PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Physical Chemistry Chemical Physics
CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS
ARTICLE TYPE www. rsc.org/xxxxxx | XXXXXXXX

A Computational Study of the Self-Assembly of the RFFFR Peptide

Morten Slyngborg,[a] and Peter Fojan,*[a]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

The β-amyloid peptide sequence, LVFFA, inspired the investigation of the fiber formation potential of the RFFFR peptide. The self-assembly was studied *in silico* by coarse grained-, atomistic molecular dynamics simulations and semi-empirical quantum mechanical calculations. The fiber formation was found to occur according to a three step process starting with the emergence of small aggregates that join together and form fiber segments that eventually form one continuous fiber. From a series of simulations the critical fiber concentration was determined to be in the interval between 70 mM and 100 mM. To obtain more structural information of the stable fiber, the final coarse grained configuration was backtransformed to atomistic detail. Based on this structure a 10 ns atomistic simulation was performed, which suggests that fiber is stabilized by hydrogen bonds and water mediated hydrogen bonds. These stabilizing bonds are, however, reduced by competitive protein-water hydrogen bonds. Hence, π-stacking is suspected to play a larger role in fiber stabilization. The π-stacking of intermolecular Phe residues are found to favor a T-shaped stacking mode, while intramolecular π-stacking interactions assume a broad variety of modes from the parallel displaced to the T-shaped stacking mode and modes in between, with equal probability. Selected snapshots from the atomistic simulation was geometry optimized by semi-empirical quantum mechanical methods to validate the fiber stability and π-stacking configuration. An average Cα-RMSD was determined to 2.68 $\AA$. These findings indicate that the fiber may be used as novel model system for the study of amyloid fibers or self-assembled conductive biowires, respectively.

1 Introduction

Self-assembled protein nanostructures have attracted much interest due to their involvement in more than 20 degenerative diseases such as Alzheimers, Parkinsons and Prion diseases.[1,2] However, recent research has also been devoted to its applications such as nanomaterial engineering, nanolithography, regenerative medicine, biosensors, and drug delivery.[3–6]

Typically these model peptides originate from larger natural occurring peptides or proteins. One such example is the peptide LVFFA, derived from the β-amyloid peptide which then plays a crucial role in Alzheimers disease.[7] The LVFFA peptide was later modified towards the analogous diphenylalanine (FF). This peptide has been extensively studied and is known to self-assemble into a large variety of structures ranging from nanotubes[8], nanowires[9], films[9,10], vertical aligned wires[10] and sponge-like structures[10] depending on pH conditions.[11] Recently, the related peptide triphenylalanine (FFF) was investigated which forms plate-like structures with lengths of several micrometers.[12] Furthermore, FFF forms nanospheres without any void space when the N-terminus is protected by a t-butyloxycarbonyl (t-Boc) group.[4]

Even though the FF and FFF peptides are derived from the core recognition motif of the Alzheimer's β-amyloid peptide it is still debated whether they share any structural properties with the amyloid fibrils.[8,12,13]

The self-assembly process is difficult to study experimentally, thus computational methods are an attractive approach to study self-assembly processes.[14] The self-assembly process typically occurs on a timescale, which is outside the obtainable timescales for atomistic molecular dynamics (MD) and monte carlo simulations as well. In these cases coarse grained (CG) force fields represent an appealing alternative as the total amount of atoms, in general, is reduced by one third, allowing much larger systems and timescales to be simulated. The MARTINI CG force field[15] is amongst the most popular and versatile force fields. A number of complex molecules has been simulated, such as lipids[15], sterols[16], DNA[17], sugars[18], polymers[19], nanoparticles[20], proteins[21–23] and different solvents[15,21,24] as well.

The MARTINI CG force field[15] has some limitations compared to atomistic force fields. The structure of large biomolecules is not reproduced accurately over time in the standard version of the MARTINI force field[15]. Furthermore, the secondary structure has been fixed during the course of the simulation which significantly restricts the phenomena that can be studied. Additionally, the CG simulations yield less structural information than similar atomistic simulations as some atoms are not explicitly included. The latter limitation is circumvented in some CG models by using multiscaling, where parts of the simulation are in atomistic detail while the

[a] *Department of Physics and Nanotechnology, Aalborg University, 9220 Aalborg st, Denmark.*
* *Fax: XX XXXX XXXX; Tel: XX XXXX XXXX; E-mail: xxxx@aaa.bbb.ccc*

rest is represented by a CG model.[25] Another option is to translate the final CG structure into the equivalent atomistic structure and continue the simulation for a short period. Both options have been implemented in the MARTINI model[15], but it has been of limited use so far, due to the comprehensive workload required. However, a recent algorithm has made the implementation of the back-translation function more readily accessible.[26]

The present study investigates, for the first time, the effects of modifying the FFF peptide to the amino acid sequence RFFFR. The motivation for the modification is to direct the self-assembly process towards nano-fibers, which may better resemble amyloid fibrils than the FF or FFF structures. The arginine residues are intended to restrict the interaction of Phe residues in one direction only.

2 Methods

The MARTINI CG force field[15] used in the present study, joins in general four heavy atoms into one CG spheres/bead. Each bead is assigned the united properties of the enclosed atoms, substantially reducing the amount of interactions to be calculated. In this model, the protein backbone is represented by a single bead while the individual amino acid side chains are represented by one or more beads, depending on the amino acid. Non-bonded interactions in the MARTINI force field are parameterized from experimental data such as the partitioning free energies between amino acid side chains and the oil/water interface. The bonded interactions in the MARTINI force field are determined from the distribution of bond lengths, angles and dihedral angles derived from protein structures from the Protein Data Bank and comparison with atomistic force fields.

The initial self-assembly is simulated with the MARTINI force field[15] succeeded by a distance restrained MARTINI simulation. This final structure is back-transformed into an atomistic structure. Starting from this structure a 20 ns atomistic simulation is performed with the OPLS-aa force field[27–29]. From the atomistic simulation, structural information of the peptide and the π-stacking interaction of phenylalanine (Phe) residues have been extracted. Furthermore, a series of CG simulations are performed to determine the minimum peptide concentration necessary to form a single stable nano-fiber.

All simulations were performed with GROMACS v.4.6[30]. The initial RFFFR structure was created and geometry optimized for an extended strand structure within YASARA[31] The resulting structure was converted to a MARTINI CG peptide (Fig. 1) using the martinize.py v.2.4 script. The initial MD configuration was constructed by adding 27 CG peptides at random positions in a box with dimensions 12nm x 5nm x 5 nm (x,y,z). Water was then added at any sterically allowed position, amounting to 2149 MARTINI polarized water molecules. After each modification, a steepest decent minimization with 10.000 steps was used to minimize the potential energy of the system.

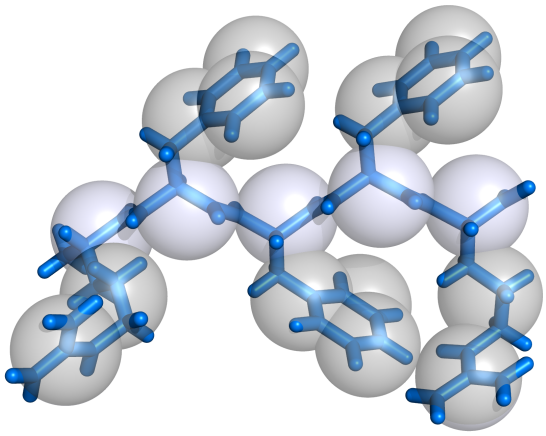
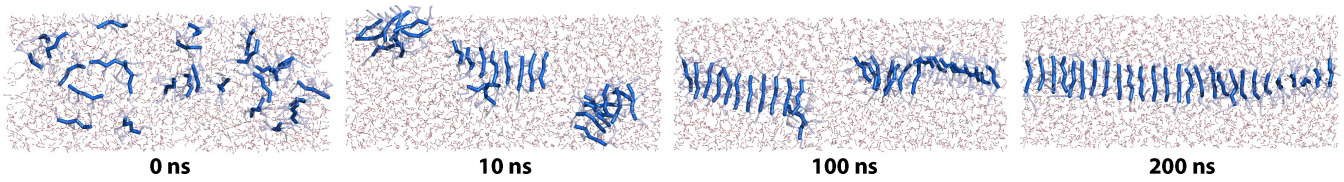


Fig. 1 Atomistic structure of the peptide (RFFFR), created in YASARA (blue) and the corresponding MARTINI structure. In The MARTINI model amino acid backbones are represented by one bead (white), while side chains of arginine and phenylalanine have two and three bead, respectively (gray).

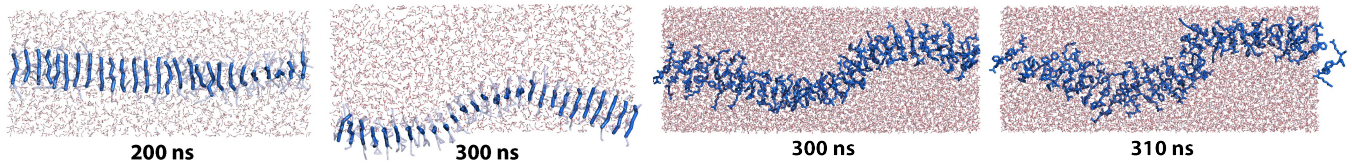
Using time steps of 20 fs, an isotropic simulation of 200 ns was performed with the MARTINI force field v.2.2P[15]. The final configuration was used in a 100 ns continued distance restrained CG simulation where the Phe residues were locked in a trans-configuration, due to a observed very slow equilibration of the cis-trans configuration. The initial velocities were adopted from the last frame of the 200 ns long CG simulation.

The final CG structure from the distance restrained simulation was converted to an atomistic structure using the Backward script[26]. Water was removed in order to center the fiber in the simulation box, after which 8517 SPC water molecules, 108 chloride ions and 54 sodium ions were added to ensure system charge neutrality. Then an energy minimization was performed with a steepest decent minimization with 100.000 steps. This was followed by a 100 ps long semi-isotropic simulation with the OPLS-aa force field[27–29] to equilibrate the system. The resulting structure was used in a 10 ns long isotropic simulation performed with the OPLS-aa force field with time steps of 1 fs.

The initial velocities of the first CG and atomistic simulations were assigned a Maxwell distribution at 323 K and 300 K, respectively. During the CG simulations the temperature and pressure were kept constant at 323 K and 1.013 bar with the v-rescale[32] and berendsen[33] algorithm, respectively. The elevated temperature was adopted in order to prevent the system from getting trapped in a local conformational energy minima during the self-assembly. The atomistic simulation was kept at 300 K and 1.013 bar with the Nose-Hoover[34]



(a) Snapshots from the initial CG simulation



(b) Snapshots from the distance restraint CG simulation **(c)** Snapshots from the atomistic simulation

Fig. 2 Water molecules and CG side chains are made transparent for simplicity sakes. (a) Snapshots from the initial 200 ns CG simulation where the self-assembly of the nano-fiber occurs. (b) Snapshots from the continued distance restraint 100 ns CG simulation where the cis-configuration of the peptide is prevented. (c) Snapshots from the 10 ns atomistic simulation.

and Parrinello-Rahman[35] algorithm, respectively. This close to room temperature was adopted as the elevated temperature was no longer needed. In both the CG and atomistic simulations periodic boundary conditions (PBC) were applied in all directions and electrostatic interactions were calculated by the particle mesh ewald[36] (PME) algorithm with a real space cut-off at 1.5 nm in the CG simulations and 1 nm in the atomistic simulation.

Selected snapshots from the all atom simulation were geometry optimized by semi-empirical quantum mechanical (SQM) calculations using the MOPAC[37] package. The MOPAC package[37] developed by Stewart was used as the implemented module MOZYME[37] allows for the calculation of more than 1000 atoms, which is the approximate limit of standard SQM calculations. However, MOZYME[37] utilizes a localized molecular orbital method that allows for processing a system of 15,000 atoms in a very fast timescale compared to other solutions. The newest MOPAC version allows for the use of the new PM7[38] method which is additionally well known for its very high accuracy that borders DFT-D calculations.[38–40]

The OPLS-aa snapshots that was geometry optimized with PM7[38] was as follows (starting from the beginning of the OPLS-aa simulation): 2500 ps, 3750 ps, 5000 ps, 5625 ps, 6250 ps, 6875 ps, 7500 ps, 8125 ps, 8750 ps, 9375 ps and 10,000 ps. Explicit solvent from the MD snapshots were removed and solvent effects were accounted for by the COSMO implicit water model[41] with a dielectric constant of 78.4. The geometry optimization convergence criterion was set to a maximum gradient of 10.0 kcal/mol/Å, after which a second calculation was performed with a convergence criterion of 5.0 kcal/mol/Å. Finally a single SCF calculation was performed to correct any error in heat of formation.

As PBC are not possible to implement with the SQM approach, on larger systems with many charges, a second simulation series was performed. This series was performed in the same way as described above, but with frozen alpha carbon atoms of the arginine residues in the peptides constituting the ends of the fiber, enforcing the fiber to remain in a stretched configuration. This series will be denoted the restricted PM7 calculations henceforth. As a reference a 10 ns all atom MD simulation was performed containing only four peptides in a large simulation box. Hence PBC were not important during this simulation and subsequent PM7 calculations yielded relatively low RMSD values.

Additionally, a series of CG simulations with varying peptide concentrations were performed to determine the minimum concentration at which peptides assembled into stable fibers. These simulations were all carried out with the same parameters as the non-distance restrained CG simulation. To improve validity of the results, more simulations were performed with concentrations close to the critical fiber concentration. The simulations had different initial velocities in order to cover a wider conformational space. Furthermore, the critical fiber concentration was confirmed in another MD simulation series where a stable fiber was inserted into a simulation box corresponding to the concentrations above and below the critical fiber concentration. In this way it was possible to study fiber stability or disassembly.

Simulation time of CG and atomistic simulations do not scale 1:1 and a direct translation is often difficult. The MARTINI time compared to atomistic time is in general scaled by

a factor of 4 [16,42], but several other factors have been reported earlier [43,44]. For this reason actual simulation rather than effective time has been used throughout this paper.

3 Results and Discussion

3.1 Peptide Self-Assembly

Snapshots of the CG self-assembly simulation at different time intervals are shown in Fig. 2a. Initially all peptides are randomly distributed in the simulation box. After 10 ns three small fiber segments with significant defects have been formed. 100 ns later, these smaller fibers self-assemble into one fiber with defects at the fiber ends. After 200 ns a continuous fiber has been formed that spans the entire PBC. The peptides in the fiber from the initial CG simulation were mainly found to have an antiparallel configuration where Phe-2 is located in proximity to Phe-1 and Phe-3 of a neighboring peptide (Fig. 3a). However, some peptides adopt a cis-configuration for a short period of time during the initial CG simulation. Compared to atomistic simulations of the fiber, the amount of cis-configurations is found to be relatively high (data not shown).

For this reason the final configuration of the 200 ns long CG simulation was used as initial configuration for a 100 ns long distance restraint CG simulation (Fig. 2b) where the peptide cis-configuration was excluded. This resulted in a nearly defect free fiber where almost all peptides were oriented antiparallel.

To regain atomistic details, the final configuration of the 100 ns long distance restraint CG simulation was converted to an atomistic configuration and continued for 10 ns (Fig. 2c). During the simulation, the Phe residues assume distinct orientation modes at the expense of backbone-backbone angle distribution which becomes more disperse.

3.2 The Phenylalanine Cis-Configuration

The existence of cis-configurations in non Proline containing proteins is rather scarce. [45] However, a statistical investigation of the protein data bank reveal that the majority of cis-configured residues in non Proline containing proteins are involved in an interaction with an aromatic residue. [46] Hence we expect some Phe residues to assume a cis-configuration during the MD simulation.

Initially a high level of cis-peptide configurations is observed (Fig. 4). However, the amount of cis-configurations is reduced during the initial CG simulations, indicating that cis-configurations are important for the formation of the small initial segments while larger fibers are more stable without the cis-peptide configuration.

Another indication hereof is that the continuos fiber from the distance restrained CG simulation remains stable without the cis-peptide configuration (Fig. 4), hence they are not crucial for the stability of the fiber in the MARTINI model [15]. Furthermore, it is evident from the self-assembly, that the sum of the attractive intermolecular forces amounts to a larger force than the long range electrostatic repulsive forces of the peptides N-termini with a charge of +2. Hence it is likely that the attractive forces are also sufficiently strong to induce the cis-configuration in naturally occurring fibers.

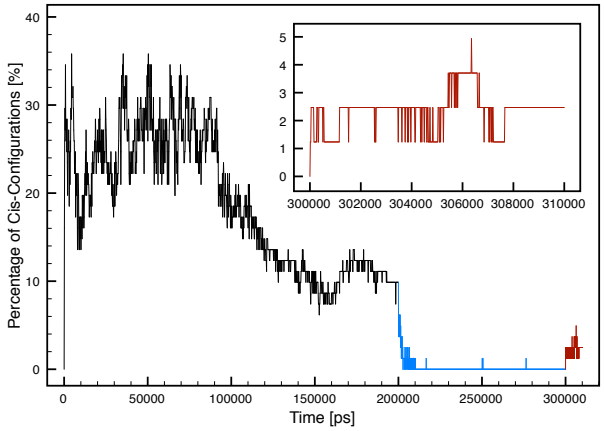
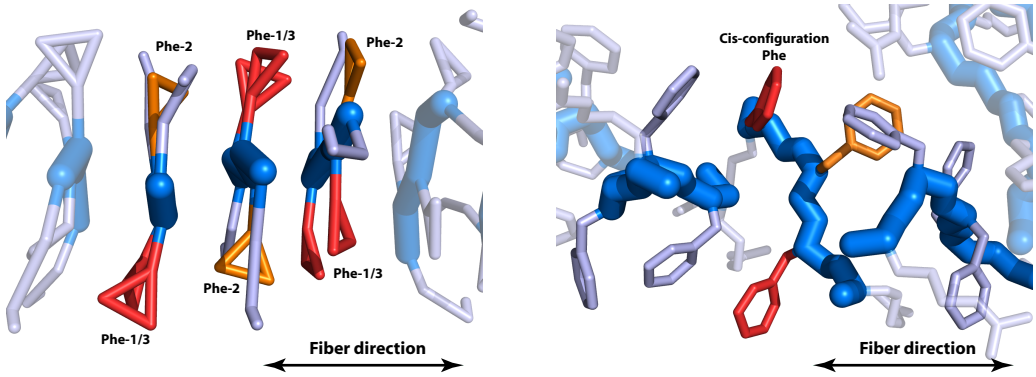


Fig. 4 Number of Phe residues in a cis-configuration as a function of time over the three simulations; Initial CG simulation (black), CG simulation with distance restraints (blue) and atomistic simulation (red). For clarity a zoom in on the atomistic simulation is depicted in the insert.

Indications hereof are observed in the atomistic simulation where some peptides resume a cis-peptide configuration. This may be the result of the disperse backbone-backbone angle distribution as it causes some intermolecular Phe side chains to move apart and it becomes energetically favorable to interact with another Phe side chain in close proximity in the opposite direction (Fig. 3b). Hence the cis-configuration increases the stability of the fiber, but the net stability is presumably reduced by the increased disperse backbone-backbone angle distribution found during the atomistic simulation.

3.3 Structural Clusters

During the atomistic simulation a number of peptides adopt similar molecular structures. These structures were grouped together in clusters based on the single linkage method [47]. According to this method a structure belongs to a certain cluster if its RMSD compared to any other molecule in the cluster is smaller than a certain cut-off value. The center structures of the clusters found during the atomistic simulation using a cut-off of 0.5 Å are shown in Fig. 5. Furthermore, the occurrence and number of conformers belonging to each cluster is also



(a) Snapshot from the initial CG simulation **(b)** Snapshot from the final atomistic simulation

Fig. 3 Water molecules are removed in both snapshots while Phe-1 and Phe-3 residues has been highlighted in red and Phe-2 residues in orange. Hydrogen and oxygen atoms are removed in the atomistic snapshot. (a) Close-up snapshot from the initial CG simulation that illustrates the fiber configuration of three representative peptides in the CG simulations. It is apparent that peptides interact in an antiparallel mode as Phe-2 interact with Phe-1 and Phe-3 residues of other peptides. (b) Close-up snapshot from the atomistic simulation illustrating a peptide with a Phe residue in a cis-configuration. It is apparent that the backbone of the peptide with the cis-configurated Phe residue is not aligned with the residual peptides. Hence the cis-Phe residue is forced to assume a cis-configuration as it is too far away to interact with any other Phe residues.

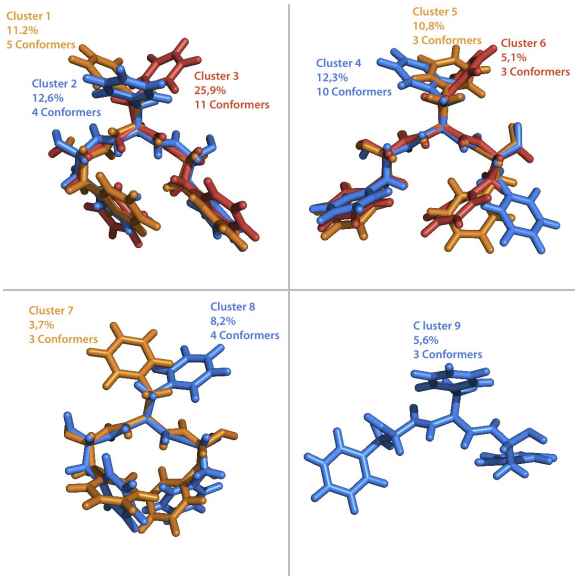


Fig. 5 Overview of the center structure of the nine clusters found during the atomistic simulation. The total occurrence and number of conformers in each cluster are also listed.

listed in Fig. 5. Structural information of Arg residues were not included in the analysis as they had a high degree of conformational freedom.

Based on a Newman projection 27 main clusters should be present. However due to the influence of neighboring peptides in the fiber, a degeneracy of the peptide configuration is found, leading to the representation of only nine clusters in the stable fiber (Fig. 5). These correspond to 95.4 % of the structures any given peptide adopts during the simulation. The remaining structures are special structures such as the cis-configuration or short lived structures that peptides adopt for less than 50 ps during the simulation. It is noteworthy that nine major clusters were identified which implies a rather strict conformational flexibility of the single peptides in the fiber. Furthermore, very few different conformers participate in the clusters. On average a single conformer participates in 1.7 different clusters, indicating that few cluster transitions occur. The two largest clusters, cluster 3 and 4, constitute together 21 different peptides, hence most peptides adopt a structure belonging to one of these clusters during the simulation.

It should be mentioned that cluster 4 is different from the other clusters, as the Phe-3 residues assumed different orientations. These structures were considered as one single cluster as they had a large distance between the Phe-1 and Phe-3 residues in common, rendering π-stacking interactions negligible and hence had larger Phe side chain orientation fluctuations.

3.4 Secondary Structure

Usually secondary structure changes of large proteins require very long atomistic simulations in order to be reliable.[48,49] For peptides, shorter time scales are sufficient to simulate the folding of a disordered structure into a near native structure.[50,51]

During the present atomistic simulation the secondary structure did not change significantly (Fig. 6). It was found

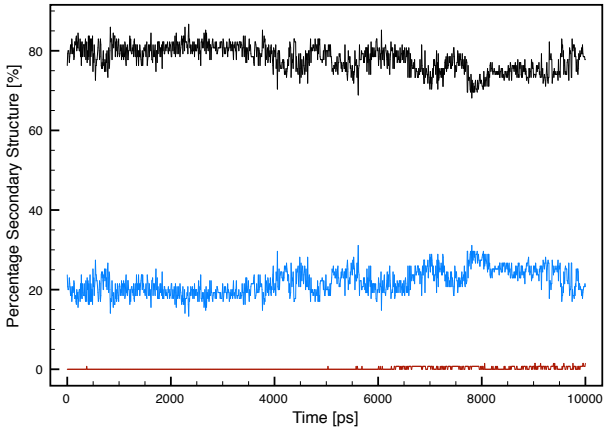


Fig. 6 The peR_{cen}tage random coil (black), beta-strand (blues) and alpha-helix (red) secondary structure the peptides assume during the atomistic simulation as a function of time.

that 80 % of the residues have random coil, 20 % beta-strand and occasionally assumed an alpha-helix like configuration. Compared to another atomistic simulation of initially pre-assembled FF and FFF structures, a beta-strand content of below half of what was determined in the current study, was reported.[12] However, experimental studies indicate that nanostructures self-assembled from the FF or FFF peptide consist of a high degree beta-sheet.[8,12] Presumably this deviation is due to the limited simulation time and number of peptides in the simulation systems. It is likely that the same applies to the present study and a much higher amount of beta-strand content would be observed experimentally for the RFFFR peptide structures.

3.5 Hydrophobic Effects and H-bonds

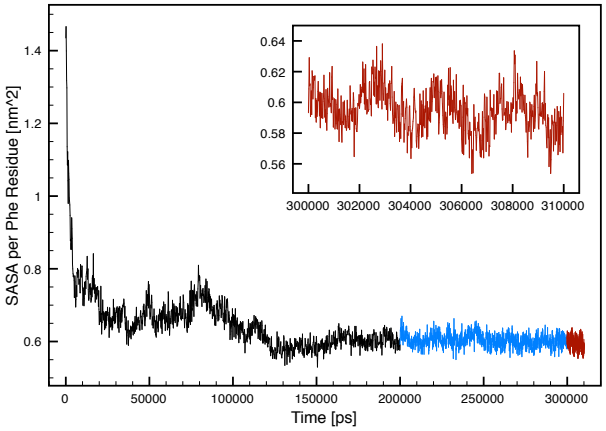
The hydrophobic effect is a major driving force for association of apolar substances in aqueous solutions.[52] It can be enthalpy or entropy driven due to the exclusion of water molecules from the apolar surfaces.[53] As RFFFR consists of a hydrophilic shell and a hydrophobic core, peptide association is expected as a result of the hydrophobic effect. From the solvent accessible surface area (SASA) of the Phe residues as a function of time (Fig. 7a) it is apparent that it decreases drastically during the self-assembly and remains quite stable after equilibrium has been reached. This indicates that hydrophobic effects contributes to the self-assembly. A similar tendency and conclusion was reached in a MARTINI simulation study of the FFF peptide.[54]

The stabilizing influence of hydrogen bonds in the resulting fibers has been studied by converting the CG system to atomistic details and performing a10 ns MD simulation with the OPLS-aa force field. To validate the obtained results, semi-empirical quantum mechanical calculations with PM7 (section 3.7) were compared to the OPLS-aa force field results. These simulations resulted in on an average 3.3 hydrogen bonds formed per peptide between two peptides (Fig. 7b). This is approximately 1.3× more hydrogen bonds than what was reported for FF.[55] Which leads to the conclusion that RFFFR fibers might be stabilized more by hydrogen bonds than FF structures. FF peptides form simple head (NH_3^+) to tail (COO^-) hydrogen bonds while RFFFR form a relative elaborate network (Fig. 7b). A similar complex network was observed in another all atom simulation of FFF.[12] Concerning RFFFR, only 0.6 head to tail hydrogen bonds per peptide are formed while approximately 1 hydrogen bond is formed between Phe main chain to Phe main chain and Arg side chain to Arg main chain, respectively.

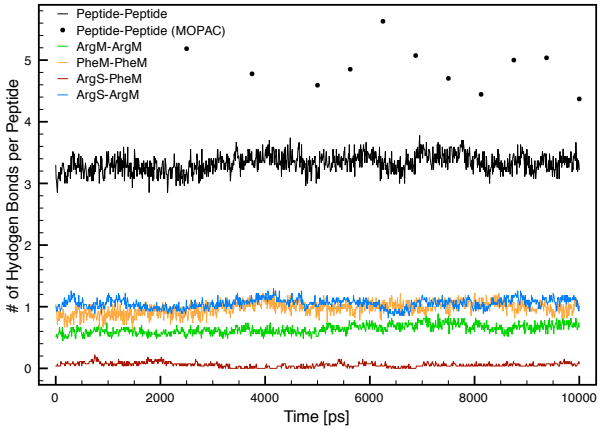
Being an amphiphile, RFFFR has more contact with water molecules than FF which forms more compact sheltered structures. For this reason a high amount of hydrogen bonds are also formed to water molecules (Fig. 7c). On average 16.4 hydrogen bonds per peptide are formed to water molecules, where hydrogen bonds to Arg main chain and side chain constitute by far the largest part. These bonds weaken the peptide-peptide hydrogen bonds as water competes for the hydrogen bond interaction. Furthermore, it is generally believed that supramolecular structures cannot be formed in water solely based on hydrogen bonds, on account of competitive hydrogen bond with water.[56] However, hydrophobic regions avert, to some extend, this competition and allows for the self-assembly based on hydrophobic effects and hydrogen bonds.[57,58] The Phe residues in RFFFR may induce such a compartmentalization as indicated by the relative low water hydrogen bond competition of Phe residues (Fig. 7c). This also explains why far less peptide-water hydrogen bonds (0.5 per peptide) are observed in FF simulations, due to the lac of hydrophilic residues.

However, in depth analysis of the peptide-water hydrogen bonds, reveal that a large amount of these compose of peptide-water-peptide hydrogen bonds (Fig. 7d). These water mediated hydrogen bonds actually add to the stability of the fiber.[59] Each peptide forms 4.1 water mediated hydrogen bonds whereof 2.6, 0.9 and 0.6 water mediated hydrogen bonds are between Arg main chain, Arg side chain and Phe main chain, respectively. Hence Arg residues do not contribute significantly to the stability through peptide-peptide hydrogen bonds, but in return stabilizes the fiber through water mediated hydrogen bonds.

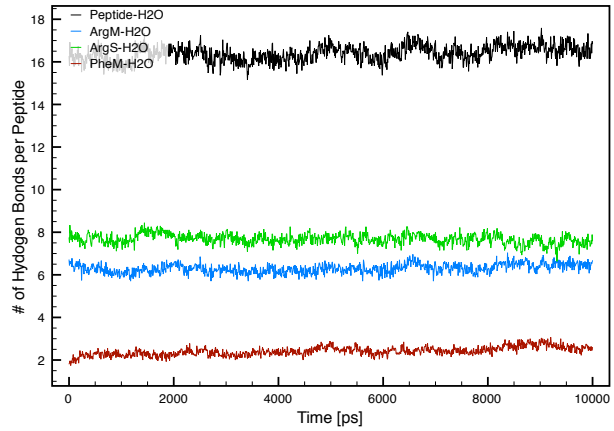
Compared to the PM7 geometry optimized structures, the total amount of peptide-peptide hydrogen bonds are consistently low (Fig. 7b). In addition to the underestimation of the hydrogen bond strength of empirical force fields[29], this might imply that the OPLS-aa force field does not accurately account for hydrogen bonds. However, this is not expected to



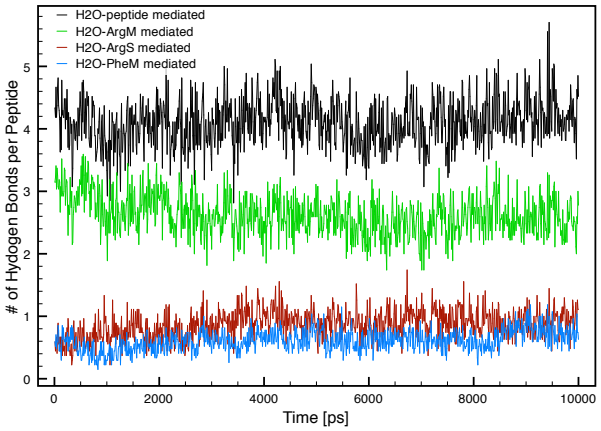
(a) SASA per Phe residue as a function of time. Initial CG simulation (black), CG simulation with distance restraints (blue) and atomistic simulation (red).



(b) Peptide-peptide hydrogen bonds as a function of time.



(c) Water-peptide hydrogen bonds as a function of time.



(d) Water mediated peptide-peptide hydrogen bonds as a function of time.

Fig. 7 Overview of the hydrophobic effect as a function of time and hydrogen bonds as a function of time. A hydrogen bond is counted if the doner-acceptor distance is less than 0.35 nm and the hydrogen-donor-acceptor angle is greater than 150°. Where ArgM is Arg main chain; PheM is Phe main chain; ArgS is Arg side chain; PheS Phe side chain.

alter the obtained supramolecular fiber structures significantly, as indicated by the RMSD values obtained from the PM7 calculations (see section 3.7). Hence this deviation is expected to only increase the stability of the self-assembly compared to what is observed with the OPLS-aa force field.

3.6 Phe-Phe π-Stacking

Hydrogen bonds are important to stability, but π-stacking might be even more important. These interactions are believed to play an important role in a wide range of phenomena including the stereo-chemistry of organic reactions[60], protein folding[61,62], protein self-assembly[55,63] and DNA and RNA base-pairing[64]. High level quantum mechanical calculations indicate that π-stacking energies are comparable to hydrogen bonds.[65] Since, more π-stacking interactions are observed than hydrogen bonds in the RFFFR fiber, π-stacking is expected to be the dominating stabilizing force in the resulting fiber assembly.

Standard empirical force fields, such as GROMOS, AMBER, CHARMM and OPLS-aa, account for π-stacking effects by modeling partial charges and the Lennard-Jones 12-6 potential function.[66] These force fields are rather limited by being completely devoid of any electronic structure, hence any charges are assigned to the nuclear center. Furthermore, since atomic charge is not an observable it is difficult to assign partial charges, but the force fields are parameterized to fit experimental or quantum mechanical calculated data.

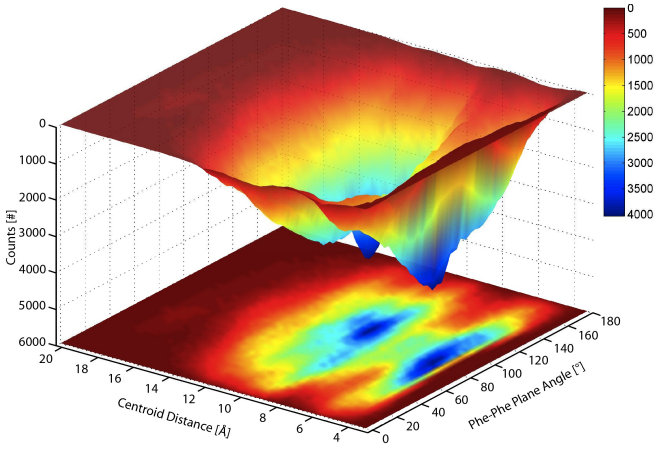


Fig. 8 Intermolecular Phe-Phe plane angles versus centroid distance (R_{cen}) of the atomistic simulation. Data from the first nanosecond is not included.

A recent benchmark of different force fields found that the OPLS-aa force field was amongst the most accurate force fields to account for non-bonded interactions such as π-stacking interactions.[29] Despite of the underestimated hydrogen bonds the OPLS-aa force field exceeded even DFT calculations.[29] Since no partial charges are used in aromatic amino acid residues in the MARTINI force field[15] no intermolecular Phe-Phe stacking mode is dominant during the present CG simulations.

It has been suggested that a small partial charge ($<\pm 0.153$), assigned to the C and H atoms belonging to Phe benzene rings, favors the parallel displaced (PD) stacking mode, while a high partial charge ($>\pm 0.3$) favors the T-shaped stacking mode.[67] The OPLS-aa force field applies a small partial charge (± 0.115), but the PD stacking mode was not found to be the preferred stacking mode during the simulation. Our findings conform with the majority of experimental and computational studies indicating that the T-shaped mode is more stable than the PD mode in proteins.[68] However, we also observe a more complex stacking behavior concerning the interplay between intra- and intermolecular π-stacking interactions. This may well be related to findings from a statistical investigation of the protein data bank that shows that over 80 % of the aromatic residues in the surveyed proteins interact with more than one π-π pair, rendering an exact stacking model difficult to formulate.[69]

The intermolecular Phe-Phe plane angles from neighboring Phe residues versus centroid distance (R_{cen}) is illustrated in Fig. 8. Two maxima are found at Phe-Phe plane angles of 87° and 100° and Phe-Phe separations of 5.5 Å and 10 Å, respectively. Thus the T-shaped mode (60-120°) is dominant, a hybrid mode (30-60° and 120-150°) in between the T-shaped and PD mode is common while the frequency of the PD mode (0-30° and 150-180°) is low. However, it is noteworthy that the Phe-Phe plane angles of the individual Phe pairs tend to remain stable during the atomistic simulation.

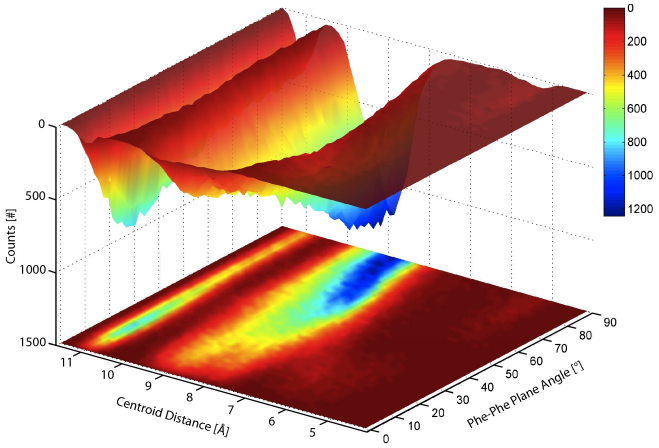
From Fig. 8 it is apparent that the R_{cen} may be as small as 4 Å in the PD stacking mode, while the R_{cen} between two Phe benzene rings in a T-shaped stacking mode may only come as close as 4.5 Å. This phenomenon is observed in other simulation studies of Phe stacking as well and is attributed to sterical hindrance.[55,62]

In the 6.8 Å$<R_{cen}<$8.7 Å range, the Phe-Phe plane angles are restricted to the T-shaped mode in the angle interval of 72-113°. Hence Phe ring pairs moving apart or closer need to adopt a T-shaped mode through this saddle point. The π-stacking interaction cut-off value of 7.5 Å, generally applied, is based on a statistical study of 505 non-homologous proteins from the protein data bank.[62] This cut-off value coincides with the saddle point found in the Phe-Phe interaction angle distribution plot (Fig. 8). The peptides with a Phe-Phe plane angle maximum of 100° and a separation distance of 10 Å shown in Fig. 8, preferentially adopt a perpendicular orientation towards each other. The maximum found at a separation distance of 4-5 Å, suggests a similar intermolecular π-stacking behavior of the RFFFR peptide compared to the FF and FFF peptide, as they too preferably assume T-shaped stacking modes.[12,13,54,55]

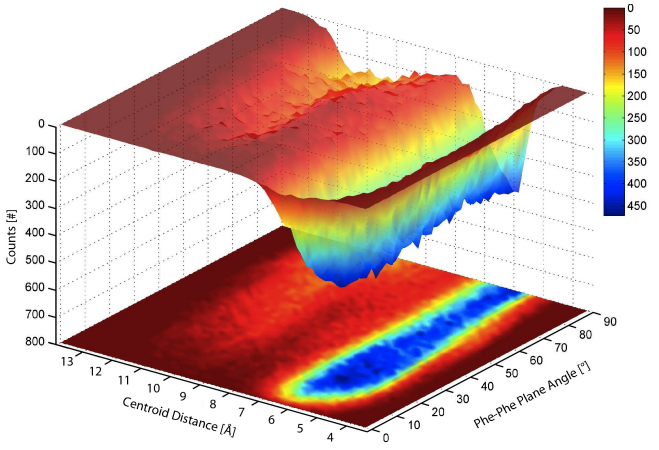
PM7 geometry optimized structures conform with these results and only minor differences were observed in population distribution over the angle interval (see section 3.7). A very similar saddle point and the same maximum were obtained (Fig. 2S). A minor difference between minimum R_{cen} was discovered but this may be attributed the absence of PBC. It was also observed that the propensity for Phe residues to assume a T-shape stacking mode is slightly overestimated in the OPLS-aa force field compared to PM7 calculations.

Concerning intramolecular π-stacking interactions it is found that the R_{cen} between Phe-1/Phe-3 and Phe-2 internally in the peptides are larger than the cut-off length, rendering these interactions negligible (Fig. 9a). Hence the two distinct modes found in the angle distribution analysis are a consequence of the energetically favorable peptide configurations previously discussed. However, it is noteworthy that no direct transitions between the two modes are observed, indicating that a large energy barrier exists between these two different peptide configurations.

The intramolecular Phe-Phe angle distribution of Phe-1 and Phe-3 residues is found to slightly favor the PD π-stacking mode, but a quite even distribution across all three regions (0-30°, 30-60° and 60-90°) is found (Fig. 9b) which is coherent with the cluster analysis (Fig. 5). This indicates that the Phe side chains have a sufficient high flexibility to enable them to assume an intermolecular T-shaped stacking mode. Whereas intramolecular Phe residues do not favor



(a) Intramolecular Phe-Phe benzene angles between Phe-2 and Phe-1 or Phe-3 residues only versus centroid distance (R_{cen}) of the atomistic simulation.



(b) Intramolecular Phe-Phe benzene angles between Phe-1 and Phe-3 residues only versus centroid distance (R_{cen}) of the atomistic simulation.

Fig. 9 Intramolecular Phe-Phe benzene angle distributions from the atomistic simulation. Data from the first nanosecond is not included.

any specific stacking mode in RFFFR, the T-shaped mode is dominant both inter- and intramolecular in FF and FFF peptide structures.[12,13,54,55] This difference arises from the design of RFFFR, as Phe residues may only interact in one direction, rendering it geometrically impossible that both inter- and intramolecular Phe residues stack in the T-shaped mode. Furthermore, the intermolecular Phe residues are able to move closer than intramolecular Phe residues (evident from Fig. 8 compared to Fig. 9b). Hence π-stacking interactions are stronger for inter- than intramolecular Phe residues, which may explain why intramolecular Phe residues adjust their stacking mode in such a way that the intermolecular Phe residues are able to stack in a T-shaped mode.

Within the limitations imposed to the simulations and static fixation of parameters relevant to the interactions, it is shown that intramolecular Phe-Phe stacking orientation do not contribute the stability of the fiber. Had the opposite been the case, the fiber may not have remained stable during the final atomistic simulation or the PM7 geometry optimization. However, a somewhat over exaggerated, due to the missing PBC, average Cα RMSD value of 2.68 $\AA$ was found, indicating that the two models produce relative comparable results (see section 3.7). Reference simulations where PBC were unimportant, resulted in an average Cα RMSD value of 1.31 $\AA$ indicating the high accuracy of the OPLS-aa force field.

3.7 Semi-Empirical Quantum Mechanical Calculations

Root Mean Square Deviation of the Different Calculations
Since no PBC were implemented during the PM7 calculations a large structural difference was expected compared to the MD simulation snapshots calculated with the OPLS-aa force field. The PM7 calculations mimic a free fiber with a finite length in solution, while the OPLS-aa simulations mimic a continues infinite fiber in solution. As indicated by the RMSD values (Fig. 10a), the RMSD values were found to fluctuate a lot. This fluctuate is related to the contraption and expansion of the fiber with finite length calculated by PM7, while the continues fiber remain rather the same length during the MD simulation. This effect was circumvented in the restricted PM7 calculations and yielded much more stable and in general much lower RMSD values (Fig. 10b).

As a reference, four peptides in the fiber from the OPLS-aa simulation were cut out and inserted into a large simulation box (10 nm x 10 nm x 10 nm) after which the box was filled with water. This configuration was energy minimized and served as starting point in a 10 ns long MD simulation performed with the OPLS-aa force field using the same parameters as the OPLS-aa simulations described in the paper. This small fiber segment would never interact significantly with its periodic mirror counterpart due to the large simulation box. PM7 calculations were performed on selected snapshot of this simulation. This resulted in low RMSD values between the OPLS-aa snapshots and the PM7 geometry optimized snapshots (Fig. 10c).

Comparing the RMSD values of the simulation series (Fig. 10) it is apparent that the PM7 calculations deviate a lot from the OPLS-aa calculation. However, the restricted PM7 calculations yield some low RMSD values, indicating that the large values of the PM7 calculations is likely to originate form the missing PBC. Hence the Phe-Phe configuration data from the PM7 calculations apply to another situation than the one in the OPLS-aa MD simulation. The restricted PM7 calculations have an average RMSD (Protein-H) value of 3.35

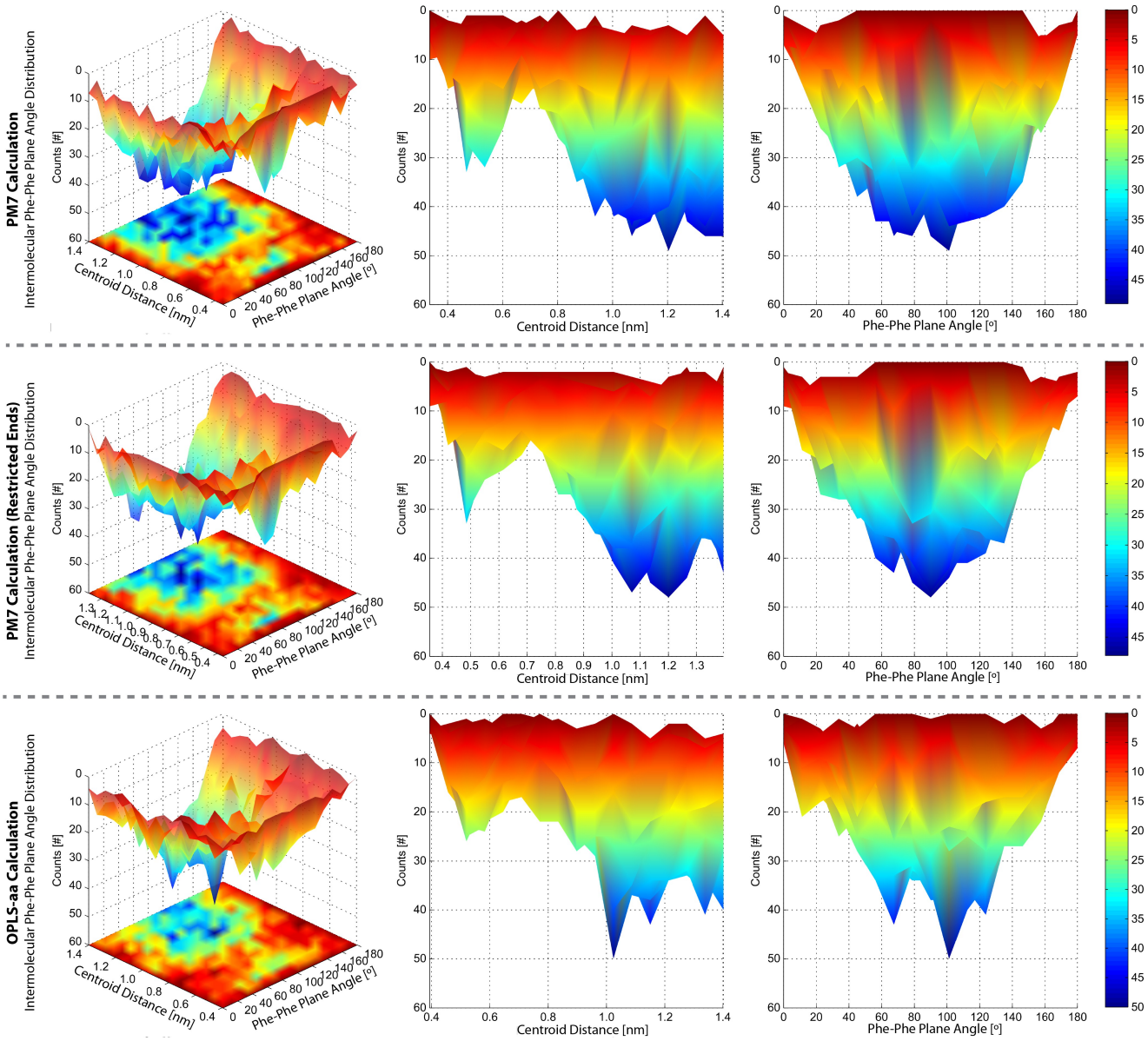


Fig. 11 Intermolecular Phe-Phe plane angle distributions versus centroid distance (R_{cen}) determined from PM7, restricted PM7 and OPLS-aa calculations of selected snapshots from a 10 ns long MD simulations performed with the OPLS-aa force field. π-π stacking with a R_{cen} above 7.5 *Å* is negligible, hence maximums above this distance yields information of the peptide configurations rather than π-π stacking.

Å compared to the OPLS-aa calculations. As the average RMSD (Protein-H) value of the reference simulation with four peptides is 1.96 Å, the restricted PM7 calculations are acceptable considering the imposed imitation of PBC.

The average Ca RMSD values of the PM7, restricted PM7 and small four peptide PM7 calculations are 5.99 Å, 2.68 *Å* and 1.31 Å, respectively. A CHARMM simulation study [70] of eight different proteins showed average RMSD (Ca) values (compared to the crystal structure) in the range of 1.06 *Å* - 3.58 Å. The same eight proteins had average RMSD (Ca) values ranging from 3.16 - 4.15 *Å* when the MARTINI force field was used. [71] Reports of RMSD values from other studies using different all-atom force fields (AMBER, CHARMM, GROMOS, OPLS-aa) range from 0.94 - 4 Å. [72,73] Considering this the obtained OPLS-aa results are in very good agreement with the restricted PM7 calculations.

Interestingly, the Phe residues obtain the lowest RMSD values of all the analyzed groups (Fig. 10). Only the Ca

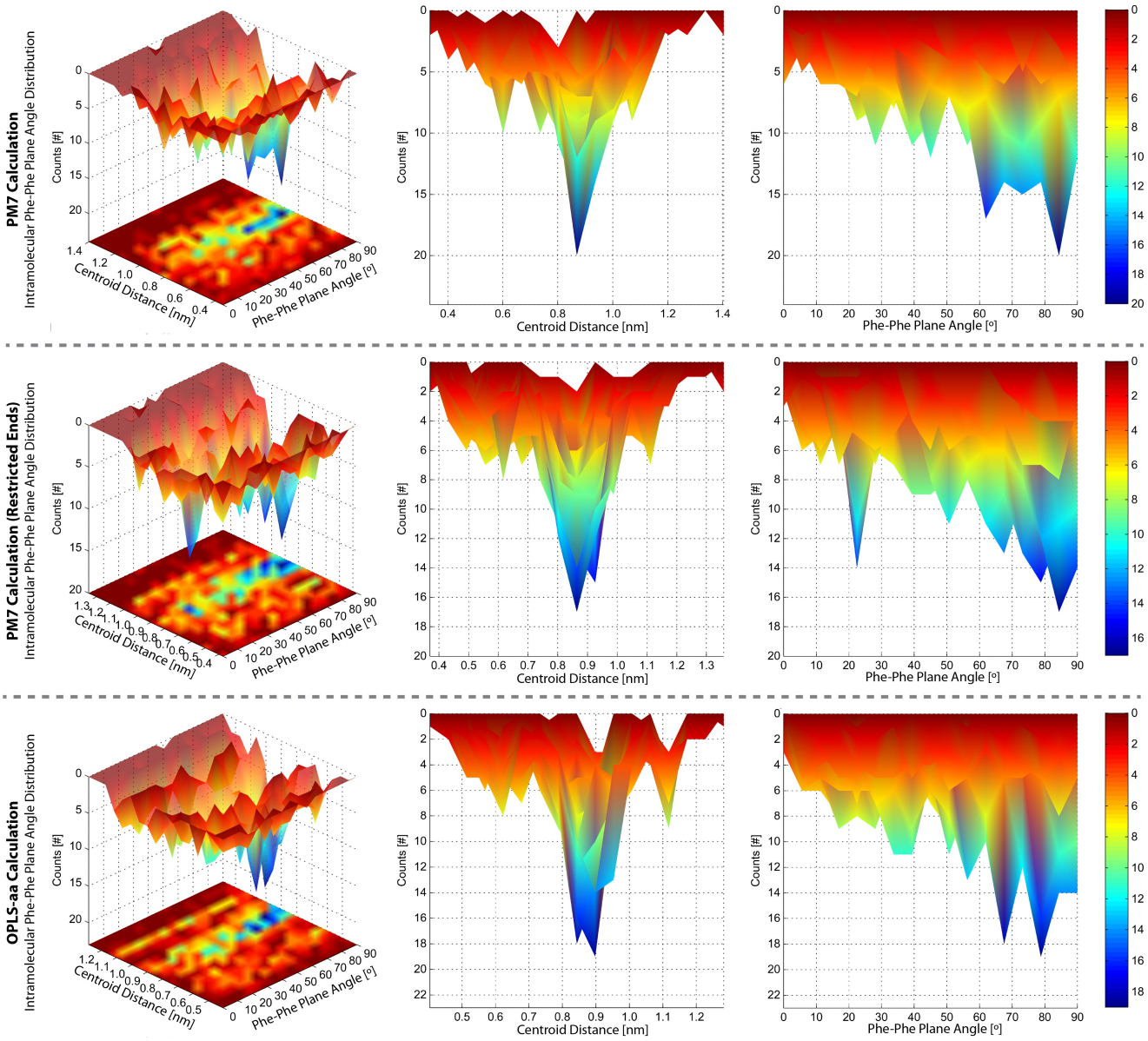
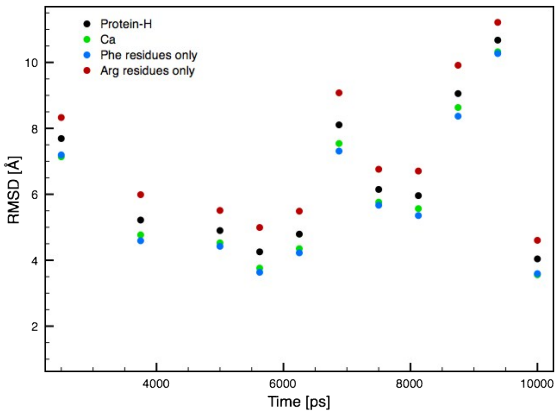


Fig. 12 Intramolecular Phe-Phe plane angle distributions versus centroid distance (R_{cen}) determined from PM7, restricted PM7 and OPLS-aa calculations of selected snapshots from a 10 ns long MD simulations performed with the OPLS-aa force field. π-π stacking with a R_{cen} above 7.5 $\AA$ is negligible, hence maximums above this distance yields information of the peptide configurations rather than π-π stacking.

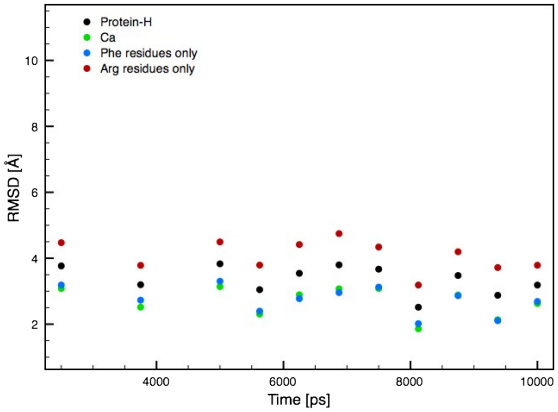
group from the calculations of the small simulation consisting of four peptides obtain a lower RMSD than the Phe residues groups. Hence the OPLS-aa Phe configurations are in very good agreement with the configurations obtained by PM7. The RMSD values of the Ca atoms are only marginal higher than the RMSD values for Phe residues, which means that the secondary structure is not expected to be significant different. In all simulation series, arginine residues yield the highest RMSD value, which is not surprisingly as this group is the most flexible group. It could also be an indication of a discrepancy between the two water models used, as implicit solvent was used in the PM7 calculations while explicit solvent was used in the OPLS-aa calculations.

Intermolecular Phe-Phe Plane Angle distribution

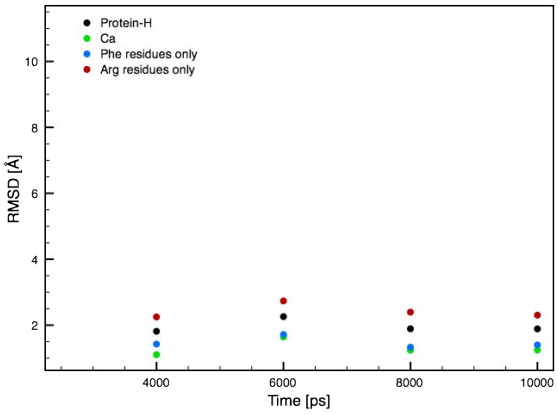
It is noteworthy that the intermolecular Phe-Phe plane angle distributions in Fig. 11, was calculated based on all Phe-Phe interactions within 1.4 nm. Hence the results are not directly



(a) RMSD values of OPLS-aa snapshots compared to the PM7 geometry optimized snapshots.



(b) RMSD values of OPLS-aa snapshots compared to the restricted PM7 geometry optimized snapshots.



(c) RMSD values of OPLS-aa snapshots, from a simulation with four peptides in a big simulation box, compared to the comparable PM7 geometry optimized snapshots.

Fig. 10 Overview of the RMSD values of the OPLS-aa snapshots compared to the PM7 geometry optimized counterpart.

comparable to Fig. 8, where only the Phe-Phe interactions of neighbor peptides were included in the analysis.

From Fig. 11, it is apparent that Phe residues preferably stack in the T-shaped mode in both PM7 calculation series and the OPLS-aa calculations. Furthermore, all calculations result in a saddle-point at approximately 7 $\AA$ that is close to the before mentioned limit (7.5 Å) where Phe-Phe stacking is pertinent. Hence both methods are equally accurate in determining the cut-off value, even when simulated under different situations. The small discrepancy between the experimental value and the simulated might be explained by the fact that 7.5 $\AA$ was determined from an investigation of the protein data bank.[62] Hence the analysis was based on intramolecular interactions rather than intermolecular interactions, which is the case in this study.

Overall the surface-shapes are very comparable with only few discrepancies. In the R_{cen} area where Phe-Phe stacking is pertinent ($R_{cen} < 7.5$ Å), it is apparent that the Phe side-chains may approach each other more in both PM7 calculation series than in the OPLS-aa calculations. From the PM7 calculations the closest observed R_{cen} was 3.41 Å, in the restricted PM7 calculations the smallest observed R_{cen} was 3.58 Å, while it was 3.97 $\AA$ during the OPLS-aa calculations. A similar tendency is observed for Phe residues that stack in a T-shape mode, though they in general are further apart by a small measure, due to steric hindrance (Fig. 11). It is likely that this difference is related to the absence of PBC in the PM7 calculations, however it does not explain the discrepancy between the restricted PM7 calculations and the OPLS-aa calculations. Hence a minor inconsistency might be introduced by the OPLS-aa force field compared to the PM7 method.

In the range 0 $\AA < R_{cen} <$ 7.0 $\AA$ of the Phe-Phe plane angle distribution, a more narrow distinct local maximum is observed for both PM7 calculation series compared to the OPLS-aa calculations (Fig. 11). However, a larger population of Phe-Phe residues are found to be in a T-shaped stacking mode in the OPLS-aa calculations (Table 1). Though the restrictive angle range of the T-shaped mode in the PM7 calculations seems rather imposed, Table 1 might imply that the propensity for Phe residues to stack in a T-shaped mode might be slightly exaggerated in the OPLS-aa force field.

Above the R_{cen} where Phe-Phe stacking is pertinent, the Phe-Phe plane angle distribution populate a slightly broader angle range in the PM7 calculations compared to the restricted PM7 calculations and even more so, compared to the OPLS-aa calculations (Table 2). This may be related to the absence of PBC in the PM7 calculations or an effect from the potential energy surface related to the specific force fields. However, the models conform rather well, which is an indication of the similarity in the peptide configurations rather than a similar π stacking configuration.

Table 1 Overview of the peR_{cen}tage population of three stacking modes, parallel, semi-parallel and T-shaped found from the calculation series performed with PM7, restricted PM7 and OPLS-aa of the selected snapshots. The population it only determined for intermolecular Phe pairs with a R_{cen} less than 7 Å.

$0\,Å < R_{cen} < 7.0\,Å$	PM7	Restricted PM7	OPLS-aa
Parallel stacking mode	18.1 %	18.6 %	20.1 %
Hybrid stacking mode	33.7 %	35.4 %	20.4 %
T-shaped stacking mode	48.2 %	46.0 %	59.5 %

Table 2 Overview of the peR_{cen}tage population of three stacking modes, parallel, semi-parallel and T-shaped found from the calculation series performed with PM7, restricted PM7 and OPLS-aa of the selected snapshots. The population it only determined for intermolecular Phe pairs with a R_{cen} greater than 7 $Å$ and less than 14 Å.

$7\,Å < R_{cen} < 14.0\,Å$	PM7	Restricted PM7	OPLS-aa
Parallel stacking mode	14.6 %	13.3 %	14.0 %
Hybrid stacking mode	37.7 %	37.5 %	34.6 %
T-shaped stacking mode	47.7 %	49.2 %	51.4 %

Intramolecular Phe-Phe Plane Angle Distribution

The intramolecular Phe-Phe plane angle distributions in Fig. 12 were calculated by the same method as Fig. 9. Hence the results are directly comparable, but for simplicity and to ensure that the observed PM7 results are not a consequence of selecting specific snapshots, the analysis was also conducted on the same OPLS-aa snapshots.

As was the case with the thorough investigation in Fig. 9, the amount of Phe-Phe interaction in the range where π-π stacking is pertinent is very low compared to the amount of interactions above this range. The data set in Fig. 12 is too low to obtain a detailed overview of the Phe-Phe plane angle distribution in the range within 0-7.5 Å. However, some minor peaks appear all over the angle range (0-90°), indicating the no preferable stacking mode exists in the R_{cen} range of 0 - 7.5 Å. In all calculations the minor peaks are centered on approximately 6 Å.

Above the R_{cen} where π-π stacking is pertinent the Phe residues preferably stack in a T-shaped mode with a R_{cen} of approximately 9 $Å$ in all calculations. However, in Fig. 9 a larger population of Phe pairs were found to stack in a parallel mode with a separation of 11 Å. This configuration is eliminated during PM7 global geometry optimization, while it appears in both the restricted PM7 and OPLS-aa calculation series. As was the case in Fig. 9, this mode is very restricted to a Phe-Phe separation of 11 $Å$ in the restricted PM7 and OPLS-aa calculation with very little R_{cen} flexibility and few transition possibilities. Due to the absence of this configuration in the PM7 calculation series, it is likely that the configuration is linked to a stretched fiber. This seems likely, as Phe residues in a peptide in a stretched fiber would need to spread wider (a larger intermolecular Phe pair separation would render the interactions negligible). Furthermore, it is know from Fig. 9 that this configuration originates from Phe-2 in relation to Phe-1 and Phe-3 and not from Phe-1 in relation to Phe-3. Hence the Phe residues in the configuration point in opposite directions and would yield a wider peptide.

Opposed to the analysis in Fig. 9, no favored stacking mode is observed at a R_{cen} of 11 Å. It is possible that this is due to the limited data-set. Or it could be a random consequence of the selected snapshots missing this configuration. However, it is noteworthy that an extra major local maximum is observed in the restricted PM7 calculation series with an Phe-Phe plane angle of 22° and a R_{cen} of 9 Å. This maximum was not observed in any other analysis. The origin of this mode is unknown but it is possible that it is the missing maximum at 20° and a R_{cen} of 11 $Å$ from Fig. 9a, that has shifted due to the allowance of the Phe-Phe pairs to move closer. However the mode is very distinct and seems very restrictive indicating a very fixed peptide configuration.

3.8 Critical Fiber Concentration

The critical fiber concentration of RFFFR was determined from a series of self-assembly MD simulations. Starting with randomly positioned peptides a number of simulations with the same amount of peptides, but at different concentrations, were performed to monitor if fibers or smaller segments thereof were formed. Several, factors such as pH, temperature and simulation system size also influences the critical fiber concentration, which was not investigated in the present study.

An overview of the performed simulations is given in Table 3. The simulations with peptide concentrations of 120 mM and 170 mM resulted in continuous fibers rapidly.

At a peptide concentration of 100 mM, the peptides assembled into fiber segments and in some cases a single fiber with disconnected ends was formed. As it might have been a matter of time before these fibers assembled into one continuous fiber, the simulations were extended to 1 μs, after which continuous fibers assembled in 2 out of 7 simulations. The extended timescale and the low ratio of simulations where fibers assembled, indicates that 100 mM is close to the critical fiber concentration.

The simulations with peptide concentration of 70 mM and 20 mM did not form single or continues fibers, not even on a 1 μs timescale. Instead the peptides assembled into many small fiber or micelle like structures, indicating that these are below the critical fiber concentration. Hence it is likely that under these simulation conditions the critical fiber concentration of RFFFR is between 70 mM and 100 mM.

Table 3 Overview of simulations performed to investigate at which peptide concentrations fibers self-assemble from initial randomly positioned peptides. All simulations contain the same amount of peptides.

Concentration (mM)	Fibers/Simulations	Durations (ns)
170	1/1	200
120	1/1	200
100	2/7	1000
70	0/5	1000
20	0/1	200

To verify the results an additional series of simulations was performed where the single continuous fiber from the simulation of the peptide concentration 120 mM was inserted into a larger simulation box yielding a peptide concentration of 100 mM and 70 mM. An overview of these simulations is illustrated in Table 4.

It is apparent that the fiber remains stable in all simulations at 100 mM peptide concentration but disassembles in all simulations at 70 mM peptide concentration. Hence these results are coherent with the conclusion from the self-assembly simulation series and the critical fiber concentration is found to be in between 70 mM and 100 mM.

Table 4 Overview of simulations performed where the single continues fiber from the simulation with 120 mM were inserted into bigger simulation boxes in order to verify that the the critical fiber concentration is between 70 mM and 100 mM.

Concentration (mM)	Fiber Remained Stable	Durations (ns)
100	5/5	200
70	0/5	200

4 Conclusions

The RFFFR peptide was designed to form nano-fibers and the self-assembly potential has been investigated through a combination of CG-, atomistic MD simulations and semi-empirical quantum mechanical calculations. Arg residues were designed to restrict the interactions of Phe-Phe in one direction and ensure solubility. Thereby the self-assembly of the peptide preferred fiber formation over random aggregation as has been reported for FF and FFF peptides.[4,9–12]

Above a critical fiber concentration determined to be in the interval of 70 mM - 100 mM, the peptides assemble into fibers according to a three-step process. Initially randomly dispersed peptides aggregate into small clusters, as these grow larger, they assume small fiber segment structures. These small fibers grow into one large fiber spanning the PBC.

Hydrophobic effects might play an important role in the self-assembly and the final fiber is found to be stabilized by a large amount of intermolecular and water mediated hydrogen bonds. However, the amount is rather underestimated compared to semi-empirical quantum mechanical calculations, indicating that hydrogen bonds might play an even larger role.

π-stacking interactions between Phe residues are also found to be important for the self-assembly process. Intermolecular Phe residues favorably stack in a distinct T-shaped mode, while intramolecular Phe residues, within the range where π-stacking is pertinent, are found to stack in no distinct mode. Semi-empirical quantum mechanical calculations verified these results with only minor differences between the OPLS-aa and PM7 calculations. Among the most significant deviations were a slightly increased propensity for Phe residues to assume T-shaped stacking modes.

As π-stacking has been proven to promote amyloid formation and is believed to be the most important factor, the RFFFR peptide has proven to be a novel suitable model system for investigation of the formation, stability and disassembly of amyloids as well.[74–77] Furthermore, the structure formed by RFFFR has unique properties that can be exploited in other applications such as biological nanowires with conductive properties facilitated through charge transport between overlapping delocalized aromatic π-orbitals as well.

References

1 G. Bhak, Y. Choe and S. R. Paik, *BMP reports*, 2009, **42**, 541–551.
2 M. Stefani and C. M. Dobson, *Journal of Molecular Medicine*, 2003, **81**, 678–699.
3 E. Gazit, *Chemical Society Reviews*, 2007, **36**, 1263–1269.
4 T. H. Han, T. Ok, J. Kim, D. O. Shin, H. Ihee, H.-S. Lee and S. O. Kim, *Small*, 2003, **6**, 945–951.
5 L. Liu, K. Busuttil, S. Zhang, Y. Yang, C. Wang, F. Besenbachera and M. Dong, *Physical Chemistry Chemical Physics*, 2011, **13**, 17435–17444.
6 J. V. González-Aramundiz, M. V. Lozano, A. Sousa-Herves, E. Fernandez-Megia and N. Csaba, *Expert Opinion on Drug Delivery*, 2012, **9**, 183–201.
7 C. Sotoa, M. S. Kindyc, M. Baumannb and B. Frangione, *Biochemical and Biophysical Research Communications*, 1996, **226**, 672–680.
8 M. Reches and E. Gazit, *Science*, 2003, **300**, 625–627.
9 T. H. Han, W. J. Lee, D. H. Lee, J. E. Kim, E.-Y. Choi and S. O. Kim, *Advanced Materials*, 2010, **22**, 2060–2064.
10 M. Hashemi, P. Fojan and L. Gurevich, *Journal of Self-Assembly and Molecular Electronics*, 2013, **1**, 195–208.
11 X. Yan, P. Zhu and J. Li, *Chemical Society Reviews*, 2010, **39**, 1877–1890.
12 P. Tamamis, L. Adler-Abramovich, M. Reches, K. Marshall, P. Sikorski, L. Serpell and E. Gazit, *Biophysical Journal*, 2009, **96**, 5020–5029.
13 C. H. Görbitz, *Chemistry - A European Journal*, 2001, **7**, 5153–5159.
14 N. Thota, Z. Luo, Z. Hu and J. Jiang, *Journal of Physical Chemistry B*, 2013, **117**, 9690–9698.
15 S. J. Marrink, A. H. de Vries and A. E. Mark, *Journal of Physical Chemistry B*, 2004, **108**, 750–760.
16 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.
17 J. Corsi, R. W. Hawtin, O. Ces, G. S. Attard and S. Khalid, *Langmuir*, 2010, **26**, 12119–12125.

18 C. A. López, A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger and S. J. Marrink, *Journal of Chemical Theory and Computation*, 2009, **5**, 3195–3210.

19 H. Lee, A. H. de Vries, S. J. Marrink and R. W. Pastor, *The Journal of Physical Chemistry B*, 2009, **113**, 13186–13194.

20 J. Wong-Ekkabut, S. Baoukina, W. Triampo, I.-M. Tang, P. D. Tieleman and L. Monticelli, *Nat Nano*, 2008, **3**, 363–368.

21 S. J. Marrink and D. P. Tieleman, *Chemical Society Reviews*, 2013, **42**, 6801–6822.

22 J. S. rensen, X. Periole, K. K. Skeby, S. J. Marrink and B. S. tt, *Biophysical Chemistry*, 2011, **2**, 2385–2390.

23 M. Baaden and S. J. Marrink, *Structural Biology*, 2013, **23**, 878–886.

24 K. Prasitnok and M. R. Wilson, *Physical Chemistry Chemical Physics*, 2013, **15**, 17093–17104.

25 A. J. Rzepiela, M. Louhivuori, C. Peter and S. J. Marrink, *Physical Chemistry Chemical Physics*, 2011, **13**, 10437–10448.

26 T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink and D. P. Tieleman, *Journal of Chemical Theory and Computation*, 2014, **10**, 676–690.

27 W. L. Jorgensen and J. Tirado-Rives, *Journal of the American Chemical Society*, 1988, **110**, 1657–1666.

28 G. A. Kaminski and R. A. Friesner, *Journal of Physical Chemistry B*, 2001, **105**, 6474–6487.

29 R. S. Paton and J. M. Goodman, *Journal of Chemical Information and Modeling*, 2009, **49**, 944–955.

30 S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.

31 E. Krieger, G. Koraimann and G. Vriend, *Proteins: Structure, Function, and Bioinformatics*, 2002, **47**, 393–402.

32 G. Bussi, D. Donadio and M. Parrinello, *The Journal of Chemical Physics*, 2007, **126**, –.

33 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684–3690.

34 A. Cheng and K. M. Merz, *The Journal of Physical Chemistry*, 1996, **100**, 1927–1937.

35 M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182–7190.

36 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577–8593.

37 J. J. P. Stewart, *"MOPAC"*, 2012, http://openmopac.net.

38 J. J. P. Stewart, *Journal of Molecular Modeling*, 2013, **19**, 1–32.

39 M. Korth, M. Pitoáňk, J. Řezáč and P. Hobza, *Journal of Chemical Theory and Computation*, 2010, **6**, 344–352.

40 N. D. Yilmazer and M. Korth, *Computational and Structural Biotechnology Journal*, 2015, **13**, 169–175.

41 A. Klamt and G. Schuurmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, **5**, 799–805.

42 D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

43 R. Baron, A. H. de Vries, P. H. Hnenberger and W. F. van Gunsteren, *J. Phys. Chem. B*, 2006, **110**, 8464–8473.

44 R. Baron, A. H. de Vries, P. H. Hnenberger and W. F. van Gunsteren, *J. Phys. Chem. B*, 2006, **110**, 15602–15614.

45 G. Ramachandran and A. K. Mitra, *Journal of Molecular Biology*, 1976, **107**, 85–92.

46 A. Jabs, M. S. Weiss and R. Hilgenfeld, *Journal of Molecular Biology*, 1999, **286**, 291–304.

47 R. Sibson, *The Computer Journal*, 1973, **16**, 30–34.

48 D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan and W. Wriggers, *Science*, 2010, **330**, 341–346.

49 P. L. Freddolino, F. Liu, M. Gruebele and K. Schulten, *Biophysical Journal*, 2008, **94**, L75–L77.

50 S. Chowdhury, M. C. Lee, G. Xiong and Y. Duan, *Journal of Molecular Biology*, 2003, **327**, 711–717.

51 C. Simmerling, B. Strockbine and A. E. Roitberg, *Journal of the American Chemical Society*, 2002, **124**, 11256–11259.

52 D. Chandler, *Nature*, 2005, **437**, 640–647.

53 A. Biela, N. N. Nasief, M. Betz, A. Heine, D. Hangauer and G. Klebe, *Angewandte Chemie International Edition*, 2013, **52**, 1822–1828.

54 C. Guo, Y. Luo, R. Zhou and G. Wei, *Nanoscale*, 2014, **6**, 2800–2811.

55 C. Guo, Y. Luo, R. Zhou and G. Wei, *Nano*, 2012, **6**, 3907–3918.

56 *An Introduction to Hydrogen Bonding*, ed. G. Jeffrey, Oxford University Press, Oxford, 1997.

57 *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. G. Fasman, Plenum, New York, 1990.

58 H. Fenniri, P. Mathivanan, K. L. Vidale, D. M. Sherman, K. Hallenga, K. V. Wood and J. G. Stowell, *Journal of the American Chemical Society*, 2001, **123**, 3854–3855.

59 D. Xu, C. J. Tsai and R. Nussinov, *Protein Engineering*, 1997, **10**, 999–1012.

60 D. A. Evans, K. T. Chapman, D. T. Hung and A. T. Kawaguchi, *Angewandte Chemie International Edition in English*, 1987, **26**, 1184–1186.

61 J. Vondrek, L. Bendov, V. Klusk and P. Hobza, *Journal of the American Chemical Society*, 2005, **127**, 2615–2619.

62 G. B. McGaughey, M. Gagné and A. K. Rappé, *The Journal of Biological Chemistry*, 1998, **273**, 15458–15463.

63 D. Ranganathan, V. Haridas, R. Gilardi and I. L. Karle, *Journal of the American Chemical Society*, 1998, **120**, 10793–10800.

64 C. A. Hunter, *Philosophical Transactions: Physical Sciences and Engineering*, 1993, **345**, 77–85.

65 P. Jureka and P. Hobza, *Journal of the American Chemical Society*, 2003, **125**, 15608–15613.

66 G. Chessari, C. A. Hunter, C. M. R. Low, M. J. Packer, J. G. Vinter and C. Zonta, *Chemistry A European Journal*, 2002, **8**, 2860–2867.

67 S. Sun and E. R. Bernstein, *Journal of Physical Chemistry*, 1996, **100**, 13348–13366.

68 C. Chipot, R. Jaffe, B. Maigret, D. A. Pearlman and P. A. Kollman, *J. Am. Chem. Soc.*, 1996, **118**, 11217–11224.

69 S. K. Burley and G. A. Petsko, *Science*, 1985, **229**, 23–28.

70 M. Feig, *Journal of Chemical Theory and Computation*, 2008, **4**, 1555–1564.

71 J. Gu, F. Bai, H. Li and X. Wang, *Int J Mol Sci.*, 2012, **13**, 14451–14469.

72 D. J. Price and C. L. Brooks, *Journal of Computational Chemistry*, 2002, **23**, 1045–1057.

73 L.-B. R, S. LG, P.-F. L and V. H, *PLoS ONE*, 2015, **10**,.

74 E. Gazit, *FEBS Journal*, 2005, **272**, 5971–5978.

75 D. Zanuy and R. Nussinov, *Journal of Molecular Biology*, 2003, **329**, 565–584.

76 A. Caflisch, *Current Opinion in Chemical Biology*, 2006, **10**, 437–444.

77 A. P. Pawara, K. F. DuBay, J. Zurdo, F. Chiti, M. Vendruscolo and C. M. Dobson, *Journal of Molecular Biology*, 2005, **350**, 379–392.