beta$ -sheet ( $d_{st}^n$ , where  $n$  ranges from 1 to 6 and refers to the number of intra-strand distances within a sheet, indicating that each of the six intra-sheet distances listed in Table S1 corresponds to the average for a sheet made with seven molecules), the distance between adjacent sheets laterally associated ( $d_s^n$ , where  $n$  ranges from 1 to the maximum number of pairs than can be constructed for a given model considering

laterally assembled sheets) and the distance between adjacent aligned sheets ( $d_L^n$ , where  $n$  ranges from 1 to the maximum number of pairs that can be constructed for a given model considering aligned sheets). Both  $d_{st}^n$  and  $d_s^n$  were calculated by averaging all  $C^\alpha \cdots C^\alpha$  distances between strands belonging to the same or to adjacent sheets, respectively, while  $d_L^n$  was computed as the distance between the two close and aligned strand edges considering the center of masses of the Fmoc group and/or the carbon atom of the C-terminal carboxylic group.

Details about  $d_{st}^n$ ,  $d_s^n$  and  $d_L^n$  averages and their fluctuations, which are expressed through the standard deviations, are provided in Tables S1 and S2 for 10 and 20 ns trajectories, respectively. Although the P-TT2, P-HT2 and P-RD models (Figures 2a, 2d and 2e, respectively) remain relatively stable during the first 10 ns of simulation, as is reflected by the acceptable geometric parameters (Table S1), the three assemblies are disrupted when the trajectories are extended to 20 ns. The P-TT2 model undergoes some punctual but drastic intra- and inter-sheet deformations (*e.g.* average  $d_{st}^3$ ,  $d_s^2$  and  $d_L^2$  values are  $8.01 \pm 5.79$ ,  $16.96 \pm 2.45$  and  $27.57 \pm 12.98$  Å, respectively, after 20 ns). Although in the P-HT2 model the strands remain associated within the  $\beta$ -sheets, some  $d_s^n$  and  $d_L^n$  grow to  $\sim 20$  and  $\sim 40$  Å, respectively, evidencing the complete loss of the regularity in the inter-sheet association. Finally, the P-RD model exhibits an intermediate behaviour in which both the intra-sheet and inter-sheet distances grow more moderately, reaching values that range from 6 and 8 Å and from 11 to 15 Å, respectively. This behaviour provokes a loss of the initial assembly after 20 ns, even though the distribution of the sheets in final structure suggests some kind of organization (discussed in next sub-section). All these features are graphically

illustrated in Figure 3, which compares the  $d_{st}^n$ ,  $d_s^n$  and  $d_L^n$  distributions calculated for the P-TT2 and P-RD models considering the 0-10 ns and 0-20 ns time frames. The antiparallel assemblies AP-TT2, AP-HT2 and AA-RD (Figures 2b, 2c and 2f, respectively) resulted very unstable. In the former and the latter the average inter-sheet distances are consistent with a loss of the association between  $\beta$ -sheets, while in the AP-HT2 the three geometric parameters grow very rapidly showing a complete loss of the intra- and inter-sheet interactions.

Although results displayed in this section indicate that 3D models derived from stable 2D building blocks are not able to hold regular associations, accurate analysis of the recorded trajectories allowed us to extract some important conclusions. Head-to-tail assemblies of  $\beta$ -sheets do not favour the formation of sufficiently regular structures, which is a necessary condition to build fibril supramolecular structures. On the other hand,  $\beta$ -sheets featuring antiparallel disposition of strands were totally unfeasible. A very relevant feature is that, when laterally assembled  $\beta$ -sheets are constituted by parallel oriented strands (*i.e.* P-RD model), the organization looks to remain partially regular after 20 ns. However, even in the best case of the three MD replicas, the final outcome for the P-RD model was more like an amorphous than a regular aggregation (Figure 2e). Careful visualization of the three replicated trajectories for the P-RD model clearly suggested that we underestimated the supramolecular chemistry associated to the fibril formation by constructing parallel models through simple translation of the building blocks. This feature is reflected in Figure 4, which depicts the starting coordinates and one of the final snapshots considering different points of view. More specifically, visualization of the P-RD trajectories suggested that the disorganization of the supramolecular 3D starting structure was driven by the Fmoc groups that tried to

assemble. In the next sub-section the compatibility of this behaviour with previously explored association modes is analysed.

### **Self-assembly into fibril supramolecular structures driven by hydrophobic Fmoc groups**

Fmoc-RGDS  $\beta$ -sheets were built using 24  $\beta$ -strands and, subsequently, laterally assembled as in P-RD models considering blocks of 4, 5 and 6 sheets. The initial disposition of the sheet for these new 3D models, which have been denoted P-RD# with # = 4, 5 and 6, is compared in Figure 5 with the structure recorded after 20 ns of MD production (three replicas for each model, starting configurations exhibiting small differences in intra- and inter-sheet distances). The objective of these simulations is to examine if the results in the previous subsection, in which the P-RD regular association evolved towards a disordered structure, was due to the fact that the number of  $\beta$ -strands considered in each sheet was relatively low (*i.e.* only 7).

The temporal evolution of the geometric parameters,  $d_{st}^n$  and  $d_s^n$ , for a representative trajectory of each model is displayed in Figures S8-S10, while the average values and their fluctuations are listed in Table S3. The results clearly indicate that the maximum stabilization comes from the lateral association of four sheets. Thus, the average  $d_{st}^n$  and  $d_s^n$  values for P-RD4 range from  $4.83 \pm 0.28$  to  $4.94 \pm 0.27$  Å and from  $8.46 \pm 0.33$  to  $9.15 \pm 0.33$  Å, respectively. Moreover, the final organization displayed in Figure 4a would point towards the correct supramolecular association. Regarding to P-RD5 and P-RD6, although intra-sheet  $d_{st}^n$  values remain close to those mentioned for P-RD4, inter-sheet  $d_s^n$  grow to  $11.20 \pm 1.82$  and  $12.75 \pm 3.33$  Å, respectively. On the other hand, the

energy contribution associated to intermolecular non-bonding interactions, which include Fmoc...Fmoc stacking, hydrogen bonds, salt bridges and van der Waals, was  $-176.7 \pm 2.9$ ,  $-173.7 \pm 2.0$  and  $-178.5 \pm 3.8$  kcal/(mol-molecule) for P-RD4, P-RD5 and PRD-6, respectively. These results, which were calculated considering the last 5 ns of each production trajectory, indicates that enlargement of the number of sheet does not result in any clear benefit in terms of energetic stability.

Overall, these results suggest that the P-RD4 model can be considered as a good building block for the construction of organized fibrils. Thus, analysis of the three replicated trajectories recorded for the P-RD4 packing suggested that fibrils could be based on the aligned assembly of two P-RD4 building blocks. Preliminary calculations (*i.e.* energy minimization and very short MD runs; not shown) using an implicit solvation model rather than explicit solvent molecules indicated that an assembly like that schematically displayed in Figure 6a, which has been denoted 2(P-RD4), provides a fairly stable fibril organization.

To gain more information about the stability of the 2(P-RD4) model, 20 ns production MD runs (three replicas) were performed considering 24 strands for each sheet immersed in a dielectric medium (*i.e.* implicit solvation model). Figure 6b represents recorded snapshots at different time intervals, while the temporal evolution and the average values of the intra- and inter-sheet geometric parameters with their fluctuations are provided in Figures S11 and Table S4, respectively. Despite the huge dimensions of this model and the large number of geometric parameters to keep under control, all intra- and inter-sheet distances remained stable during the whole trajectory, fluctuations being infrequent and small. Interestingly, average inter-sheet distances,  $d_s^n$  and  $d_L^n$ , exhibit two preferential values. Thus,  $d_s^n$  values are grouped around  $\sim 7.7$  and

$\sim 9.3$  Å while  $d_L^n$  adopts values close to  $\sim 25.5$  and  $\sim 29.9$  Å. This feature has been attributed to the twist of the aggregated  $\beta$ -sheets around a common helical axis that coincides with the fiber axis, which occurs spontaneously during the first nanoseconds of simulation. Aggregation of these twisted  $\beta$ -sheets results in a coil along the fiber axis, helping create additional favorable intra- and inter-sheet interactions.<sup>49</sup> Thus, the twist of  $\beta$ -sheets in the 2(P-RD4) packing model allows a further aggregated and stably condensed structural alignment. This supramolecular organization around the fibre axis without breaking apart  $\beta$ -strands within the same  $\beta$ -sheet results in a bimodal distribution of the inter-sheet parameters. This is reflected in the corresponding distribution profiles displayed in Figure 7. Similar observations have been reported in the assembly of other small amphiphilic peptides.<sup>50</sup>

The apparent success of the 2(P-RD4) model should be considered with caution since the competition between the different types of hydrogen bonding interactions (*e.g.* strand $\cdots$ strand and strand $\cdots$ water) was not taken into account once solvent was reduced to a simple continuum dielectric medium. Also, it is worth noting that the description of interactions provided by classical force-fields is less accurate than that offered by quantum mechanical calculations.<sup>51</sup> Despite of these limitations, which are general to all force-fields, it should be noted that Amber parameters were explicitly developed to aid in the process of manipulating peptides and proteins and, therefore, they are considered among the more accurate for the simulation of this kind of biomolecules. The next subsection is devoted to explore the stability of the 2(P-RD4) supramolecular in explicit water considering production trajectories of up to 200 ns.

### **Atomic structure and hierarchical assembly of Fmoc-RGDS fibril: Testing the stability of the 2(P-RD4)**

MD simulations in explicit water of the 2(P-RD4) supramolecular structure (3 replicas) were performed considering sheets with 9  $\beta$ -strands. Inspection of the intermediate and final snapshots (Figure S12) clearly indicates that the stability of these packing model is not altered by specific peptide···water interactions, supporting our previous observations using implicit solvation. This is corroborated by the temporal evolution and the averages of the intra- and inter-sheet geometric parameters (Figures S13 and Table S5, respectively). The bimodal distribution of averaged inter-sheet distances remains in explicit water, even though the coiling around the fiber axis is, obviously, much less evident in this case than in the system with 24  $\beta$ -strands per sheet.

As a final test of stability, the production trajectories of two of the three replicas conducted for the 2(P-RD4) model in explicit water were enlarged from 20 to 200 ns. Our analyses showed that the hierarchical assembly associated to this supramolecular structure is highly stable in this long trajectory. Figure 8, which represents snapshots separated by 40 ns intervals, indicates the absence of changes in both intra- and inter-sheet interactions, the only movement shown by this supramolecular organization being associated to the dynamical breathing typically observed for stable systems in solution. These breathing fluctuations essentially affect the interaction between the two P-RD4 building blocks through the hydrophobic core formed by the Fmoc units. Thus, the hydrophobic region periodically expands and contracts, as is evidenced by the snapshots extracted at 80, 120, 160 and 200 ns. The dynamical behavior of this stable supramolecular structure is reflected in Table S6, which lists the average intra- and inter-sheet distances with their standard deviations. Thus, the deviations of a few of

such geometric parameters, especially  $d_L^1$  and  $d_L^4$ , grows to  $\sim 3$  Å. These relatively high deviations are not detected for averages performed over intervals of 20 ns only.

Detailed structural analysis of the 2(P-RD4) model has been performed considering both the  $C^\alpha \dots C^\alpha$  and  $C^{\text{FMOC}} \dots C^{\text{FMOC}}$  partial radial distribution function ( $g_{C^\alpha-C^\alpha}$  and  $g_{C^{\text{FMOC}}-C^{\text{FMOC}}}$ , respectively), where  $C^\alpha$  refers to the  $\alpha$ -carbon atoms of the RGDS sequence and  $C^{\text{FMOC}}$  corresponds to the carbon atoms of the Fmoc groups. Figures 9a and 9b plot the  $g_{C^\alpha-C^\alpha}(r)$  and  $g_{C^{\text{FMOC}}-C^{\text{FMOC}}}(r)$  functions, respectively, calculated using the 200 ns production trajectories. The X-ray diffraction patterns of a regular cross- $\beta$  structure typically show an equatorial reflection between 9.5 and 10.5 Å that corresponds to the average distance between sheets to form lamellae (*i.e.*  $\beta$ -sheet stacking spacing) and a meridional reflection with a primary peak at 4.5-5.5 Å, which is associated to the distance between  $\beta$ -strands. For Fmoc-RGDS these spacings appeared at 4.48 and 9.92 Å, respectively.<sup>11</sup> As can be seen, the main peaks observed at  $g_{C^\alpha-C^\alpha}(r)$  and  $g_{C^{\text{FMOC}}-C^{\text{FMOC}}}(r)$  functions closely correspond to such values, indicating an excellent correspondence between the 2(P-RD4) cross- $\beta$  model and experimental data. Other important spacings at 6.76 and  $11.6 \pm 0.1$  Å are also clearly detected in Figure 9.

## DISCUSSION

This work provides a unique series of simulations devoted to understanding and characterizing at the atomic level the hierarchical assembly of Fmoc-RGDS molecules. For this purpose, a bottom-up approach has been applied to progressively ascertain how the constituent Fmoc-RGDS molecules, which adopts a straight  $\beta$ -strand conformation, organizes into  $\beta$ -sheets, 2D packed  $\beta$ -sheets and, finally, 3D fibrils.

Previous simulations on a single  $\beta$ -sheet<sup>11</sup> or two associated  $\beta$ -sheets<sup>24</sup> indicated that the antiparallel configuration of  $\beta$ -strands within the same sheet and the antiparallel association of sheets are unstable for the Fmoc-RGDS peptide sequence. These two features are fully consistent with the results obtained in this work. Thus, supramolecular arrangements constructed using antiparallel  $\beta$ -sheets (AA-RD) and considering an antiparallel disposition of motifs made with parallel  $\beta$ -sheets (AP-TT2 and AP-HT2) have been found to be highly unstable (*i.e.* initial assemblies are completely lost in less than 10 ns). Within this context, it should be noted that both antiparallel and parallel  $\beta$ -sheets have been found in fibrils and amyloid-like crystals formed by short fragments of full-length amyloid-forming proteins.<sup>52-56</sup> Although the antiparallel alignment may produce more favourable electrostatic interactions, either parallel or antiparallel alignments can produce favourable intermolecular hydrophobic interactions.<sup>57</sup>

Self-assembly in the proposed 2(P-RD4) model is based on the formation of  $\pi$ - $\pi$  stacking interactions between hydrophobic Fmoc groups. This interaction apparently has a dual role: it allows the hydrophobic moieties to collapse and escape from the polar solvent (Figure 10a), whereas the  $\pi$ -stacked aromatic rings settle and anchor each sheet organization. The global stability is then completed by the formation of inter-strand hydrogen bonds between the Asp main chain amide group and the guanidine group belonging to Arg set on neighbouring sheets (see *Inter II* and *Inter III* in Figure 10a). The origins of this singular interaction are discussed below.

The organization presented above revolves around a hydrophobic core in which the Fmoc-groups of  $\beta$ -strands belonging to the same  $\beta$ -sheet form columns of  $\pi$ -stacked aromatic rings arranged in parallel (Figure 10b, interactions labelled *intra I*). The face-to-face distance between adjacent interacting Fmoc groups within a column is around

~4.0 Å. As mentioned above, the eight columns of  $\pi$ -stacked Fmoc groups pack laterally forming a compact and dense hydrophobic core, in which two central columns are surrounded by three adjacent columns at each side. This efficient packing mode protects the aromatic groups from penetrating water molecules. In the last decade, it has been shown that the addition of aromatic groups, as Fmoc, to the *N*-terminus of some small peptides allows them to form stable hydrogels.<sup>22,58-62</sup> The morphology and dimensions of the obtained nanostructures were found to depend on the force that drives the own self-assembly process<sup>22,59,60</sup> and the peptide sequence.<sup>61,62</sup> Many of these short peptides, as for example Fmoc-FF<sup>22</sup> and Fmoc-FF-Pyr<sup>+</sup> (where Pyr<sup>+</sup> represents a pyridinium moiety at the *C*-terminus),<sup>63</sup> are arranged as antiparallel organizations of  $\beta$ -sheets with the Fmoc groups acting like a zipper to bring neighbouring sheets together.

From a topological point of view, the 2(P-RD4) model proposed for Fmoc-RGDS closely resembles the cross- $\beta$  core structure of amyloids. This is also composed of arrays of  $\beta$ -sheets with  $\beta$ -strands running perpendicular to the long axis of the fibrils, ~4.7 Å apart, and forming hydrogen bonds that run approximately parallel to the fibril axis.<sup>63,64</sup> Thus, in addition to the discussed Fmoc...Fmoc interactions at the hydrophobic core stabilizing the two stacked layers of four assembled  $\beta$ -sheets, the hierarchically assembly of the constituent  $\beta$ -strands involves a rich variety of intermolecular interactions. Each  $\beta$ -sheet of the 2(P-RD4) model is stabilized by a network inter-backbone hydrogen bonds involving the parallel in-register Fmoc-RGDS  $\beta$ -strands (Figure 10b, backbone...backbone hydrogen bonds are labelled *intra* II). The Fmoc groups appetite for their own class drags the strands together and limits the spatial arrangements that charged side chains can adopt once in axial disposition of strands and sheets is appointed. As a consequence, both Arg and Asp side chains

belonging to the same chains form intra-strand salt bridges (Figure 10c, interactions corresponding to label *Strd* I). The preference towards intra-strand salt bridges before inter-strand interactions was observed in other cross- $\beta$  based fibrils, in derivatives of natural peptides.<sup>65</sup>

Due to the shorter length of the Asp side chain, there is an alteration on the  $\beta$ -sheet conformation of each strand at this site. The conformation adopted by this amide group is set perpendicular to the fibril growth and in this orientation forms dipole-dipole interactions with the Asp carboxyl group (Figure 10c, set of interaction shown as *Strd* II). This interaction prevents the formation of a fourth spine of amide $\cdots$ amide hydrogen bonds that this peptide should present if all its residues adopted a standard  $\beta$ -sheet conformation. However, the odd orientation of the central amide allows their carbonyl groups to form inter-sheet polar interactions previously mentioned (*inter* II and III in Figure 10a), which aided anchoring the four strands lateral pack.

According to this microscopic study, hydrogen bonding, salt bridges and  $\pi$ - $\pi$  stacking interactions coexist in the highly ordered packing network 2(P-RD4) proposed for Fmoc-RGDS amphiphile. However, analysis of the overall results, including all the examined unstable models, clearly demonstrate that specific hydrogen bonds and salt bridges between adjacent molecules located at the same and neighbouring  $\beta$ -sheets, respectively, are not as important as Fmoc $\cdots$ Fmoc interactions in driven and stabilizing the self-assembly of Fmoc-RGDS in the supramolecular structure. Thus, simple lateral packing modes, in which Fmoc hydrophobic groups were located at the solvent accessible surface, resulted highly unstable. Taken together, the results obtained in this and previous works<sup>11,24</sup> suggest that extensive hydrogen bonds and salt bridges play less of a role in forming and stabilizing the fibril supramolecular structure and a greater role

in determining the nature of the higher order aggregation in intermediate states, as the building blocks. Hierarchical assembly in amphiphilic peptide fibrils was proposed one decade ago by Aggeli *et al.*,<sup>66</sup> who suggested that initially assembled  $\beta$ -sheets form higher-order structures driven by the balance between hydrophobicity and hydrophilicity of the intermediate structures. In recent years this concept has been extended to other biological materials.<sup>67-71</sup> According to this discussion, we propose that amphiphilic Fmoc-RGDS self-assemble into the 2(P-RD4) packing model, giving place to fibril structures of width  $\sim 5$  nm. These fibrils further aggregate in couples into the higher-order supramolecular aggregated, such as the fibre-like structures observed by cryo-TEM and small angle X-ray scattering (diameter  $\sim 10$  nm).<sup>11</sup>

With the aim to quantify the above referred interactions among the  $\beta$ -strands of the  $\beta$ -sheets: hydrogen bonding, salt bridges, and  $\pi$ - $\pi$  stacking, a quantum chemical analysis has complemented the above MD calculations. In particular, we took a snapshot from the MD simulation of the 2(P-RD4) and we considered a model of the 8  $\beta$ -sheets with 3  $\beta$ -strands each one. All quantum mechanics calculations were performed at this geometry, without further relaxation. Table 2 includes the interaction energies between different pairs of  $\beta$ -strands, either belonging to different  $\beta$ -sheets, thus giving rise to lateral interactions, or to the same  $\beta$ -sheet, thus evaluating the stacking interactions. This interaction energy is further decomposed by means of an energy decomposition analysis (EDA, see Methods section below). The corresponding numbering of the  $\beta$ -sheets as well as a 3D plot of the distribution of the eight  $\beta$ -sheets is depicted in Figure 11.

First we focus on the lateral interactions of one P-RD4 building block, *i.e.*  $\beta$ -sheets 1 to 4. The interaction energies,  $\Delta E_{\text{int}}$ , amount from -13.1 (1t-2t) to -16.7 (2t-3t) and to -

19.9 (3t-4t) kcal/mol. As can be seen from Figures 9a and 10b, the two latter lateral P-RD4 units adopt a geometry in which the three rings of the Fmoc groups are oriented in parallel, whereas the Fmoc of  $\beta$ -sheet 1 is oriented perpendicular to that of 2, thus justifying the lower  $\Delta E_{\text{int}}$  value. If we go into the corresponding decomposition of  $\Delta E_{\text{int}}$  into Pauli, electrostatic, orbital, and dispersion interactions, it is clearly observed how the main contribution comes from the dispersion interaction. This mainly arises from the  $\pi$ - $\pi$  stacking of the aromatic rings of the Fmoc groups. Afterwards, the most relevant component is the electrostatic that can be attributed to the inter-strand hydrogen bonds between the Asp main chain amide group and the guanidine group of Arg on the neighboring  $\beta$ -sheets, as discussed below (see *Inter II* and *Inter III* in Figure 10a). There are some orbital interactions mainly originated from the same hydrogen bonds. The EDA of different hydrogen bonds show that the attractive interactions are about 30-50% orbital interactions depending on the type of the hydrogen bond and the rest are electrostatic interactions.<sup>72</sup> Our results show that the network of hydrogen bonds formed between 2t and 3t or 3t and 4t is stronger than that of 1t and 2t.

With the aim to better know the role played by each part of the  $\beta$ -strand in such interactions, we have cut the bond between the Fmoc and the peptide chains and obtained two radical doublet fragments. With these fragments we have carried out the same EDA analysis (see Table S1 in the Supporting Information) for 2t-3t and 3t-4t  $\beta$ -strands separately for Fmoc and the peptide chains fragments. The interaction energy appears to be larger between the peptide chains (-29 – -35 kcal/mol) than between Fmoc units (*ca.* -23 kcal/mol), thus proving the larger contribution of the peptide chains in the inter-strand interactions. If we go into the EDA components, in both cases the largest contribution comes from the orbital interactions. These orbital interactions are

artificially high because the radical character of the Fmoc and peptide chain units. The Pauli repulsions between the Fmoc groups make the most important difference, being larger in this case than between the peptide chains. On the other hand, the dispersion interaction between the aromatic groups in the Fmoc groups is also much more important than the dispersion in the peptide chains. As a summary,  $\pi$ - $\pi$  stacking interactions between Fmoc groups and hydrogen bond interactions in peptide chains makes the largest contribution to the lateral  $\beta$ -strand interactions.

The determinant interaction among the different  $\beta$ -strands comes from the  $\pi$ - $\pi$  stacking interactions between the aromatic rings of Fmoc groups, as stated above. In this case, the  $\Delta E_{\text{int}}$  between two stacked  $\beta$ -strands of the same  $\beta$ -sheet is calculated to be  $-40.6 \text{ kcal mol}^{-1}$ , an interaction that is two or three times larger than that of the inter-strand ones. The determinant component is again the dispersion interaction, but now both electrostatic and orbitals interactions are of the same magnitude. It is worth mentioning that not only the  $\pi$ - $\pi$  stacking contributes to this  $\Delta E_{\text{int}}$ , as the hydrogen bond and salt-bridges intra-strand interactions are also present (see Figures 10b and 10c). On the other hand, if two alternated stacked  $\beta$ -strands are considered, the interaction energy is almost zero as can be expected from the separation between the two  $\beta$ -strands of about  $8 \text{ \AA}$ . Overall, these quantum chemical calculations confirm the decisive role played by the  $\pi$ - $\pi$  stacking interactions between the rings of the Fmoc groups, although both inter-strand and intra-strand hydrogen bond (as seen in the  $\Delta E_{\text{oi}}$  and  $\Delta V_{\text{elstat}}$  terms) and salt bridges (as seen in the  $\Delta V_{\text{elstat}}$  component) interactions do also play a non-negligible role.

Finally, Table 2 also shows how the interaction energy clearly diminishes for the crossed interactions between the top  $\beta$ -strand of one  $\beta$ -sheet and the middle ( $\Delta E_{\text{int}}(1t-$

2m) = -4.0 kcal/mol) or bottom ( $\Delta E_{\text{int}}(1\text{t}-2\text{b}) = -1.3$  kcal/mol)  $\beta$ -strand of the  $\beta$ -sheet next to it, due to the longer distance between them. Something similar happens when the top  $\beta$ -strands of two  $\beta$ -sheets belonging to the two building blocks of 2(P-RD4) fibril are analysed, obtaining values between -0.2 (3t-6t) to -3.2 (3t-5t) kcal mol<sup>-1</sup>, depending on the distance between their Fmoc groups, as their peptide chains cannot interact. These crossed interactions, which are weak but more frequent than stacked and lateral  $\beta$ -strands interactions, correspond basically to dispersion interactions.

## CONCLUSIONS

Here we used theoretical calculations to provide atomistic details of the hierarchical assembly in Fmoc-RGDS organized fibrils. MD simulations considering a large number of models indicate that Fmoc-RGDS fibrils consist of cross- $\beta$  a structure containing arrays of  $\beta$ -sheets with  $\beta$ -strands running perpendicular to the long axis of the fibrils and forming hydrogen bonds that run approximately parallel to the fibril axis. This organization is achieved around a hydrophobic core involving columns of  $\pi$ -stacked Fmoc rings arranged in parallel. Accordingly, both MD simulations on realistic models and quantum mechanical calculations on reduced models indicate that the fibrils are stable because of: 1) the  $\pi$ - $\pi$  stacking interactions between hydrophobic Fmoc groups; and 2) inter-strand interactions between the Asp main chain amide group and the guanidine group belonging to Arg set on neighbouring sheets. Fmoc $\cdots$ Fmoc aromatic interactions play a crucial role in stabilizing the supramolecular self-assembly while hydrogen bonds and salt bridges are essential to stabilize the intermediate states formed by aggregated building blocks. All these interactions result in a very compact and dense packing mode with a hydrophobic core protecting from penetrating water molecules.

Theoretical calculations such as those reported in this work are powerful tools not only to provide complementary information at the molecular level that cannot be reached experimentally, but also for the bottom-up design of novel nanomaterials, as for example new fibrils for therapeutic use.

## ACKNOWLEDGEMENTS

This work has been supported by MICINN and FEDER funds (projects number MAT2012-34498 and CTQ2011-23156/BQU), Generalitat de Catalunya (XRQTC, 2014SGR931, and ICREA Academia 2014 prize for M.S.). We are grateful to the Consorci de Serveis Universitaris de Catalunya (CSUC) for providing excellent services and computational resources. J.P. thanks the Netherlands Organization for Scientific Research (NWO-CW, NWO-EW, and NWO-ALW) for financial support.

## REFERENCES

1. P. Nguyen and P. Derreumaux, *Acc. Chem. Res.*, 2014, **47**, 603.
2. F. Baftizadeh, X. Biarnes, F. Pietrucci, F. Affinito and A. Laio, *J. Am. Chem. Soc.*, 2012, **134**, 3886.
3. D. Thirumalai, G. Reddy and J. E. Straub, *Acc. Chem. Res.*, 2012, **45**, 93.
4. E. L. Bakota, O. Sensoy, B. Ozgur, M. Sayar and J. D. Hartgerink, *Biomacromolecules*, 2013, **14**, 1370.
5. E. Brandenburg, H. von Berlepsch, U. I. M. Gerling, C. Bottcher and B. Kokschi, *Chem. Eur. J.*, 2011, **17**, 10651.
6. O. S. Lee, S. I. Stupp and G. C. Schatz, *J. Am. Chem. Soc.*, 2011, **133**, 3677.

7. Y. Raz, B. Rubinov, M. Matmor, H. Rapaport, G. Ashkenasy and Y. Miller, *Chem. Commun.*, 2013, **49**, 6561.
8. N. Haspel, D. Zanuy, C. Alemán, H. Wolfson and R. Nussinov, *Structure*, 2006, **14**, 1137.
9. I. Buch, B. R. Brooks, H. J. Wolfson and R. Nussinov, *NanoLett.*, 2009, **9**, 1096.
10. P. W. J. M. Frederix, W. J. M. Pim, G. G. Scott, Y. M. Abul-Haija, D. Kalafatovic, C. G. Pappas, N. Javid, N. T. Hunt, R. V. Ulijn and T. Tuttle, *Nature Chem.*, 2015, **7**, 30.
11. V. Castelletto, C. M. Moulton, G. Cheng, I. W. Hamley, M. R. Hicks, A. Rodger, D. E. Lopez-Perez, G. Revilla-Lopez and C. Aleman, *Soft Matter*, 2011, **7**, 11405.
12. E. Ruoslahti and M. D. Pierschbacher, *Science*, 1987, **238**, 491.
13. R. O. Hynes, *Cell*, 1992, **69**, 11.
14. M. Tirrell, E. Kokkoli and M. Biesalski, *Surf. Sci.*, 2002, **500**, 61.
15. D. L. Hern and J. A. Hubbell, *J. Biomed. Mater. Res.*, 1998, **39**, 266.
16. M. S. Hahn, L. J. Taite, J. J. Moon, M. C. Rowland, K. A. Ruffino and J. L. West, *Biomaterials*, 2006, **27**, 2519.
17. K. M. Eckes, X. J. Mu, M. A. Ruehle, P. Y. Ren and L. J. Suggs, *Langmuir*, 2014, **30**, 5287.
18. A. Saha, S. Bolisetty, S. Handschin and R. Mezzenga, *Soft Matter*, 2013, **43**, 10239.
19. X. J. Mu, K. M. Eckes, M. M. Nguyen, L. J. Suggs and P. Y. Ren, *Biomacromolecules*, 2012, **13**, 3562.
20. R. Orbach, I. Mironi-Harpaz, L. Adler-Abramovich, E. Mossou, E. P. Mitchell, V. T. Forsyth, E. Gazit and D. Seliktar, *Langmuir*, 2012, **28**, 2015.

21. M. Hughes, L. S. Birchall, K. M. Zuberi, L. A. Aitken, S. Debnath, N. Javid and R. V. Ulijn, *Soft Matter*, 2012, **8**, 11565.
22. A. M. Smith, R. J. Williams, C. Tang, P. Coppo, R. F. Collins, M. L. Turner, A. Saiani and R. V. Ulijn, *Adv. Mater.*, 2008, **20**, 37–41.
23. V. Castelletto, G. Cheng, B. W. Greenland, I. W. Hamley and P. J. F. Harris, *Langmuir*, 2011, **27**, 2980.
24. D. E. López-Pérez, G. Revilla, I. W. Hamley and C. Alemán, *Soft Matter*, 2013, **9**, 11021.
25. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179.
26. H. C. Andersen, *J. Comput. Phys.*, 1983, **52**, 24.
27. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.
28. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
29. A. Onufriev, D. Bashford, and D. A. Case. *J. Phys. Chem.*, 2000, **104**, 3712
30. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781.
31. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684.
32. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999.
33. G. J. Martyna, D. L. Tobia and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177.

34. G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders and T. Ziegler, *J. Comput. Chem.*, 2001, **22**, 931.
35. S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463.
36. S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787.
37. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, 132, 154104.
38. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, 32, 1456.
39. A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.
40. C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.
41. K. Morokuma, *Acc. Chem. Res.*, 1977, **10**, 294.
42. T. Ziegler and A. Rauk, *Theor. Chim. Acta*, 1977, **46**, 1.
43. T. Ziegler, A. Rauk and E. J. Baerends, *Theor. Chim. Acta*, 1977, 43, 261.
44. F. M. Bickelhaupt, A. Diefenbach, S. P. de Visser, L. J. de Koning and N. M. M. Nibbering, *J. Phys. Chem. A*, 1998, **102**, 9549.
45. F. M. Bickelhaupt and E. J. Baerends in *Reviews in Computational Chemistry*, Vol. 15 (Eds.: K. B. Lipkowitz and D. B. Boyd), Wiley-VCH, New York, **2000**, pp. 1-86
46. M. von Hopffgarten and G. Frenking, *WIREs-Comput. Mol. Sci.*, 2012, **2**, 43.25.
47. J.-P. Colletier, A. Laganowsky, M. Landau, M. Zhao, A. B. Soriaga, L. Goldschmidt, D. Flot, D. Cascio, M. Sawaya and D. Eisenberg, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 16938.
48. J. Madine, H. A. Davies, C. Shaw, I. W. Hamley and D. A. Middleton, *Chem. Commun.*, 2012, **48**, 2976.
49. P.-N. Cheng, J. D. Pham and J. S. Nowick, *J. Am. Chem. Soc.*, 2013, **135**, 5477.

50. C. Hauser, R. deng, A. Mishra, Y. Loo, U. Khoe, F. Zhuang, D. W. Cheong, A. Accardo, M. B. Sullivan, C. Riekel, J. Y. Ying and U. A. Hauser, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 1361.
51. R. S. Paton and J. M. Goodman, *J. Chem. Inf. Model.*, 2009, **49**, 944.
52. J. T. Nielsen, M. Bjerring, M. D. Jeppesen, R. O. Pedersen, J. M. Pedersen, K. L. Hein, T. Vosegaard, T. Skrydstrup, D. E. Otzen and N. C. Nielsen, *Angew. Chem. Int. Ed.*, 2009, **48**, 2118.
53. M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riekel and D. Eisenberg, *Nature*, 2007, **447**, 453.
54. R. B. Wickner, F. Dyda and R. Tycko, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 2403.
55. W. Qiang, W. M. Yau, Y. Q. Luo, M. P. Mattson and R. Tycko, *Proc. Natl. Acad. Sci. U.S.A.*, 2012, **109**, 4443.
56. R. Tycko, K. L. Sciarretta, J. Orgel and S. C. Meredith, *Biochemistry*, 2009, **48**, 6072.
57. R. Tycko and R. B. Wickner, *Acc. Chem. Res.*, 2013, **46**, 1487.
58. R. V. Ulijn and A. M. Smith, *Chem. Soc. Rev.*, 2008, **37**, 664.
59. Z. M. Yang, H. W. Gu, Y. Zhang, L. Wang and B. Xu, *Chem. Commun.*, 2004, 208.
60. M. Reches and E. Gazit, *Nat. Nanotechnol.*, 2006, **1**, 195.
61. V. Jayawarna, M. Ali, T. A. Jowitt, A. E. Miller, A. Saiani, J. E. Gough and R. V. Ulijn, *Adv. Mater.*, 2006, **18**, 611.
62. S. Debnath, A. Shome, D. Das and P. K. Das, *J. Phys. Chem. B*, 2010, **114**, 4407.

63. A. Fitzpatrick, G. T. Debelouchina, M. J. Bayro, D. K. Clare, M. A. Caporini, V. S. Bajaj, C. P. Jaronec, L. Wang, V. Ladizhansky, S. A. Müller, C. E. MacPhee, C. A. Waudby, H. R. Mott, A. De Simone, T. P. J. Knowles, H. R. Saibil, M. Vendruscolo, E. V. Orlova, R. G. Griffin, and C. M. Dobson *Proc. Natl. Acad. Sci.*, 2013, **110**, 5468.
64. S.-Y. Ow and D. E. Dunstan, *Prot. Sci.*, 2014, **23**, 1315.
- 65 D. Zanuy, N. Haspel, H. H. Tsai, B. Ma, K. Gunasekaran, H.J. Wolfson and R. Nussinov. *Phys. Biol.*, 2004, **1**, 89.
66. A. Aggeli, I. A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. B. McLeish, A. N. Semenov and N. Boden *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 11857.
67. G. M. Whitesides and B. Grzybowski, *Science*, 2002, **295**, 2418.
68. M. J. Buehler and Y. C. Yung, *Nat. Mater.*, 2009, **8**, 175.
69. T. P. J. Knowles, T. W. Oppenheim, A. K. Buell, D. Y. Chirgadze and M. E. Welland, *Nat. Nano*, 2010, **5**, 204.
70. A. W. P. Fitzpatrick, G. T. Debelouchina, M. J. Bayro, D. K. Clare, M. A. Caporini, V. S. Bajaj, C. P. Jaronec, L. Wang, V. Ladizhansky, S. A. Müller, C. E. MacPhee. C- A- Waudy, H. R. Mott, A. De Simone, T. P. J. Knowles, H. R. Saibil, M. Vendruscolo, E. V. Orlova, R. G. Griffin and C. M. Dobson, *Proc. Natl. Acad. Sci. U.S.A.*, 2013, **110**, 5468.
71. S.-Y. Qin, Y. Pei, X.-J. Liu, R.-X.; Zhuo and X.-Z. Zhang, *J. Mater. Chem. B*, 2013, **1**, 668.
72. J. Poater, X. Fradera, M. Solà, M. Duran and S. Simon, *Chem. Phys. Lett.*, 2003, **369**, 248.

**Table 1.** Main characteristics of the systems simulated in this work: dimensions of the simulation box, number of Fmoc-RGDS strands and Na<sup>+</sup> counterions, and number of water molecules for simulations with explicit solvent representation.

Models	Initial box size (a × b × c, in Å <sup>3</sup> )	Number of strands and Na <sup>+</sup> ions <sup>a</sup>	N <sub>w</sub> <sup>b</sup>
<b>Initial Models</b>			
AARR / AP-HT2 / AP-TT2	100×100×100	42	30464
P-HT2 / P-RD / P-TT2	100×100×100	42	30464
<b>P-RD based models</b>			
P-RD/4	300×300×300	96	-
P-RD/5	360×360×360	120	-
P-RD/6	360×360×360	144	-
<b>Final model</b>			
2(P-RD4)	100×100×100	36	31444

<sup>a</sup> Each strand holds a net charge of -1. Thus, all systems present the same number of Fmoc-RGDS chains than Na<sup>+</sup> ions, in order to achieve electrostatic neutrality.

<sup>b</sup> Number of water molecules.

**Table 2.** Interaction energies between two  $\beta$ -strands and the corresponding energy decomposition analysis (in kcal/mol). The numbers stand for  $\beta$ -sheets (see Figure 11), whereas “t” stands for top  $\beta$ -strand of the referred  $\beta$ -sheet, “m” for middle and “b” for bottom.

<b><math>\beta</math>-strand pair</b>	$\Delta E_{\text{Pauli}}$	$\Delta V_{\text{elstat}}$	$\Delta E_{\text{oi}}$	$\Delta E_{\text{disp}}$	$\Delta E_{\text{int}}$
<i>lateral <math>\beta</math>-strands</i>					
1t-2t	9.1	-6.4	-4.4	-11.3	-13.0
1t-2m	3.8	-2.2	-1.3	-4.3	-4.0
1t-2b	0.0	-1.0	-0.1	-0.2	-1.3
2t-3t	14.3	-10.7	-4.8	-15.4	-16.6
3t-4t	12.3	-13.0	-5.4	-13.8	-19.9
4t-5t	2.2	-0.8	-0.8	-3.3	-2.7
2t-5t	0.6	-0.2	-0.2	-1.4	-1.2
3t-5t	4.4	-1.8	-1.2	-4.6	-3.2
3t-6t	0.0	-0.1	0.0	-0.0	-0.1
<i>stacked <math>\beta</math>-strands</i>					
3t-3m	89.3	-36.2	-39.0	-54.7	-40.6
3t-3b	0.1	0.7	-0.3	-1.1	-0.6

## CAPTIONS TO FIGURES

**Figure 1.** Scheme displaying the relative orientation of the  $\beta$ -sheets as well as the position of the Arg and Asp side groups in the six 3D models (right) constructed to study the assembly of Fmoc-RGDS  $\beta$ -strands. These models were derived from four 2D building blocks (left). Each arrow (black and grey) represents the top view of a  $\beta$ -sheet involving 7 explicit peptide molecules. The round end in each arrow refers to the position of the Fmoc groups in the sheet. Round ends at the AA-RD model and building block have been omitted for clarity, avoiding superposition with the arrow-heads. The positions of the Arg and Asp side groups for the peptides containing in each sheet are indicated by blue and red lines, respectively. The origin and labels of the 3D models represented in (a)-to-(d) is described in the text.

**Figure 2.** Structure of the 3D packing models derived from 2D building blocks (see Figure 1) for Fmoc-RGDS at the beginning, after 10 ns and after 20 ns (only parallel models) of MD simulations in explicit water: (a) P-TT2; (b) AP-PT2; (c) AP-HT2; (d) P-HT2; (e) P-RD; and (f) AA-RD. Fmoc groups are depicted in green while the  $\beta$ -strands are represented using yellow arrows. All the structures were depicted using the same point of view. For the P-TT2 (a), P-HT2 (d) and P-RD (e), an additional view is provided by rotating  $90^\circ$  with respect one of the axis. Models were constructed considered 7 strands per  $\beta$ -sheet.

**Figure 3.** Distribution of the indicated geometric parameters ( $d_{st}^n$ ,  $d_s^n$  and  $d_L^n$ ) calculated for the (a) P-TT2 and (b) P-RD models considering the 0-10 ns and 0-20 ns time frames (top and bottom, respectively). For all graphics, each colour profile refers to each of the average distance displayed in Tables S1 and S2.

**Figure 4.** For the P-RD packing model, comparison from different points of view between the disposition of the Fmoc-RGDS strands at the beginning (left) and at the end (right) of the 20 ns production trajectory. The disorganization of the starting lateral packing of  $\beta$ -sheets seems to be provoked by the tendency of Fmoc groups to assembly in a hydrophobic core. Fmoc groups are depicted in green while the  $\beta$ -strands are represented using yellow arrows.

**Figure 5.** Structure of the three models derived from the P-RD one (see Figure 1) at the beginning and after 20 ns of MD simulation. The three models were constructed considering the lateral packing of 4, 5 or 6  $\beta$ -sheets with 24 Fmoc-RGDS  $\beta$ -strands each one: (a) P-RD4; (b) P-RD5; and (c) P-RD6. Fmoc groups are depicted in green while the  $\beta$ -strands are represented using yellow arrows. All the structures were depicted using the same points of view.

**Figure 6.** (a) Scheme displaying the supramolecular assembly, 2(P-RD4), constructed using the P-RD4 model as building block. (b) Structure of the 2(P-RD4) supramolecular assembly at the beginning, after 10 ns, and after 20 ns of MD simulation using an implicit solvation model. Each of the 8  $\beta$ -sheets included in the model was constructed considering 24 Fmoc-RGDS  $\beta$ -strands. Fmoc groups are depicted in green while the  $\beta$ -strands are represented using yellow arrows. All the structures were depicted using the same points of view.

**Figure 7.** Distribution of the indicated geometric parameters ( $d_s^n$  and  $d_L^n$ ) calculated for the 2(P-RD4) supramolecular model model. For both graphics, each colour profile refers to each of the average distance displayed in Table S4.

**Figure 8.** Structure of 2(P-RD4) fibril supramolecular structure at different time intervals of the 200 ns production trajectory. Each of the 8 sheets included in the supramolecular packing was constructed considering 9 Fmoc-RGDS  $\beta$ -strands. Fmoc groups are depicted in green while the  $\beta$ -strands are represented using yellow arrows. All the structures were depicted using the same points of view.

**Figure 9.** Radial distribution functions of (a)  $C^\alpha \dots C^\alpha$  and (b)  $C^{\text{FMOC}} \dots C^{\text{FMOC}}$  pairs, where  $C^\alpha$  refers to the  $\alpha$ -carbon atoms of the RGDS sequence and  $C^{\text{FMOC}}$  to the carbon atoms of the Fmoc groups. The profiles were calculated for the 200 ns production trajectories of the 2(P-RD4) model.

**Figure 10.** Schematic representation of the interactions network responsible of the fibril organization and its stability: (a) Inter-sheet interactions mainly correspond to lateral  $\pi$ - $\pi$  interactions (labeled *Inter I*), combined with intermolecular hydrogen bonds (*Inter II* and *III*, respectively); (b) Each sheet is doubly stabilized by inter-strand  $\pi$ - $\pi$  interactions among consecutive Fmoc rings (*intra I*) and by amide...amide hydrogen bonds (*Intra II*); (c) Each strand within a  $\beta$ -sheet forms intra-strand salt bridges (Arg...Asp side chains, labeled *Strd I*), and dipole...dipole interactions between accessible polar groups of the same strand (*Strd II*, see text for details)

**Figure 11.** (a) Scheme with the corresponding numbering of the supramolecular assembly 2(P-RD4); (b) Structure of the 2(P-RD4) with only the top  $\beta$ -strands of each  $\beta$ -sheet and the corresponding numbering.

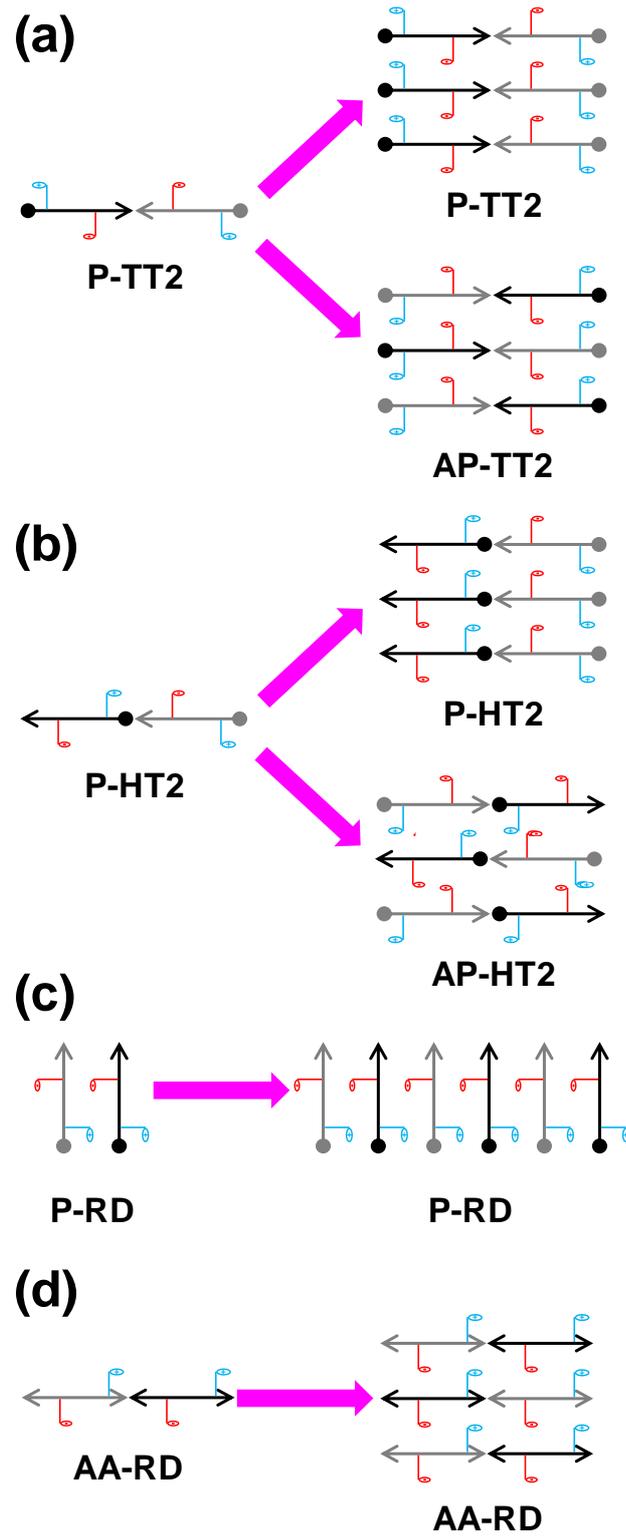


Figure 1

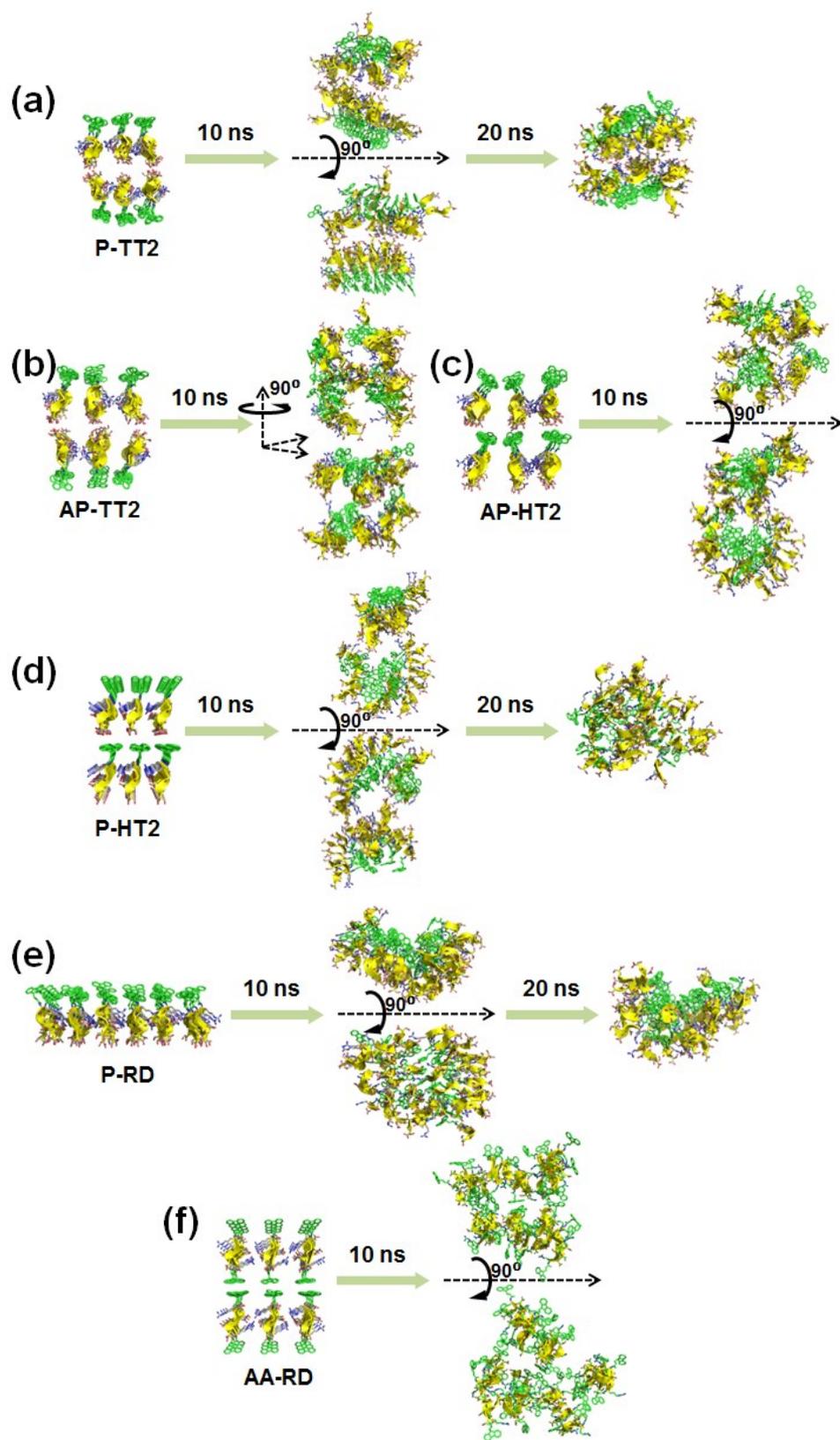


Figure 2

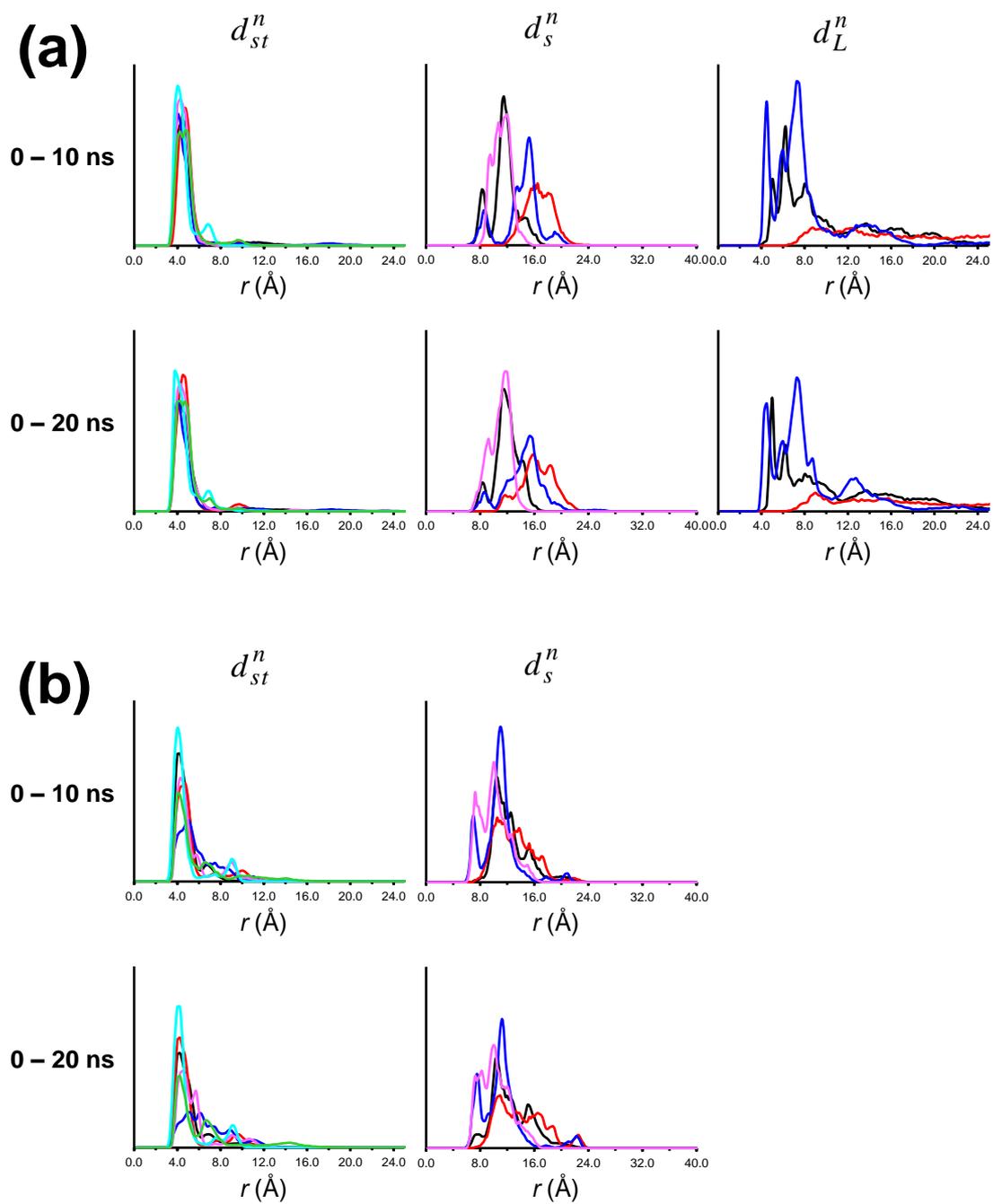


Figure 3

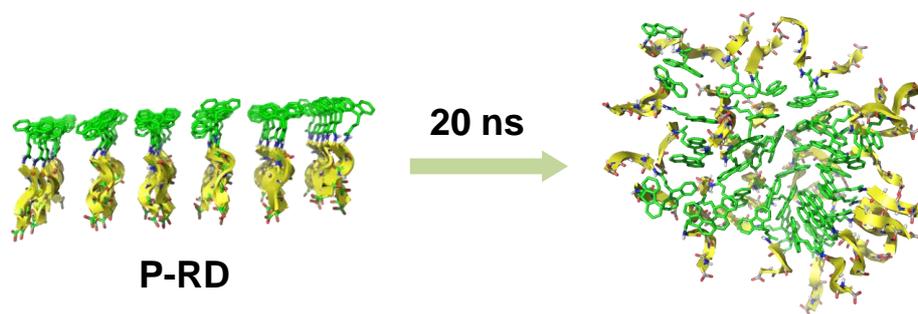


Figure 4

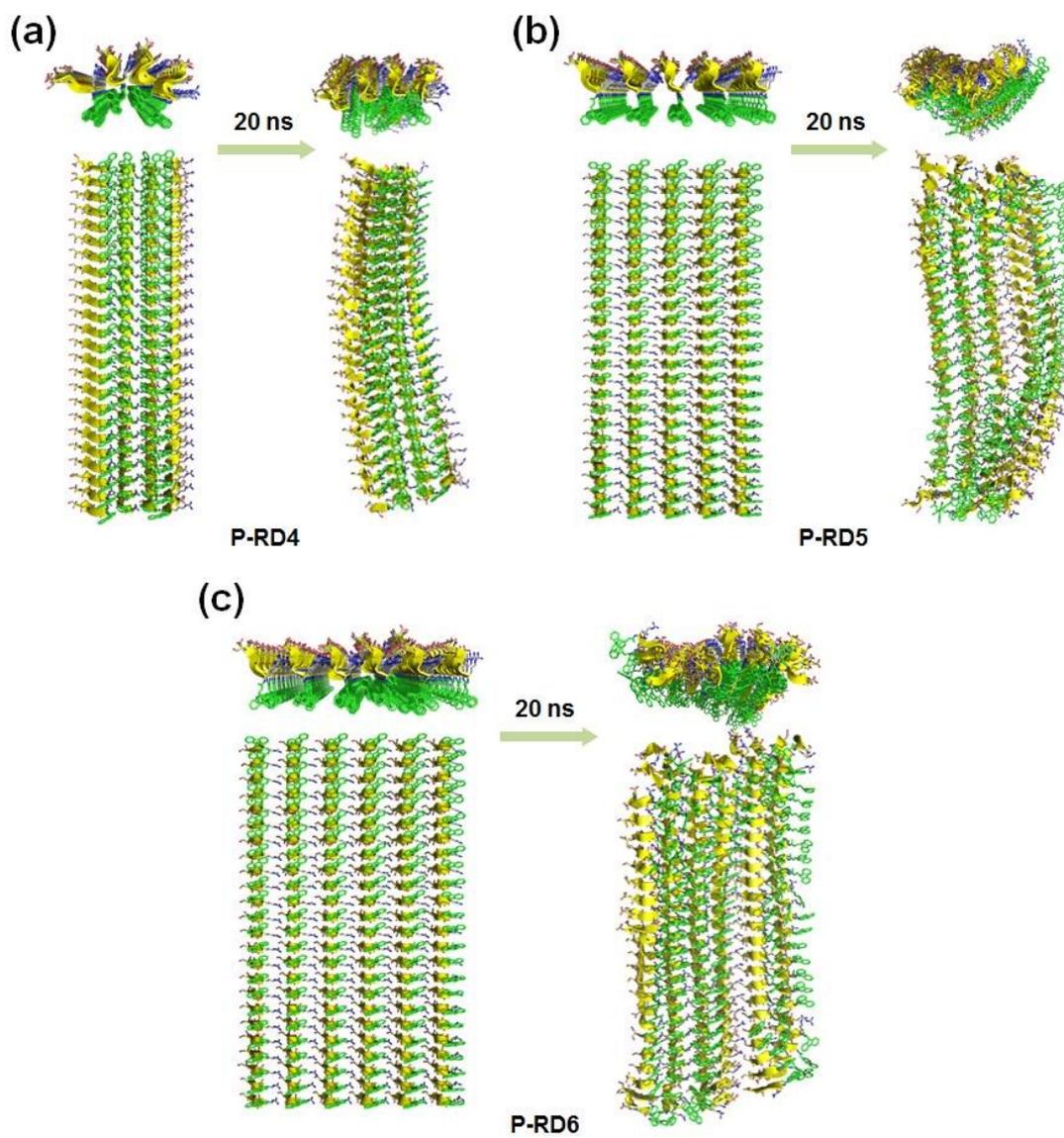


Figure 5

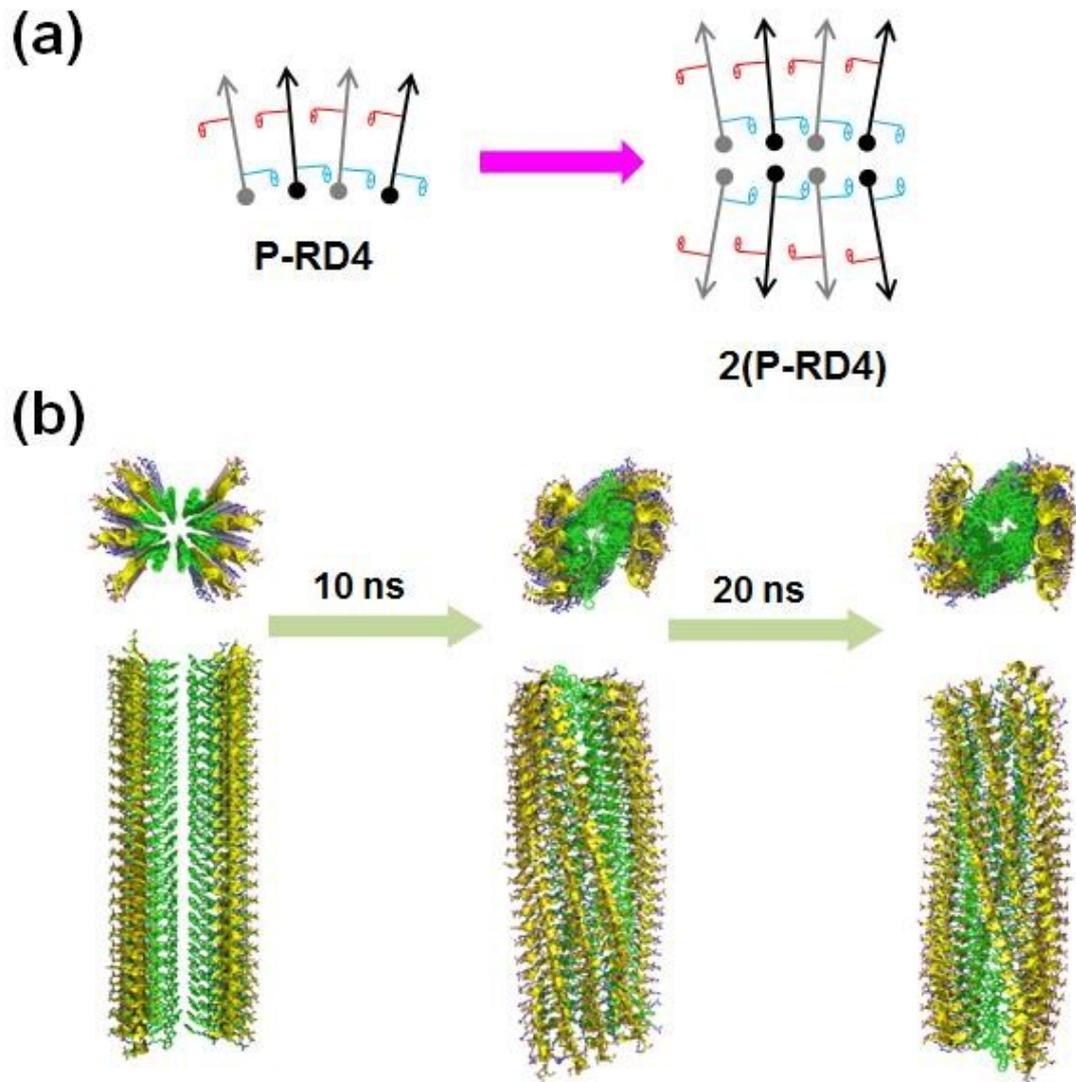


Figure 6

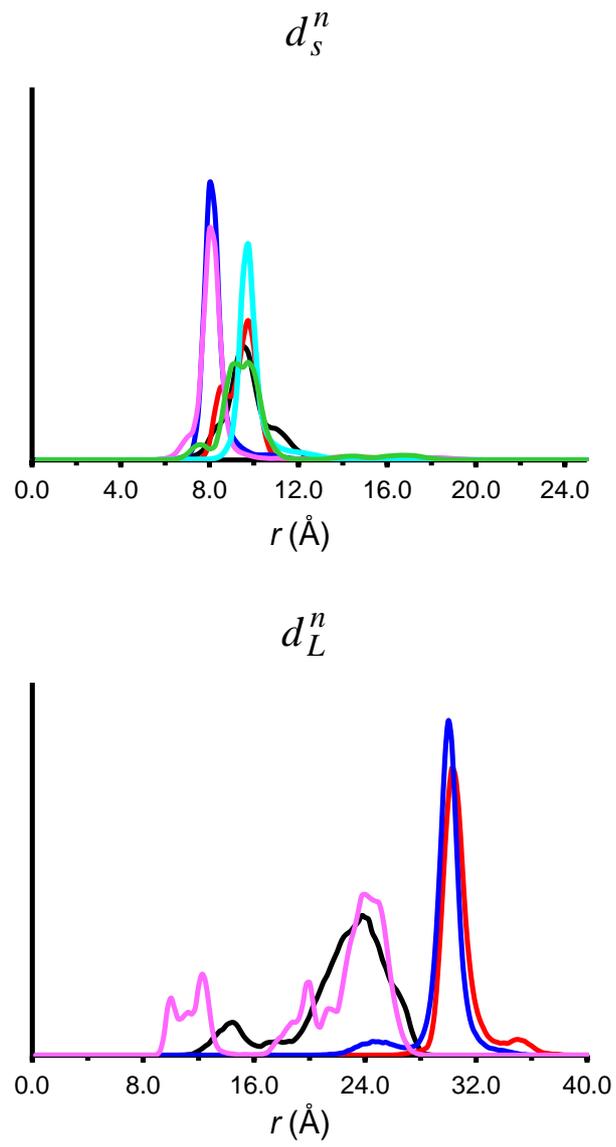


Figure 7

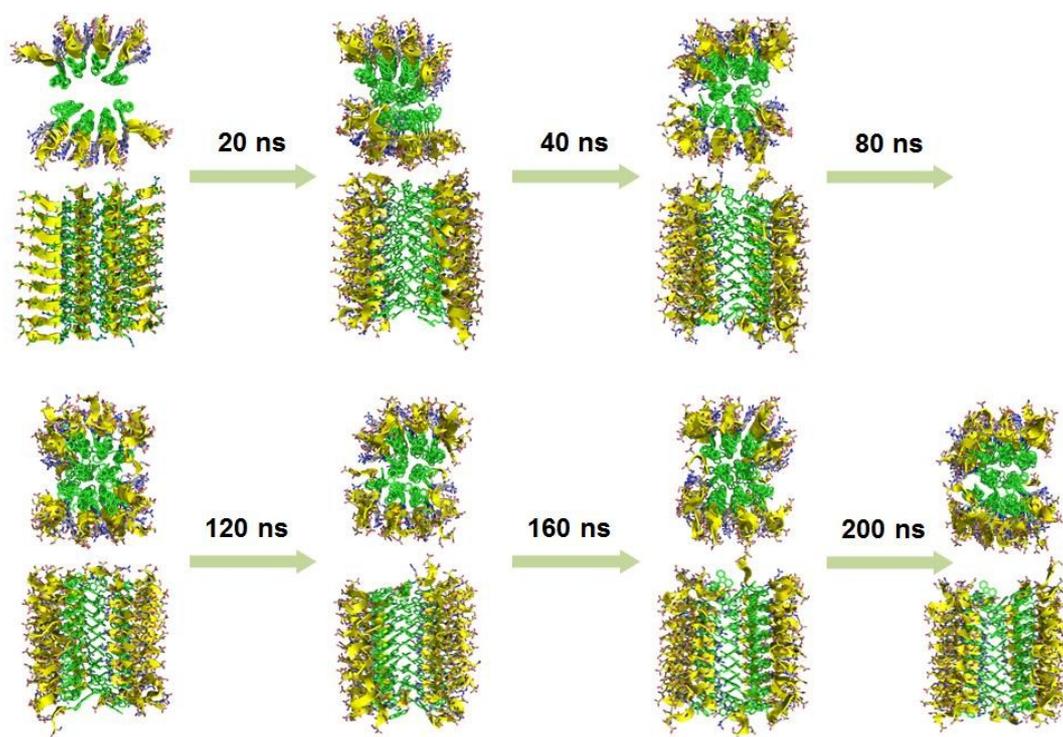


Figure 8

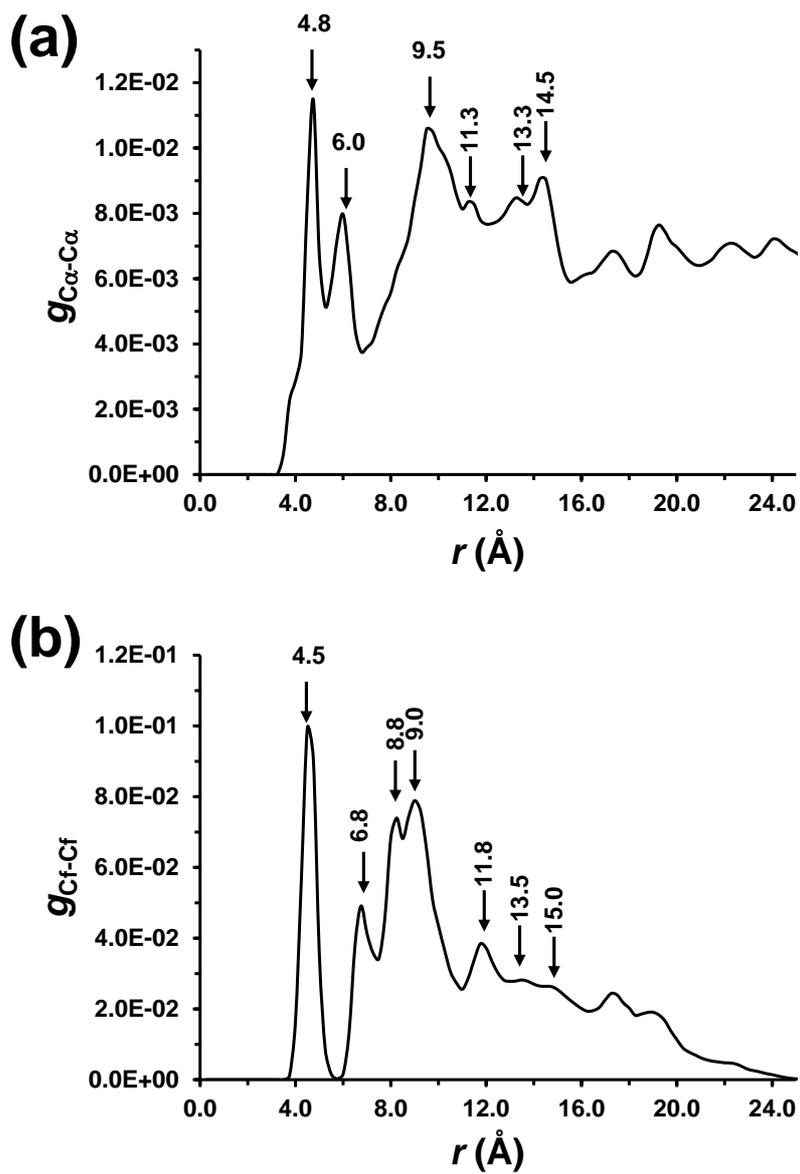


Figure 9

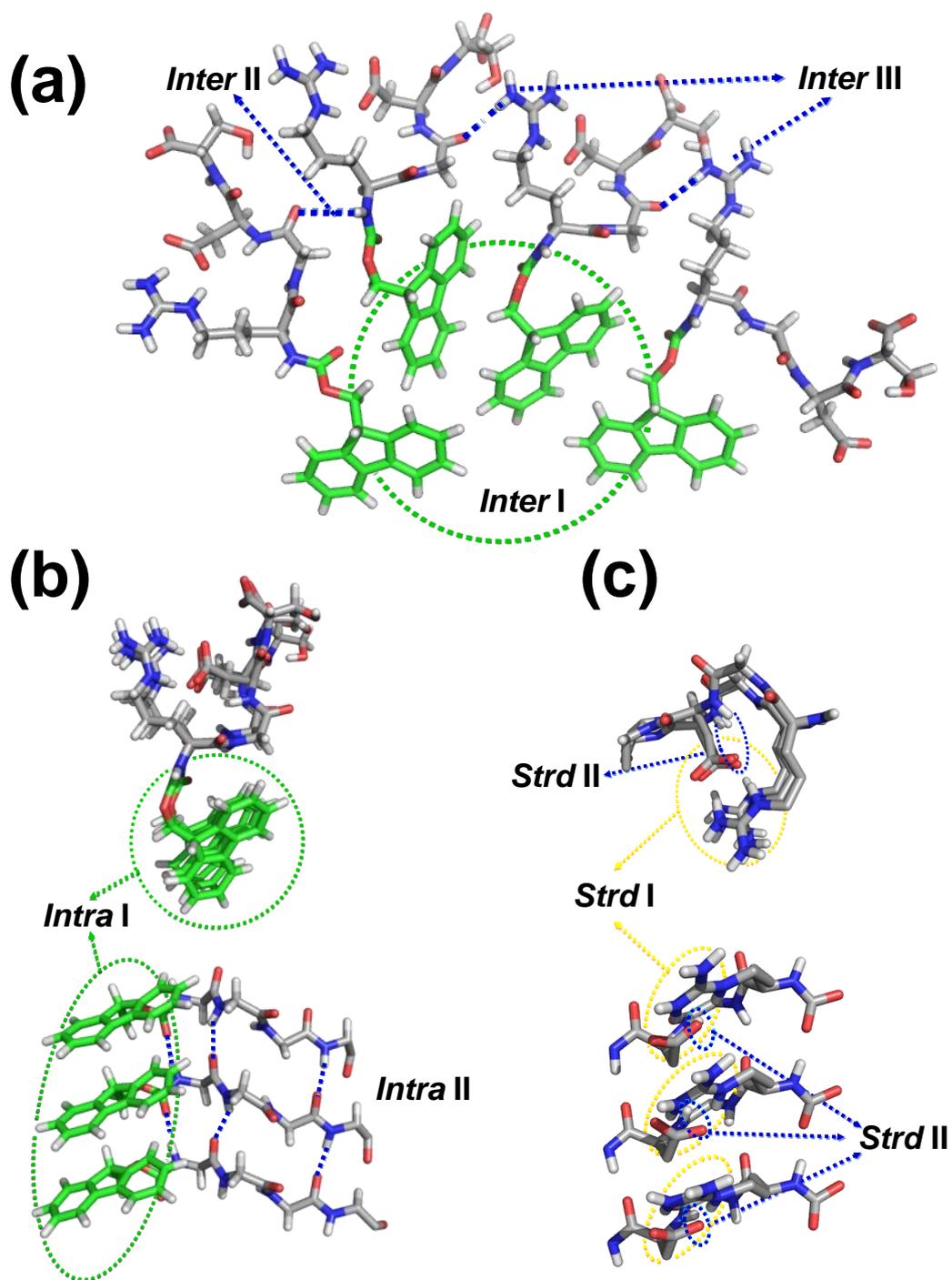


Figure 10

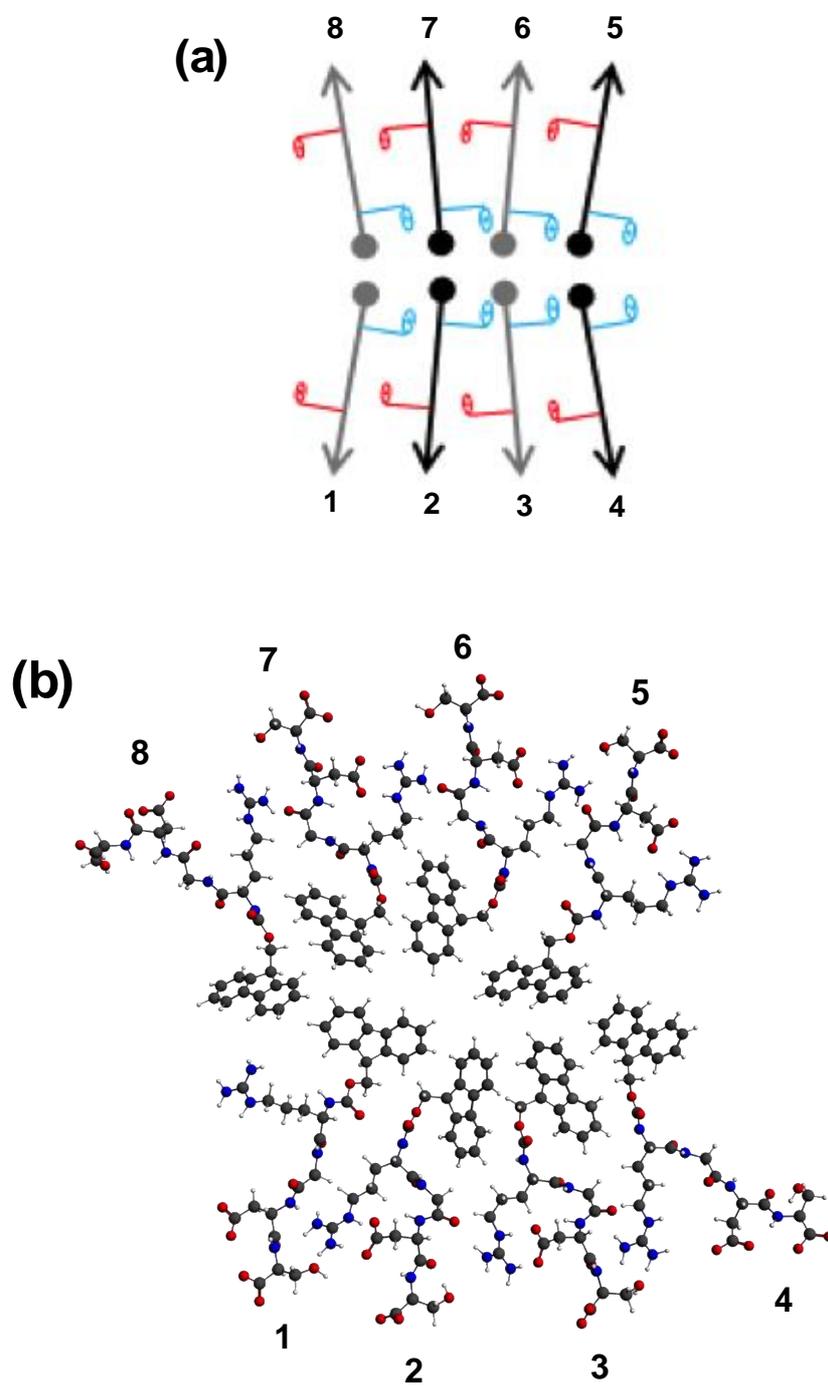


Figure 11

## Graphical Abstract

