

Computational prediction of tripeptide-dipeptide co-assembly

Inês P. Moreira,^a Gary G. Scott,^a Rein V. Ulijn^b and Tell Tuttle^a

^a*WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, UK*

^b*Advanced Science Research Center (ASRC), City University of New York, 85 St Nicholas Terrace, New York NY10031, USA*

Corresponding author: Prof. Tell Tuttle, tell.tuttle@strath.ac.uk

Dedication: In memory of Dieter Cremer – supervisor and mentor (TT).

Computational prediction of tripeptide-dipeptide co-assembly

In this work, we describe the development of a computational screening approach for tripeptide-dipeptide co-assembly. Studies are carried out both in water and in oil-water mixtures, to evaluate possible candidates that give rise to hydrogels or more stable emulsions, respectively, through nanofibre formation. The results give rise to design rules for the identification of promising systems for numerous types of soft materials. The possibility of achieving innovative functional materials through the co-assembly of tripeptides and dipeptides is studied. In particular, coarse-grained simulations allowed for the extraction of some promising dipeptides that, together with H-aspartyl-phenylalanyl-phenylalanine-OH (DFF), are able to act as hydrogelators or emulsifiers with superior characteristics relative to DFF on its own.

Keywords: co-assembly; peptides; molecular dynamics

Introduction

Co-assembly of two or more types of molecules is a powerful technique to produce supramolecular structures with enhanced complexity and functionality.¹⁻⁴ For example, certain Fmoc-peptide based gelators/surfactants have been shown to co-assemble into nanofibres with tunable surface functionality.⁵⁻⁷ The use of co-assembly to control the physical properties has been demonstrated through the co-assembly of diphenylalanine dipeptide nanotubes with the amine analogue Boc-FF.⁸ In addition to morphological control through molecular design of different triggers, co-assembly allows for an extra control over the assembly pathway.^{3,9} Different research studies have been published on the design and use of co-assembled systems,¹⁰ such as the co-assembly between peptide amphiphiles and fatty acids¹¹ or Fmoc-protected peptides co-assembled with melamine.¹²

In a previous report, Fleming *et al.* carried out a design study to elucidate different co-assembly models: orthogonal, cooperative and disruptive.⁷ The first was shown to occur when the hydrogelator and surfactant molecules establish their own preferred interactions, not significantly affecting the aromatic stacks or β -sheet type arrangements formed by the other partner molecule. Cooperative co-assembly happens when the building blocks assemble into a common preferred mode of self-assembly, and disruptive co-assembly when intercalation of groups compromises the integrity of the desired supramolecular arrangements.⁷

A similar nomenclature was also used by Adams' group, who studied the combination of two variable low molecular weight (LMW) gelators and their mode of assembly to control the mechanical properties of gels.¹³ They demonstrated that, apart from co-assembly, self-sorting can occur, by which there is independent assembly of each constituent into different fibre networks.^{3,14} Orthogonal self-assembly (different from orthogonal co-assembly), which is the independent but simultaneous assembly of different supramolecular structures within the same system, has recently received increased attention. Different constituents, such as gelators and surfactants, are able to self-assemble independently into, *e.g.*, fibrillar networks with micelles, and co-exist, each with their own characteristics.^{15,16} Such systems can potentially create novel and more complex architectures that neither system could achieve individually.¹⁵ The disruptive model can also be perceived as a *perturbed* co-assembly, since the nanofibrous networks are not fully disrupted and they may lead to the formation of other equally interesting structures.

To further enhance the properties of short peptide assemblies, co-assembly can be attractive but, to date, the discovery of co-assembled systems has either been through serendipity or on the basis of simple design rules (*e.g.*, charge complementarity).

Computational screening methods can be helpful to provide insight into the design of different materials.^{17,18} Wei and Gazit used a large set of coarse-grained molecular dynamics (CG-MD) simulations to study the co-assembly between the dipeptide FF and the tripeptide FFF, using mixtures of varying FFF mass percentage (P_{FFF}). In this work, the authors demonstrated the formation of distinct structures in the co-assembled state, with nanotoroids achieved for a P_{FFF} 0.167-0.33 as opposed to the cylindrical or spherical nanoassemblies formed with a P_{FFF} of 0 or 1, respectively.¹⁹

The ability of unprotected tripeptides to act as LMW gelators and surfactants has previously been predicted computationally, using atomistic^{20,21} and coarse-grained methodology.^{18,22} In these latter works, CG-MD simulations were used to analyse the assembly tendencies, providing further information about the most favourable assembly for each peptide system in a water box or in a water/octane box.²² It was found that certain tripeptides, such as H-aspartyl-phenylalanyl-phenylalanine-OH (DFF), create bilayers in aqueous medium, while others, such as H-lysyl-tyrosyl-phenylalanine-OH (KYF), are able to self-assemble into fibres.²² When in a water/oil system, the first case forms a weakly stabilising monolayer at the interface between water and oil, and the second creates strong interfacial networks capable of stabilising oil-in-water droplets for extended periods.²²

The aim of this study is thus to investigate whether coarse-grained molecular dynamics screening methods can be used to identify candidates for co-assembly of di- and tri-peptides to give desired structures. The objective is not to predict or select the best option for a specific application, but to explore and determine the design rules for modulating the assembly of DFF, selected as an example of a non-fibrous forming peptide (*vide infra*). Thus, a screening of all 400 possible dipeptides in combination with a tripeptide aqueous solution of DFF is carried out to assess the ability of nanofibre

formation. The behaviour of these co-assembled systems is also analysed in a water/octane biphasic system to test if interfacial nanofibrous networks could give rise to more stable emulsions.

Methods

A full screening of all the possible dipeptides in combination with DFF was carried out using GROMACS molecular dynamics package, version 4.5.3²³ and MARTINI force field (version 2.2).²⁴ Peptide structures were converted to the CG representation by using martinize.py script.²⁵

To start with, 300 molecules of DFF (in its zwitterionic form) were randomly placed in a 12.5 x 12.5 x 12.5 nm³ box with a minimum distance of 3 Å between them and the box was solvated with MARTINI CG standard water. For the DFF in a biphasic system, 1000 molecules of octane were added before the solvation in order to achieve a density of water approximate to the experimental value (999 kg.m⁻³). Octane was used to represent a very apolar phase, in order to simulate a highly unfavourable water/oil mixture. The system was then neutralised, by adding 300 sodium ions to compensate for the negative aspartic acid charges, and energy minimised using the steepest descent integrator. The Berendsen algorithms were used to keep temperature at 303 K and pressure at 1 bar.²⁶ All the simulations were ran for 100 ns using a 25 fs time step, which equates to an effective 400 ns of atomistic simulation time due to the smoothness of the CG potentials.²⁴

For the screening of all dipeptides with DFF in water, 150 molecules of each dipeptide (in their zwitterionic form) were added to the same size cubic box with 150 molecules of DFF, separately. The boxes were then solvated, neutralised (depending on the side chains of each dipeptide), minimised in the same way and ran for 100 ns. For

the screening of all dipeptides with DFF in biphasic systems, the same protocol as in water was used but 1000 molecules of octane were added before solvation.

The conformations obtained for each of the simulated systems were analysed based on their aggregation propensity (AP). The calculation of the solvent-accessible surface area (SASA) in Å² after assembly and its comparison to the initial SASA is a way to quantify the level of aggregation of the different peptides.¹⁷ The AP score is defined as the ratio between the SASA in the randomised initial state of the simulation and the SASA in the final configuration of the simulation, according to Equation 1:

$$\text{AP score} = \frac{\text{SASA}_{\text{initial}}}{\text{SASA}_{\text{final}}} \quad (1)$$

The SASA values were calculated using the VMD scripting tools.²⁷ The typical rolling sphere radius of 1.4 Å was used, as it approximates the radius of a water molecule. The AP score is calculated for the whole system (AP_{total}), for only DFF molecules (AP_{DFF}) and for only dipeptide molecules (AP_{dip}). This is achieved by using a specific selection of atoms/beads for the SASA determination.

The hydrophilicity-corrected AP_{total} score (AP_H) is used to generate design rules for hydrogelators as hydrophobic peptides can be insoluble and a positive bias is needed for hydrophilic peptides, analogous to previous work.¹⁸ This is done by introducing a measure of the hydrophilicity, by using the change in free energy when the dipeptide (the DFF is not taken into consideration as it remains for all the systems) is transferred from water to n-octanol ΔG_{w,o}. This is obtained from Equation 2:

$$\text{AP}_H = (\text{AP}_{\text{total}})^\alpha \cdot (\Delta G_{w,o})'_{\text{dip}} \quad (2)$$

The AP_{total} score is calculated from Equation 1 accounting for all the atoms in the system except solvent and ions. When AP_H' is calculated, AP_{total} is normalised between

0 and 1 by following Equation 3:

$$AP_{\text{total}}' = \frac{AP_{\text{total}} - (AP_{\text{total}})_{\text{min}}}{(AP_{\text{total}})_{\text{max}} - (AP_{\text{total}})_{\text{min}}} \quad (3)$$

In turn, α is a coefficient that can be varied according to the desired weight of the normalised AP score to the AP_H score, where $\alpha = 2$ is used in this case to obtain a good compromise between the hydrophilicity of each dipeptide and the AP score. The normalised $(\Delta G_{w,o})_{\text{dip}}$ value is obtained by Equation 4:

$$(\Delta G_{w,o})'_{\text{dip}} = \frac{(\Delta G_{w,o})_{\text{dip}} - (\Delta G_{w,o})_{\text{min}}}{(\Delta G_{w,o})_{\text{max}} - (\Delta G_{w,o})_{\text{min}}} \quad (4)$$

where $(\Delta G_{w,o})_{\text{dip}}$ is calculated by the sum of the free energies of transfer of the two constituent amino acids from water to n-octanol (kcal.mol^{-1}). The $\Delta G_{w,o}$ values used were those reported by Wimley and White.²⁸ The values for the charged side chain amino acids as used in this work are taken, except for histidine, where the value corresponding to the neutral amino acid was used. $(\Delta G_{w,o})_{\text{min}}$ represents the most hydrophobic dipeptide ($-2.09 \times 2 = -4.18$ for WW) and $(\Delta G_{w,o})_{\text{max}}$ the most hydrophilic dipeptide ($3.64 \times 2 = 7.28$ for DD), normalising each dipeptide in a 0-1 scale.

Results and discussion

Aqueous systems

DFE begins to self-assemble into bilayer-type structures in water by the end of 400 ns simulation (See Figure 1b), which is in agreement with the previously observed bilayers after 9.6 μs of a CG-MD simulation.²² Even though equilibrated structures are probably not reached after 400 ns, this simulation time was shown to be sufficient for an initial

screening and for calculating and extracting conclusions regarding the aggregation propensity AP.¹⁷ The AP score for DFF in aqueous medium was found to be 2.25, which corresponds to the hydrophobic collapse and the corresponding decrease in solvent-accessible surface area (SASA) throughout the course of the simulation.

The 400 dipeptides based on the 20 canonical amino acids, when combined with DFF in a water box in a 1:1 ratio, show different behaviours over the course of the CG-MD simulation, which is reflected in the analysed total aggregation propensity score (AP_{total}) (Figure 2a), when plotted against the dipeptides' water-octanol partition free energy ($\Delta G_{\text{w/o}}$).

The highest AP score (*i.e.*, the most significant decrease in SASA upon assembly) is observed for DFF/SW ($AP_{\text{total}} = 2.4$), for which a fibre structure is formed and the final surface area is small (Figure 2b) in comparison to the initial state, giving a higher aggregation score. As seen in Figure 3b and Table S1 in Supporting Information, the lowest AP score is achieved for the non-assembling DFF/EK ($AP_{\text{total}} = 1.3$), as the SASA is high. In addition to the AP_{total} , the AP_{DFF} and AP_{dip} scores are also analysed, which take into account only the aggregation of molecules of the same type (*i.e.*, the aggregation between DFF and DFF, or dipeptide and dipeptide – green 'surface' in Figure 2c). The different ranges of values and tendencies are observed as a function of the dipeptide hydrophilicity (Figure 2a). Differently from the dipeptide that reached the highest AP_{total} when combined with DFF (represented in green), SK reached the maximum AP_{DFF} (cyan marks) and FF the maximum AP_{dip} score (orange). The final snapshots (after 400 ns) of the simulations that achieved the highest AP scores are presented in Figure 2b and the lowest AP scores in Figure 3b.

The scores vary according to what they account for, where the AP_{DFF} and AP_{dip} scores are independently not as high for the system that achieved the highest AP_{total} due

to the intercalated nature of DFF and the dipeptide, *e.g.*, SW. This type of assembly is able to strengthen the interactions to form fibrous structures, related back to the previously stated nomenclature as cooperative co-assembly. A schematic drawing of the different co-assembly behaviours believed to occur for these systems reaching the different maximum AP scores is thus presented in Figure 2c and Figure 4.

The highest AP_{DFF} , achieved for SK ($AP = 2.38$), is higher than AP_{DFF} for DFF on its own, which suggests the formation of a bilayer-type structure, with the aspartic acids from the DFF (in blue) facing the outside to interact with some SK dipeptides. As represented in Figure 2c, the dipeptides coat the fibre/bilayer in arrangements such as the one observed for SK, which increases the whole surface accessible area and thus do not present the highest AP_{total} scores. Since there are two independent modes of assembly, this is considered to be an example of orthogonal assembly. In turn, dipeptides that have high AP_{dip} scores such as diphenylalanine (FF) self-assemble quite independently (*perturbing self-assembly*), as shown before for dipeptides only in aqueous medium.¹⁷ FF molecules also interact with the phenylalanines from DFF but do not improve the independent assembly of DFF, actually accounting for one of the lowest AP_{DFF} scores ($AP_{DFF} = 1.47$ seen in Table S1 from Supporting Information), showing it can be detrimental to the tripeptide structure (*i.e.*, *perturbing co-assembly*).

The highest total aggregation propensity values are reached by the most hydrophobic dipeptides, with at least one aromatic amino acid (see Table 1). This presents no surprise as the hydrophobic effect is known to dominate aggregation in water, while hydrophilic dipeptides show a lower tendency to aggregate with DFF. However, hydrogel formation requires favourable interactions between nanofibres formed and the solvent, thus requiring inclusion of hydrophilic groups,²⁹ where the peptide sequence plays a key role in the control of the self-assembling hydrogel

properties.³⁰ To account for this, a corrected AP score system that favours hydrophilic residues (AP_H) was previously developed for tripeptides,¹⁸ which we use for the total system as explained in the Methods. When correcting this indicator by adding a positive bias towards hydrophilic peptides, hydrophilic dipeptides are now ranked higher than more hydrophobic dipeptides with the same AP_{total} score (Figure 5). There were only 29 systems, out of 400, for which the AP_{total} was higher than 2 (red diamonds in Figure 5). This cut-off value has been shown to be a reasonable indicator for the selection of the best dipeptides for self-assembly.¹⁷ When comparing the two scoring systems, there are seven dipeptides (KH, SK, KW, RF, KF, FR and WK) that fall into both categories of $AP_{total} > 2$ and AP_H top 30 (blue squares) and these all have intermediate hydrophilicity values.

Previously, we have shown the sequence dependence for the aggregation propensity in the case of tripeptides, where it was possible to draw conclusions about the relative positioning of the amino acids in the tripeptide to encourage self-assembly and hydrogelation.¹⁸ Therefore, we applied a similar process of analysing the average contribution of each amino acid to the AP_H scores obtained after 400 ns, by comparing the aggregation propensities when a certain amino acid is placed in position 1 or 2 (N-terminus or C-terminus, respectively). Using this approach, we could determine which dipeptide should be used for the production of the best interconnection between dipeptide and DFF, which produces nanofibres and, consequently, hydrogels. As observed, aromatic (Figure 6a) and cationic (Figure 6c) amino acids more strongly promote aggregation, contributing to higher AP_H values. Aromatic amino acids are more favourable in the C-terminus position than at the N-terminus, except for tyrosine, where the difference is not significant. In contrast, higher AP_H scores are reached when cationic amino acids are placed in the N-terminus. A similar behaviour was previously

observed for tripeptides,¹⁸ where the cationic and H-bond donor amino acids K, R, S and T are advantageous in position 1 of the chain. In this case, the relative positioning of hydrophilic amino acids (Figure 6b), including T, show no strong preference for the C- or N-terminus, with the exception of S. There is a slight preference for negatively charged amino acids (E and D) to be positioned at the C-terminus (Figure 6d), while there is no sequence dependence for small hydrophobic side chains (Figure 6e). In general, there was no particular preference for the position of uncharged hydrophilic/hydrophobic amino acids within the dipeptide. The relative position of the two amino acids in some of the cases was shown to be an important factor for the ability to aggregate as for tripeptides or larger peptides/proteins. The positioning of the amino acids can indeed determine the mode of interaction between each dipeptide and the tripeptide DFF, which is demonstrated by the different structures formed when using similar dipeptides (see RF *versus* FR in Table 1). A number of dipeptides are shown to assemble in combination with DFF, forming fibres instead of bilayers. The combination of a cationic amino acid at the N-terminus and an aromatic amino acid such as F or W in the C-terminal position of the dipeptide promotes cooperative co-assembly with the amphiphilic DFF.

Biphasic systems

When in a biphasic system of water and octane, octane droplets are formed with DFF assembled at the interface as a monolayer by the end of 400 ns (Figure 1d), with the hydrophilic anionic aspartic acid preferentially exposed to the water and phenylalanine residues facing the organic solvent in the interior of the droplet, as previously observed.²² In this biphasic system, AP_{total} only scored 1.63, as opposed to the 2.25 achieved for the aqueous system. This is due to the formation of DFF interfacial monolayers at the dispersed octane droplets in water, which increases the solvent-

accessible surface area, in comparison with the aggregation of DFF molecules into a bilayer.

The co-assembly behaviour was subsequently applied to the biphasic systems, based on the assumption that co-assembled fibres could be formed at the interface between two immiscible liquids as was previously observed for tripeptide gelators.²² The rationale used for the selection of the best dipeptide that, in combination with DFF, could become a good emulsifier was thus based on the water simulations and on the ability of DFF and dipeptide to undergo cooperative co-assembly to form nanofibres. When AP_{DFF} is high ($AP_{DFF} > 2.2$), nanofibres or bilayers are formed with the dipeptide coating their interface in an orthogonal manner, such as in the case of SK (Figure 2b). These types of structures are not considered to be sufficiently stable to coat the octane droplets' surface once used in a biphasic system. The systems where both DFF and dipeptides cooperate to reach a co-assembled fibrous structure in water would be of interest to stabilise a biphasic medium. These nanofibres would assemble at the interface between water and oil, reducing their interfacial tension, but also at the aqueous environment to increase the viscosity of the medium and delay droplet coalescence.

The selection of the systems that possibly undergo a cooperative assembly between DFF and the dipeptide at the interface involves the study of the compromise between the AP_{DFF} and AP_{dip} values. In order to exclude dipeptides that tend to aggregate by themselves and undergo perturbed assembly (*i.e.*, the hydrophobic ones), an empirical measure was determined based upon the inspection of the simulated systems. The absolute value of the difference between the AP scores for DFF in the presence of the dipeptide (AP_{DFF}) and the DFF alone in water (AP_{DFF})₀ should be less than 0.7. On the other hand, to discard the previously described orthogonal co-assembly

systems and account only for interconnecting fibre formation through cooperative co-assembly, the absolute value of the difference between the AP scores for dipeptide with DFF (AP_{dip}) and dipeptide alone (using as $(AP_{\text{dip}})_0$ the reported AP values from dipeptide screening)¹⁷ should be greater than 1, as in Equation 5.

$$\begin{cases} |(AP_{\text{DFF}}) - (AP_{\text{DFF}})_0| < 0.7 \\ |(AP_{\text{dip}}) - (AP_{\text{dip}})_0| > 1. \end{cases} \quad (5)$$

The selected systems that meet the defined criteria (SW, SF, TW, CW, IF and LF) are analysed through CG simulations of a water/octane system and positioned in a similar region within the whole range of AP values (See Figure 7).

The biphasic CG simulations of DFF with those six dipeptides are visually analysed to assess whether nanofibres are formed that can stabilise droplets in a different way (Figure 8). The system that reached the highest AP_{total} in the aqueous simulations, when SW is used in combination with DFF, presents a very well-covered stable water/octane interface. Even though it is not possible to clearly observe nanofibres at the interface due to the size limitations of the computational model, the interactions between the peptides affect the arrangement of the peptides at the interface, rather than simply responding to the environment of the octane and water as in the pure surfactant-type adsorption observed for DFF (Figure 1d). This observation implies that a more efficient emulsification may be achieved. The other selected systems present a similar behaviour, which coincides with the established rules for better emulsifiers.

The total AP score (AP_{total}) was more closely analysed for all the biphasic systems in order to study the tendencies for co-assembly and investigate any trends in the data (Figure 9). It is possible to observe that aromatic amino acids are prone to

assemble when in a water/octane system, even though there is no significant preference for either the N- or C-terminal position within the dipeptide.

The co-assembly of SW, SF, TW and CW with DFF in a water/octane system were able to convert a surfactant-type emulsification of the tripeptide DFF into a more stabilising interfacial network. This validates the design rules for hydrogelators and also the tendencies studied for biphasic systems, as the selected dipeptides are formed by a hydrophilic and an aromatic amino acid, more specifically at the N-terminus and C-terminus, respectively.

Conclusions

This work has shown that the co-assembly of different dipeptides with the DFF tripeptide can result in systems that have variable characteristics (act as better hydrogelators or emulsifiers) and while different systems may have some similarities – e.g., the inclusion of an aromatic amino acid in the dipeptide sequence greatly favours co-assembly – there are also clear differences that can be controlled by dipeptide sequence. In the case of hydrogelators, we found that the inclusion of a cationic amino acid at the N-terminus led to better fiber formation, when co-assembled with DFF in a purely aqueous system. However, to form emulsions a balance between the hydrophobic environment of the oil (octane) and the hydrophilic environment of the water is critical, in this case charged amino acids at the N-terminus were apparently too hydrophilic and H-bond donors such as S, T, and C were found to yield the best results. The experimental verification of these proposed systems is the focus of future work and will be reported subsequently.

In the current work, we have focussed on the ability of the primary sequence of the dipeptide to modify the nanomorphology of the resulting system and as such the

properties of the bulk material. As a result of this focus we have controlled other aspects of the system design such as the pH, temperature and stoichiometry of the reaction mixture. Indeed, these are all physical levers that can easily be manipulated to help vary the resulting structures that can be obtained through co-assembly. For example, in a similar manner to the approach taken by Wei and Gazit¹⁹ the ratio of the dipeptide and tripeptide of the systems found to co-assemble in this work could be varied to further tune the material's properties. Indeed, we hope that the identification of promising systems that can potentially co-assemble into hydrogels while in water and into more effective emulsifiers while in a biphasic system can trigger further investigation due to their responsive nature.

Acknowledgements. Results were obtained using the EPSRC funded ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk; EPSRC grant no. EP/K000586/1). This work was supported by the H2020 Marie Curie Actions via the European Initial Training Network SMARTNET.

1. The tables with the full results are provided as supplementary information.

References

- ¹ Z. Yu, F. Tantakitti, T. Yu, L.C. Palmer, G.C. Schatz, and S.I. Stupp, *Science* (80-.), **351**, 497 (2016).
- ² K.E. Inostroza-Brito, E. Collin, O. Siton-Mendelson, K.H. Smith, A. Monge-Marcet, D.S. Ferreira, R.P. Rodríguez, M. Alonso, J.C. Rodríguez-Cabello, R.L. Reis, F. Sagués, L. Botto, R. Bitton, H.S. Azevedo, and A. Mata, *Nat. Chem.* **7**, 897 (2015).
- ³ E.R. Draper, E.G.B. Eden, T.O. McDonald, and D.J. Adams, *Nat. Chem.* **7**, 848 (2015).
- ⁴ D.K. Smith, *Chem. Soc. Rev.* **38**, 684 (2009).
- ⁵ Y.M. Abul-Haija, S. Roy, P.W.J.M. Frederix, N. Javid, V. Jayawarna, and R. V. Ulijn, *Small* **10**, 973 (2014).
- ⁶ E. V. Alakpa, V. Jayawarna, A. Lampel, B. Pé, R. V. Ulijn, and M.J. Dalby, *Chem* **1**, 298 (2016).
- ⁷ S. Fleming, S. Debnath, P.W.J.M. Frederix, N.T. Hunt, and R. V. Ulijn, *Biomacromolecules* **15**, 1171 (2014).
- ⁸ L. Adler-Abramovich, P. Marco, Z.A. Arnon, R.C.G. Creasey, T.C.T. Michaels, A. Levin, D.J. Scurr, C.J. Roberts, T.P.J. Knowles, S.J.B. Tandler, and E. Gazit, *ACS Nano* **10**, 7436 (2016).
- ⁹ B. Adhikari, J. Nanda, and A. Banerjee, *Soft Matter* **7**, 8913 (2011).
- ¹⁰ P. Makam and E. Gazit, *Chem. Soc. Rev.* **47**, 3406 (2018).
- ¹¹ Z. Yu, A. Erbas, F. Tantakitti, L.C. Palmer, J.A. Jackman, M. Olvera de la Cruz, N.-J. Cho, and S.I. Stupp, *J. Am. Chem. Soc.* **139**, 7823 (2017).
- ¹² P. Xing, P. Li, H. Chen, A. Hao, and Y. Zhao, *ACS Nano* **11**, 4206 (2017).
- ¹³ C. Colquhoun, E.R. Draper, E.G.B. Eden, B.N. Cattoz, K.L. Morris, L. Chen, T.O. McDonald, A.E. Terry, P.C. Griffiths, L.C. Serpell, and D.J. Adams, *Nanoscale* **6**, 13719 (2014).
- ¹⁴ K.L. Morris, L. Chen, J. Raeburn, O.R. Sellick, P. Cotanda, A. Paul, P.C. Griffiths, S.M. King, R.K. O'Reilly, L.C. Serpell, and D.J. Adams, *Nat. Commun.* **4**, 1480 (2013).

- ¹⁵ A. Brizard, M. Stuart, K. Van Bommel, A. Friggeri, M. De Jong, and J. Van Esch, *Angew. Chemie - Int. Ed.* **47**, 2063 (2008).
- ¹⁶ A. Heeres, C. Van Der Pol, M. Stuart, A. Friggeri, B.L. Feringa, and J. Van Esch, *J. Am. Chem. Soc.* **125**, 14252 (2003).
- ¹⁷ P.W.J.M. Frederix, R. V. Ulijn, N.T. Hunt, and T. Tuttle, *J. Phys. Chem. Lett.* **2**, 2380 (2011).
- ¹⁸ P.W.J.M. Frederix, G.G. Scott, Y.M. Abul-Haija, D. Kalafatovic, C.G. Pappas, N. Javid, N.T. Hunt, R. V. Ulijn, and T. Tuttle, *Nat. Chem.* **7**, 30 (2015).
- ¹⁹ C. Guo, Z.A. Arnon, R. Qi, Q. Zhang, L. Adler-Abramovich, E. Gazit, and G. Wei, *ACS Nano* **10**, 8316 (2016).
- ²⁰ A. V. Vargiu, D. Iglesias, K.E. Styan, L.J. Waddington, C.D. Easton, and S. Marchesan, *Chem. Commun.* **52**, 5912 (2016).
- ²¹ A.M. Garcia, D. Iglesias, E. Parisi, K.E. Styan, L.J. Waddington, C. Deganutti, R. De Zorzi, M. Grassi, M. Melchionna, A. V. Vargiu, and S. Marchesan, *Chem* (2018).
- ²² G.G. Scott, P.J. McKnight, T. Tuttle, and R. V. Ulijn, *Adv. Mater.* **28**, 1381 (2016).
- ²³ B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).
- ²⁴ S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, and A.H. De Vries, *J. Phys. Chem. B* **111**, 7812 (2007).
- ²⁵ Martinize.Py, [Http://Www.Cgmartini.Nl/Index.Php/Downloads/Tools/204-Martinize](http://Www.Cgmartini.Nl/Index.Php/Downloads/Tools/204-Martinize), 2017.
- ²⁶ H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
- ²⁷ W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graph.* **14**, 33 (1996).
- ²⁸ S.H. White and W.C. Wimley, *Biochim. Biophys. Acta - Rev. Biomembr.* **1376**, 339 (1998).
- ²⁹ I. Ramos Sasselli, P.J. Halling, R. V. Ulijn, and T. Tuttle, *ACS Nano* **10**, 2661 (2016).
- ³⁰ E.F. Banwell, E.S. Abelardo, D.J. Adams, M.A. Birchall, A. Corrigan, A.M. Donald,

M. Kirkland, L.C. Serpell, M.F. Butler, and D.N. Woolfson, *Nat. Mater.* **8**, 596 (2009).

Table 1. List of the top 10 dipeptides as ranked by AP_{total} score, also presenting their AP_{DFF} and AP_{dip} scores, together with the observed structure from the simulation final snapshot. Phenylalanine from DFF is represented in white and aspartic acid from DFF in blue VDW particles, all dipeptides are represented in red, while water beads are omitted for clarity.

Figure 1. Snapshots of the DFF control system in a water box in the beginning (a) and end of the simulation (400 ns) (b) and snapshots of the DFF control system in a water/octane box in the beginning (c) and end of the simulation (400 ns) (d). Phenylalanine is represented in white and aspartic acid in blue VDW particles. Octane is in yellow, while water beads are omitted for clarity.

Figure 2. (a) Positioning of the DFF in combination with SW, SK and FF amongst the whole range of AP_{total} , AP_{DFF} and AP_{dip} for all the systems, plotted as a function of the dipeptide hydrophilicity ($\Delta G_{\text{w,o}}$); (b) Final snapshots (400 ns) of the systems that achieved the maximum AP scores, DFF/SW for AP_{total} , DFF/SK for AP_{DFF} and DFF/FF for AP_{dip} ; (c) Schematic representation of the cooperative co-assembly, where dipeptides cooperate with the DFF to form fibres (representing *e.g.*, DFF/SW), of the orthogonal co-assembly, where dipeptides coat the DFF fibres (representing *e.g.*, DFF/SK) and of the perturbing co-assembly, where dipeptides tend to aggregate between themselves independently from the DFF (representing *e.g.*, DFF/FF). DFF is represented in blue (including Asp and Phe) and dipeptides in red, while water beads are omitted for clarity. The shaded rectangles reflect the fact that the dipeptides or the tripeptides DFF are not “visible” to the probe when accounting for AP_{DFF} or AP_{dip} , respectively. The green line represents the solvent accessible surface area for each of the scoring systems (AP_{total} , AP_{DFF} or AP_{dip}).

Figure 3. (a) Positioning of the DFF in combination with EK, FK and AA amongst the whole range of AP_{total} , AP_{DFF} and AP_{dip} for all the systems, plotted as a function of the dipeptide hydrophilicity ($\Delta G_{\text{w,o}}$)_{dip}; (b) Final snapshots (400 ns) of the systems that achieved the minimum AP scores, DFF/EK for AP_{total} , DFF/FK for AP_{DFF} and DFF/AA for AP_{dip} for the aqueous simulations.

Figure 4. (a) Schematic representation of the cooperative co-assembly, where dipeptides cooperate with the DFF to form fibres (representing *e.g.*, DFF/SW); (b) Schematic

representation of the orthogonal co-assembly, where dipeptides coat the DFF fibres (representing *e.g.*, DFF/SK); (c) Schematic representation of the perturbing co-assembly, where dipeptides tend to aggregate between themselves independently from the DFF (representing *e.g.*, DFF/FF). DFF is represented in blue (including Asp and Phe) and dipeptides in red, while water beads are omitted for clarity. The shaded rectangles reflect the fact that the dipeptides or the tripeptides DFF are not “visible” to the probe when accounting for AP_{DFF} or AP_{dip} , respectively. The green line represents the solvent accessible surface area in each case and the highlighted squares represent the maximum of each of the scoring systems (AP_{total} , AP_{DFF} or AP_{dip}).

Figure 5. Aggregation propensity from the aqueous simulations of DFF/dipeptide plotted as a function of the dipeptide hydrophilicity ($\Delta G_{w,o}$). Black diamonds represent the AP_{total} , which also includes the red triangles, that represent the dipeptides for which $AP_{total} > 2$. The green circles represent the top 30 hydrophilicity-corrected AP_H and the blue squares the overlap between the highest from the two methods.

Figure 6. Average AP_H ' scores of the dipeptides, when in combination with DFF, in a water medium, with the corresponding amino acid on the x-axis in the N-terminus or C-terminus. Amino acids are grouped per: (a) Aromatic; (b) Hydrophilic; (c) Cationic; (d) Anionic; (e) Small/hydrophobic side chains.

Figure 7. Positioning of the DFF in combination with each one of the six filtered dipeptides according to the criteria presented above for emulsifiers (SW, SF, TW, CW, IF and LF) amongst the whole range of: (a) AP_{total} ; (b) AP_{DFF} ; (c) AP_{dip} for all water/octane simulations.

Figure 8. Final snapshots (400 ns) of different dipeptides in combination with DFF in a water/octane system. These 6 were the filtered dipeptides according to the criteria presented for emulsifiers. Phenylalanine from DFF is represented in white and aspartic acid from DFF in blue VDW particles, all dipeptides are represented in red, octane in yellow and water beads are omitted for clarity.

Figure 9. Average AP_{total} ' scores of the dipeptides, when in combination with DFF, in a water/octane medium, with the corresponding amino acid on the x-axis in the N-terminus and C-terminus. Amino acids are grouped per: (a) Aromatic; (b) Hydrophilic; (c) Cationic; (d) Anionic; (e) Small/hydrophobic side chains.