Journal Name

ARTICLE

Using experimental and computational energy equilibration to understand hierarchical self-assembly of Fmoc-dipeptide amphiphiles

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

I. R. Sasselli,[a] C. Pappas,[a,b] E. Matthews,[a] T. Wang,[c] N. T. Hunt,[d] R. V. Ulijn[a,b,e,*] and T. Tuttle[a,*]

Despite progress, fundamental understanding of the relationships between molecular structure and self-assembly configuration of Fmoc-dipeptides is still in its infancy. In this work we provide a combined experimental/computational approach that make use of free energy equilibration of a number of related Fmoc-dipeptides to arrive at an atomistic model of Fmoc-threonine-phenylalanine-amide (Fmoc-TF-NH_2) which forms twisted fibres. By using dynamic peptide libraries where closely related dipeptide sequences are dynamically exchanged to eventually favor the formation of the thermodynamically most stable configuration, the relative importance of C-terminus modifications (amide versus methyl ester) and contributions of aliphatic versus aromatic amino acids (phenylalanine F vs leucine L) is determined (F>L and NH_2>OMe). The approach enables the comparative interpretation of spectroscopic data, which can then be used to aid the construction of the atomistic model of the most stable structure (Fmoc-TF-NH_2). The comparison of the relative stabilities of the models using molecular dynamic simulations and the correlation with experimental data using dynamic peptide libraries and a range of spectroscopy methods (FTIR, CD, fluorescence) allows for the determination of the nanostructure with atomistic resolution. The final model obtained through this process is able to reproduce the experimental observed formation of intertwining fibres for Fmoc-TF-NH_2, providing information of the interactions involved in the hierarchical supramolecular self-assembly. The developed methodology and approach should be of general use for the characterization of supramolecular structures.

Introduction

Self-assembling peptide amphiphiles represent a promising minimalistic approach for the formation of dynamic materials with tuneable properties and potential applications in biomedicine and nano/biotechnology.[1] There are now many examples of dipeptides protected at their N-terminus with a variety of aromatic groups,[2] most commonly the Fmoc (9-fluorenylmethyloxycarbonyl) moiety (Scheme 1) (as well as other aromatic ligands).[3] A number of these Fmoc-peptides form nanostructures through a combination of aromatic stacking and H-bonding interactions.[2-4] It is now clear that a variety of nanoscale structures may be obtained depending on the chemical nature of the building blocks used. Despite progress,[2-3] fundamental understanding of the relationships between molecular structure and self-assembly configuration is still in its infancy.

The dipeptide sequence has been shown to have important effects on the nanostructure shape even for small variations in the amino acid side chain.[5] The modification of the C-terminus has also previously been shown to have an influence on the self-assembled structures formed.[2] In particular, Ryan *et al.* demonstrated an enhanced self-assembly propensity for fluorinated Fmoc-phenylalanine (Fmoc-F) with an amidated C-terminus compared to the free acid terminated version.[6]

The relative importance of the supramolecular interactions (*e.g.,* π-stacking versus H-bonding) which give rise to various structures formed (*e.g.,* fibres, spheres, sheets)[2] are evident, but the underlying changes in supramolecular interactions are not well understood. Thus, for a more rational approach to the design of aromatic peptide amphiphiles for specific applications it is imperative that an understanding of the exact impact of different chemical groups on intermolecular interactions in the nanostructure is obtained. Therefore, the supramolecular arrangements in Fmoc-dipeptide nanostructures should ideally be understood with atomistic detail.

Typically, combinations of spectroscopic techniques have been used to elucidate the role of the different interactions in the nanostructure formation, but they do not give direct information of the groups involved or their conformations.[7] Furthermore, although X-ray diffraction techniques can provide structural information, it is still not clear if the results obtained

[a] Department of Pure & Applied Chemistry, WestCHEM, University of Strathclyde. 295 Cathedral Street, Glasgow, G1 1XL, UK. Email: tell.tuttle@strath.ac.uk
[b] Advanced Science Research Center (ASRC), City University of New York (CUNY), 85 St Nicholas Terrace, New York, New York 10031, USA.
[c] Imaging Facility of CUNY, Advanced Science Research Center (ASRC), 85 St Nicholas Terrace, New York, New York 10031, USA.
[d] Department of Physics, University of Strathclyde, SUPA, 107 Rottenrow East, Glasgow, G4 0NG, UK.
[e] Hunter College, Department of Chemistry and Biochemistry, 695 Park Avenue, New York, New York 10065, USA. Email: rein.ulijn@asrc.cuny.edu
Electronic Supplementary Information (ESI) available: ESI and simulations videos. See DOI: 10.1039/x0xx00000x

in the dried sample are comparable with systems that are usually 95-99 wt% of water.[8]

All-atom molecular dynamics (MD) simulations have been applied for the study of different supramolecular systems to gain understanding of the intermolecular arrangements and mechanisms of formation of the nanostructures.[9] In some of these cases, the relative stability of all possible arrangements were compared using small systems of 4 to 8 molecules and short simulations of around 15 – 25 ns.[9a-d] However, one could expect the results obtained to be influenced by the limited size of the systems studied as the objects formed are typically composed of many millions of molecules (≈ 3.5·106 molecules in a 20 nm width, 1 micron length fibre). While these sizes are unfeasible using currently available computational approaches, it is clear that the system should contain at least sufficient molecules to approach a short section of a complete fibre. Upon simulating larger systems, the computational costs increases and not all conformations can be tested. Hence a starting structure is required, which is usually proposed on the basis of experimental studies and subsequently tested for stability over time.

Several different models have been proposed for Fmoc-dipeptide nanostructures based on experimental observations. Smith *et al.* presented an early model for Fmoc-FF-OH that was informed by spectroscopy, x-ray diffraction and high-resolution electron microscopy data. In that paper it was proposed that nanostructures were formed by antiparallel H-bonded peptide stacks interlocked in a zipper-like fashion by extended π-stacking interactions involving both the fluorenyl groups and the F side chains.[10] The interlocked stacks resulted in elliptical, chiral fibres which were in qualitative agreement with spectroscopic characterization results, although interpretation of the FT-IR results as antiparallel beta sheets has since been shown not to be diagnostic of an antiparallel arrangement.[11] Similar H-bonded stacks interlocked via π-stacking interactions were used to show the formation of 2D nanostructures from Fmoc-serine-phenylalanine-methyl ester (Fmoc-SF-OMe).[4c, 12] However, to date MD simulations have not been able to demonstrate the stability of these arrangements.

Mu *et al.* composed a side by side study of a number different supramolecular starting points for Fmoc-di-alanine (Fmoc-AA), using arrangements with the Fmoc group in different dispositions but always placed in the core of the fibre.[9e] The models suggest that Fmoc-AA-OH nanostructures are mainly the result of the hydrophobic effect that positions the fluorenyl groups in the core of the fibres, in parallel stacks, where they can establish π-stacking interactions and expose the more hydrophilic part of the molecule (typically the N-terminal carboxylic acid) to the solvent.[9e] MD simulations were used to demonstrate the stability of these arrangements; however the resulting structures did not contain H-bonding interactions, which is not consistent with the FT-IR data or the supramolecular ellipticity observed using circular dichroism (CD). In addition, Eckes *et al.* replaced the amide group between the two alanines to show that the system can still self-assemble into nanostructures, suggesting that amide H-bonding is not critical for these nanostructures to form.[9f]

The relative contributions of non-covalent interactions acting cooperatively determine the thermodynamic tendency of an Fmoc-peptide to self-assemble. However, it is known that many supramolecular gels do not represent a thermodynamic equilibrium state and are instead kinetically trapped in local minima.[8a, 13] Often there are kinetic effects which can lead to structures that differ from the thermodynamically favoured one, giving rise to metastable gel polymorphs (which represent local minima).[14] Therefore, in order to elucidate molecular level insights that are of relevance to thermodynamic modelling, it is important to control the self-assembly process to ensure the formation of the thermodynamically favoured nanostructure, which can be reproducibly accessed.

In this work we study four closely related Fmoc-dipeptides that form nanostructures: Fmoc-TF-NH_2, Fmoc-TF-OMe, Fmoc-TL-NH_2 and Fmoc-TL-OMe (Fig. 1b). These are formed enzymatically in situ by the condensation of the two precursors (Fig. 1a): Fmoc-T (A) and B_i (F-NH_2, F-OMe, L-NH_2 and L-OMe);[5a, 15] the enzymatic self-assembly ensures the thermodynamic control of the self-assembly process and allows for direct comparison of their thermodynamic stability using a dynamic library approach[7d, 15b, 16]. The four molecules have the Fmoc group and the first amino acid (T) in common, and the variations are the phenylalanine/leucine substitution (F/L) and the C-terminus substitution (amidated, NH_2/methyl ester, OMe). F and L differ in the aromaticity of the former and in its higher hydrophobicity (logP=1.25) compared with L (logP=0.91).[17]

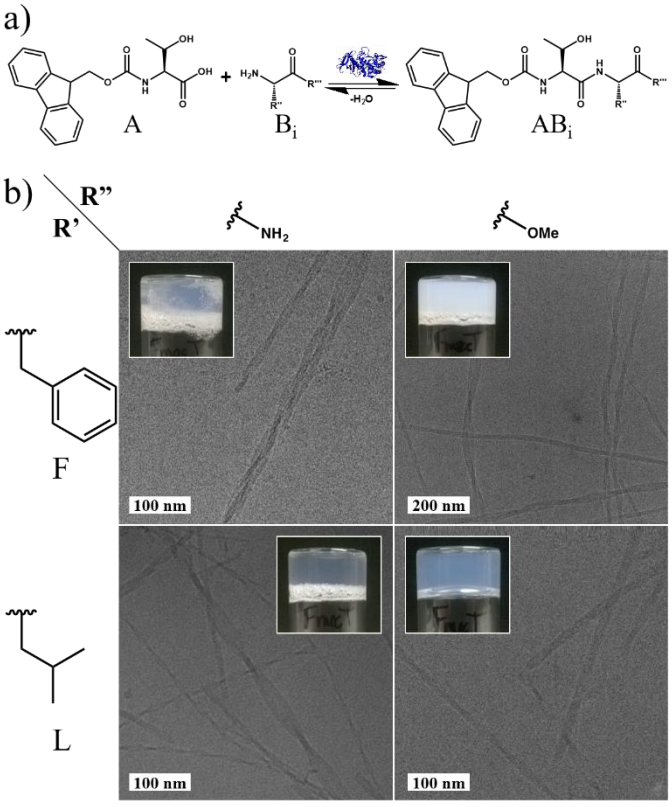

Fig. 1. (a) Condensation reaction of Fmoc-T (A) and C-protected amino acid (B_i) via reverse hydrolysis in the presence of thermolysin to form the Fmoc-dipeptide (AB_i); and, (b) amino acid side chains (R') and C-termini (R") for the four Fmoc-dipeptide under study and the corresponding cryo-TEM images and gel pictures inset.

The experimental information of the supramolecular interactions (obtained from FT-IR, fluorescence and CD) and shape (TEM) is then used to build two atomistic models of Fmoc-TF-NH_2 fibres. These models are compared using MD simulations and the analysis of the interactions through the simulations is subsequently correlated with the experimental data in order to determine which model fits better with the actual molecular arrangement in the nanostructures. The final model is then able to accurately describe the experimental observed formation of intertwined fibres forming nanoscale twisted structures for Fmoc-TF-NH_2.

Results and discussion

Experimental results

In order to establish the relative self-assembly propensities of four related Fmoc-dipeptides, they were compared side-by-side and in direct competition using biocatalytic self-assembly to ensure assembly was fully reversible (under thermodynamic control).[15b] The Fmoc-dipeptide nanostructures were first formed separately by in situ condensation of Fmoc-T with F-NH_2, F-OMe, L-NH_2 or L-OMe to form Fmoc-TF-NH_2, Fmoc-TF-OMe, Fmoc-TL-NH_2 or Fmoc-TL-OMe, respectively. The yields of the Fmoc-dipeptide formation in the isolated systems are ~81%, ~90%, 42% and 77%, respectively (Fig. S1).

The macroscopic appearance of the gels after 6h (at which the yields are: ~79%, ~88%, 36% and 80%, respectively, Fig. S1) differ in their level of transparency (Fig. 1b insets), which is due to the entanglement of the structures resulting in different light scattering.[18] We note that the yields of each isolated system are slightly different at 6h relative to the yields after 48h. Thus, while the systems are largely equilibrated after 6 h, there remains some variation (small increases or decreases in total yields) over a longer time period due to kinetic effects such as nucleation, which can retard the equilibration process. Nonetheless, the macroscopic features of the gel are not altered during this latter equilibration phase.

The cryo-TEM images (Fig. 1b) show that the four systems are similarly structured at the nanoscale, forming twisted fibres, for the four systems (Fig. 1b: ~20 nm, Fmoc-TF-NH_2; ~15-50 nm, Fmoc-TF-OMe; ~10-30, Fmoc-TL-NH_2; and ~20 nm, Fmoc-TL-OMe) (further images are supplied in Supporting Information – Figures S2 and S3). Some images show patterns in the ribbons of certain diameter changes (Fig. 1b: ~10 nm, Fmoc-TF-NH_2; ~20 nm, Fmoc-TF-OMe; ~20 nm in Fig. S2b; ~10 nm in Fig. S3a and S3d), which suggest that the final ribbons might be formed by lateral aggregation of fibres. We note that the similarity between these structures provides an interesting starting point – previous reports have focused on small changes in the peptide resulting in dramatic changes in nanostructures;[5a, 19] however, the current systems all four peptides give rise to closely related structures while their molecular structures vary in H-bonding (OMe/NH_2) capability and aromatic stacking (F/L).

The peptide condensation yields observed for the direct competition dynamic peptide library (DPL) experiment (Fig. 2a) provide the relative self-assembling tendency of the four Fmoc-

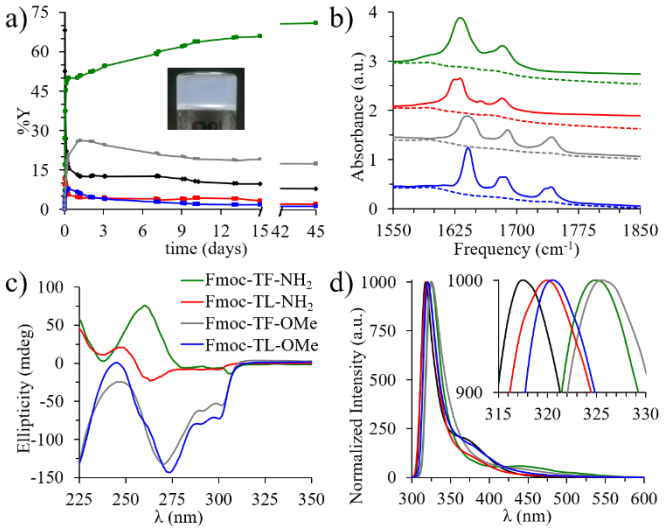

Fig. 2. (a) DPL results (sample picture inset); (b) FT-IR (precursors in dashed lines and gels in solid lines), (c) CD and (d) fluorescence spectra (λ_{max} zoomed inset) of the four gelators. Fmoc-dipeptides colour code is provided in (c), the black lines in (a) and (d) are the Fmoc-T precursor.

dipeptides (as conceptually demonstrated previously)[15b, 16a]: Fmoc-TF-NH_2 (~70%), Fmoc-TF-OMe (~18%), Fmoc-TL-NH_2 (~2%) and Fmoc-TL-OMe (~1%). The dramatic differences obtained in competition compared to the condensation yields in single peptide systems demonstrate that these differences are not a reflection of thermolysin's selectivity or the individual stability of the gelators formed (Fig. S4), but reflect the relative thermodynamic stability of the resulting nanostructures from the peptide amphiphiles. These results show Fmoc-TF-NH_2 is formed in a yield that is four times higher than that of the second most stable product (Fmoc-TF-OMe).

The preference for the TF dipeptides over TL indicates that the presence of the aromatic group is the main differentiating effect driving the self-assembly, which can be either due to the enhanced hydrophobicity, due to the extra π-stacking interactions, or, most likely, due to a combination of both effects. Fmoc-TL-NH_2 and Fmoc-TL-OMe are formed in negligible amounts in the competition experiment. The secondary effect, which governs the relative ranking of the systems with the same peptide unit, is the presence of the amide group, which is likely to be due to the extra hydrogen bonds that this group is able to form upon self-assembly. Although the F containing molecules also show higher yields in the isolated systems, the DPL yields and their order differ to those obtained in the isolated systems, which suggests that the percentage yields obtained from the pure systems cannot always be used to assess the relatively stabilities of the nanostructures, as has been shown before.[5a] To further verify the reversibility of the system, demonstrating that it is indeed possible to reversibly access thermodynamic minima, a sequential competitive experiment showed that Fmoc peptides (competition between Fmoc-TL-OMe and Fmoc-TL-NH_2, where L-NH_2 is added after 48 h) with better self-assembling propensity can be sequentially accessed (Fig. S5) i.e. it is

possible to navigate the free energy landscape experimentally using the DPL approach.

Having established the relative stability of the four Fmoc-peptides studied, we subsequently investigated their self-assembly propensities side-by-side. The Fmoc-peptides were produced enzymatically to ensure the structures formed can be reproducibly accessed and represent thermodynamically favourable structures. The FT-IR spectra in the region 1600-1800 cm^{-1} are analysed to study the H-bonding patterns of the amide region of the molecules. The presence of narrow and shifted amide and carbamate peaks in this region evidence the formation of ordered H-bonded stacks but does not provide information of specific arrangements.[11] All the systems show the peaks for the amide (1620-1650 cm^{-1}) and the carbamate (1675-1690 cm^{-1}) vibrations (Fig. 2b), which demonstrate the presence of extended hydrogen bonding stacks.[10-11, 20] Beyond distinction between carbamate and amide carbonyls, the results do not give a clear indication of which groups, or residues, are interacting through hydrogen bonds, but it is clear that there are differences in the stacks between the molecules with different C-termini. As can be expected, the –OMe gelators give rise to an extra peak around 1745 cm^{-1} due to the vibration of the carbonyl in the terminal ester group (Fig. 2b grey and blue lines), but also the amide and carbamate peaks appear at higher frequencies for OMe containing molecules than for the amidated ones. The position of the OMe containing gelators' peaks positions are consistent with previous work, which also mentioned the important role of the H-bonded stacks and the lack of influence of the inclusion of the aromatic side chain in these stacks.[5a, 19b] The amide peak for the terminally amidated Fmoc-dipeptides appears around 1630 cm^{-1} and around 1640 cm^{-1} for the methyl ester molecules; while the carbamate peak is observed around 1682 cm^{-1} for the NH_2 and around 1687 cm^{-1} for the OMe. This shift to lower frequencies indicates a more effective coupling of the carbonyl vibrations in the amidated Fmoc-dipeptides. The better coupling in the NH_2 containing nanostructures suggests that vibration of this terminal group is closer in frequency to the other carbonyl containing vibrations. A more ordered stack due to the extra H-bonding donor could also explain the better coupling. Most likely a combination of both effects plays a role here. Therefore, the differences in the FT-IR, and hence, in the H-bonds cannot explain the preference observed in the DPL for F containing molecules, but can explain the secondary preference, the higher stability of the nanostructures formed by amidated Fmoc-dipeptides.

The FT-IR spectra obtained from gels containing L (Fig. 2b red and blue lines) show broadened peaks compared to the F peptides that appear as double peaks: the amide vibration peak of the Fmoc-TL-NH_2 (red) and the carbamate and methyl ester group vibrations of the Fmoc-TL-OMe (blue). These peaks were shown to be due to kinetically trapped states in a time course experiment using Fmoc-TL-OMe (Fig. S6). In this experiment these peaks did not appear upon slow enzymatic controlled formation of the gel, but they appeared upon disruption and quick reformation of the gel, which suggests such structures represent kinetic products.

The fluorenyl group does not contain a stereo centre, and hence the CD signal shown in the fluorenyl absorption region (250 – 310 nm, Fig. S8) is due to the supramolecular ellipticity originating from the formation of a chiral nanostructure (Fig. 2c), as previously described and demonstrated by the appearance of chiral structures in TEM.[7b-d, 21] The CD spectra of the four systems show a degree of supramolecular ellipticity which is more intense in the systems with a methyl ester C-terminus. This dependence on the C-terminus substitution, suggests a correlation between the hydrogen bonds and ellipticity of the structure. The positive CD signal that is observed between 250 and 275 nm for the Fmoc-TF-NH_2 is due to the phenylalanine. This signal is masked in the case of the other F containing molecule, Fmoc-TF-OMe, due to the strong supramolecular ellipticity of the Fmoc group which appears in the same region.

The fluorescence λ_{max} red shift has been previously related to the formation of π-stacking interactions, with a red-shift in emission typically associated with enhanced stacking.[5a, 7a, 10, 19b, 20, 22] In our study, to address and compare the red shift in the four gels we use the Fmoc-T precursor solution, which appears at 317.5 nm, as reference (Fig. 2d). It can be seen that the maximum red shift is shown by Fmoc-TF-OMe, 8nm, and the second, by Fmoc-TF-NH_2, 7.5 nm. The red shift for Fmoc-TL-OMe and Fmoc-TL-NH_2 is clearly lower, 3 nm and 2.5 nm, respectively. This observation suggests that the presence of F strengthens the Fmoc – Fmoc π-stacking, or introduces extra F – Fmoc interactions, which is consistent with previous studies.[5a, 19b] The second structural feature which shows an effect on the π-stacking is the presence of the OMe C-terminus. Although this effect is small (0.5 nm), it is consistent in the systems and opposite to the effect of the higher hydrophobicity of the environment compared to the amidated Fmoc-dipeptides.

The fluorenyl emission spectra show additional peaks that are commonly used to assess changes in the supramolecular structure, which are the peaks at 365 nm and at 450 nm. The first peak is due to the fluorescence of the excimer and the latter one has been associated with phosphorescence of the excimer in fluorenyl derivatives[22] and has been correlated with formation of extended fibrils thought to involve formation of stacked Fmoc-aggregates.[2, 10] Structurally, the presence of the 365 nm peak is attributed to the presence of parallel Fmoc arrangements in micellar aggregates,[5a, 10, 19b, 20b] which is supported in this study by the presence of this peak for Fmoc-T (Fig. 2d black). Although the gels also show some emission on this area (especially Fmoc-TL-OMe, Fig. 2d blue), this is likely due to the fact that the conversion of this system reaches only ~70%, and hence, there is still a substantial amount of free Fmoc-T molecules which remains as micellar aggregates.

The fluorescence results can provide insight into the preference, observed in the DPL experiments, for TF-containing molecules over TL peptides to self-assemble. That is, the improvement of the π-stacking interactions between the Fmoc moieties is a key component in stabilizing these systems. However, fluorescence results cannot explain the secondary preference for amidated molecules, but the FT-IR shows a clear difference between amidated and methyl ester molecules,

which suggests that it is a more robust measure of the influence of the terminal group. Although the higher hydrophobicity of F is expected to influence its enhanced self-assembling tendency, the results show that in the C-terminus substitution the trends are opposite to the hydrophobicity increase. Hence, other intermolecular interactions play a significant role, where the improvement in the π-stacking is the main contributor governing the relative stability of the self-assembled structures. However, the hydrogen bonds are relevant as the extra amide hydrogen bonds compensate the small improvement that the methyl ester group contributes to the π-stacking interaction.

Computational results

Given the experimental information, which confirms the presence of H-bonded stacks, π-stacking interactions and chiral organization of fluorenyl moieties two models were developed. As the atomistic models were built to try to maximize the interactions suggested experimentally, the molecules which showed the best self-assembling tendency (Fmoc-TF-NH_2), and hence the most favourable interactions, was chosen for this process (Fig. 3a-b).

The models were built with an antiparallel arrangement, consistent with the model proposed by Smith *et al.* for the Fmoc-FF-OH system,[10] This was done because the parallel arrangements do not allow for the extended structures required to build fibres (Fig. S10a-b). The type of antiparallel arrangement (Fig. S10d) was selected to allow the formation of a hydrophobic core containing the F side chains and a hydrophilic surface, with the T side chains (Fig. S12a and S13a). Also, the hydrophobic core allows the F side chains to form interactions, which would help explain the observed F preference over L to self-assemble.

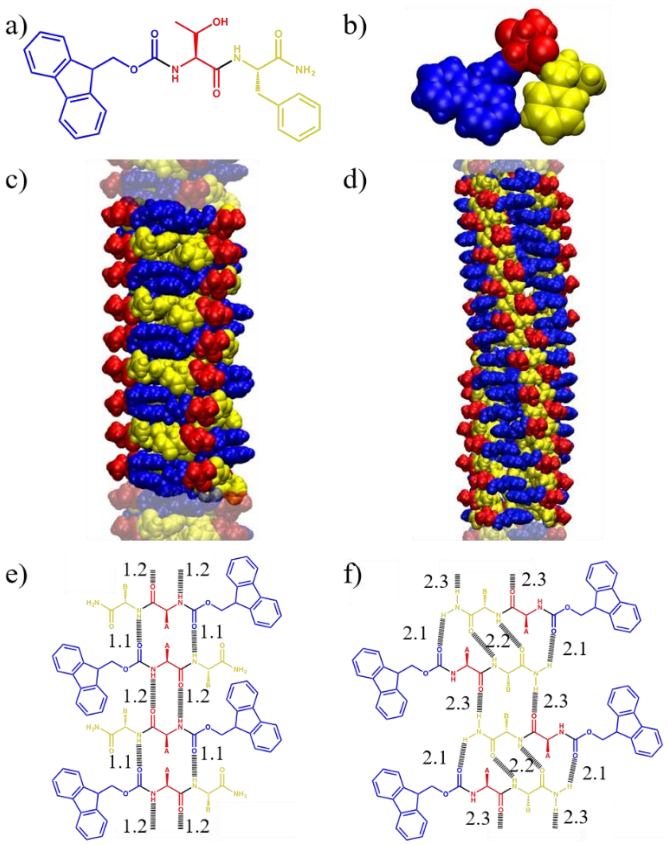

Fig. 3. Fmoc-TF-NH2 in (a) 2D and (b) 3D-vdw representations. (c, e) Fibre Model 1 and (d, f) Model 2: (c-d) side view and (e-f) scheme of the H-bonded conformation. H-bonds labelled: (e) 1.1 corresponds to Fmoc – F and 1.2 to T – T; (f) 2.1 corresponds to Fmoc – F, 2.2 to F – F, and 2.3 to T – F.

The two models (Fig. 3c-d) differ in the H-bonds involved (Fig. 3e-f), and it can be seen that these interactions impact on the ability to form extended stacks of fluorenyl groups: Model 1 shows the fluorenyl groups in dimers (Fig. 3c) while Model 2 forms extended fluorenyl stacks (Fig. 3d). Model 2 shows a significant twisting of the fluorenyl groups which correlates to the observed ellipticity in the CD signal for this system. Moreover, this twisting results from the H-bonding arrangements in this model and as such directly correlates the H-bonding network to the ellipticity, which was observed experimentally. In comparison with previous models, Model 2 is the most similar to the model presented in 2008 by Smith *et al.* except for the interlocked phenyl groups in the stacks, which in the current models are situated in the core of the fibre.

The two proposed fibre models were simulated for 150 ns in a water box surrounded by extra molecules (33 in Model 1 simulation and 60 in Model 2) to improve the stability of the fibres (Fig. S12 and S13) and to reflect the dynamic nature of the self-assembled system. The size of the fibres, and hence the number of molecules forming each fibre, is delimited to connect both edges through the periodic simulation box (The fibre in Model 1 is composed of 27 molecules and of 60 in Model 2). The final simulation systems are composed of 60 molecules for Model 1 and by 120 for Model 2. Further details of the methods employed and the construction of the systems are included in the Supporting Information.

The simulations' snapshots (Fig. 4a,c; and Figures S10 and S11) qualitatively show that after 150 ns of the simulation Model 2 maintains the overall fibre shape while Model 1 does not. An analysis of the specific backbone H-bonds between the different residues (Fmoc, T and F) was carried out to quantitatively assess the relative stabilities of the two models and their specific interactions (Fig. 4b,d). Both H-bonds graphs show that the number of interactions decreases substantially in the early stage of the simulations: in the first 25 ns the total number of H-bonds changes from ~0.25 H-bonds/molecule to ~0.1 and from ~0.45 H-bonds/molecule to ~0.2, for Model 1 and Model 2, respectively. This is due to the instability of the systems, which is in part due to the limited size of the systems. However, specific analysis of the H-bonds help to determine the relative stability of both models.

After construction of the models, a temperature is introduced involving gradual heating from 0 to 298K, in order to bring the system to experimental conditions. Of the main backbone H-bonds in Model 1, Fmoc – F (1.1 in Fig. 3) and T – T (1.2), only the former persist through the heating step (see Methods Section in the ESI) and they do not persist as main interactions after 30 ns of the simulation (Fig. 4b). After these 30 ns and for the rest of the simulation, the main backbone H-bonds that persist are Fmoc – F, F – F and T – F, which match those in Model 2 (2.1, 2.2 and 2.3, respectively). In other words, as the

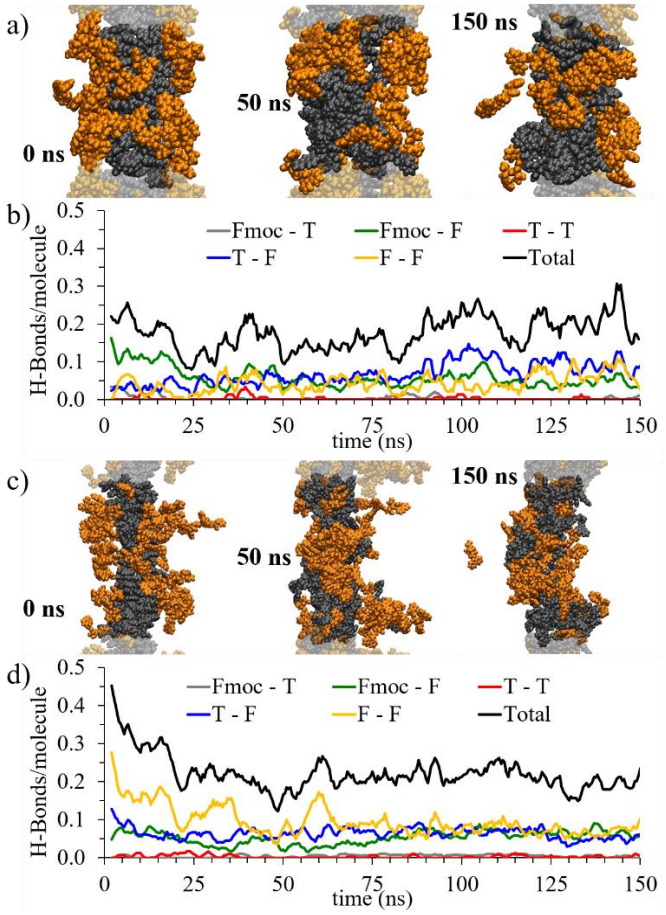

Fig. 4. (a-b) Model 1 and (c-d) Model 2 (a, c) simulation snapshots and (b, d) backbone H-bonds evolution with time between the different residues (Fmoc, T and F) and the sum of all (Total). Fibre model molecules are represented in grey and extra molecules in orange. Additional snapshots for the systems are provided in Figure S14 and S15 of the ESI.

simulation progresses Model 1 starts to resemble the H-bonding patterning initially created for Model 2 implying that model 2 is more stable. In the case of Model 2, the graph shows that the loss of total H-bonds is mainly due to the reduction of F – F H-bonds (2.2), suggesting that the model overestimates these interactions. This is in agreement with the low influence of the internal amide group on the self-assembling tendency observed previously in Fmoc-AA-OH,.[9f] Despite this observation, the main interactions in this model: Fmoc – F (2.1), F – F (2.2) and T – F (2.3); maintain the same relative importance for the duration of the simulation and with much smaller fluctuations than observed in Model 1. Furthermore, after 75 ns, these interactions have significantly reduced fluctuations, suggesting that an equilibrated structure has been reached. This does not happen at any point in the Model 1 simulation.

Therefore, taking into account that in the Model 1 simulation the H-bonds evolve to resemble those consistent with Model 2 H-bonds, and that in the Model 2 simulation the H-bonds show a higher stability, it can be concluded that Model 2 represents a more stable structure. Furthermore, this correlates with the experimental results that suggest an elliptical structure that is correlated with the H-bonding and with the influence of the amidated C-terminus in the H-bonding. Namely, only in Model 2 is the amidated C-terminus involved in forming additional H-bonds. The atomic detail provided by the model also provides clear insights into the structural role of the F side chains, the C-terminus and the positioning of chemical groups to ensure they are exposed at the surface of the nanostructures (Fig. 3d).

Although experimentally Fmoc-TF-NH_2 has been observed to form fibres, it has also been observed to form twisted ribbons (Fig. 5a-b) suggesting some lateral interactions between fibrils. Some cryo-TEM images suggest the formation of the ribbons by lateral aggregation of the fibres but it is not clear if the ribbons keep the fibre shapes or they evolve to a bilayer-like structure, as was suggested for previously studied Fmoc-dipeptide 2D

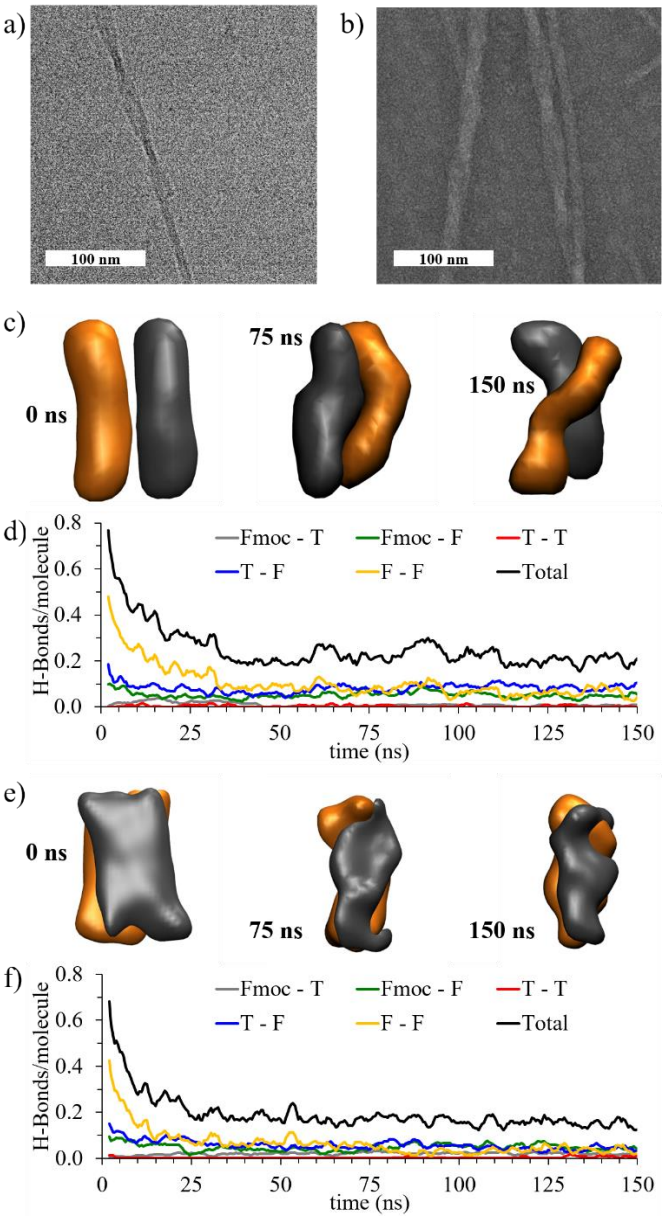

Fig. 5. Fmoc-TF-NH2 (a) Cryo-TEM and (b) TEM images. (c-d) *Two-fibres* and (e-f) *bilayer* (c, e) simulation snapshots and (d, f) backbone H-bonds evolution with time between the different residues (Fmoc, T and F) and the sum of all (Total), ignoring interfibre/layer interactions. Snapshots presented with a density surface representation for clarity with the fibres/layers in different colours. VdW snapshots in Figure S17.

nanostructures.[4c, 12] To study this, both systems were considered, *two-fibres* (Fig. 5c) and *bilayer* (Fig. 5e), and their stabilities compared. The *two-fibres* system uses two Model 2 fibres and the *bilayer* system uses the same type of stacks, with the H-bonds that have been demonstrated to produce the most stable structures, and maximizing the packing of the aromatic groups.

The snapshots of the simulation show that the *bilayer* system (Fig. 5e) does not keep its shape through the simulation and the final snapshot shows fibre type structures. However, in the final structure of the *two-fibres* (Fig. 5c, 150 ns) both fibres are still clearly visible, demonstrating that this conformation is more stable. Furthermore, both fibres become twisted around each other, forming a structure that resembles those observed by TEM (Fig. 5a-b).

Analysing the H-bonds, the higher stability of the *two-fibres* system, when compared to the one fibre system, is clear. While in the *two-fibres* system the predominant bonds are the same through the simulation (with the same reduction of importance of the F – F H-bonds), the preferred H-bonds in the *bilayer* system are not clear and an additional interaction (Fmoc – T) becomes as important as the starting three. This interaction also appears at the beginning of the *two-fibres* system but disappears again before 50 ns, after which point the simulation can be considered equilibrated.

Further analysis (Fig. 6) shows that the *two-fibres* system takes more time than the *bilayer* system to have H-bonds between the backbones (Fig. 6c, e) of different fibres/layers, further indicating that the starting structure for the *two-fibres* is more stable. For the *bilayer* system the interactions of the T-side chain with the other layer (Fig. 6f) indicates a lack of stability of the layers, as the T-side chains are initially orientated to face towards the opposite sides of the layer (Fig. 6b, red) and therefore a disruption of the structure is necessary to allow these interactions.

This analysis (Fig. 6) shows that the lateral aggregation of the two fibres is mainly due to the H-bonds of the T-side chain (Fig. 6a, red) with the Fmoc moiety (Fmoc – T(s)), probably with the ether oxygen of the carbamate, which is not involved in any H-bond in Model 2. In addition, the T-side chains form H-bonds with the F backbone and amidated C-terminus are involved in the lateral aggregation (F – T(s)). Unexpectedly, the interactions between the T-side chains of the different fibres (T(s) – T(s)) do not play an important role, probably due to the ellipticity of the fibres.

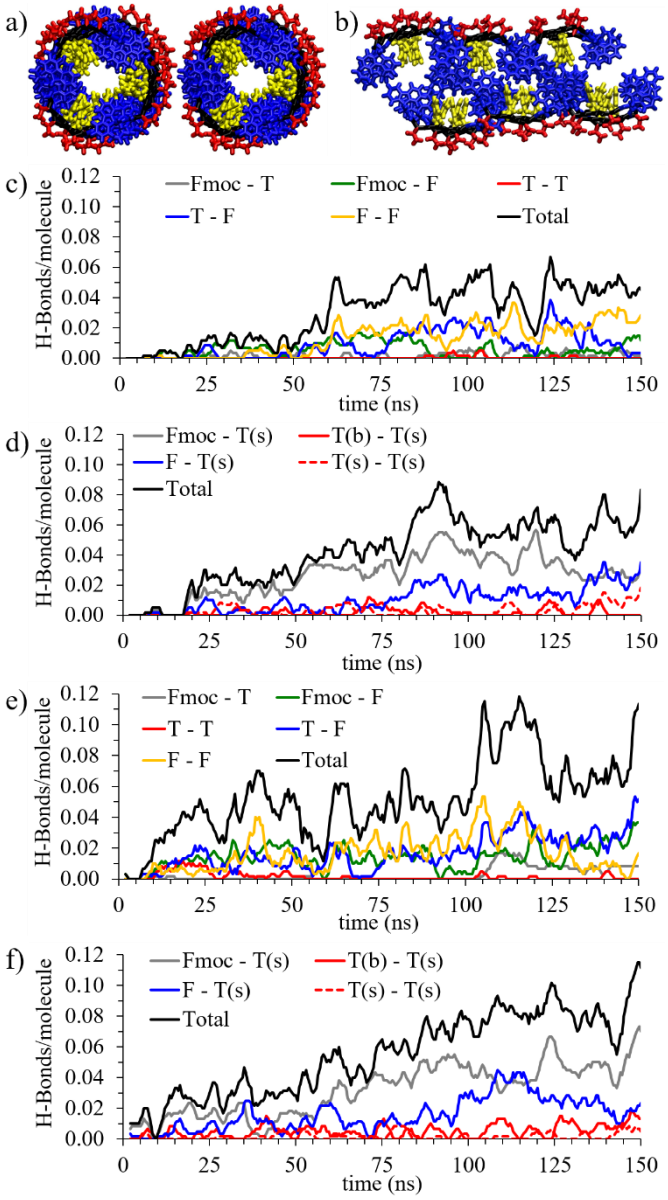

Fig. 6. (a, c-d) *Two-fibres* and (b, e-f) *bilayer* (a-b) structure top view and (c-f) interfibre/layer H-bonds evolution with time between the different residues (Fmoc, T and F) and the sum of them (Total). (a-b) Fmoc in blue, peptide backbone in black, T-side chain in red and F-side chain in yellow. Interfibre/layer H-bonds (c, e) between different residues backbones (Fmoc, T and F) and their sum (Total); and (d, f) involving the T-side chain (T(s)) with the different residues backbones and other T-side chains. T(b) in these graphs refers to H-bonds formed with the T-back bone.

Conclusions

In conclusion, the DPL experiments have shown the higher tendency of F and amidated C-terminus containing Fmoc-dipeptides to self-assemble into nanostructures. This was found for molecules with similar nanostructures and hence the resulting preferences offered the opportunity to provide unique information into the subtle effect of these substitutions on the intermolecular level arrangements. The F preference is due to an improvement in the π-stacking interactions (fluorescence) and the C-terminus preference, due to better H-bonding (FT-IR). The ellipticity (CD) is correlated with the H-bonded stacks through the dependence on the C-terminus substitution. This demonstrates that, by using DPLs on closely related systems, more structural detail can be obtained and the role of the different interactions can be better understood than by independent characterization of the different molecules.

Despite the lack of structural detail given by spectroscopy and electron microscopy experiments, different atomistic models can be built based on the interactions present and the observed preferences from the DPL experiments. The models can then be validated by assessing their relative stabilities using MD

simulations. The final model is consistent with the experimental results: a clear correlation between ellipticity and the formation of more H-bonds; a dependence on H-bonds involving the C-terminus; an improved self-assembling tendency for F relative to L, due to the core π-stacked arrangements; and the preference for a model with extended Fmoc – Fmoc π-stacking. Furthermore, the model demonstrated its validity by reproducing the formation of a twisted fibre from two straight fibres. The analysis of these results demonstrated that the formation of this type of superstructures is due to the lateral aggregation of fibres which is due to the higher stability of the fibres in the twisted ribbon conformation. The analysis also showed that this lateral aggregation involves interactions between the T-side chain and the carbamate group of Fmoc, which was not expected to be involved in this process.

Thus, the validated models and the analysis of their simulations provide detailed information of the structural role of the different chemical groups and moieties in the formation of nanostructures. In addition, the importance of the surface exposed groups of the nanostructure for the formation of higher levels of arrangement, twisted fibres (superstructures) has also been highlighted. This is a next step towards understanding the hierarchical self-assembly of peptide amphiphiles with atomistic resolution and the full understanding of the interactions involved, which is essential for the design of new peptide-based materials. It is likely that solid state nmr and high resolution cryo-TEM methodology will add further to elucidation of supramolecular nanostructures.

Acknowledgements

The authors gratefully acknowledge the financial support by the EC 7th Framework Programme Marie Curie Actions via the European ITN SMARTNET No. 316656 and EMERgE/ERC No. 258775. Results were obtained using the EPSRC funded ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk). EPSRC grant no. EP/K000586/1.

Notes and references

1. (a) T. Aida, E. Meijer and S. Stupp, *Science*, 2012, **335**, 813; (b) G. M. Whitesides and B. Grzybowski, *Science*, 2002, **295**, 2418; (c) A. R. Hirst, B. Escuder, J. F. Miravet and D. K. Smith, *Angew. Chem.*, 2008, **47**, 8002; (d) M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. McRee and N. Khazanovich, *Nature*, 1993, **366**, 324; (e) E. Gazit, *Chem. Soc. Rev.*, 2007, **36**, 1263; (f) J. Boekhoven and S. I. Stupp, *Adv. Mater*, 2014, **26**, 1642; (g) J. Naskar, G. Palui and A. Banerjee, *J. Phys. Chem. B*, 2009, **113**, 11787; (h) Y. Gao, F. Zhao, Q. Wang, Y. Zhang and B. Xu, *Chem. Soc. Rev.*, 2010, **39**, 3425; (i) J. D. Hartgerink, E. Benlash and S. L. Stupp, *Science*, 2001, **294**, 1684.
2. S. Fleming and R. V. Ulijn, *Chem. Soc. Rev.*, 2014, **43**, 8150.
3. K. Tao, A. Levin, L. Adler-Abramovich and E. Gazit, *Chem. Soc. Rev.*, 2016.
4. (a) Y. Zhang, H. Gu, Z. Yang and B. Xu, *J. Am. Chem. Soc.*, 2003, **125**, 13680; (b) A. Mahler, M. Reches, M. Rechter, S. Cohen and E. Gazit, *Adv. Mater*, 2006, **18**, 1365; (c) H. Shao and J. R. Parquette, *Chem. Commun.*, 2010, **46**, 4285; (d) R. Vegners, I. Shestakova, I. Kalvinsh, R. M. Ezzell and P. A. Janmey, *J. Pept. Sci.*, 1995, **1**, 371.
5. (a) M. Hughes, P. W. J. M. Frederix, J. Raeburn, L. S. Birchall, J. Sadownik, F. C. Coomer, I. H. Lin, E. J. Cussen, N. T. Hunt, T. Tuttle, S. J. Webb, D. J. Adams and R. V. Ulijn, *Soft Matter*, 2012, **8**, 5595; (b) D. M. Ryan, S. B. Anderson and B. L. Nilsson, *Soft Matter*, 2010, **6**, 3220.
6. (a) D. M. Ryan, T. M. Doran, S. B. Anderson and B. L. Nilsson, *Langmuir*, 2011, **27**, 4029; (b) J. Gao, H. Wang, L. Wang, J. Wang, D. Kong and Z. Yang, *J. Am. Chem. Soc.*, 2009, **131**, 11286.
7. (a) Z. Yang, H. Gu, D. Fu, P. Gao, J. K. Lam and B. Xu, *Adv. Mater*, 2004, **16**, 1440; (b) L. Chen, S. Revel, K. Morris, L. C. Serpell and D. J. Adams, *Langmuir*, 2010, **26**, 13466; (c) L. Chen, K. Morris, A. Laybourn, D. Elias, M. R. Hicks, A. Rodger, L. Serpell and D. J. Adams, *Langmuir*, 2010, **26**, 5232; (d) L. Chronopoulou, S. Sennato, F. Bordi, D. Giannella, A. Di Nitto, A. Barbetta, M. Dentini, A. R. Togna, G. I. Togna and S. Moschini, *Soft Matter*, 2014, **10**, 1944.
8. (a) D. J. Adams, K. Morris, L. Chen, L. C. Serpell, J. Bacsa and G. M. Day, *Soft Matter*, 2010, **6**, 4144; (b) K. A. Houton, K. L. Morris, L. Chen, M. Schmidtmann, J. T. A. Jones, L. C. Serpell, G. O. Lloyd and D. J. Adams, *Langmuir*, 2012, **28**, 9797.
9. (a) M. Caruso, E. Gatto, E. Placidi, G. Ballano, F. Formaggio, C. Toniolo, D. Zanuy, C. Alemán and M. Venanzi, *Soft Matter*, 2014, **10**, 2508; (b) H. X. Xu, A. K. Das, M. Horie, M. Shaik, A. M. Smith, Y. Luo, X. Lu, R. Collins, S. Y. Liem, A. Song, P. L. A. Popelier, M. L. Turner, P. Xiao, I. A. Kinloch and R. V. Ulijn, *Nanoscale*, 2010, **2**, 960; (c) V. Castelletto, C. Moulton, G. Cheng, I. Hamley, M. R. Hicks, A. Rodger, D. E. López-Pérez, G. Revilla-López and C. Alemán, *Soft Matter*, 2011, **7**, 11405; (d) D. E. Lopez-Perez, G. Revilla-Lopez, I. W. Hamley and C. Aleman, *Soft Matter*, 2013, **9**, 11021; (e) X. J. Mu, K. M. Eckes, M. M. Nguyen, L. J. Suggs and P. Y. Ren, *Biomacromolecules*, 2012, **13**, 3562; (f) K. M. Eckes, X. Mu, M. A. Ruehle, P. Ren and L. J. Suggs, *Langmuir*, 2014, **30**, 5287.
10. 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.
11. S. Fleming, P. W. J. M. Frederix, I. R. Sasselli, N. T. Hunt, R. V. Ulijn and T. Tuttle, *Langmuir*, 2013, **29**, 9510.
12. M. Hughes, H. X. Xu, P. W. J. M. Frederix, A. M. Smith, N. T. Hunt, T. Tuttle, I. A. Kinloch and R. V. Ulijn, *Soft Matter*, 2011, **7**, 10032.
13. (a) J. R. Moffat and D. K. Smith, *Chem. Commun.*, 2008, 2248; (b) P. Terech, N. M. Sangeetha and U. Maitra, *J. Phys. Chem. B*, 2006, **110**, 15224; (c) J. Raeburn and D. J. Adams, *Chem. Commun.*, 2015, **51**, 5170.
14. (a) N. A. Dudukovic and C. F. Zukoski, *Soft Matter*, 2014, **10**, 7849; (b) J. J. de Jong, L. N. Lucas, R. M. Kellogg, J. H. van Esch and B. L. Feringa, *Science*, 2004, **304**, 278; (c) S. Hecht, *Small*, 2005, **1**, 26.
15. (a) Z. Yang, G. Liang and B. Xu, *Acc. Chem. Res.*, 2008, **41**, 315; (b) R. J. Williams, A. M. Smith, R. Collins, N. Hodson, A. K. Das and R. V. Ulijn, *Nat. Nanotechnol.*, 2009, **4**, 19.
16. (a) A. K. Das, A. R. Hirst and R. V. Ulijn, *Faraday Discuss.*, 2009, **143**, 293; (b) N. Sreenivasachary and J.-M. Lehn, *Proc. Natl. Acad. Sci.*, 2005, **102**, 5938; (c) J. Li, P. Nowak and S. Otto, *J. Am. Chem. Soc.*, 2013, **135**, 9222; (d) J. M. A. Carnall, C. A. Waudby, A. M. Belenguer, M. C. A. Stuart, J. J.-P. Peyralans and S. Otto, *Science*, 2010, **327**, 1502; (e) J. M. Lehn, *Chem.-Eur. J.*, 1999, **5**, 2455.
17. S. H. White and W. C. Wimley, *Biochim. Biophys. Acta Rev. Biomemb.*, 1998, **1376**, 339.
18. (a) Z. Yang, G. Liang, M. Ma, Y. Gao and B. Xu, *J. Mater. Chem.*, 2007, **17**, 850; (b) A. Z. Cardoso, A. E. A. Alvarez, B. N. Cattoz, P. C. Griffiths, S. M. King, W. J. Frith and D. J. Adams, *Faraday Discuss.*, 2014, **166**, 101; (c) 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.

19. (a) M. Hughes, L. S. Birchall, K. Zuberi, L. A. Aitken, S. Debnath, N. Javid and R. V. Ulijn, *Soft Matter*, 2012, **8**, 11565; (b) C. Tang, R. V. Ulijn and A. Saiani, *Langmuir*, 2011, **27**, 14438.
20. (a) S. Fleming, S. Debnath, P. W. J. M. Frederix, T. Tuttle and R. V. Ulijn, *Chem. Commun.*, 2013, **49**, 10587; (b) C. Tang, R. V. Ulijn and A. Saiani, *Eur. Phys. J. E*, 2013, **36**, 1.
21. G. Gottarelli, S. Lena, S. Masiero, S. Pieraccini and G. P. Spada, *Chirality*, 2008, **20**, 471.
22. (a) H. K. Kang, D. E. Kang, B. H. Boo, S. J. Yoo, J. K. Lee and E. C. Lim, *J. Phys. Chem. A*, 2005, **109**, 6799; (b) F. L. Minn, J. P. Pinion and N. Filipescu, *J. Phys. Chem.*, 1971, **75**, 1794.