Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels

Pim W. J. M. Frederix,[†‡] Gary G. Scott,[†] Yousef M. Abul-Haija,[†] Daniela Kalafatovic,[†] Charalampos G. Pappas,[†] Nadeem Javid,[†] Neil T. Hunt,[‡] Rein V. Ulijn,[†∥*] and Tell Tuttle[†*]

[†]WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, UK; [‡]SUPA, Department of Physics, University of Strathclyde, 107 Rottenrow East, Glasgow, G4 0NG, UK; [∥]Advanced Science Research Center and Hunter College, City University of New York, 85 St Nicholas Terrace, New York, NY10031, USA. *rein.ulijn@asrc.cuny.edu, tell.tuttle@strath.ac.uk.

TOC Text. Peptides that self-assemble into nanostructures are of interest for cosmetics, food, biomedical and nanotechnological applications. Here, we demonstrate computational tools, which enable the peptide sequence space to be rapidly searched for supramolecular properties giving rise to the first unprotected tripeptide hydrogelators.

Abstract. Peptides that self-assemble into nanostructures are of tremendous interest for biological, medical, photonic and nanotechnological applications. The enormous sequence space that is available from 20 amino acids likely harbours many interesting candidates, but it is currently not possible to predict supramolecular behaviour from sequence alone. Here, we demonstrate computational tools to screen for the aqueous self-assembly propensity in all of the 8,000 possible tripeptides, and evaluate these by comparison with known examples. We applied filters to select for candidates that simultaneously optimize the apparently contradicting requirements of aggregation propensity and hydrophilicity, which resulted in a set of design rules for self-assembling sequences. A number of peptides were subsequently synthesized and characterised, including the first reported tripeptides that are able to form a hydrogel at neutral pH. These tools, which enable the peptide sequence space to be searched for supramolecular properties, enable minimalistic peptide nanotechnology to deliver on its promise.

Molecular self-assembly of oligopeptides into nanostructures holds much promise for a range of potential applications in biomedicine, food science, cosmetics and nanotechnology (see e.g. refs. 1,2). This class of materials is highly versatile because of the combinatorial complexity achieved by combining 20 amino acids into peptide building blocks with a wide range of chemical functionality. The use of very short peptides, pioneered by Gazit,[3] is especially attractive, enhancing opportunities for rational design combined with robustness, scalability and cost reduction. Two main challenges are currently limiting the expansion of this field. Most examples of short peptides (<5 amino acids) that have been discovered since diphenylalanine (FF[3] – in this work we employ the standard single letter amino acid code in the naming of peptides) contain only hydrophobic amino acids. This is no surprise as hydrophobic interactions dominate self-assembly in water, but it also limits their aqueous solubility and restricts potential applications. Secondly, in spite of two decades of intensive research since the first examples of short self-assembling peptides,[4,5] most cases have been either discovered by serendipity or by mapping onto known sequence design rules from biological systems.[6–8] We aim to address both issues by generating design rules that indicate a peptide's suitability for creating nanostructures under aqueous, pH neutral conditions. We stress that the approach makes no assumptions about the origins of self-assembly and therefore allows for unbiased discovery and selection.

Experimentally, a small set of tripeptides has been reported to assemble into nanostructures in (mainly) aqueous environments, e.g. CFF forms nanospheres, FFF forms fibrous and plate-like assemblies;[6,7] VFF, FFV and LFF form heterogeneous nanostructures;[9–11] micelle formation was discovered in VYV[12] and KFG,[8] which in the latter case could reversibly be converted to nanotubes by lowering the pH; disordered aggregates were found upon drying of a solution of DFN.[13] One common approach to alter the self-assembly properties of short peptides is protection of the terminal amine or acid groups, with acetyl,[14–18] *t*-butyloxycarbonyl[19,20] or large aromatic groups,[21,22,23,24] reducing charge repulsions and introducing π-stacking/hydrophobic contributions to favour self-assembly and gelation. Simple rules have been described for assembling peptides based on repeating sequences based on biological systems.[25,26] However,

gelators based on unprotected tripeptides are still elusive and only a small section of the available sequence space has been explored. Unprotected peptides are inherently biodegradable to natural metabolites and therefore of interest to the design of materials that interface with biological systems. In terms of applications, they complement protected peptides, with unprotected variants more likely to be acceptable for applications in food science, cosmetics and biomedicine.

A number of researchers have recently studied molecular self-assembly in a supramolecular materials' context using computational approaches.[27,28] In previous work, we have shown that the propensity of dipeptides (two amino acids) to aggregate can be predicted using coarse-grain (CG) Molecular Dynamics (MD).[29] Several other studies comprising short peptide fragments of biological relevance, such as NFGAIL[14,30] (a fragment of human islet amyloid polypeptide) or FF,[29,31] FFF[32] and KLVFFAE[33] (parts of amyloid β_{16-22}), have shown the usefulness of CG-MD for studying peptide self-assembly. However, in all of these cases, the focus was on systems that were experimentally known to self-assemble.

In the current work, we provide a design-oriented approach by screening all possible combinations of amino acids in tripeptides (20^3 = 8,000 different sequences) and subsequently identifying those with the best predicted properties for experimental investigation, thus demonstrating the development and experimental validation of a methodology to predict the self-assembly properties of unprotected tripeptides. On an atomistic level, Smadbeck *et al*. have recently developed a computational protocol for the discovery of self-assembling protected tripeptides from a library based on conservative mutations of template molecule Ac-IVD (128 peptides).[18] However, we have chosen to utilize the coarse-grain MARTINI force field, which has been extensively parameterized for amino acids (see Fig. 1a).[34–37] This force field provides a speed up compared to atomistic force fields by approximately 3 orders of magnitude, allowing access to a sufficient simulation size and length to study all possible tripeptides without bias towards a certain structure. The ranking of the output of molecular simulations according to descriptors such as propensity to aggregate and hydrophilicity allows the selection of a set of candidates to form nanostructures in water.

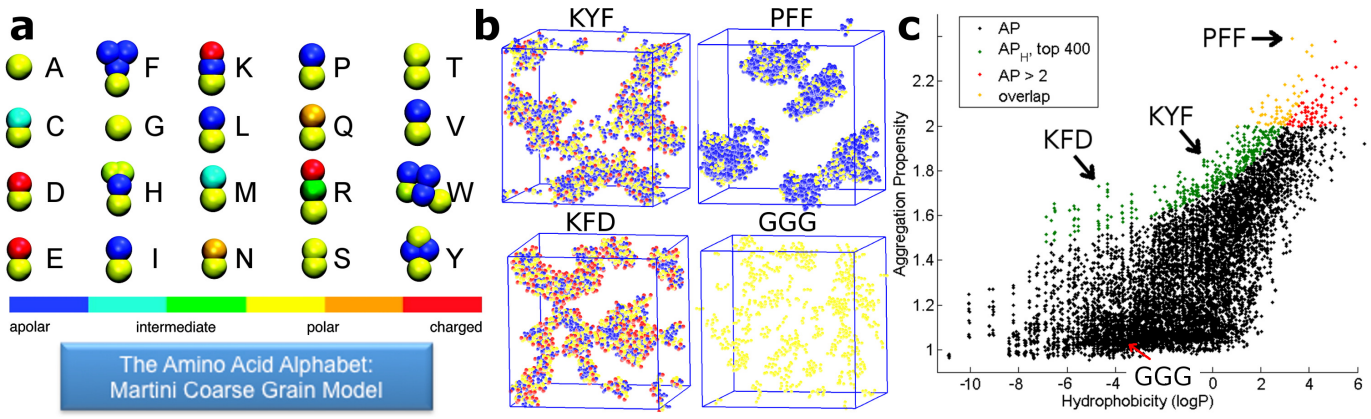


Figure 1. Screening for self-assembling tripeptides. a, Representation of all 20 gene-encoded amino acids in the MARTINI force field. Different colours represent different types of beads, as indicated by the legend. Image adapted with permission from ref. [34]. Copyright 2008 American Chemical Society. **b,** 50 ns MD simulation results of GGG, PFF, KFD and KYF tripeptides, showing various levels of aggregation. **c,** Aggregation propensity (AP) as a function of hydrophobicity for all 8,000 tripeptides. Red diamonds represent all tripeptides with AP > 2. Green diamonds represent the top 400 tripeptides from the AP_H score (AP_H = hydrophobicity-corrected AP and is proportional to the AP and the partitioning coefficient logP through equation 2 in the text) with the overlapping candidates shown in orange. The arrows point to the data points for GGG, PFF, KFD and KYF.

This paper is structured as follows: a simulation protocol for rapidly selecting candidates with self-assembly propensity is presented, which is subsequently adapted to favour candidates that include hydrophilic residues. The results of these simulations are analyzed in terms of aggregation propensity, structural features and general design rules and subsequently verified against the experimental results from literature. Finally, a number of new peptides with promising scores were selected from the simulation results, synthesized and analyzed with regards to their assembly behaviour by diffusion ordered NMR spectroscopy (DOSY), transmission electron microscopy (TEM), fourier transform infrared (FTIR) spectroscopy and dynamic light scattering (DLS).

Results and Discussion

Initial screening phase

50 ns simulations were performed for all 8,000 tripeptides studied. From the last frame of these simulations, the Aggregation propensity (AP, see *Experimental Section*) was measured. A list of high (AP > 2) scoring peptides can be found in the Supplementary Information and the results are displayed in Supplementary Figs. 2 and 3 (analogous to Fig. 2a). On average, hydrophobic tripeptides have higher AP scores and W, F, Y give rise to relatively high contribution to the AP. Remarkably, T and S also stand out as contributing strongly compared to other amino acids with similar hydrophobicity. However, a wide range of AP scores were observed with intermediately hydrophilic peptides as shown in Fig. 1c, which displays the AP scores as a function of the total hydrophobicity (logP) of the tripeptide. It is noted that only a weak correlation exists between the total hydrophilicity of the tripeptide and its propensity to aggregate: while hydrophobic peptides (logP > 3) always display a relatively high score and hydrophilic peptides (logP < -7) have a low tendency to aggregate, intermediately hydrophilic or amphiphilic peptides exhibit a wide range of AP scores from ~1 (no aggregation) to ~2.4 (strong aggregation). This confirms that aggregation is not a process that can be predicted solely on basis of a peptide's hydrophilicity and the simulations presented here are needed to distinguish between good and poor candidates for self-assembly.

Strongly hydrophobic peptides are often insoluble in water. However, our screening approach does not, *a priori*, exclude any peptide based on known practical solubility limitations. As such, hydrophobic peptides produce high AP scores. However, the simulations presented here are of insufficient size and detail to distinguish between the processes of aggregation, precipitation, crystallization or self-assembly. Therefore, in order to select a more appropriate subset of peptides for practical use we have developed a hydrophilicity-corrected score, AP_H, which introduces a positive bias towards hydrophilic peptides (for details see experimental section). The inclusion of hydrophilic residues has previously been shown to transform insoluble nanofibres to a hydrogel network.[38]

Fig. 2a shows the normalized AP_H score for all 8,000 tripeptides. The amino acids on the axes are ordered from hydrophobic to hydrophilic according to the Wimley–White scale.[39,40] It is clear that a different set of peptides is indicated as having self-assembly propensity combined with hydrophilicity when compared to Supplementary Fig. 2. Contributions from peptides containing charged (K, R, D, E) and hydrogen bonding amino acids (mainly T and S) stand out more using the AP_H score, which will be further discussed in the next section. The top 400 peptides selected by this protocol are indicated in Fig. 1c and were subjected to an extended screening phase.

Generation of design rules

The simulation results of all 8,000 tripeptides were analyzed by studying the contributions of specific amino acids to the AP_H score. This allows for the determination of design rules, i.e., placement of certain amino acids in a particular position within the peptide chain to promote self-assembly. Fig. 2c shows the average AP_H score of all peptides with a certain amino acid in position 1 (N-terminus), 2 or 3 (C-terminus). Several interesting observations were made: (I) Aromatic amino acids (F, Y and to a lesser extent W) are more favourable in position 2 and 3 compared to the N-terminal amino acid, (II) negatively charged amino acids (E and especially D) are strongly favoured in position 3 and (III) positively charged and hydrogen bond donating amino acids (K, R; S, T) promote self-assembly when located at the N-terminus. Also proline is favoured in position 1, which could be because of its unique conformational properties allowing better packing of the short peptides; the 'kink' in the backbone chain leading to more ordered self-assembly is possibly related to the observation by Marchesan *et al*. that changing the stereochemistry of the N-terminal amino acid from L to D improves gelation for FFV, VFF and LFF through steric effects.[9–11]

It is not straightforward finding a physical rationale for these observations, indicating the importance of an unbiased screening approach. We propose that positive and negative residues near the N- and C-terminus, respectively, create a 'doubly charged' head group, which could drive alignment of the molecules by repulsion of equal charges and strong intermolecular salt bridge formation. Moreover, rules I and II are congruent with several reports in literature: for protected tripeptides it has been hypothesized that two sequential aromatic residues can provide a strong tendency to aggregate because of the conformation of the backbone and aromatic interactions, while negative residues at the end of the peptide provide the amphiphilicity to provide an ordered self-assembled structure (see Table 1 and Supplementary Tables 4 and 5).[18] It is also interesting to compare these rules with those created for searching for aggregation-prone sequences in peptides and proteins by Dobson and co-workers.[41,42] Although their results are directly based on various experimentally aggregating protein fragments, there is a good correlation between their ranking of amino acids and the results found here, especially for the middle amino acids in the tripeptides. Moreover, as it has been shown that patterning of amino acids is very important in secondary structure and amyloid formation,[42,43] the aggregation potential of sequences of (three) amino acids measured here could be as important as that for single amino acids in identifying aggregation-prone regions in longer peptides, although care has to be taken with the charged N- and C-termini of the

tripeptides. The design rules found here were used to select a number of tripeptides from across the range of AP and AP_H scores to study experimentally (see Table 1 and below).

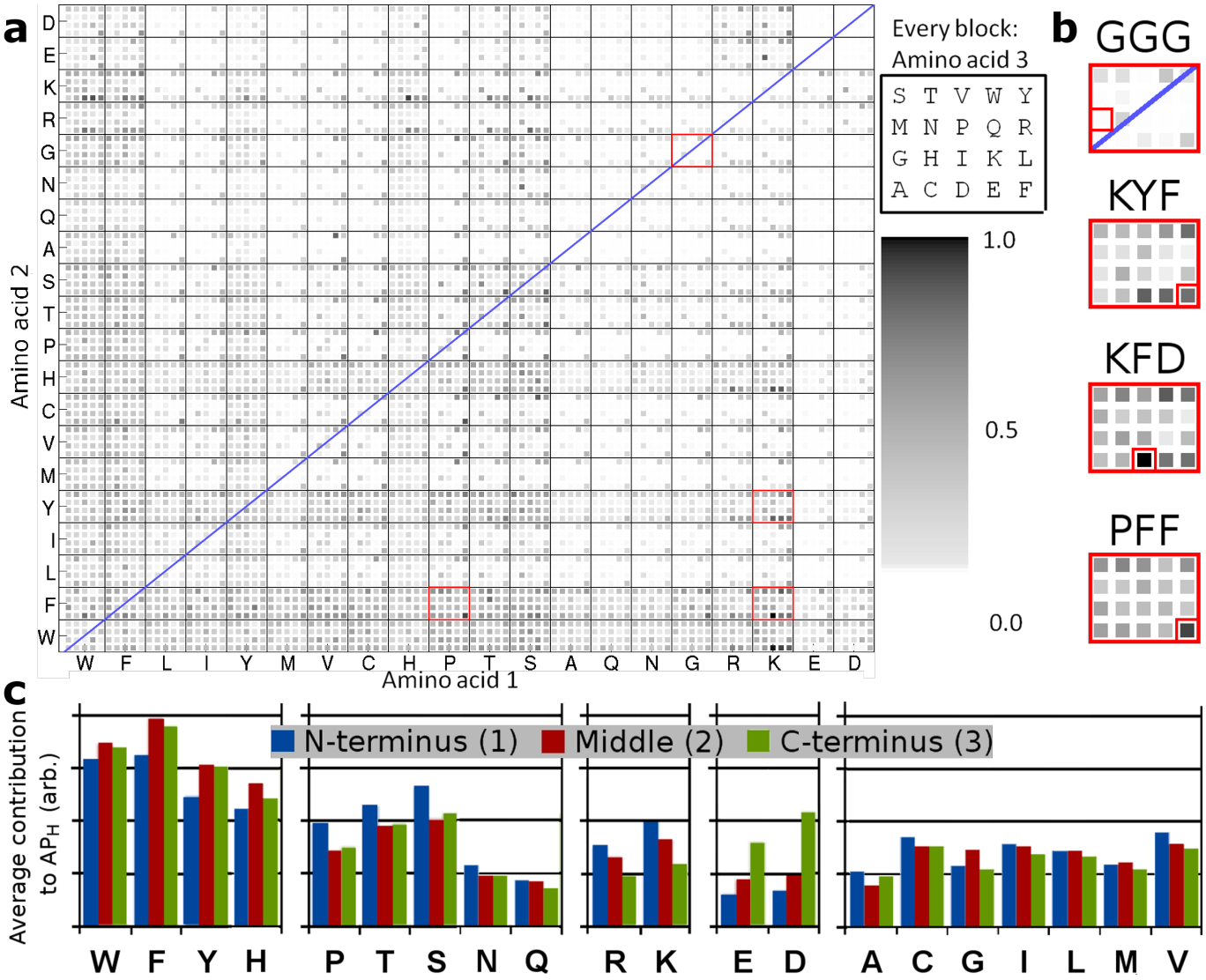


Figure 2. From screening to design rules. a, Normalized AP_H score for all 8000 combinations of three amino acids after a 50 ns simulation. Within every rectangle, the third amino acid is represented by the position of the colored square at the locations indicated in the legend on the right. A darker shade indicates a larger degree of aggregation. An equivalent graph with amino acid 2 and 3 on the x- and y-axes can be found in Supplementary Fig. 3. **b,** Expansion of the highlighted areas in **a** with four peptide entries indicated. **c,** Average AP_H scores of tripeptides with the specific amino acid on the x-axis in the N-terminal (blue), middle (red) and C-terminal position (green). A higher score indicates a higher propensity to aggregate. Amino acids are grouped by aromatic, hydrophilic, cationic, anionic and small/hydrophobic side chains.

Evaluation of simulation length, water model and force field

The emergence of structural features in peptide self-assembly takes place on timescales that are longer than the 50 ns used in the initial screening phase.[44] The simulation time was extended to 400 ns for peptides with either a high AP_H score (top 5% of 8,000) or with an AP score > 2 (124 peptides, 1.55%, including 53 peptides that fall in both categories), as highlighted in Fig. 1c. This cut-off has been previously shown to be reasonable for the AP score for self-assembly candidates in the study of all dipeptides, although a direct comparison with dipeptide AP scores is not possible because of differences in relative surface area.[29] Moreover, for these simulations the standard coarse-grain water was replaced by 'polarizable water' coarse-grain beads (PW), which have been shown to represent the polarizability and charge

screening of water more accurately.[45] Full lists of the AP and AP_H scores of the selected peptides can be found in the Supplementary Information.

Supplementary Table 1 shows the AP scores of the experimentally studied peptides at 50 ns in normal CG water and the score at 50 and 400 ns in PW. Changing the water model has a small influence on the reported AP scores after 50 ns: for strongly hydrophobic tripeptides the AP score generally increases slightly (average ~0.1), while for peptides containing hydrophilic and charged residues the AP scores similarly decrease, as can be expected when charge–charge interactions are more effectively screened. Although the AP_H scores would also be impacted, the use of PW significantly enhances the computational cost (by a factor of 2–5 on our system) and thus the limited additional detail provided was not deemed appropriate for our simulations. However, we recommend the use of PW for calculations where detailed structure and dynamics are important and as such PW was employed after the initial screening phase.

When comparing the values at 50 and 400 ns (both in PW) it becomes apparent that AP scores do increase slightly when extending the simulation. The 50 ns initial screening time was chosen in order to allow for rapid scanning of a large number of peptides. This screening time was not optimized which may result in some slow-nucleating peptides being missed during the initial screen, but the rapid reduction of the search space makes the problem of discovering new self-assembling peptide nanostructures more tractable. As no peptides were observed to change from 'not aggregating' to 'strongly aggregating' or *vice versa* and their relative scores remain broadly constant, the data suggests that 50 ns is a sufficient simulation length and standard CG water is appropriate for an initial screening phase for these systems.

It should be noted that when the MARTINI force field is used, the secondary structure of the peptides is constrained and, as a consequence, some structural processes such as interpeptide β-sheet formation are not captured accurately.[34,46] Although it has been shown that these dynamics can be improved using pseudodihedral potentials,[46] we did not want to bias our simulations towards a specific secondary structure. The results found here are therefore somewhat limited in terms of the reliability of the intermolecular architecture, as will be further discussed below. However, De Jong *et al*. have shown that the interaction strength between side chain beads and backbone beads closely match atomistic reference models[36,37] and Singh and Tieleman have shown that the partitioning coefficients for most amino acids are in good agreement with experimental values,[35] so we expect that the level of aggregation is well-represented in the simulations.

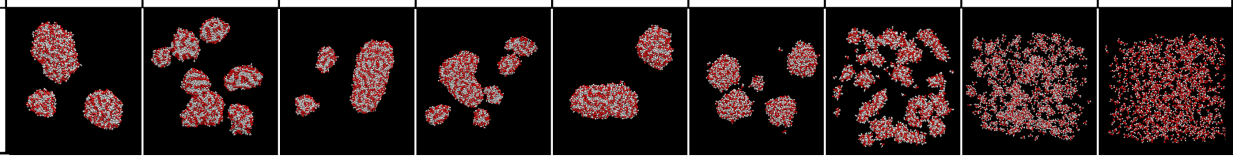
Evaluation of examples from the literature.

The top half of Table 1 shows all tripeptides that were examined for their assembling properties found in the literature. CFF, FFF, LFF, VFF and FFV were reported to form extended nanostructures under (mainly) aqueous conditions.[6,7,9,10] All these tripeptides were found to have an AP score > 2 after the 50 ns simulation, which supports that this is a reasonable cut-off for assembling peptides. Interestingly, Marchesan *et al*. noted that VFF (AP = 2.3) showed more evidence for nanostructure formation than structural isomer FFV (AP = 2.0),[9] which agrees with our observation that aromatics in the 2nd and 3rd position improve aggregation. KFG, reported to form vesicles and nanotubes has an AP of 1.6, but does in part obey our design rules with K at the N-terminus and F in the second position.

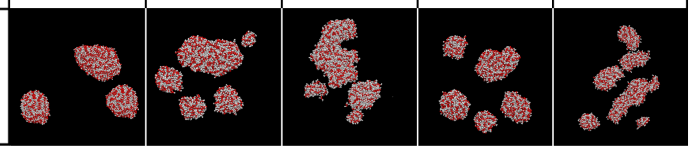
To study the formation of nanostructures in the simulations, the simulation time was extended further using a larger peptide box (1,200 ns, 1,200 peptides, 0.14 M in CG water) for the tripeptides studied experimentally in the literature. The last frames of the MD simulations are displayed in Table 1. The aspect ratio and rough shape of the nanostructures were determined from the moments of inertia along the principal axes of the largest cluster of peptides (see Supplementary Table 2). The shapes observed are in agreement with the experimentally observed nanostructures in Table 1: FFF (experiment: spheres, plates), CFF, VYV (experiment: spheres) and FFV and LFF (experiment: fibres) match the computational result. Longer simulations on FFF were reported to produce nanospheres and nanorods, similar to the structures observed here.[32] Tripeptides DFN and ECG, which were experimentally not found to exhibit assembly in solution,[13,47] were also calculated to have a low propensity to aggregate. It should be noted that the size of the aggregates in these simulations does not compare to the experimental size of the aggregates due to the limited number of molecules in the simulation. It has become apparent that while the proposed computational method is mainly valuable in determining a good set of candidates for self-assembling nanostructures after 50 ns, structural information can be obtained as well with much longer, extended, simulations as was reported previously.[29,31,32]

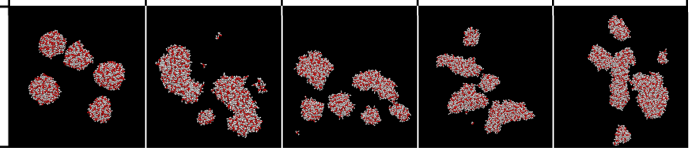
Table 1. Aggregation propensity and structural results from CG MD simulations. Top: comparison of MD results with tripeptides studied in the literature; bottom left: tripeptides experimentally studied in this work; and bottom right: a list of the top twenty peptides as ranked by their AP and AP_H scores. The diffusion constant D as determined by DOSY NMR spectroscopy is given as an experimental measure of size of the aggregates (smaller aggregates have a higher diffusion constant). The ranking on the AP score out of 8,000 tripeptides is given in brackets beside the AP value. 'N.S.' indicates the tripeptide was not water-soluble. The 'Structure' rows show the output of the 1200 ns CG MD simulation.

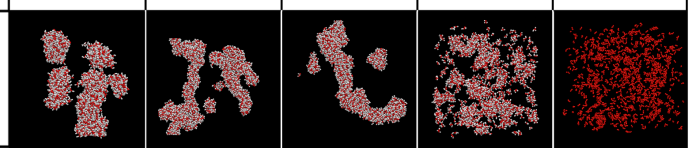
Literature

Peptide	VFF	FFF [a]	LFF	CFF [a]	FFV	VYV	KFG	DFN	ECG
Ref.	9	7,32	10	6	9	12	8	13	47
Assembly observed?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
AP (rank)	2.33 (4)	2.26 (7)	2.07 (58)	2.05 (69)	2.03 (89)	1.88 (377)	1.56 (737)	1.17 (4720)	1.01 (7590)
AP_H rank	159	2422	2390	703	1319	223	2001	4867	7660
Structure									

This work

Peptide	PFF	FYI	FHF	YFI	KFF
D (10^{-10} m^2/s)	N.S.	N.S.	N.S.	3.8	3.8
AP 50 ns (rank)	2.39 (1)	2.22 (10)	2.09 (44)	1.97 (190)	1.92 (264)
AP_H rank	4	107	301	1041	30
Structure					

Peptide	IYF	KYF	RYF	KYY	KYW
D (10^{-10} m^2/s)	N.S	3.7	3.8	3.7	3.7
AP 50 ns (rank)	1.89 (347)	1.85 (478)	1.80 (648)	1.79 (662)	1.78 (728)
AP_H rank	1572	28	212	23	146
Structure					

Peptide	FYK	KFD	KHD	KLL	GGG
D (10^{-10} m^2/s)	3.9	4.0	4.5	3.9	6.3
AP 50 ns (rank)	1.73 (941)	1.73 (959)	1.63 (1480)	1.31 (3751)	1.07 (6278)
AP_H rank	207	1	7	3918	6095
Structure					

Top scoring peptides

#	AP score	AP_H score
1	PFF	KFD
2	WFL	KWD
3	MFF	HKD
4	VFF	PFF
5	FFM	KWE
6	FWF	WKE
7	FFF	KHD
8	WWF	PCF
9	FWI	KWF
10	FYI	KFW
11	VFW	KHE
12	PWF	TSF
13	IFF	SCW
14	LCF	WKD
15	WLL	KYD
16	SFW	KEH
17	IFW	GFF
18	WFF	VAW
19	FFW	SSF
20	IMW	KYE

[a]: in the presence of hexafluoroisopropanol solvent.

Predictive value of the model

The computational method was validated in the previous section by comparison with experimental work from the literature. However, the added value of the proposed procedure lies predominantly in identifying new self-assembling peptides. We tested its predictive value by synthesizing and characterizing various tripeptides indicated as good candidates for nanostructure formation.

The bottom half of Table 1 contains the tripeptides that were selected for synthesis and experimental characterization. The rationale behind selecting certain tripeptides was that they were representative examples of peptides ranking highly on either table (PFF, FYI, FHF, KFD, KHD) or followed the design rules set out (PFF, KYF, KFD, KHD). Because the KYF peptide formed a hydrogel, similar peptides were then chosen to check if they followed the same trends (KYW, KFF, KYY, RYF). The other peptides were chosen to provide non-assembling and scrambled sequence comparisons. These peptides were dissolved in water at pH 7 and their aggregation behaviour was studied in order to test the predictive value of our method. For KYF, KYY, KFF and KYW this afforded the formation of translucent self-supporting hydrogels, while PFF, FHF and RYF gave suspensions, IYF and FYI precipitated and the other tripeptides gave clear solutions.

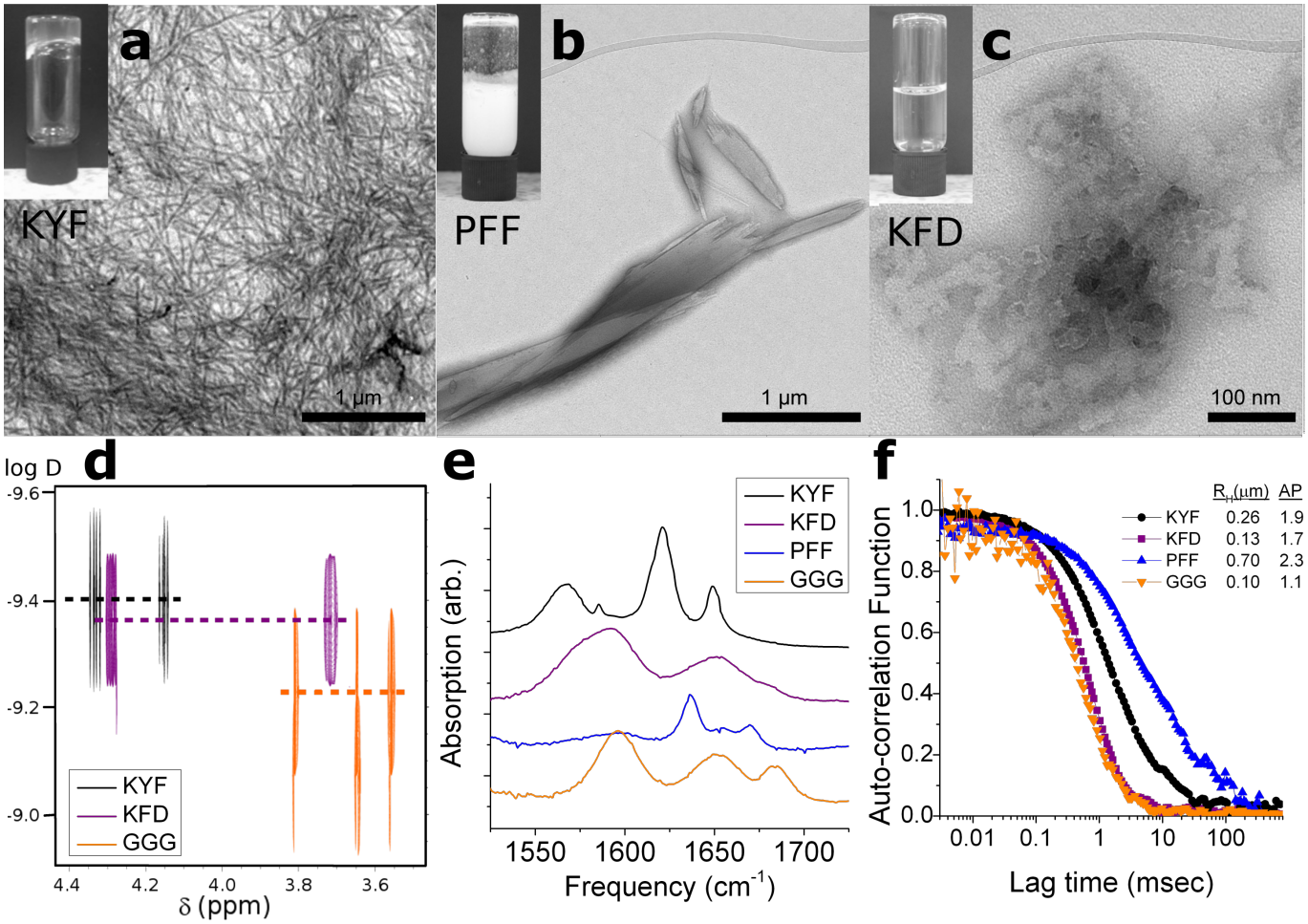


Figure 3. The characterization of selected tripeptides. a-c, TEM images of KYF, KFD and PFF tripeptides. Insets show photos of the gel, solution and suspension, respectively. **d,** α-proton region of the DOSY spectra for KYF, KFD and GGG tripeptides at 10 mM at pH 7. Horizontal lines indicate the value for the logarithm of the diffusion constant D and show slow diffusion in water for KYF and KFD (i.e. larger aggregates) compared to reference peptide GGG. Note that PFF is not included because of its low solubility. **e,** FTIR absorption spectra in the amide I region of tripeptides (30 mM in D_2O at pH 7): KYF, KFD, PFF and GGG. Narrowing and red-shifting of the amide modes of KYF and PFF indicate the presence of well-ordered aggregates for these peptides. Spectra have been vertically offset for clarity. **f,** DLS auto-correlation functions for KYF, KFD, PFF and GGG (10 mM at pH 7). The inset shows the hydrodynamic radii (R_H in μm) and AP scores at 50 ns for the four peptides, which indicate the relative sizes of the peptide aggregates.

To correlate the AP scores calculated with experimentally observed aggregation, the presence and size of aggregates were determined by diffusion ordered NMR spectroscopy (DOSY). This method is able to determine the diffusion coefficient *D* of supramolecular aggregates and therefore their relative size, as described by the Stokes–Einstein equation.[48,49] Diffusion constants for all the studied tripeptides can be found in Table 1. Fig. 3d displays the α-proton region of the DOSY spectra of KYF, KFD and GGG and full DOSY spectra of all peptides can be found in Supplementary Figs. 4-14. The poor solubility of PFF, FHF, IYF and FYI prevented reliable extraction of aggregate size for these peptides. For the remaining peptides, DOSY results are in good agreement with the predicted aggregation propensity at 50 ns. As expected, the negative control GGG has the highest diffusion constant. KFD has a relatively low AP and shows an intermediate diffusion rate. KYF has the lowest diffusion constant, consistent with its tendency to form an aggregated

hydrogel structure. It can be seen from Fig. 1c and Table 1 that KYF ranks high on the AP_H score (#28 out of 8,000), again demonstrating the value of taking the hydrophilicity of peptides into account (KYF only ranks #478 in the AP score). Note that peptide RYF (#212) did not give a gel (see Supporting Information) and SYF (#380) was insoluble in water (data not shown).

Further studies were employed to elucidate the nanostructural components of KYF (gel), KFD (#1 from AP_H score), PFF (#1 from AP score) and GGG (negative control) specifically, using FTIR spectroscopy, DLS and TEM.

The conformation of the peptide backbones can be probed by infrared (IR) spectroscopy.[50,51] When aggregation takes place via intermolecular hydrogen bonding of the amide groups, the 1650-1655 cm^{-1} amide I absorption observed for free peptides in solution typically narrows and shifts to lower frequency, while the 1595 cm^{-1} peak, assigned to carboxylate groups, broadens or decreases in intensity because of protonation or salt bridge formation. This was clearly observed for PFF in Fig. 3e, as the amide I absorption shifts to 1637 cm^{-1} and significantly narrows compared to non-aggregating peptide GGG, consistent with a β-sheet like arrangement of the amide groups. KFD shows no significant narrowing of the amide absorption compared to GGG, indicating no extended H-bonding network formation. Note that small absorption maxima in the 1670-1685 cm^{-1} region of GGG are assigned to residual trifluoroacetic acid absorption and the 1570-1580 cm^{-1} band for KFD is assigned to the aspartic acid side chain carboxylate group.

For KYF (and the other gelating peptides, see Supplementary Figure 15), a dramatic change in the amide region was noticed upon gelation with intense IR absorption peaks at 1621 and 1649 cm^{-1} for the amide groups and 1568 cm^{-1} for salt-bridged carboxylate groups. This indicates strong intramolecular hydrogen bonding between amide modes and strong interactions of the N-terminus or lysine side chains with the C-terminus, both suggesting a well-ordered peptide nanostructure.[50,51] FTIR spectra and analysis for other peptides can be found in Supplementary Fig. 15.

To further confirm the presence (and size) of the peptide aggregates formed, dynamic light scattering (DLS) experiments were performed on the KYF gel, PFF suspension and KFD and GGG solutions. Fig. 3f shows the intensity auto-correlation function and the average hydrodynamic radius of the aggregates determined from the extracted diffusion constant. The relative size of the aggregates (0.26, 0.13, 0.70 and 0.10 μm for KYF, KFD, PFF and GGG, respectively) correlates reasonably well with the relative AP scores for these peptides as predicted by the MD simulations (Table 1), although it was surprising to see such large aggregates for GGG. DLS results for other peptides can be found in Supplementary Fig. 26.

For the peptides synthesized here, TEM studies reveal an entangled fibrous network for the KYF hydrogel (Fig. 3a). Although KYF was not observed to form fibres after 1200 ns in the MD simulation, further extension of the simulation to 4800 ns or an increase in the peptide concentration resulted in an elongated, fibrous nanostructure (see Supplementary Fig. 27). This simulated fibre proved stable through hydrogen bonding, salt bridges and hydrophobic interactions when the resulting structure was backmapped to an atomistic structure and equilibrated (see Supplementary Information, Supplementary Figs. 28 and 29 and Supplementary Table 3). TEM images of PFF (Fig. 3b) reveal short crystalline nanostructures with a large aspect ratio, while KFD (Fig. 3c) was observed to form strongly curved fibres together with more amorphous regions. Additional TEM images can be found in the Supplementary Information (Supplementary Figs. 16-25). These results confirm the presence of nanostructures for these peptides as predicted for all assembling peptides.

Conclusions

In this paper we have presented a coarse-grain MD protocol for screening peptides for their aggregation behaviour and applied this to the set of 8,000 gene-encoded tripeptides. After an initial 50 ns screening phase a subset of peptides was selected based on their aggregation propensity (AP) and hydrophilicity-corrected AP. This set was then used in extended simulations to study the tripeptide dynamics and nanostructure formation.

The simulation results indicate only a weak correlation between hydrophilicity and AP, confirming that self-assembly propensity is not simply a measure of hydrophobicity and clearly illustrating the usefulness of MD simulations for selection of self-assembling peptides. Furthermore, a set of design rules that promote aggregation was described, where aromatic amino acids are most favorable in positions 2 and 3 in a tripeptide, while positive and H-bonding residues favour position 1 (N-terminus) and negative residues position 3 (C-terminus).

The results of the simulations were validated by comparison with experimental results from literature and by synthesis and characterization of a set of tripeptides which are indicated by our method and design rules to be promising candidates for self-assembly. Interestingly, this led to the discovery of the first unprotected full-L tripeptides (KYF, KYY, KFF and KYW) that form a hydrogel in the absence of organic solvents. More generally, good agreement was observed between predicted AP scores and experimental behavior. These results warrant further exploitation of the protocol for screening larger peptides in the search for new nanostructures.

Finally, in this current work we have presented a selection method based on the hydrophilicity of the tripeptides. However, it is clear that depending on the desired properties of the self-assembling peptide new filters can be developed. Automated screening based on the shape of nanostructure,[52] peptide foldability,[53] the presence of certain

amino acid interactions or other molecular properties could be equally useful and constitute future directions for this research.

Methods

Experimental procedures for CG MD, shape determination, backmapping, peptide synthesis, HPLC, DOSY NMR spectroscopy, FTIR spectroscopy, DLS and TEM can be found in the Supplementary Information.

Scoring method

Aggregation propensities (AP) for every tripeptide were calculated as the ratio of solvent accessible surface areas (SASAs) at the start and finish of an MD run in VMD[54] according to equation 1, analogous to previous work.[29]

$$AP = \frac{SASA_{initial}}{SASA_{final}} \quad (1)$$

The newly proposed hydrophilicity-corrected scores (AP_H) were calculated from AP and logP scores using the empirically founded equation 2:

$$AP_H = (AP')^{\alpha} \cdot logP' \quad (2)$$

where the apostrophe indicates the normalization of the respective variable between 0 and 1. As *logP(trip)* is a unitless number linearly proportional to $\Delta G_{water\text{-}oct}$ and is normalized in Eq. 1, it was chosen to define it simply as the sum of the Wimley-White whole-residue hydrophobicities[39,40] $\Delta G_{water\text{-}oct}$ (kcal/mol) for the tripeptide, given by:

$$logP = \sum_{i=1}^{3} \Delta G_{water-octanol} \quad (3)$$

In equation 2, α is an arbitrary coefficient that can be used to determine the weight of the normalized AP score to the AP_H value. For this study, α = 2 was used to give a good compromise between AP and hydrophilicity. Depending on the desired properties, the exponent can be decreased (increased) to include more (less) hydrophilic peptide. For a selection with α = 1 and α = 3 see Supplementary Information Fig. 1.

Computational power allowed only the top 400 scoring peptides indicated after 50 ns by this protocol to be used in extended simulations.

Preparation of tripeptide samples

Tripeptides were dissolved in water by sonication and vortexing, followed by bringing the pH to 7.2 ± 0.1 by dropwise addition of 0.25 M NaOH solution to a final concentration of 30 mmol/L. These were diluted to 10 mmol/L for DLS and NMR experiments. Note that the KYF gel could also be prepared by direct dissolution into a 0.1 M phosphate buffer at pH 7. Characterization of the resulting samples took places at least 24 hours after sample preparation. Samples for FTIR and NMR analysis were prepared using D_2O (99.9% D) and NaOD solution instead of H_2O and NaOH.

Acknowledgements

We would like to acknowledge C. Irving for assistance with DOSY NMR spectroscopy and M. Mullin (Glasgow University) for help with TEM. P.W.J.M.F., N.T.H. and R.V.U. acknowledge the financial support of the European Research Council (No. 334949: SPRITES-H2). R.V.U. acknowledges funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / EMERgE/ ERC grant agreement No. 258775. G.G.S. acknowledges the financial support by Macphie of Glenbervie. C.G.P. acknowledges the financial support by Linn Products. Y.M.A. acknowledges the financial support by FP7 Marie Curie Actions of the European Commission, via the initial training network ReAd (No. 289723). Results were obtained using the EPSRC funded ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk), EPSRC grant no. EP/K000586/1.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.V.U. and T.T.

Author Contributions

P.W.J.M.F. was responsible for computational work and IR spectroscopy. Y.M.A., D.K., C.G.P. and G.G.S. performed peptide synthesis and characterization. N.J., Y.M.A. and G.G.S. performed TEM. C.G.P. performed DOSY NMR spectroscopy. N.J.

and D.K. performed DLS. N.T.H., R.V.U. and T.T. contributed to experimental design. All authors commented on the manuscript, P.W.J.M.F., R.V.U. and T.T. wrote the paper.

Competing financial interests

The University of Strathclyde has filed a patent application on technology related to the processes described in this article. Several authors are listed as inventors on the patent application.

REFERENCES

1. Zhao, X. *et al*. Molecular self-assembly and applications of designer peptide amphiphiles. *Chem. Soc. Rev*. **39,** 3480–3498 (2010).
2. Zelzer, M. & Ulijn, R. V. Next-generation peptide nanomaterials: molecular networks, interfaces and supramolecular functionality. *Chem. Soc. Rev*. **39,** 3351–3357 (2010).
3. Reches, M. & Gazit, E. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science* **300,** 625–627 (2003).
4. Ghadiri, M. R., Granja, J. R., Milligan, R. A., McRee, D. E. & Khazanovich, N. Self-assembling organic nanotubes based on a cyclic peptide architecture. *Nature* **366,** 324–327 (1993).
5. Zhang, S., Holmes, T., Lockshin, C. & Rich, A. Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc. Natl. Acad. Sci. USA* **90,** 3334 –3338 (1993).
6. Reches, M. & Gazit, E. Formation of Closed-Cage Nanostructures by Self-Assembly of Aromatic Dipeptides. *Nano Lett*. **4,** 581–585 (2004).
7. Tamamis, P. *et al*. Self-Assembly of Phenylalanine Oligopeptides: Insights from Experiments and Simulations. *Biophys. J*. **96,** 5020–5029 (2009).
8. Moitra, P., Kumar, K., Kondaiah, P. & Bhattacharya, S. Efficacious Anticancer Drug Delivery Mediated by a pH-Sensitive Self-Assembly of a Conserved Tripeptide Derived from Tyrosine Kinase NGF Receptor. *Angew. Chem. Int. Ed*. **53,** 1113–1117 (2014).
9. Marchesan, S., Easton, C. D., Kushkaki, F., Waddington, L. & Hartley, P. G. Tripeptide self-assembled hydrogels: unexpected twists of chirality. *Chem. Commun*. **48,** 2195–2197 (2012).
10. Marchesan, S. *et al*. Unzipping the role of chirality in nanoscale self-assembly of tripeptide hydrogels. *Nanoscale* **4,** 6752–6760 (2012).
11. Marchesan, S. *et al*. Chirality effects at each amino acid position on tripeptide self-assembly into hydrogel biomaterials. *Nanoscale* **6,** 5172–5180 (2014).
12. James, J. & Mandal, A. B. The aggregation of Tyr-Phe dipeptide and Val-Tyr-Val tripeptide in aqueous solution and in the presence of SDS and PEO–PPO–PEO triblock copolymer: Fluorescence spectroscopic studies. *J. Colloid Interf. Sci*. **360,** 600–605 (2011).
13. Reches, M., Porat, Y. & Gazit, E. Amyloid Fibril Formation by Pentapeptide and Tetrapeptide Fragments of Human Calcitonin. *J. Biol. Chem*. **277,** 35475–35480 (2002).
14. Hauser, C. A. E. *et al*. Natural tri- to hexapeptides self-assemble in water to amyloid beta-type fiber aggregates by unexpected alpha-helical intermediate structures. *Proc. Natl. Acad. Sci. USA* **108,** 1361–1366 (2011).
15. Lakshmanan, A. & Hauser, C. A. E. Ultrasmall peptides self-assemble into diverse nanostructures: morphological evaluation and potential implications. *Int. J. Mol. Sci*. **12,** 5736–5746 (2011).
16. Lakshmanan, A. *et al*. Aliphatic peptides show similar self-assembly to amyloid core sequences, challenging the importance of aromatic interactions in amyloidosis. *Proc. Natl. Acad. Sci. USA* **110,** 519–524 (2013).
17. Cao, M., Cao, C., Zhang, L., Xia, D. & Xu, H. Tuning of peptide assembly through force balance adjustment. *J. Colloid Interf. Sci*. **407,** 287–295 (2013).

18. Smadbeck, J. *et al*. De Novo Design and Experimental Characterization of Ultrashort Self-Associating Peptides. *PLoS Comput Biol* **10,** e1003718 (2014).
19. Das, A. K., Bose, P. P., Drew, M. G. B. & Banerjee, A. The role of protecting groups in the formation of organogels through a nano-fibrillar network formed by self-assembling terminally protected tripeptides. *Tetrahedron* **63,** 7432–7442 (2007).
20. Subbalakshmi, C., Manorama, S. V. & Nagaraj, R. Self-assembly of short peptides composed of only aliphatic amino acids and a combination of aromatic and aliphatic amino acids. *J. Pept. Sci.* **18,** 283–292 (2012).
21. Smith, A. M. & Ulijn, R. V. Designing peptide based nanomaterials. *Chem. Soc. Rev.* **37,** 664–675 (2008).
22. Zhang, Y., Gu, H., Yang, Z. & Xu, B. Supramolecular Hydrogels Respond to Ligand–Receptor Interaction. *J. Am. Chem. Soc.* **125,** 13680–13681 (2003).
23. Yang, Z., Liang, G., Ma, M., Gao, Y. & Xu, B. Conjugates of naphthalene and dipeptides produce molecular hydrogelators with high efficiency of hydrogelation and superhelical nanofibers. *J. Mater. Chem.* **17,** 850–854 (2007).
24. Chen, L., Revel, S., Morris, K., C. Serpell, L. & Adams, D. J. Effect of Molecular Structure on the Properties of Naphthalene–Dipeptide Hydrogelators. *Langmuir* **26,** 13466–13471 (2010).
25. DeGrado, W. F. & Lear, J. D. Induction of peptide conformation at apolar water interfaces. 1. A study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **107,** 7684–7689 (1985).
26. DeGrado, W. F. Design of Peptides and Proteins. *Adv. Protein Chem.* **39,** 51–124 (1988).
27. McCullagh, M., Prytkova, T., Tonzani, S., Winter, N. D. & Schatz, G. C. Modeling Self-Assembly Processes Driven by Nonbonded Interactions in Soft Materials†. *J. Phys. Chem. B* **112,** 10388–10398 (2008).
28. Lee, O.-S., Cho, V. & Schatz, G. C. Modeling the Self-Assembly of Peptide Amphiphiles into Fibers Using Coarse-Grained Molecular Dynamics. *Nano Lett.* **12,** 4907–4913 (2012).
29. Frederix, P. W. J. M., Ulijn, R. V., Hunt, N. T. & Tuttle, T. Virtual Screening for Dipeptide Aggregation: Toward Predictive Tools for Peptide Self-Assembly. *J. Phys. Chem. Lett.* **2,** 2380–2384 (2011).
30. Wu, C., Lei, H. & Duan, Y. Formation of Partially Ordered Oligomers of Amyloidogenic Hexapeptide (NFGAIL) in Aqueous Solution Observed in Molecular Dynamics Simulations. *Biophys. J.* **87,** 3000–3009 (2004).
31. Guo, C., Luo, Y., Zhou, R. & Wei, G. Probing the Self-Assembly Mechanism of Diphenylalanine-Based Peptide Nanovesicles and Nanotubes. *ACS Nano* **6,** 3907–3918 (2012).
32. Guo, C., Luo, Y., Zhou, R. & Wei, G. Triphenylalanine peptides self-assemble into nanospheres and nanorods that are different from the nanovesicles and nanotubes formed by diphenylalanine peptides. *Nanoscale* **6,** 2800–2811 (2014).
33. Thirumalai, D., Klimov, D. & Dima, R. Emerging ideas on the molecular basis of protein and peptide aggregation. *Curr. Opin. Struct. Biol.* **13,** 146–159 (2003).
34. Monticelli, L. *et al*. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **4,** 819–834 (2008).
35. Singh, G. & Tieleman, D. P. Using the Wimley-White Hydrophobicity Scale as a Direct Quantitative Test of Force Fields: The MARTINI Coarse-Grained Model. *J. Chem. Theory Comput.* **7,** 2316–2324 (2011).
36. De Jong, D. H., Periole, X. & Marrink, S. J. Dimerization of Amino Acid Side Chains: Lessons from the Comparison of Different Force Fields. *J. Chem. Theory Comput.* **8,** 1003–1014 (2012).
37. De Jong, D. H. *et al*. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **9,** 687–697 (2013).
38. Zaccai, N. R. *et al*. A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.* **7,** 935–941 (2011).
39. White, S. H. & Wimley, W. C. Hydrophobic interactions of peptides with membrane interfaces. *Biochim. Biophys. Acta BBA - Rev. Biomembr.* **1376,** 339–352 (1998).

40. Wimley, W. C., Creamer, T. P. & White, S. H. Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host–Guest Pentapeptides†. *Biochemistry (Mosc.)* **35,** 5109–5124 (1996).
41. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide andprotein aggregation rates. *Nature* **424,** 805–808 (2003).
42. Pawar, A. P. *et al*. Prediction of 'Aggregation-prone' and 'Aggregation-susceptible' Regions in Proteins Associated with Neurodegenerative Diseases. *J. Mol. Biol.* **350,** 379–392 (2005).
43. West, M. W. *et al*. De novo amyloid proteins from designed combinatorial libraries. *Proc. Natl. Acad. Sci.* **96,** 11211–11216 (1999).
44. Ash, W. L., Zlomislic, M. R., Oloo, E. O. & Tieleman, D. P. Computer simulations of membrane proteins. *Biochim. Biophys. Acta BBA - Biomembr.* **1666,** 158–189 (2004).
45. Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D. & Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput Biol* **6,** e1000810 (2010).
46. Seo, M., Rauscher, S., Pomès, R. & Tieleman, D. P. Improving Internal Peptide Dynamics in the Coarse-Grained MARTINI Model: Toward Large-Scale Simulations of Amyloid- and Elastin-like Peptides. *J. Chem. Theory Comput.* **8,** 1774–1785 (2012).
47. Lyon, R. P. & Atkins, W. M. Self-Assembly and Gelation of Oxidized Glutathione in Organic Solvents. *J. Am. Chem. Soc.* **123,** 4408–4413 (2001).
48. Cohen, Y., Avram, L. & Frish, L. Diffusion NMR Spectroscopy in Supramolecular and Combinatorial Chemistry: An Old Parameter—New Insights. *Angew. Chem. Int. Ed.* **44,** 520–554 (2005).
49. Pouget, E. *et al*. Elucidation of the Self-Assembly Pathway of Lanreotide Octapeptide into beta-Sheet Nanotubes: Role of Two Stable Intermediates. *J. Am. Chem. Soc.* **132,** 4230–4241 (2010).
50. Barth, A. & Zscherp, C. What Vibrations Tell About Proteins. *Q. Rev. Biophys.* **35,** 369–430 (2002).
51. Fleming, S. *et al*. Assessing the Utility of Infrared Spectroscopy as a Structural Diagnostic Tool for β-Sheets in Self-Assembling Aromatic Peptide Amphiphiles. *Langmuir* **29,** 9510–9515 (2013).
52. Fuhrmans, M. & Marrink, S.-J. A tool for the morphological analysis of mixtures of lipids and water in computer simulations. *J. Mol. Model.* **17,** 1755–1766 (2011).
53. Georgoulia, P. S. & Glykos, N. M. On the Foldability of Tryptophan-Containing Tetra- and Pentapeptides: An Exhaustive Molecular Dynamics Study. *J. Phys. Chem. B* **117,** 5522–5532 (2013).
54. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

TOC figure