Chem Soc Rev

Review Article

Guiding principles for peptide nanotechnology through directed discovery†

Received 1st December 2017,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

A. Lampel,[a] R. V. Ulijn*[a,b,c] and T. Tuttle*[d]

Life's diverse molecular functions are largely based on only a small number of highly conserved building blocks- the twenty canonical amino acids. These building blocks are chemically simple, but when they are organized in three-dimensional structures of tremendous complexity, new properties emerge. This review explores recent efforts in the directed discovery of functional nanoscale systems and materials based on these same amino acids, but that are not guided by copying or editing biological systems. The review summarises insights obtained using three complementary approaches of searching the sequence space to explore sequence-structure relationships for assembly, reactivity and complexation, namely: (i) strategic editing of short peptide sequences; (ii) computational approaches to predicting and comparing assembly behaviours; (iii) dynamic peptide libraries that explore the free energy landscape. These approaches give rise to guiding principles on controlling order/disorder, complexation and reactivitiy by peptide sequence design.

1. Introduction

Across all life forms, DNA encodes for just twenty, chemically simple amino acids. Instructed by DNA sequence, these amino acids combine, *via* amide bonds, into polymers of hundreds of amino acid residues, giving rise to the functional proteins that are responsible for fundamental chemical processes of life: molecular recognition, catalysis & reactivity and self-assembly. It is now well established that much shorter oligopeptides can also have these functions, and they may be used for the fabrication of customizable supramolecular materials with a wide variety of applications.[1, 2] In addition, functional roles of short peptides in biological contexts are increasingly appreciated and documented.[3] Compared to proteins, peptides are readily synthesized and can be scaled, they also have low structural complexity, which enables rational and systematic studies using both experiment and computation leading to connections between peptide sequence and supramolecular functionality to be established.

The identification of conserved amino acid sequence patterns to help identify certain modes of folding or assembly is well established in assessment of protein structures.[4-7] For example, the identification of a binary pattern of nonpolar and polar amino acids in proteins[8] inspired the design of a number of oligopeptides with alternating nonpolar/polar amino acids that self-assemble into β-sheet structures which form hydrogels[9-13]. Typically, these peptides contain nonpolar amino acids such as phenylalanine or valine and one charged polar amino acid such as lysine[11] or arginine[9], or alternation of oppositely charged amino acids such as glutamic acid and lysine[13]. Complementary to the search of simple conserved patterns in protein sequences, a reductionist approach, where

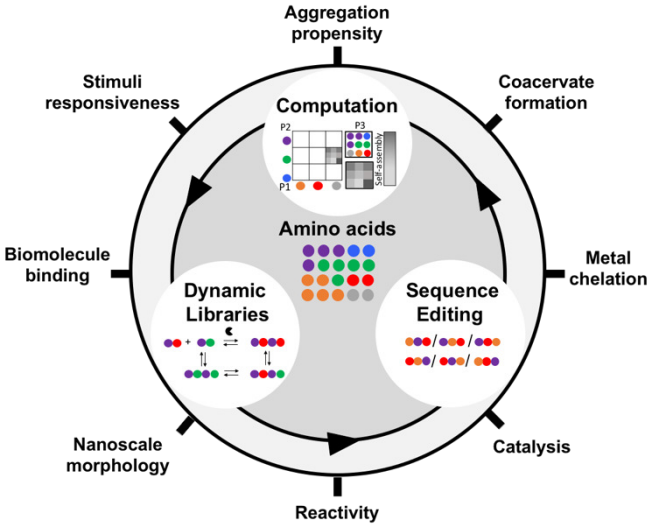

Fig. 1. Main approaches for directed discovery of peptide nanostructures: sequence editing (section 2) computational search (section 3), and dynamic libraries (section 4). Each search starts with elected pool of amino acids (colored beads) and desired properties are applied for selection of nanostructures. Suggested properties for selection including reported examples (blue) are presented around the outer circle. Each one of the properties can be applied by each of the search methods.

[a] Advanced Science Research Center (ASRC) at the Graduate Center of the City University of New York (CUNY), New York, USA.
Email: rein.ulijn@asrc.cuny.edu
[b] Department of Chemistry and Biochemistry, Hunter College of CUNY, 695 Park Ave., New York, NY 10031, USA
[c] PhD programs in Biochemistry and Chemistry, The Graduate Center of the City University of New York, NY 10016, USA.
[d] WestCHEM/Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, UK.
Email: tell.tuttle@strath.ac.uk

† Electronic Supplementary Information (ESI) available. See DOI: 10.1039/x0xx00000x

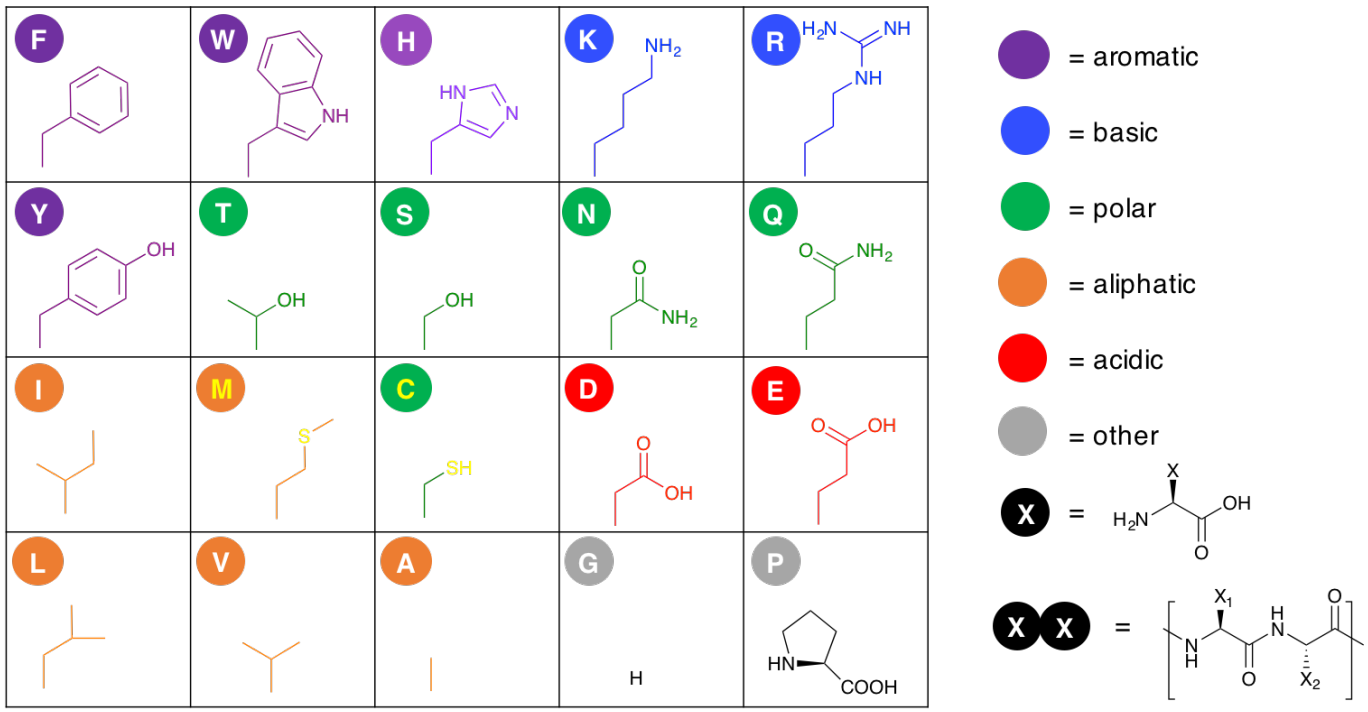

Fig. 2. Twenty gene encoded amino acid side chains color coded based on aromatic (purple), basic (blue), polar (green) aliphatic (orange), acidic (red) and other (grey).

protein length is dramatically reduced, led to the discovery of simple self-assembling peptides derived from amyloid proteins.[14, 15 16] These approaches gave rise to sequence-to-structure relationships for simple, engineerable peptide-based materials (discussed in section 2).

Alongside copying and simplifying natural designs, a well-established synthetic approach is to equip short peptides with ligands that aid their assembly. For example, there is now a significant body of work focused on peptide amphiphiles, short peptides that are modified with either aliphatic[17] or aromatic[18] ligands. This area has been extensively reviewed by us[18] and others[19-22] and here we discuss those examples where the focus was systematic, side-by-side study of peptide sequences in an attempt to derive sequence-to-structure correlations (section 2.2).

Alongside these approaches that are inspired or informed by naturally occurring self-assembling systems, or that involve chemical modification of short peptides to enhance their assembly propensity, there are open questions around which other amino acid patterns may give rise to structure and function. The combinatorial space, even for short peptides, is simply too vast to be fully explore experimentally. Consequently, there is a need for the development of generally usable, automated ways to map or search the entire peptide sequence space for new functions. While combinatorial peptide libraries, including those derived from phage display[23] or those that rely on chip-based screening[24] are suited to identify binders or substrates for catalytic reactions, these methods are not suited for materials discovery as peptides are not able to assemble and interactions are restricted due to surface/phage anchoring. Thus, as the promise of peptide-based nanostructures for a growing number of applications becomes increasingly clear, new methods are required to map and search the sequence space for function.

In such unbiased searching of the short peptide sequence space for supramolecular structures and functions, a key question is then how to identify those sequences, that possess the properties that are sought. Here, we discuss complementary experimental and computational approaches aimed at searching the short peptide sequence space for function, a set of approaches we collectively refer to as *directed sequence discovery* (Fig. 1) .

Directed discovery of peptide sequences differs from rational design in that no aspect of the specific solution to the problem, such as molecular configuration or chemical composition, is pre-determined at the outset. Rather, the desired properties of the final system are taken into account during the selection of *input* building blocks (subsets of short peptides or amino acids) based on their ability to deliver these properties (this constitutes the “directed” component of the process). The various potential systems that can result from these building blocks are then systematically searched and evaluated against the desired properties to determine if any matches are found (the “discovery” phase of the process).

Each of the three methods discussed provide different ways of searching the free energy landscape, and assessing peptide candidates for ‘fitness’ for a desired property. We will summarise what has been learned from different methodologies, focusing on short peptides (typically eight amino acids, or less) as well as short peptides functionalized with aromatic or aliphatic ligands. First, self-assembling peptides sets that include systematic sequence and compositional variation will be discussed, (Section 2), then computational directed search approaches (Section 3), and last,

the dynamic library approach to search for peptide nanostructures (Section 4). Finally, conclusions from the presented studies and future trends in directed discovery will be discussed.

We will focus on searching of the chemical space provided by the canonical (gene encoded) amino acids, which be denoted by single letter codes. These include aromatic, basic, polar, aliphatic, acidic, and other residues (Fig. 2) that, in combination, offer a tremendous variety of non-covalent interactions giving rise to functional supramolecular structures.

2. Systematic Sequence Editing of Self-Assembling Peptides

Due to the vast numbers of possible self-assembling peptides, experimental combinatorial screening approaches, *i.e.* synthesizing and verifying structures formed by large numbers of sequences, are realistic only when focused on sub-sets of the sequence space. Therefore, a central experimental approach to search for guiding principles or sequence/structure relationships in self-assembling peptides consists of studies of systematically chosen peptide sets, rather than exploring the entire available chemical space. Approaches include the (reductionist) truncation and sequence modification of amyloid polypeptide sequences, systematic editing of individual residues in a given sequence, or keeping composition the same but varying of amino acid order in short peptides and peptide conjugates (sequence isomers). The section is structured according to peptide length, starting with dipeptides.

2.1 Sequence variations of established self-assembling motifs

Dipeptides

The first systematic studies of sequence-dependent supramolecular organization of short peptides was focused on single crystal x-ray diffraction structures of dipeptide crystals, by Görbitz[16]. In the crystal structures of LL, LF, FL and FF, obtained by crystallisation from aqueous solutions, tube-like morphologies were observed, consisting of hydrophilic channels filled with water molecules, lined with H-bonding of the peptide backbone. While LL, LF and FL crystals formed small channels (2.5 X 6.0 Å for LL and LF, 4.0 X 6.0 Å for FL) enclosed by interactions between four peptide molecules, FF crystal contained much larger hydrophilic cavities with a diameter of 10 Å that are formed by six molecules.[16] Crystal structures of the hydrophobic dipeptides and sequence isomers AI, IA, VI, IV, VV,[25] VA and AV[26] were later reported by the same group. Nanotube pore size was found to be sequence-dependent, where the pore diameter is inversely correlated with the steric bulk (number of carbon atoms) of the amino acid side chains. Thus, AI and VV formed pores with larger diameters than VI and IV.[25] No significant change in the structural properties of the resulting crystals was found for dipeptide sequence isomers. More recently, Erdogan *et al*.[27] showed solvent-dependent formation of a wide range of nano- and microstructures by VA, including highly ordered β-sheet plate-like structures in a non- aqueous solvent, whereas the sequence isomer AV formed rod-like assemblies in the same solvent.[27]

By applying a reductionist approach, starting from β-amyloid, Gazit and Reches discovered that the minimal self-assembling motif, FF, formed well-ordered nanotubes (Fig. 3a-c).[14] The nanotubes are formed through H-bonding of the peptide backbone into β-sheets and aromatic stacking of the phenyl groups. This discovery is in agreement with the FF crystal structure reported by Görbitz (Fig. 3d-f), as both the nanotube and the crystal have the same XRD structure.[16, 28, 29] Moreover, CG simulations by us and others[30] showed over-time formation of nanotubes with similar dihedral angles to those found experimentally[31] (Fig. 3g-j) (see section 3.1).

Sequence variation of FF, to other aromatic dipeptides FW, WF, WY and WW were studied under the same conditions,[14] and only FW was shown to form nanotubes. Recently, Zhang and co-workers investigated the aqueous assembly of both FW and WF and reported that WF can form spherical nanoparticles, which show a remarkable shift in fluorescence emission to the visible range, but without clear evidence of an organized H-bonded structure, suggesting aromatic stacking interactions is a main driver in assembly of these structures.[32] No clear differences in morphology, structure, and fluorescence of particles formed by FW were observed. In addition, Zhang confirmed that WW failed to form ordered structures as was previously reported by Gazit[14].

The discovery that aromatic dipeptides form discrete nanostructures subsequently lead to dipeptide designs that combine non-aromatic and aromatic amino acids. Ventura and co-workers reported that the dipeptide analogue IF assembled into a network of nanofibrils forming a hydrogel while the dipeptide VF did not form structures,[33] showing that minor modifications in polar/nonpolar balance, such as a single methylene group, have substantial impact on self-assembly. Overall, these experimental studies on dipeptide sequence variants[25, 32, 33] (confirmed by our computational CG mapping[31] as shown in section 3.1) strongly suggest that the position of amino acids within a dipeptide has little effect on self-assembly propensity in aqueous solutions. An exception of this observation was shown for self-assembly in non-aqueous solvents.[27]

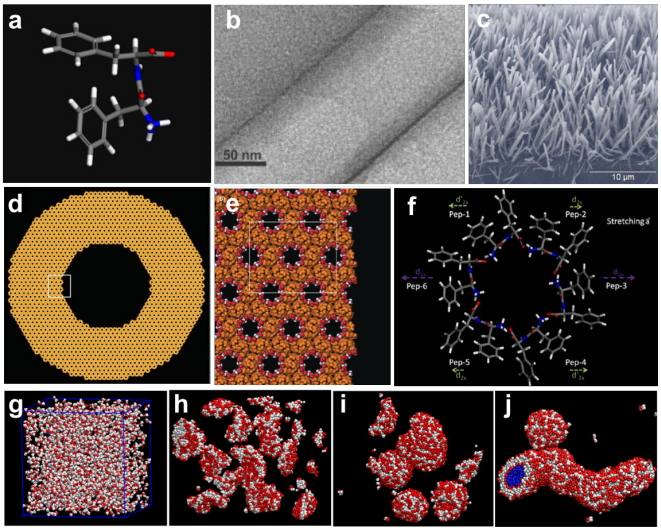

Fig. 3. Structural models of FF nanotubes. **a**. Model of hollow FF nanotube with a 110 nm outer diameter and a 50 nm inner diameter. **b**. Peptide-channel interface. Main chain atoms are colored according to atom type, F side chains are in orange. [1] **c**. Structural model reported by Azuri *et al*.,[4] showing multiple unit cells (a single one in white frame) on *xy* plane based on DFT calculation and Görbitz crystal structure. **d-g**. MD simulations of FF nanotube formation over time: (d) 0 µs; random monomers. The periodic box (blue lines) is indicated. (e) 0.2 µs; sheet-like aggregates formation. (f) 0.5 µs; sheet folding into vesicles. (g) 1.5 µs; fused vesicles forming a hollow tube where the end of the tube is cut off to show water beads inside (blue). Backbone beads in red, side chain beads in white. Water beads are omitted.[6]

Tripeptides

Following the discovery of FF, a logical approach was to systematically append additional amino acids to introduce functionality and assess assembly propensity. Thus, systematic sequence editing led to the identification of a number of self-assembling tripeptides: CFF forms nanospheres[34]; FFF forms fibrous and plate-like assemblies by β-sheets structures[35]; VYV assembles into micelles,[36] and KFG shows a concentration-dependent transition from micellar structures to β-sheet-like fibers.[37] Rapaport and co-workers reported on a series of self-assembling FXF tripeptides.[38] FEF assembled into elongated ribbons at acidic pH by antiparallel β-sheet packing and π-π stacking of the phenyl groups and formed a self-supporting hydrogel in hexafluoroisopropanol (HFIP) diluted in water. The peptides FTF, FCF and Ac-FEF-NH_2 formed hydrogels at higher concentrations than that of FEF (4% compared to 0.1%, respectively) and FKF formed a gel in 0.1 M KCl following thermal annealing and cooling. The reported assembly of these peptides occurs at acidic pH, likely requiring the C termini to be protonated. The non-aromatic VEV did not form hydrogels or β-sheet structures, emphasizing the important role of aromatic stacking in stabilizing the β-sheet structures.

Marchesan and co-workers showed that changing the chirality from *L* to *D* of the first amino acid within VFF and FFV giving rise to β-sheet-like nanotapes and twisted fibers, respectively in peptides that do not assemble when all *L* configuration.[39] A dramatic change in supramolecular structures was observed upon systematically changing the chirality of each of the amino acids within FFV.[40] This study showed that the chirality of the first amino acid is key for self-assembly, as only tripeptides in which the N-terminal amino acid had the opposite chirality than that of the two other amino acids (^{D}FFV and F^{D}F^{D}V) form networks of extended anti-parallel β-sheet fibers.

In a recent extensive comparative study, Hauser and co-workers studied 54 selected tripeptides derived from systematic variations of seven sequences[41] found by a computational screen of N-acetylated tripeptides.[42] Although clear sequence design rules could not be established based on this work,[41] it was shown that substitutions of single amino acids in specific positions dramatically changed the self-assembly behaviour, whereas other positions were less susceptible to such changes. For example, L1 but not L2 within the peptide Ac-LLE was susceptible to substitution to I, and self-assembly of the peptide Ac-YYD[42] could be 'turned off' or on (to form fibrils) by substituting Y1 or Y2 to F, respectively.

Following computational mapping of the tripeptide sequence space to search for self-assembling peptides (section 3.1),[43] we selected a number of sequences with high predicted assembly scores, but that also include polar components to favour interactions with water. This was done in an effort to produce supramolecular hydrogels, which require a balance between high assembly propensity and favourable solvent interactions.[44] We found that KYF, KYY, KFF and KYW assembled into translucent gels, with strong intermolecular H-bonding observed for KYF, KFF and KYY using FTIR upon gelation.[43] These findings exemplify three guiding principles for self-assembly: aqueous self-assembly is favoured by trripeptides containing (i) *two aromatics*; further enhanced when these aromatic residues are direct neighbours, so (ii) *paired aromatics* and the inclusion of charged residues at the termini enhances self-assembly propensity when paired with the same-charged terminus (so lysine at the N-terminus, asparatic acid at the C-terminus, leading to (iii) *paired charges*. Table 1 summarizes the literature results obtained for tri- (and some tetra-) peptides with guiding principles illustrated (Figure 5).

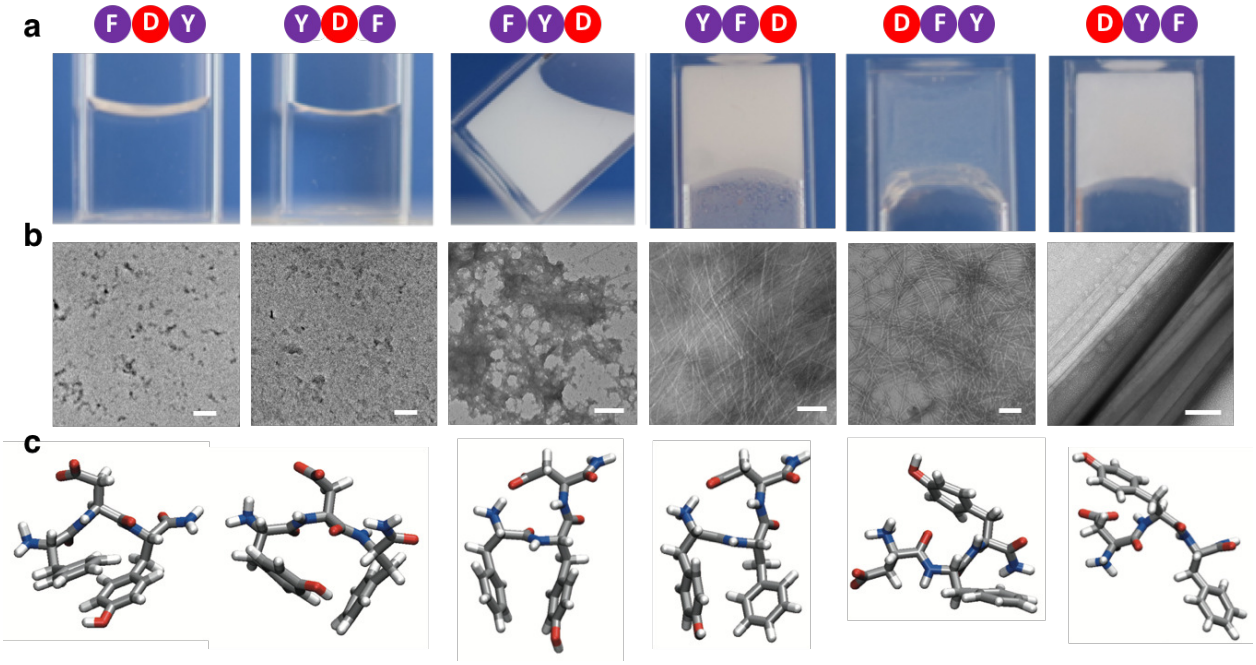

Fig. 4. Differential assembly of tripeptide sequence isomers. **A**. Macroscopic images of the materials formed by self-assembly of sequence isomers containing F (dark purple), Y (light purple) and D (red). (20 mM in phosphate buffer, pH 8). **B**. TEM micrographs of nanostructures formed by self-assembly of the six isomers. Scale bars are 100 nm. **C**. Preferred conformation of each peptide obtained by atomistic molecular dynamics simulations.

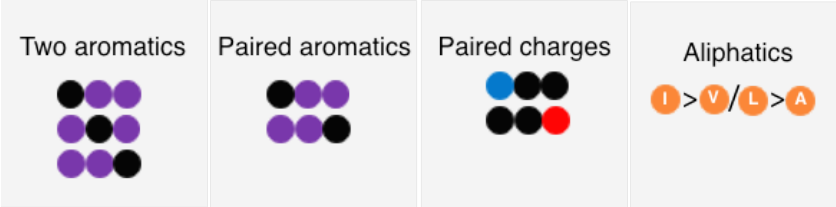

Fig. 5. Guiding principles for short self-assembling peptides.

Peptide Sequence	Nanostructure	Concentration	Conditions	Guiding Principle	Reference
FFF	Planar β-sheets structures	2 mg mL^{-1}	fluorinated alcohol/water	Paired aromatics	35
PFF	Crystalline fibres	12.3 mg mL^{-1}	Water	Paired aromatics	43
CFF	Nanospheres	2 mg mL^{-1}	HFIP/water	Paired aromatics	17
FFD	Nanofibres	12.8 mg mL^{-1}	PBS, pH 7.4	Paired aromatics; Paired charge	128
FFD	Bilayers	17 mg mL^{-1}	Oil/water	Paired aromatics; Paired charge	45
DFF	Aggregates	12.8 mg mL^{-1}	PBS, pH 7.4	Paired aromatics	128
DFF	Bilayers	17 mg mL^{-1}	Oil/water	Paired aromatics	45
FDF	Bilayers	N.A.	Computationally	Two aromatics	128
FEF	Ribbons	2.2 mg mL^{-1}	Aqueous, pH 5.3	Two aromatics	38
FTF	Gel	40 mg mL^{-1}	HFIP/water pH 3-5	Two aromatics	38
FCF	Gel	40 mg mL^{-1}	HFIP/water pH 3-5	Two aromatics	38
VEV	No self-assembly	0.7 mg mL^{-1}	HFIP/water	Non	38
VYV	Micelles	>0.015 mg mL^{-1}	Aqueous, pH N.A	Non	36
VFF, FFV	No self-assembly	6.6 mg mL^{-1}	Aqueous, pH switch from 12 to 7.4	Paired aromatics	39
^{D}VFF, ^{D}FFV	Gels, Nanotapes and twisted fibres	6.6 mg mL^{-1}	Aqueous, pH switch from 12 to 7.4	Paired aromatics	39
FYD-NH_2,	Aggregates	8.8 mg mL^{-1}	PB, pH 8	Paired aromatics	46
YFD-NH_2	Fibres	8.8 mg mL^{-1}	PB, pH 8	Paired aromatics	46
Ac-YYD	Fibres	40 mg mL^{-1}	Water	Paired aromatics; Paired charges	42
Ac-YFD	Fibres, gel	15 mg mL^{-1}	Water	Paired aromatics; Paired charges	41
Ac-FYD	No self-assembly	Not reported	Water	Paired aromatics; Paired charges	41
DFY-NH_2	Fibres gel	8.8 mg mL^{-1}	PB, pH 8	Paired aromatics	46
DYF-NH_2	Crystals gel	8.8 mg mL^{-1}	PB, pH 8	Paired aromatics	46
YDF-NH_2, FDY-NH_2	No self-assembly	8.8 mg mL^{-1}	PB, pH 8	Two aromatics	46
KYF, KFF, KYW	Fibrous nanostructures	17 mg mL^{-1}	Oil/water	Paired aromatics; Paired charges	45
KYF, KFF, KYW	Nanofibres gel	13-14 mg mL^{-1}	Water	Paired aromatics; Paired charges	43
KFD	Amorphous	12.2 mg mL^{-1}	Water	Paired charges	43
KFG	'Random coil' spheres	0.5 mg mL^{-1}	Aqueous, pH 7.4	Paired charges	37
KFG	β-sheet-like fibres	5 mg mL^{-1}	Aqueous, pH 7.4	Paired charges	37
KLL	Fibrous structures	11.2 mg mL^{-1}	Water	Paired charges	43
FYK	Aggregates	13.7 mg mL^{-1}	Water	Paired aromatics	43
KFFE	β-sheets fibrils	0.17 mg mL^{-1}	PB pH 7 or water	Paired aromatics; Paired charges	51
FEFK	No self-assembly	40 mg mL^{-1}	Water	Two aromatics	52
KVVE	β-sheets fibrils	0.14 mg mL^{-1}	PB pH 7 or water	Paired charges	51
KLLE, KAAE	Aggregates	0.14-0.15 mg mL^{-1}	PB pH 7	Paired charges	51
^{1}Ac-ILE	Nanofibres gel	20 mg mL^{-1}	Water	Paired charges	41
^{2}Ac-LLE	Bead-like structures	5 mg mL^{-1}	Water	Paired charges	42
Ac-LIE	Crystalline structures	5 mg mL^{-1}	Water	Paired charges	41

Table 1. Systematic sets of short self-assembling peptides and corresponding guiding principle. ^{1}Beads and crystalline structures were observed for Ac-ILE alongside fibres. ^{2}Fibres and crystalline structures were observed for Ac-LLE at 25 mg mL^{-1} and 30-40 mg mL^{-1}, respectively.

Self-assembly, but not gelation was demonstrated by PFF, which has high assembly propensity but lacks a polar component to favourably interact with solvent, resulting in crystalline structures with a strong indication of β-sheet. Likewise, KLL, assembled into fibrillar structures but did not form a gel, suggesting that paired aromatics provide a more favourable polar/assembly ratio to favour gelation.[43]

Examples of tripeptides that do not form assembled structures are useful to further confirm sequence to structure relationships: FYK, which positions the lysine at the C-terminus, instead of the N-terminus, does not form ordered nanostructures, suggesting that *paired charges* enhance self-assembly propensity, while unpaired charges reduce it. In addition, KFD, which has *paired charges* but only a single aromatic, formed disordered aggregates with no indication for intermolecular H-bonding. Thus, these findings demonstrate the critical combination of the guiding principles for self-assembly of short peptides.

Due to the amphiphilic nature of some of these tripeptides, they were explored as emulsifiers. We identified a set of tripeptides with remarkably different interfacial self-assembly behaviour, depending on their sequence.[45] The selected tripeptides form fibrous nanostructures (KYF, KFF, KYW)[43] or bilayer-like structures (DFF, FFD) in water.[45] These peptides showed two different emulsification behaviours for the K and D containing peptides: KYF, KFF and KYW stabilized oil droplets by formation of nanofiber networks, whereas DFF and FFD were induced to self-assemble in oil and act according to a more traditional surfactant model. These behaviours were in agreement with those predicted by coarse grained simulations. Emulsification strength for all five peptides was found in the following order: KYF>KYW>KFF>DFF/FFD, with K at the first position giving rise to more stable emulsifiers and F at the second position resulting in weaker emulsifiers and no clear difference in emulsification for the D-containing tripeptides.

Tripeptides sequence isomers

Sequence isomerization of a tripeptide was recently demonstrated to enable degrees of supramolecular order/disorder to be systematically controlled.[46] Based on the observation that paired aromatics enhance self-assembly in tripeptides,[43] we selected a tripeptide containing the two aromatic amino acids Y and F and the acidic amino acid D to promote ionic interactions. Terminal amides were used to promote self-assembly at pH 8, note that this means that 'paired charges' are not possible here, as the C-terminus is not charged. The sequence isomers displayed remarkably different self-assembly behaviours (Fig. 4), in overall agreement with these proposed guiding rules.[43] Thus, no assembly was observed when aromatics were unpaired, for FDY and YDF; FYD and YFD formed amorphous aggregates and opaque nanofiber gel, respectively; and DFY and DYF assembled into translucent nanofiber gel and crystalline fibres, respectively (Fig. 4a-b). The position of the paired aromatics near the C-terminus favouring aggregation and assembly is also found in computational assessment of the sequence space, see section 3.1. The variable supramolecular order could in this case be exploited- as it strongly dictated the positioning and chemical environment of tyrosine residues, which could in turn dictate the outcome of a subsequent enzymatic oxidation process. The structures formed present Y residues in varying steric and chemical contexts, which could be exploited as supramolecular templates to present tyrosine residues in different orientations and environments for enzymatic oxidation by tyrosinase, giving rise to melanin-like polymeric pigments that have sequence-encoded properties.

Tetrapeptides

It has been shown that, in general, peptides with even numbers of amino acids are more favourable for β-sheet structures formation compared to those with odd numbers,[47] due to the equal number of H-bond donors and acceptors on both directions of the molecule that facilitate unidirectional growth. A number of systematic studies focusing on tetrapeptides were reported.

Lu and co-workers showed that substituting L with I promoted formation of β-sheets within a series of non-aromatic, acetylated tetrapeptides. While Ac-L_3K assembled into nanospheres, the dominant interaction observed for Ac-I_3K, which assembled into nanofibers was H-bonding between the peptide backbones into antiparallel β-sheets.[48] These findings correlate with other studies related to shorter peptides reviewed above[33] showing that isoleucine strongly promotes β-sheets structure formation. Lu and co-workers continued studying the effect of varying numbers and composition of hydrophobic amino acids within amphiphilic peptides based on Zhang's original peptide surfactant designs.[49] They showed clear transitions from ordered β-sheet structures to disordered structures upon substitution of I for L to (acetylated and amidated) peptide sequences. Accordingly, I_3K formed β-sheet structure, LI_2K formed a mixture of β-sheet and disordered structures and L_3K formed predominantly disordered assemblies.[50] However, increasing the number of L from 3 (L_3K) to 5 (L_5K) promoted transition from disordered structures to β-sheets. Increasing the number of I within I_xK peptides reduced the diameter of the resulting fibres.

Tjernberg *et al.* reported that substituting L with V promotes self-interactions, as KFFE and KVVE, but not the L-containing counterpart KLLE and KAAE, form amyloid-like fibrils that showed a typical Congo red birefringence,[51] indicating the importance of aromatic stacking to fibril formation as well as the balance of hydrophobic/steric and polar residues. The tetrapeptides KFFK and EFFE formed similar fibrils, but only when co-assembled in equimolar concentration, owing to E/K electrostatic interactions. These findings demonstrate the role of *paired aromatics* in short self-assembling peptides.

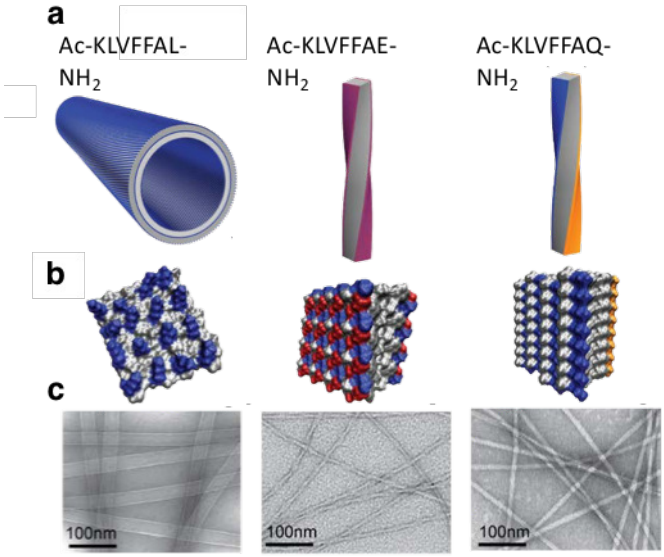

Fig. 6. ## needs copyright dealt with## Sequence variants of β-amyloid core peptide KLVFFAL. **a**. Graphical cartoons, models (**b**) and TEM micrographs (**c**) of Ac-KLVFFAL-NH_2 and E7 and Q7

Similarly, Saiani and Miller[52] showed that tetrapeptide FEFK, which

was used as a precursor for an enzymatic assembly experiment, in which the aromatics are unpaired, fail to self-assemble. Yet, the octapeptide variant of this sequence formed by enzymatic condensation (discussed in details in Section 4) assemble into a fibrous gel, suggesting that alternating nonpolar/polar amino acids pattern promote self-assembly in longer peptides (>4 amino acids long).

Penta- to octapeptides.

Several groups, including Kapurniotu,[53] Tycko,[54] Lynn,[55] and Gazit,[14, 56, 57] systematically investigated sequence/structure relationships in minimal self-assembling penta- to octapeptide fragments of amyloid polypeptides. By truncating the calcitonin-derived fragment DFNKF, Gazit and Reches reported that the pentapeptide was the minimal sequence forming ordered structures.[57] Kapurniotu identified FGAIL and NFGAIL as minimal assembling sequences within Islet Amyloid Polypeptide.[53] Using systematic alanine substitution of all residues in NFGAIL,[56] Gazit showed that substitution of F completely inhibited self-assembly, while alanine substitution of I or L reduced the kinetics, alanine substitution of N or G increased the kinetics of assembly. These patterns demonstrate the contribution of aliphatic and aromatic amino acids to peptide self-assembly into extended fibrillar structures.

Hauser and co-workers identified a series of systematically selected tri- to hexapeptides that self-assemble into amyloid-like fibrils with an intermediate α-helical structures.[58] These hexamers had sequences designed to overall have a steric cone like shape and gradient of decreasing hydrophobicity (and consequent steric bulk) from aliphatic amino acids at the (acetylated) N-terminus to polar amino acids at the C terminus. Systematic alanine substitution of the first four aliphatic amino acids within the peptide LIVAGD showed that the decreasing hydrophobicity gradient promoted self-assembly. Moreover, a sequence-dependent gelation strength was reported in the following pattern (from strong to weak gels): acidic (D and E)>neutral (S and T)>basic (K) amino acid at the C-terminus of the hexamers, in agreement with *paired charges* guiding principle, which propose that C-terminal D or E promote self-assembly.

Very recently, David Lynn's group compared the structure, but also the consequent catalytic activity of sequence variants of the β-amyloid core motif LVFF, focusing on the ability to vary sequence to control both substrate binding and reactivity in a model retro-aldol reaction.[59] The peptide Ac-KLVFFAL-NH_2 assembled into antiparallel out-of-register β-sheets which further stack to form nanotubes, where the (acetylated) N-terminal amino acid is positioned outside the H-bonded array. Solid state NMR analysis and MD simulations showed that half of the K residues are exposed to the nanotube surface (Fig. 6a-b). Substituting K1 with R did not change the size or morphology of the resulting nanotubes, however, it completely inhibited the retro-aldol catalytic activity of the K-containing peptide. Substituting L7 with E or Q resulted in formation of shallower cross-β grooves (Fig. 6a-b), where the former is due to salt bridges formed by the N-terminal K and C-terminal E in an antiparallel β-sheets arrangement, that limits substrate binding and showed no catalytic activity. Thus, the structural changes caused by single amino acid substitutions alter substrate binding and catalytic activity.

Saiani and co-workers demonstrated that replacing phenylalanine with alanine in a periodically nonpolar-polar octapeptide changed the conformation from β-sheets (by the peptides FEFEFKFK and FEFKFEFK) to α-helix (by the peptide AEAEAKAK).[13] Furthermore, the position of amino acids within the octapeptide was critical for self-assembly, as no assembly was observed for the sequence AEAKAEAK, while AEAEAKAK assembled into thick, rigid fibres.

2.2 Molecular sets of short self-assembling peptide conjugates

Here, we will focus on systematic assessment of sequence variations of sets of peptides conjugated to a synthetic group, typically an aliphatic or aromatic residue that enhances the self-assembly propensity, with an emphasis on the sequence variants rather than on the synthetic modifications used.

The interactions between π-electrons in the aromatic fluorenyl rings of the Fmoc group allows the use of short peptide building blocks as hydrogelators by N-terminus conjugation to the peptide sequence.[60] Since the discovery of the hydrogelator Fmoc-LD by Vegners,[61] Xu and co-workers reported on a series of Fmoc-dipeptides, including peptides containing combinations of A, G, S and T forming hydrogels that respond to ligand-receptor interaction.[62, 63] Hydrogels forming at physiological conditions by Fmoc-FF self-assembly were later reported by Gazit and co-workers[64] and simultaneously by our group.[65] There is now a substantial body of work on the self-assembly of aromatic peptide amphiphiles, as recently reviewed by us[18] and Gazit.[20] Here, we discuss examples that, using molecular sets of Fmoc peptides, provide evidence for the links between molecular structure and self-assembly behaviour.

Due to the well-documented, route-dependent nature of the resulting self-assembled structures,[66] a number of methods have been developed to control the kinetics of the self-assembly process, for example using gradual change in pH by using unstable esters such as gluconolactone[67] or enzymatic self-assembly processes.[68] These methods provide a convenient way to control the assembly process and produce materials that are directly comparable. For example, Adams and co-workers demonstrated pH-triggered self-assembly of Fmoc-dipeptides by hydrolysis of glucono-δ-lactone (GdL) to gluconic acid, which results in controlled pH decrease.[67]

Using enzymatic methods, two routes have been followed: (i) enzymatic hydrolysis of a blocking group that prevents precursor self-assembly, for example, we have previously reported on a series of Fmoc-dipeptides that are driven to self-assemble following enzymatic hydrolysis of a terminal methyl ester[69, 70] or phosphate ester[71] and (ii) enzymatic condensation of two non-assembling precursors into a self-assembling peptide. The advantage of the latter method is that it was found to be fully reversible[72] and therefore ensures that structures formed represent equilibrium structures. The former method can give rise to kinetic trapped structures, that are dictated by the catalyst concentration[69] so it is important to ensure that identical enzyme concentrations are used in comparative studies. In this context, thermal annealing after the enzymatic process can ensure that results are reproducible and reflect thermodynamic stabilization, rather than kinetic aspects.

Systematic sequence variations of Fmoc dipeptide-methyl esters triggered to assemble by enzymatic hydrolysis of the OMe group (via *route i*) provided insights into the relationship between sequence composition, self-assembly and materials properties.[70] Side chain length of amino acids at the second position of the peptide Fmoc-YX was found to be critical for the molecular packing.[70] Thus, T, S, and N at this position showed formation of fibrillar structures by typical β-sheet-like packing stabilized by π-stacking of the aromatic moieties (π-β assemblies), whereas Q inhibited this packing, giving rise to formation of spherical structures, presumably for steric reasons, given the only difference between Q and N is a methylene unit. We utilized the differential self-assembly of these dipeptide sequences

further to develop antimicrobial peptide nanostructures.[71] The self-assembly of phosphorylated precursors (Fmoc-Y*p*X-OH) was triggered upon removal of the phosphate (Fmoc-YX-OH) by the enzyme alkaline phosphatase, that is abundant in *E. coli* periplasmic space. As observed previously, the dephosphorylated Fmoc peptide YT, YS and YN assembled into β-sheet fibres stabilized by π-stacking interactions while YQ formed non-β-sheet spherical structures.

Sequence-structure links were found for four Fmoc dipeptide-methyl ester sequence variants formed by enzymatic condensation (via *route ii*). Addition of a side chain chiral centre at the first position by substituting S with T shifted structures morphology from planar to twisted ribbons.[73] Not surprisingly, F at the second position (Fmoc-XF-OMe) resulted in formation of thermodynamically stable structures due to interactions of the F phenyl moiety with Fmoc columns, as expressed by higher conversion yields of Fmoc-TF-OMe and Fmoc-SF-OMe and by redshifted fluorescence at 365 nm. In addition, MD simulations using a sheet-like model in a π-β bilayer configuration showed that SF and TF were more stable than SL and TL, yet, slight deviations in the β-sheet spacing, caused by the bulky phenyl group at position 2, were found by WAXS analysis.[73]

An additional study utilizing *in situ* condensation (*route ii*),[74] showed that F at the second position of Fmoc dipeptides promotes assemblies' thermodynamic stability. Both experimental and computational methods were used to study the contribution of hydrophobicity *vs.* aromaticity and C-terminal modifications to the formation of four closely related self-assembling peptides by condensation of Fmoc-T with F or L protected by either amide or methyl ester (Fmoc-TF-NH_2, Fmoc-TF-OMe, Fmoc-TL-NH_2, Fmoc-TL-OMe).[74] Interestingly, all peptides assemble into twisted fibres, with no clear difference in morphology, however, F-containing peptides formed stronger π-stacking interactions than the L-counterparts, either between the fluorenyl groups or between the fluorenyl and the phenyl groups, as discussed in section 3.2.3.

Chemical composition and sequence of aliphatic peptide amphiphiles was shown to dictate hydrogen bond alignment along fibres' extended axis and consequently influence the mechanical stiffness of the resulting materials. Pushuck *et al.* varied the number and position of V and A residues within a peptide sequence conjugated to an alkyl tail at the N-terminus and to three D residues at the C-terminus.[75] The presence of valines promoted the alignment of β-sheets to planar structures with less twists, while alanines facilitated formation of twisted β-sheets structures. These molecular arrangements correlated with the increase (by valines) or decrease (by alanines) of the formed gels stiffness.[75]

Sequence isomerization of peptide conjugates

A number of studies made use of sequence isomerization to get sequence-structure insights of peptide conjugates self-assembly. By using sequence isomers of Fmoc peptides, Hamley showed that exchanging the positions of V and K, where K is protected by N^{ε}-tert-butyloxycarbonyl (Boc), substantially change the resulting nanostructures and hydrogels properties.[76] Fmoc-KLV assembled into branched fibrils that formed stiffer gels than that formed by Fmoc-VLK. The authors suggested that when K is at the first position, cation-π interactions between Fmoc and K cause bending of the β-sheets, resulting in branched fibrils.[76]

With Saiani, we studied the self-assembly of Fmoc-GG, Fmoc-GF and Fmoc-FG and compared it with Fmoc-FF.[77] Substituting F with G decreased formation of extended β-sheet structures. Position G next to the Fmoc moiety enhanced this effect and was found critical for self-assembly, as the peptides Fmoc-GG and Fmoc-GF assembled only in their protonated form and completely arrested formation of β-sheet structures, whereas Fmoc-FG formed weak β-sheet structures at high pH. These findings indicate that the main interactions driving the formation of the fibrillar structures by these peptides are hydrophobic interactions and π-π stacking between the fluorenyl moieties with minor role for H-bonding. Interestingly, these results also suggest that the *paired aromatics* guiding rule as positioning F next to the aromatic fluorenyl moiety increased the self-assembly propensity of the peptide.

Escuder and co-workers reported on sequence isomers of benzyloxycarbonyl (Z) protected tetrapeptides containing F and D.[78] All six isomers formed gels with networks of β-sheet nanofibres by pH tuning using glucono-δ-lactone hydrolysis. No clear preference was observed for peptides with FF blocks (*paired aromatics*) compared to alternations of F and D (nonpolar/polar pattern). Yet, five out of the six sequence isomers contained paired aromatic moieties, either by a paired position of F within the peptide (Z-FFDD, Z-DDFF, Z-DFFD) or by a paired position of benzyloxycarbonyl and F (Z-FDFD, Z-FDDF). Accordingly, the peptide ZDFDF, with no paired aromatic moieties, showed weaker self-assembly propensity. In addition, peptides with *paired charges*, where the D positioned at the C-terminal, had a lower critical gel concentration and displayed increased ThT binding.

Sequence isomerisation in aliphatic peptide amphiphiles was described by Stupp and co-workers, who explored the effect of sequence isomerization on the self-assembly of tetrapeptide amphiphiles containing two valines and two glutamic acid residues conjugated to an alkyl tail at the N-terminus.[79] Four out of the six possible combinations of the tetrapeptide were examined. While all peptides exhibited a β-sheet packing, peptides with alternating nonpolar/polar sequence (VEVE and EVEV) assembled into flat nanobelt and nanoribbon structures whereas peptides with paired adjacent V and E (VVEE and EEVV) formed cylindrical nanostructures. The authors suggested that peptides with a binary hydrophobic/hydrophilic motif dimerize with no interfacial curvatures between peptide segments and alkyl tails, while changing this motif resulting in cylindrical nanofibers formation by hydrophobic collapse of the alkyl tails and intermolecular hydrogen bonding. In addition, positioning the bulkier charged glutamic acid adjacent to the alkyl tail resulted in formation of flexible nanostructures where the alkyl tails are loosely packed, presumably due to a steric effect and electrostatic repulsion among E side chains.

Summary: guiding principles from sequence editing

In summary, these studies demonstrate that a remarkable range of structure and functionality can be obtained in relatively simple short self-assembling peptides by simply altering composition or order of amino acids present. More importantly, systematic studies of sequence variants and characterization of resulting nanostructures gave rise to sequence/structure design rules or guiding principles.

A number of important conclusions can be drawn from the studies reviewed in this section (summarized in Fig. 5): (i) The majority of the reported self-assembling sequences demonstrate the three key guiding principles, *Two aromatics*, *paired aromatics* and *paired charges*, indicating that these principles can be further applied in search for additional self-assembling peptides. When comparing the *paired aromatics* principle with the alternating nonpolar/polar amino acids pattern that is well used in self-assembling peptides, we observed that the former is essential for the self-assembly of short peptides (<8 amino acids) while the latter for octapeptides or longer sequences. (ii) A distinctive trend in the self-assembly propensity of

different hydrophobic amino acids emerges from the reviewed papers in the following order: I>V/L>A. (iii) The order of amino acids within the peptide sequence has a prominent role in the self-assembly propensity of short peptide amphiphiles, while there is no such trend in unfunctionalized dipeptides. Finally, it is clear that studies in which complete sets of sequences and sequence isomers were used provide more constructive and useful information, enabling sequence-to-structure links to be established, thus facilitating the search for the next generation of nanostructures.

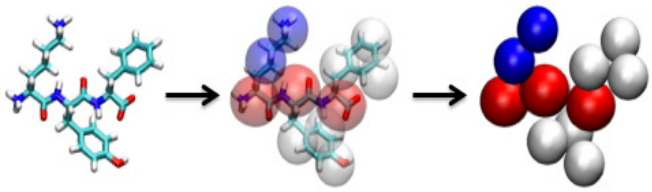

Fig. 7. Coarse-graining of the atomistic structure of the KYF tripeptids into the Martini bead representation.

3. Computational Approaches to Directed Discovery

Both computational and experimental methods can contribute to the *directed discovery* process – in this section the application of computational methods to directed discovery is discussed. The "directed" aspect of directed discovery requires that the properties of the building blocks – the individual units that form the nanostructure through non-covalent interactions – be understood. The properties in question depend on the requirements of the final system and can include, but are not limited to, the ability to form strong intermolecular interactions, the acidic or basic nature, the excitation energy, *etc*.[4, 80, 81] While some such properties (such as basicity and acidity) may be inherent to a given functional group, properties such as the HOMO-LUMO gap or intermolecular binding energy, can vary significantly based on small modifications to the system and as such it can be beneficial to evaluate the resulting structure, formed through self-assembly, for the desired property. For peptide-based systems, this has typically been achieved through the use of all-atom force field molecular dynamic simulations that study systems with known properties.[35, 73, 82-102] These systems can then be modified, for example through the systematic substitution of an amino acid in the peptide sequence, analogous to the experimental approaches discussed in the previous section, to determine how these modifications affect the structure and properties of the resulting nanostructure.[73, 103]

The alternative approach to systematic variation of a given sequence is the screening of potential combinations of building blocks to determine whether the resulting systems have the desired properties.[31, 43, 45] While this process can, in principle, result in an outright prediction of the "best" system, the nature of these predictions typically results in a small selection of systems that are predicted to be of interest for further experimental testing. Any variation between the experimental results and the computational predictions can be fed back in to refine the screening process for subsequent iterations.

3.1 Unbiased Library Searching with Coarse Grain Methods

The ability to efficiently sample the potential energy surface with coarse grain (CG) methods has led to their rapid uptake for the exploration of self-assembly processes. Coarse grained force fields typically employ "beads" to represent a collection of heavy atoms as an interaction site (Figure 7).[104-112] The loss of atomic resolution has two main advantages: (1) the larger bead masses and fewer number of particles allows for larger systems and longer timescales to be investigated and (2) the simplified interaction potentials creates a smoother potential energy surface, which can be more readily explored.[104, 113, 114] These differences in the sizes of the bead and the reduced degrees of freedom, relative to the atomistic representation, result in an increase in the speed of calculations by *ca.* three orders of magnitude. Practically, this implies that systems with several hundred short peptides in fully solvated environments can be readily equilibrated over 100's ns - µs timescales.[84] For example, a system with 300 dipeptides in a large solvent box can be simulated for 100 ns within 30 minutes on a single 12-core compute node (an equivalent calculation of a fully atomistic system would require *ca*. 1 month).[31] The disadvantages associated with CG methods are in the loss of detailed information on specific interactions that contribute to the ability of the building blocks to self-assemble and as such interpretation of the resulting structures is more ambiguous.

Despite the loss of specific interactions, CG force fields have proven to be remarkably accurate at predicting the self-assembly ability of short peptides. In 2011, the complete sequence space of the 400 dipeptides formed from the 20 canonical amino acids was screened for the ability to self-assemble.[31] This study revealed that the Martini force field was capable of discriminating between those systems that were known to self-assemble and those that did not. It is immediately clear from this figure that there is symmetry, which is

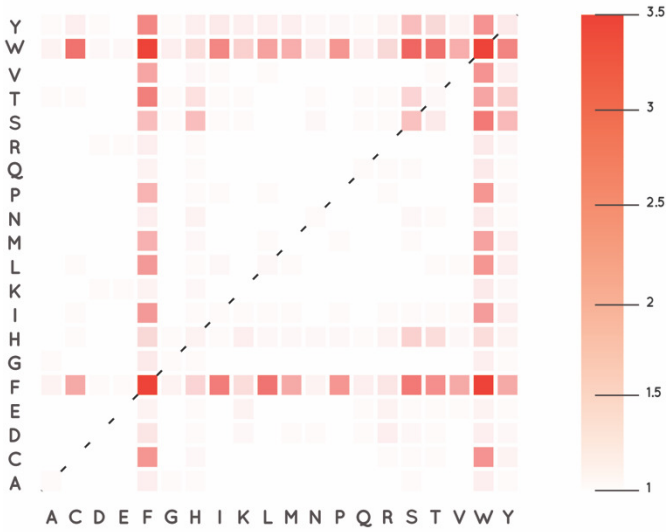

Fig. 8. Aggregation Propensity (AP) map for the 400 gene-encoded dipeptides. The N-terminus amino acids is given along the X-axis, while the C-terminus is along the Y-axis. Darker regions on the map indicate a higher propensity for the dipeptide to aggregate.

in agreement with experimental observations discussed in the previous section - in dipeptides the order of the amino acids has limited impact on self-assembly. In addition it is clear that paired aromatics favour assembly, FF and FW were both known to aggregate to form nanotubes (dark red squares in Figure 3) whereas FE and FK do not result in observable nanostructures (light squares in Figure 3).[31] In agreement with Zhang's observation of similar assembly behaviour of FW and WF, we found the same aggregation propensity (3.5) for these peptides in CG simulations.[31] Furthermore, liquid and solid interfaces;[45, 118] and the mechanism of assembly for a range of different peptide-based systems.[119-122] However, much of this work focussed on describing effects in systems that were already known to self-assemble. In 2015, the Martini force field was again employed to screen a series of peptides (the 8,000 gene-encoded tripeptides) for their ability to self-assemble into amphiphilic nanofibers that could result in a network capable of forming a self-supporting hydrogel.[43] This work resulted in the first four known examples of uncapped tripeptides that were able to form hydrogels

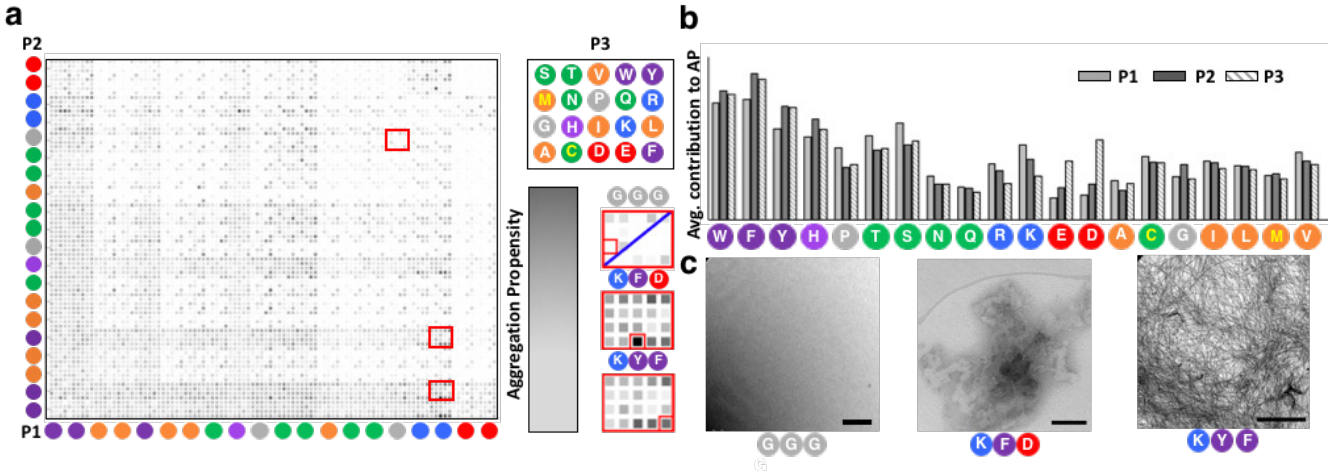

Fig. 9. Positioning of the amino acids in the tripeptide sequence that leads to the greatest chance of the peptide to aggregate. For example, W, F, and Y have a greater contribution to aggregation when positioned in the middle of the sequence, or at the C-terminus.

the experimental observation of nanoscale fibers for IF, with no assembly observed for VF discussed above, agreed with predictions, with CG simulations giving higher aggregation propensity score for IF (2.3) than for VF (1.8).[31]

In addition to providing a method to determine the aggregation propensity of the dipeptides, coarse-grain simulations have also been shown to capable of revealing the self-assembly pathway, when extended to longer simulations with larger systems of the FF dipeptide.[31] These simulations shed light on a likely mechanism by which FF organizes into nanotubes: first, randomly distributed monomers assemble into sheet-like aggregates that are then form hollow vesicles which fuse together into extended tubes (Fig. 3g-j). This study was followed by further work on the FF dipeptide that revealed the concentration dependant behaviour of the dipeptide and the ability to accurately model the appearance of different nanostructure morphologies as a consequence of the dipeptide concentration.[30] Thus, despite concerns over the accuracy of the CG approach and the lack of specific interactions, these methods, and the Martini force field in particular has been shown to provide reliable data on both the ability of short peptides to self-assemble as well as the final structures that result from this process.

Following these initial studies that focussed on the screening of all dipeptides and the concentration dependence of FF, CG force fields have been increasingly used to investigate the behaviour of peptides and peptide-amphiphiles under a range of conditions including the effect of varying temperature and electrostatics on morphological changes;[115] the role of solvent in the self-assembly process;[116, 117] the interaction of peptides at interfaces – both liquid-

described in the previous section. Moreover, it demonstrated the ability of CG methods to be applied to the discovery of new materials rather than explanation of known examples. Through analysis of this large collection of information trends and comparison with available experimental results, design rules can be extrapolated. Indeed, the tripeptide hydrogelators were found to follow several key tendencies relating to the sequence dependence of those tripeptides that were able to acts a low molecular weight hydrogelators (Figure 9).[43]

Despite the tremendous potential of CG methods in modelling the self-assembly process, the loss of atomic resolution and the inability to accurately represent specific properties of the molecules, does limit their inherent accuracy. However, the general trends about the positioning of the specific "types" of residues can arise during an initial rapid screen. In agreement with experimental observations, the data show that the preferred positioning of cationic residues in the case of the hydrogelators is at the N-terminus, whereas aromatic residues (W, F, and Y) have a greater contribution to aggregation when in the middle position, or at the C-terminus (Figure 9).[43] This does not suggest that all tripeptides with a cationic residue at the N-terminus will form nanofibers or a supramolecular hydrogel, but it does indicate a preference for this position which can help limit the search space, *i.e.*, refine the parameters used in the design effort. The emergence of these rules from large datasets based on short simulation times is increasing the scope of these methods beyond the screening of short (di-/tri-) naturally occurring peptides to functionalised peptides and peptide-based materials. The implementation of these rules in an experimental context was discussed in Section 2.1.

Co-assembly is one approach to introducing desirable properties into a system that benefits from the directed discovery approach.[123-125] Short peptide sequences within proteins are commonly required for performing specific regulatory functions such as binding, targeting, recognition, *etc.* Tripeptides in particular are a recurring sequence length that is utilised in a range of biological functions.[126] Thus, the ability to use these biologically "functional" tripeptide sequences outside of the protein environment could engender synthetic systems with the biological function without the need for the full protein machinery.

A recent example of this approach was explored in the case of the endogenous, copper binding tripeptide (GHK) which is implicated in the wound healing process.[127] GHK does not self-assemble in isolation and therefore, in order to form a hydrogel containing GHK a peptide that is able to provide structure to the hydrogel is required. Based on the screen of the previous tripeptides, the anionic structure forming peptide was expected to provide a complementary partner for the co-assembly of the functional GHK into well-ordered nanofibers (Figure 10).[128] Indeed, this was seen to occur with the "structure forming peptide" FFD creating an anionic core that the "functional peptide" GHK was able to bind. While neither peptide

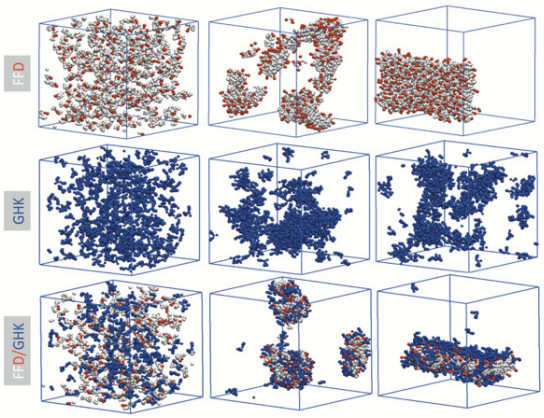

Fig. 10. Formation of a core-shell structure (bottom panel) between the structure forming peptide, FFD (upper panel), and the functional peptide, GHK (middle panel).

was able to form an ordered structure at neutral pH in isolation, the resulting co-assembly (which also included complexation of copper ions) was able to form a strong hydrogel with the functional peptide expressed on the outside of the nanofibers.

This example demonstrates the potential of the directed discovery approach. The utilisation of peptides with known functional behaviour and their incorporation into designed systems that have specific properties, such as formation of a supramolecular hydrogel, offers a pathway to exploit the existing function of peptide motifs through non-covalent incorporation into structured materials. The extraction of design rules from large-scale CG simulations of peptide-based systems will aid in matching the functional peptide with an appropriate structural counterpart. However, while CG approaches are valuable in helping to reduce the search space for the systems of interest, they achieve this by losing atomic detail that may be critical for the final system. Therefore, they are unlikely to be uniquely predictive and an efficient process will involve iteration between experimental testing and evaluation to help increase the knowledge base and refine the design rules extracted from these simulations.

3.2 Systematic Sequence Variation using Atomistic Molecular Mechanics

Atomistic molecular mechanics simulations are a natural fit for the study of peptide-based systems due to their extensively validated use for the study of proteins and other biomolecular systems.[129-132] These methods are routinely capable of simulating tens to hundreds of thousands of atoms over nanosecond timescales. These size and timescale regimes allow for an explicit consideration of the local environment that occurs during assembly. While clearly being orders of magnitude smaller than the "real" system, local effects can be revealed from these simulations. Moreover, the assembly process occurs over longer timescales than the nanosecond range and as such self-assembly into the equilibrium state is unlikely in these types of simulations.[92, 101] However, these simulations do reveal the importance of dynamic effects in the assembly process and as such have contributed significantly to the understanding of self-assembled systems. Therefore, in order to use information obtained from atomistic MD simulations in directed discovery endeavours there are two basic approaches: (i) to use a smaller model system (e.g., < 100 peptides) to study how the building blocks interact in a locally realistic environment;[35, 89, 90, 94-100, 133] or (ii) to create model systems with a presumed structure and test the stability.[46, 73, 74, 134, 135]

3.2.1 Time Evolution of Model Systems

The difficulty of using atomistic MD simulations to investigate the final structure formation of peptides has been discussed in detail in the literature. These simulations suffer from both size and time limitations that make it difficult for an equilibrium structure to be obtained.[35] Systems that spontaneously assemble from unconnected building blocks exhibit an amplified presentation of the Levinthal paradox[136, 137] – *i.e.*, that the time required to completely explore all possible conformational combinations exceeds the lifetime of the universe – as there are typically many more residues involved and fewer covalent bonds to constrain the potential conformations in self-assembled systems, relative to proteins. Therefore, a complete evolution of the system in time requires a biasing force,[88, 91] or enhanced sampling technique,[92, 101] that is able to accelerate the formation of the nanostructure. While such processes have been successfully applied, the most common approach for gaining insights for design purposes remains unbiased simulations. The unbiased evolution of the system at the reaction conditions can be important for identifying unexpected interactions that evolve in the initial structure formation and that could be critical in allowing or inhibiting a nanostructure to develop.

An example of this approach is given in a study that compared phosphorylated (Fmoc-YpL) and de-phosphorylated Fmoc-YL in the context of the enzymatic formation of Fmoc-YL nanofibers.[133] The experimental results from this study revealed that while the Fmoc-YpL formed micelles, the introduction of a phosphatase enzyme

resulted in the formation of nanofibers formed from Fmoc-YL. That is, the removal of a functional group (PO_4^{2-}) from the building block resulted in a change in the self-assembled nanostructure (Figure 6a-b).[133] This type of small perturbation of a building block is a prime example of the type of information that can be extracted from MD simulations to reveal the assembly mechanism of closely related systems.

To compare the specific interactions that occur in the micelle formation of Fmoc-YpL with the nanofiber formation that results from Fmoc-YL, MD simulations were carried out on the individual systems in an aqueous environment.[133] These simulations were carried out over 200 ns with 60 molecules of the Fmoc-YL (or Fmoc-YpL) randomly dispersed in a neutralised water box. The CHARMM force field[131] and custom parameters for the Fmoc group[97] were employed. The level of detail afforded in these simulations were able to reveal that the capping of the tyrosine sidechain with a phosphate group inhibited the nanostructure formation through disrupting stabilised H-bond formation that is required for structure formation.[133]

This above example highlights the important role that atomistic MD simulations can play in highlighting critical interactions that govern the formation of different nanostructures (micelles or nanofibers) between closely related sequences. In the context of directed discovery this approach is most closely aligned to the systematic sequence variation discussed in Section 2, where small variations in the sequence of the building block can lead to radically different structures. Understanding how these variations affect the resulting structure provide an important starting point for further modifications and the discovery of new systems.

3.2.2 Thermal Stability of Model Systems

The alternative approach to allowing a system to evolve from a randomised starting point is to provide a presumed structure (based on chemical intuition or on experimental data) as a starting point. This presumed structure can then be assessed via a quick (1 – 10 ns) simulation to determine whether the structure is stable in the proposed arrangement. The issue with this approach is that while a given arrangement may represent a local minimum it may not be the global minimum that corresponds to the self-assembled state. Therefore, the short simulation will provide some insight into the stability of the arrangement but will not be sufficiently long to allow the potential energy surface to be adequately explored.

Longer atomistic simulations based on a presumed initial structure typically reveal that the structure is dynamic and individual residues are not held fixed in their positions. For larger simulations that contain additional building blocks randomly distributed around the system, exchange of the random building blocks with those initially placed in the structure is evident.[74] This does not suggest that the initial structure is not correct, but rather the complexity of assessing the stability of a structure through a single, static arrangement.

In a combined experimental and computational study to determine the final nanostructure of Fmoc-TF-NH_2, Sasselli *et al.* employed a combination of experimental and computational energy

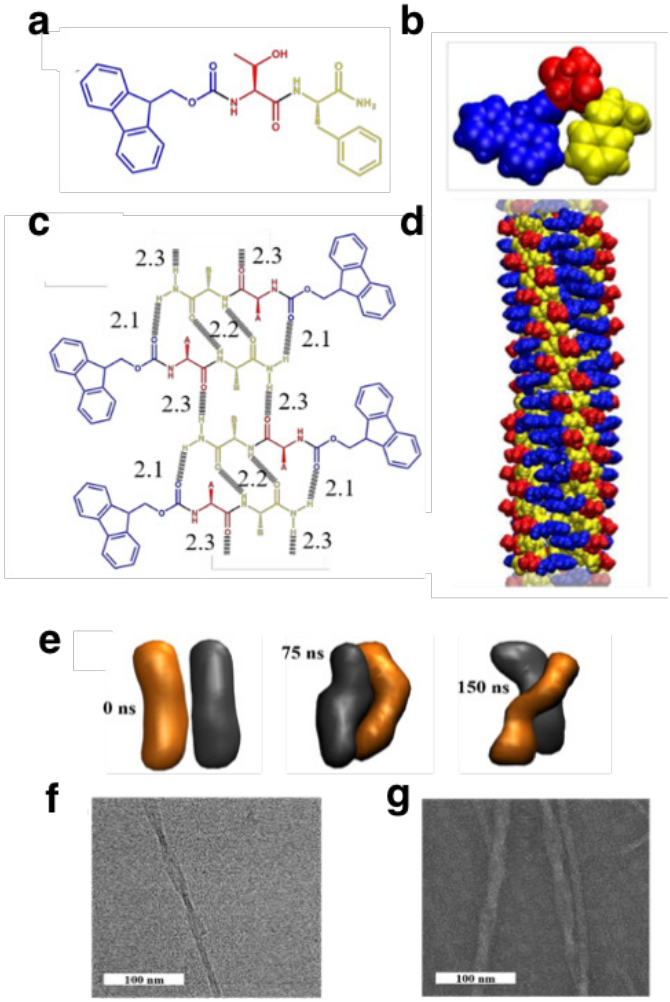

Fig. 11. Characterization of Fmoc-TL-NH_2 self-assembly. **a**. 2D and (**b**) 3D representation of Fmoc-TL-NH_2 structure. **c**. Fibre Model scheme of the H-bonded conformation: 2.1 corresponds to Fmoc-F, 2.2 to F-F, and 2.3 to T-F. **d**. Fibre Model side view. **e**. Simulation snapshots with a density surface representation showing the fibres/layers in different colors. **f-g**. Cryo-TEM (**f**) and TEM (**g**) micrographs of Fmoc-TL-NH_2 fibrils.

minimization to explore the assembly behaviour of for closely related Fmoc-dipeptides. Using comparative spectroscopic and microscopic characterisation; and extended MD simulations ultimately revealed a proposed structure (Figure 11).[74] Comparing the H-bonding patterns and chiral organization in F/L and OMe/NH_2 variants, this work demonstrated that Fmoc-TF-NH_2 formed the most stable nanostructure and more generally that F at the second position and the terminal amide promote thermodynamic stability. In order to gain molecular level insights on the stable nanostructures formed by Fmoc-TF-NH_2, we used MD atomistic models. The model revealed that the fluorenyl groups form extended stacks resulting in twisted fibres, and that the F backbone and terminal amide form H-bonds that stabilized the structure. Additionally, the observed twisted fibres were found to form by lateral aggregation of two fibres by H-bonds between the T side chain and Fmoc moiety and H-bonds of the T side chain with the F backbone and terminal amide. This work shows how MD simulations can be used to validate experimental

results on molecular sets of peptides, but also to reveal new atomic-level details, which could not be revealed otherwise, that shed light on the experimental results.[74]

The Fmoc-TF-NH_2 structure was thermodynamically more stable and the cause of this extra stability could be revealed through MD simulations. The spectroscopic (FT-IR and UV-Vis) characterisation of the systems also revealed important differences between the types of interactions that were occurring in the final nanostructures. Various models that would satisfy these interactions were built and MD simulations (150 ns with 120 Fmoc-TF-NH_2 molecules in a neutralised water box) were carried out to test the stability of the proposed structure.[74] By comparing the stability of the structures and the shift in the alignment of the building blocks as a function of time, unexpected interactions were revealed that were still capable of satisfying the experimental constraints. A further iteration led to the final structure being proposed. As a test of this structure, an additional simulation of the Fmoc-TF-NH_2 system was carried out to determine how two fibers built in this arrangement would interact. This simulation revealed the presence of twisted fibers (Figure 11e) that was consistent with the cryo-TEM (Figure 11f) and TEM studies (Figure 11g).[74]

Atomistic MD simulations have proven to be an effective approach for helping to confirm proposed nanostructures for a given system and to compare and explain the relative stability of closely related systems. The atomistic detail provided by these methods is necessary to explain the ability of atomic level differences to affect global structures. However, when comparing a more varied set of potential building blocks this approach is currently unable to rapidly screen possible systems with a high degree of confidence. Therefore, the atomistic approach offers a refinement method for understanding structures that are known to self-assemble and to provide further insight into the design rules that can be used in the development of experimental screening approaches.

4. Dynamic Peptide Libraries

While (coarse grained) molecular dynamics simulations offer a powerful approach to predict self-assembly propensities, it is currently not possible to predict structures and the resulting materials' macroscopic properties due to the complex relationship between peptide sequence, assembly propensity and properties of the material. Thus, to address this challenge our group developed a directed discovery approach based on dynamic peptide libraries (DPL), that has proven to be an effective method to search the sequence space for peptides and peptide conjugates.[138-142] The approach is based on dynamic exchange of peptide sequences using *in situ* formation and breakage of amide bonds, where the free energy involved in self-assembly of the nanostructure provides the driving force for its formation.[139, 143, 144]

The DPL system can be considered a dynamic combinatorial library (DCL) approach, where multiple subunits (or library members) are exchanged in different ways and ultimately favour the structure with the lowest free energy.[145] DCL is a tool for discovering new self-assembling structures with desired properties,[146] based on the covalent exchange of library components.[147] It allows for unexpected structures to be formed[148] based on a wide range of possible molecules in thermodynamic competition. Therefore, by using DCL, researchers can shed light on what structures are best suited to a desired purpose.[147]

For peptides, DCLs have been developed[139] using disulphide exchange of peptide components[149] that contain thiol groups, which can reversibly dimerize under reducing conditions thus forming a library of oligomeric species linked via S-S bonds.[150] For example, Otto and co-workers developed a DCL using reversible disulphide chemistry of macrocycles[151] with different size formed by threonine-containing peptide functionalized dithiol,[152, 153] where the product distribution shifts toward the self-assembling macrocycle. McLendon[154] and Marshall[155] reported on a DCL in which trimeric peptide aggregation of bidentate 2,2'-bipyridyl-peptide conjugates is stabilized by coordination with Fe^{2+} ions, where the most thermodynamically stable trimers are selected. Yet, few studies employ amide bonds exchange that links amino acids together to form new peptide sequences.[156]

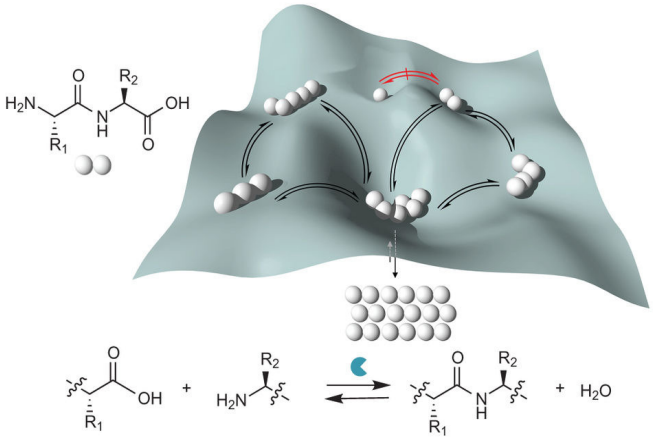

Fig. 12. DPL potential energy surface. Wells depth represents relative stability of peptides (strings of beads) formed through dynamic exchange of peptide sequence through enzymatic condensation, hydrolysis and transacylation. The deepest well shows the peptide that self-assembles into the most stable structures.

To exploit the combinatorial diversity of peptides in full, our group developed a DCL approach focused on dynamic exchange of amino acid sequences in peptides, the dynamic peptide library (DPL). In this system, the free energy gain associated with self-assembly of the peptide building blocks results in equilibrium shift in favour of the self- assembling peptides,[138, 141] shifting amide hydrolysis reactions towards amide bonds formation and thus enabling self-assembling peptides to be produced *de novo* in high yields (Fig. 12). The first example of dynamic amide bond exchange between dipeptides was reported by Swann and co-workers in 1996.[156] While incubation of YGG with FL and thermolysin formed the pentapeptide YGGFL, incubation of VA and AL with thermolysin and subsequently with the dipeptidyl peptidase Cathepsin C resulted in formation of all possible dipeptide combinations except AA and VV, with highest concentrations reported for AL. Yet, the resulting peptide products were formed in low yields (0.1% for YGGFL). Without phase separation of peptide candidates through self-assembly, product

amplification is not sufficient to allow for effective self-selection of sequences.

4.1 Fmoc-peptides

In an early work, self-assembly propensity was controlled through introduction of N-terminal aromatics. For example, non-gelling Fmoc amino acids can be linked to FF as the nucleophile through reversed hydrolysis using the non-specific endoprotease thermolysin to form Fmoc tripeptides that self-assemble into self-supporting hydrogels.[141] This work demonstrates that amide bond hydrolysis can be reversed to favour condensation, provided that the condensation product is more prone to self-assembly (and gelation) compared to the precursors. This observation clearly suggests applications in screening for stable self-assembling structures. Indeed, we developed dynamic libraries of Fmoc peptides.[138] Specifically, we demonstrated sequence selection, peptide length selection, and amplification of Fmoc-F_3 and Fmoc-L_5 using the enzyme thermolysin from reactions containing Fmoc-F/F_2 or Fmoc-L/L_2, respectively. The amplification of Fmoc-F_3 and Fmoc-L_5 accompanied with a series of spectroscopic and morphological changes showing formation of ordered structures, indicating that the selected peptides assemble into the most stable structures. When Fmoc-T and two competing nucleophiles T-OMe and F-OMe were used as chemical input, the peptide Fmoc-TF-OMe was the predominant product with 82% conversion compared to 14% for the L counterpart. Adding F-OMe sequentially to a reaction containing Fmoc-T/L-OMe in which the product Fmoc-TL-OMe reached 84% conversion, shifted product distribution in favour of Fmoc-TF-OMe formation, demonstrating that this peptide assembled into the most stable structures and are indeed reversible and under thermodynamic control. As a side note, this study also showed that the early-stage formation of structures by amplified peptides is a spatiotemporally confined process, with assembly nucleation and structure growth occur close to enzyme molecules during the initial stages of self-assembly.[138]

The assessment of additional libraries were constructed to assess di-peptide sequence/structure dependence and co-assembly.[157] Library I contained Fmoc-S and Library II Fmoc-T, with six amino acid (L, F, Y, V, G, A) esters used as competing nucleophiles. A clear trend was found in the single-nucleophile systems for both libraries with F and L as the winner nucleophiles. Accordingly, the highest conversion yields were observed for Fmoc-SF-OMe (98% yield) and Fmoc-SL-OMe (85% yield) in Library I and Fmoc-TF-OMe (96% yield) and Fmoc-TL-OMe (84% yield) in Library II, indicating that these peptides, which formed hydrogels, assembled into the most stable structures. When all six amino acid esters were directly competing in each library, the same peptides (Fmoc-SF-OMe and Fmoc-TF-OMe) were exclusively amplified. In a third library in which the two acyl donors Fmoc-S and Fmoc-T directly competed with F-OMe as the nucleophile, the peptides Fmoc-SF-OMe and Fmoc-TF-OMe were equivalently amplified, suggesting the formation of two-component nanostructures. Fluorescence spectroscopy studies indicated that higher order J-aggregates of the fluorenyl rings are stabilized by π-π stacking interactions formed by each of the two amplified peptides.

4.2 Dipeptide Exchange

To search the sequence space of unmodified self-assembling peptides of varying length, we recently developed DPLs of functionalized dipeptides with uncapped ends.[142] We narrowed the search for new self-assembling peptides by selecting a number of homo- and hetero- dyads of interest as the chemical input of the libraries. To prevent hydrolysis of the dyads' internal amide bond we used the endoprotease thermolysin, which is nonspecific for amino acids at the C terminus of the formed peptide bond, but requires an nonpolar residue at the N-terminus. Dynamic exchange of peptide sequences by enzymatic condensation, hydrolysis and trans-acylation, leading to selection of peptides that assemble into the most stable structures, was observed for all libraries.

We first showed that F_6 is the selected peptide when using the self-assembling dyad FF as chemical input, with F_4 forming at low abundance, indicating that it acts as intermediate product that is converted to the more stable F_6. W_4 was amplified in a library with W_2 as starting dyad, showing that the former assembled into stable structures. Similarly, Gazit[14] and Zhang[32] showed that W_2 does not assemble into ordered structures. To investigate non-aromatic peptide selection, we focused on aliphatic sequence library using L_2 as the starting sequence, which resulted in selection of L_6, forming a self-supporting hydrogel. FTIR, CD and TEM analyses of all three libraries showed that the selection of peptides was accompanied with formation of highly ordered structures, with W_4 formation mainly driven by the hydrophobic effect. For all libraries, we observed a clear preference towards formation of peptides with even numbers of amino acids. Using L_3 as an input dyad, resulting in exclusive formation of F_6, we showed that the enzyme thermodynamically favours addition or hydrolysis of intact dyads to form tetramers and hexamers.[142]

Next, binary and ternary mixtures of dyads as starting sequences were used and the effect of different environmental conditions was examined. An environmental selection pressure was demonstrated for peptides amplifications, depending on the solvents used, as the tetrapeptide F_2L_2 selected with the starting dyads F_2 and L_2 at 80% THF, and various sequences (F_2L_2, W_2L_2, W_4, W_2L_4) selected for the ternary mixture of dipeptides F_2, L_2 and W_2.

Finally, a binary sequence library was developed, using a number of nonpolar/polar dyads as input. FD/FS gave rise to particularly efficient condensation, and HPLC and mass spectroscopy analysis revealed that the single-sequence octamer FDFSFDFS was amplified uniquely over time, and AFM and TEM analyses revealed morphological supramolecular reconfiguration from spherical peptide aggregates to an entangled fibrillar network (Fig. 13c). The octapeptide was re-synthesised and shown to form β-sheet rich structures, assembling into a self-supporting hydrogel.[142] We have shown, with Saiani and Miller, that DPL can be utilized to drive the synthesis and gelation of ionic octapeptides from soluble solutions of the tetrapeptide FEFK. The peptide was initially hydrolyzed into the dipeptides FE and FK and subsequently an octapeptide forming the most stable nanostructures was selected and amplified.[158] A similar ionic peptide has been shown promise in regeneration of hard tissues, supporting stem cells differentiation into osteoblasts.[159]

4.3 DPL in Co-Assembly

The DPL approach can be further utilized to search for peptides that form thermodynamically stable complexes with ligand of interest. For example, we previously demonstrated the co-assembly of peptides and polysaccharides as a thermodynamic driving force for peptide sequence selection.[160] Charged Fmoc-amino acids (R, K, H, E, D, and the non-natural amino acid sulfonic acid) were used as precursors for thermolysin, F-amide as the nucleophile, and either cationic (chitosan) or anionic (heparin) polysaccharide. Low levels of Fmoc-dipeptides were formed in the absence of saccharides, while in the presence of the oppositely charged saccharide, Fmoc-dipeptide was significantly amplified. The results show that shifting the amide condensation/hydrolysis further toward condensation may be driven by free energy contributions of the electrostatic peptide/polysaccharide complexation. These findings show a selection of the oppositely charged peptide by electrostatic templating of the condensation product with the oppositely charged polysaccharide.[160]

Similarly, Williams and co-workers[161] showed that addition of the basement membrane protein, laminin, to a reaction mixture containing Fmoc-L and L_2 increased the conversion to Fmoc-L_3 from 54% (in the absence of the protein) to 80%, following exposure to immobilized thermolysin. The resulting laminin-incorporated hydrogel was subsequently utilized for tissue repair in a zebrafish model of extracellular matrix disease and found stable *in vivo* 4 days post-microinjection to the damaged tissue.

4.4 DPL for the discovery of functional materials

Taking advantage of the fact that DPL searches for the lowest free energy structures, it is an appropriate method to search for the most stable self-assembling structures formed in e.g. charge and energy transfer hydrogels. The DPL approach was thus utilized to search for peptide sequences that optimize charge transfer properties in a supramolecular peptide gel. This was done using a dynamic library of naphthalenediimide (NDI)–peptide conjugate acceptors and various π-electron donors, providing a range of functional electronic nanostructures.[162] Competitive enzymatic condensation reaction between NDI-Y and a series of amino acid (F, L, V, Y, A) amides resulted in the amplification of NDI-YF forming the thermodynamically favourable nanostructures. When screening for stable charge transfer interactions, the amplification yield of NDI-YF was highly dependent on the π-electron donor used, with significant increase from 48% to 71% observed with dialkoxy naphthalenes (1,5-DAN). Thus, selectivity in the charge transfer co-assembly process can drive self-selection and amplification of the most stable charge transfer peptide nanostructures.[162] YF was also selected in energy transfer-based DPL utilizing the well-known donor–acceptor interactions between naphthalene chromophores and dansyl chromophores.[163] Specifically, the peptide conjugate naphthoxy-YF (Nap-YF) was efficiently selected and amplified from a library of 8 competing amino acids in the presence of the water-soluble dansyl-β-alanine (DA) acceptor. The self-selection process resulted in the formation of highly efficient energy transfer gel.

4.5 Selecting for shallow thermodynamic minima

DPL can also be used for "negative selection", *i.e.* selecting peptides that have shallow minima in the free energy landscape to

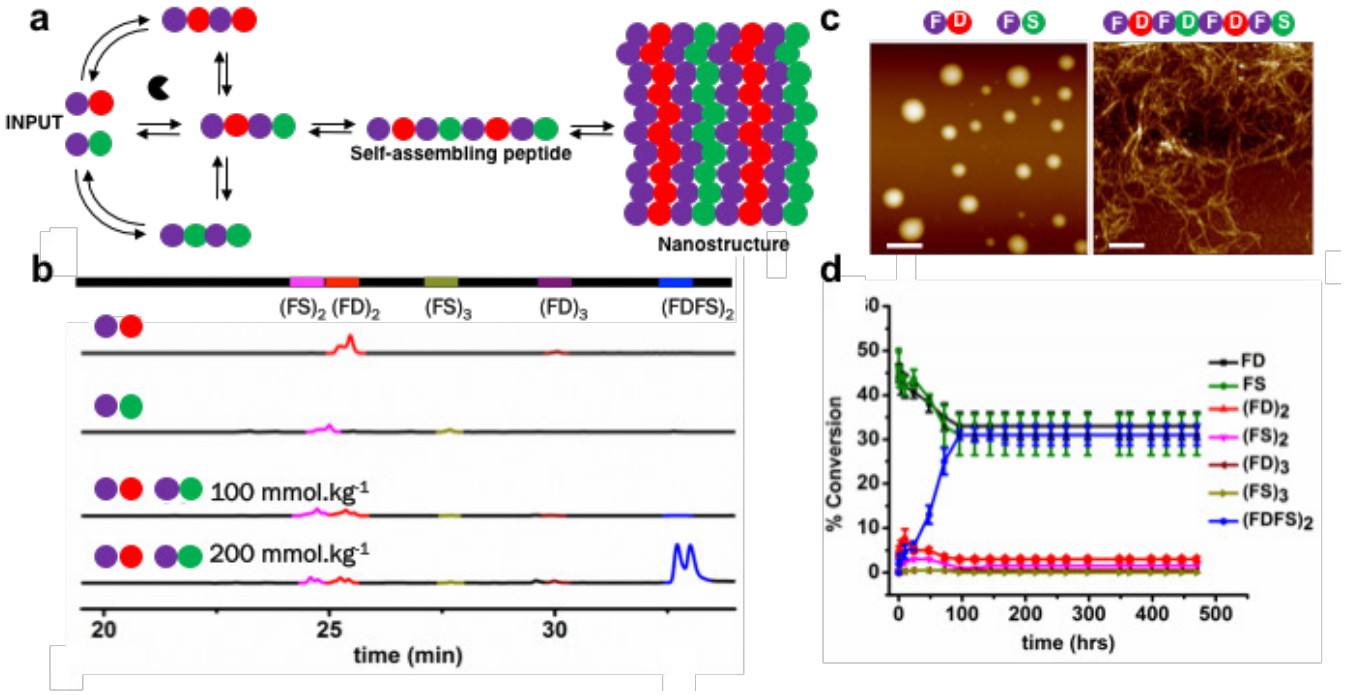

Fig. 13. DPL of binary dyads. **a**. Schematic representation of dynamic peptide libraries approach with FD/FS as chemical input. **b.** HPLC graph of: 200 mmol kg^{-1} of FD; 200 mmol kg^{-1} of FS; mixture of the dipeptides (100 mmol kg−1 each), and 200 mmol kg^{-1} each in 100 mM sodium phosphate buffer pH 8 in the presence of 1 mg ml^{-1} thermolysin. Color-coded bars above correspond to the formation of oligomers. **c**. AFM images of images of FD/FS mixtures (200 mmol kg−1 each) before (left) and 400 h after (right) enzyme addition **d**. HPLC profile of oligomer distribution over time (200 mmol kg^{-1} each). Error bars represent the s.e.m. of three different experiments.

enable their rapid exchange for rapid response materials. More specifically, while thermodynamic selection is useful in identifying stable materials, these may not be responsive to external stimuli, where a shallower assembly profile is desirable. Thus, we utilized the outcome of a DPL to establish a self-assembled system with dynamic instability. Informed by the NDI-dipeptide DPL described above,[163] we found that the Nap-YY-OMe peptide was formed in low yield, suggesting that the enzyme is able to produce the peptide, but self-assembly propensity is low. Nap-Y-OMe was used as the acyl donor with Y-NH_2, F-NH_2 or L-NH_2 to form Nap-YX-NH_2 by transacylation or Nap-Y-OH by hydrolysis of the dipeptide product over time, using either α-chymotrypsin or thermolysin.[164] Depending on the the sequence (YY vs YF), either permanent or transient gelation was observed, depending on the self-assembly propensity, and specifically the critical gelation concentration, of formed peptides. The product Nap-YF-NH_2 was formed in a concentration higher than the critical gelation concentration, giving rise to equilibrium gelation. In contrast, formation of Nap-YY-NH_2 and Nap-YL-NH_2 resulted in nonequilbrium gelation, followed by hydrolysis of the peptides into Nap-Y-OH and consequently disassembly into solutions. Microscopy analysis of the Y-NH_2 system revealed formation of entangled fibre networks at Nap-YY-NH_2 conversion yield above its critical gelation concentration. Fibres shortening was observed during Nap-YY-NH_2 hydrolysis into Nap-YY-OH in both gel and solution, and spherical aggregates formed by the completion of this process in the free-flowing solution. Refuelling the system at the complete hydrolysis stage with Nap-Y-OMe resulted in formation and breakage of gel containing fibres, up to three cycles.[164]

We exploited this concept further to develop sequence-dependent transient nanostructures, by selecting peptides that do not assemble at equilibrium.[165] In this system, the artificial sweetener aspartame (DF-OMe) was used with various amino acid amides (W, Y, F, L, V, S, T) as nucleophiles in the presence of α-chymotrypsin. This was expected to result in formation of DFX-NH_2 by transacylation, and in the case of thermodynamically disfavoured self-assembly, further hydrolysis to DF-OH. Formation of the tripeptides DFF-NH_2 and DFY-NH_2 accompanied by hydrogelation was observed for the nucleophiles F and Y, while in the presence of the other amino acid amides, aspartame hydrolysis to DF-OH was observed after 30 min. Correspondingly, the highest predicted assembly scores of F and Y were found for the second and third position within a tripeptide.[43] Gel lifetime could be tuned by using the Y *vs*. F system. A 24 h stable gel was formed by DFF-NH_2, which started to hydrolyse thereafter, while a 4 h stable gel formed by DFY-NH_2 which started to hydrolyse after 2 h and completely disappear after 10 h, indicating that DFY-NH_2 assembled into less stable structures. TEM analysis of both peptides showed dynamic instability of fibres formation and shortening, which corresponded to appearance/disappearance of hydrogen bonding networks within β-sheet-like organization, as shown by FTIR analysis. Similar to the Nap-dipeptide-NH_2 system described above, the system could be refuelled up to three cycles by addition of aspartame. A competition reaction in the presence of aspartame as the acyl donor and both F-NH_2 and Y-NH_2 as competing nucleophiles resulted in a higher conversion yield of DFY-NH_2 compared to DFF-NH_2, suggesting kinetic rather than thermodynamic selection of nanostructures. Thus, both systems show the potential in selecting peptides that form dynamically instable structures, able to assemble and disassemble, similar to natural systems.

Altogether, DPL provides is a powerful platform for directed discovery of functional materials including biomaterials[161],[158, 160] materials for electronic,[162] and energy transfer applications.[163] This can include libraries designed for self-selection of supramolecular nanostructure complexes incorporating organic[160] or inorganic[162, 163] ligands of interest. Notably, these DPLs operate under physiological conditions therefore can be implemented in biological systems. In addition, chemical synthesis of identified peptides and characterization of the self-assembly behaviour of the purified material validated the formation of functional nanostructures.[142]

5. Future Directions

While most examples discussed to date focus on supramolecular assembly and recognition, there are opportunities to go beyond self-assembly and structure formation, in searching for features such as supramolecular recognition or catalysis/reactivity. One main challenges here are the development of accurate spectroscopic activity assays, that can deal with heterogeneous mixtures. Another challenge is to adapt computational processes for these approaches. The development of appropriate diagnostics that allow for the desired properties to be efficiently screened computationally is necessary for the continued equal partnership between computation and experiment in this area. While the use of coarse grained methods has proven to be a successful strategy in the discovery of new materials, the future discovery of functional materials will require further development of these methods to ensure that the desired property of the material is adequately captured by this approach. Finally, the large datasets created through the screening process also offer a controlled set of information where, in combination with experimental results, artificial intelligence methods may be applied to extract meaningful relationships between the calculated properties of the system and the behavior of the material.

The use of computational methods in combination with experimental approaches to provide detailed information about the structure of nanoscale systems has become increasingly common and seen significant success.[74] However, while atomistic methods have traditionally been used independently to predict the stability of a given structure, the integration of atomistic simulations with the growing number of experimental spectroscopic and microscopic characterisation techniques is less common. The evaluation of complementary information from these different techniques can lead to more reliable insights into the details of the structures formed, which, in turn, helps to pave the way for the rational development of tunable materials.

For DPL to become more valuable and more generally usable, it would be beneficial to increase throughput and reduce reaction times. For efficient large scale screening using wide selection of chemical input, a number of steps should be taken. We previously observed slow reaction kinetics for days for part of the unmodified peptide libraries[142] due to heterogeneity of the mixtures. To address this challenge, temperature cycles can be applied to overcome kinetic barriers and facilitate reaction kinetics by shifting sequence exchange towards rapid hydrolysis of non-assembling sequences, expected to be more susceptible to digestion, or towards condensation into assembling structures. This can be achieved by

utilizing the enzyme thermolysin, which is stable at the temperature range for hydrogen bonding melting.

Beyond focusing on ordered systems, an emerging area is focused on function in disordered systems. Very recently, Eisenberg and co-workers[166] showed that low-complexity domains from proteins that are involved in liquid phase separation and formation of intracellular membraneless compartments form kinked β-sheets that interact weakly into fibrils and gels. Due to the biological role of these disordered proteins in condensing biomolecules through coacervation, these findings suggest that coacervation of low complexity peptides into less ordered structures, rather than the typical supramolecular ordered assemblies, is a promising process that can lead to formation of materials with emerging functionality.

6. Conclusions

Following approximately 20 years of active research in this area, it is now clear that the amino acids provide a tremendous chemical set for achieving controlled structure, molecular recognition, reactivity, catalysis, energy and electron transport. Thus, twenty amino acid building blocks represent a formidable construction set for materials and structures. However, in order for the field of peptide nanotechnology to deliver on its promise, a tremendous challenge still exist in finding governing principles or rules in peptide materials design. Progress in sequence design, designing peptide materials with desirable functions is now informed increasingly by search approaches described here.

We note that focusing only on natural amino acids has an advantage from an application perspective, as peptide-based materials can be implanted or ingested and, in principle, they are biodegradable and metabolizable to amino acids that hold GRAS (generally recognized as safe) status according to the FDA. Thus, peptide nanomaterials are of significant interest in a variety of current and potential applications, *e.g*. emulsifiers and gelators for cosmetics, food ingredients, soft biomaterials, drug delivery vehicles, *etc.*

Conflicts of interest

In accordance with our policy on Conflicts of interest please ensure that a conflicts of interest statement is included in your manuscript here. Please note that this statement is required for all submitted manuscripts. If no conflicts exist, please state that "There are no conflicts to declare".

Acknowledgements

Dr. G. G. Scott is thanked for his help with the production of some of the figures.

Notes and references

1. S. Zhang, *Nature biotechnology*, 2003, **21**, 1171-1178.
2. M. J. Webber, E. A. Appel, E. Meijer and R. Langer, *Nature materials*, 2016, **15**, 13-26.
3. I. W. Hamley, *Chem Rev*, 2017, **117**, 14015-14041.
4. R. V. Ulijn and D. N. Woolfson, *Chem Soc Rev*, 2010, **39**, 3349-3350.
5. J. M. Fletcher, R. L. Harniman, F. R. Barnes, A. L. Boyle, A. Collins, J. Mantell, T. H. Sharp, M. Antognozzi, P. J. Booth and N. Linden, *Science*, 2013, **340**, 595-599.
6. W. DeGrado and J. Lear, *Journal of the American Chemical Society*, 1985, **107**, 7684-7689.
7. L. E. O'leary, J. A. Fallas, E. L. Bakota, M. K. Kang and J. D. Hartgerink, *Nat Chem*, 2011, **3**, 821-828.
8. S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik and M. H. Hecht, *Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information*, 1993, **262**, 1680-1702.
9. S. Zhang, T. Holmes, C. Lockshin and A. Rich, *Proceedings of the National Academy of Sciences*, 1993, **90**, 3334-3338.
10. H. Xiong, B. L. Buckwalter, H.-M. Shieh and M. H. Hecht, *Proceedings of the National Academy of Sciences*, 1995, **92**, 6349-6353.
11. J. P. Schneider, D. J. Pochan, B. Ozbas, K. Rajagopal, L. Pakstis and J. Kretsinger, *Journal of the American Chemical Society*, 2002, **124**, 15030-15037.
12. J. Gao, C. Tang, M. A. Elsawy, A. M. Smith, A. F. Miller and A. Saiani, *Biomacromolecules*, 2017, **18**, 826-834.
13. A. Saiani, A. Mohammed, H. Frielinghaus, R. Collins, N. Hodson, C. Kielty, M. Sherratt and A. Miller, *Soft Matter*, 2009, **5**, 193-202.
14. M. Reches and E. Gazit, *Science*, 2003, **300**, 625-627.
15. E. Gazit, *The FASEB Journal*, 2002, **16**, 77-83.
16. C. H. Görbitz, *Chem-Eur J*, 2001, **7**, 5153-5159.
17. H. Cui, M. J. Webber and S. I. Stupp, *Biopolymers*, 2010, **94**, 1-18.
18. S. Fleming and R. V. Ulijn, *Chem Soc Rev*, 2014, **43**, 8150-8177.
19. G. Fichman and E. Gazit, *Acta biomaterialia*, 2014, **10**, 1671-1682.
20. K. Tao, A. Levin, L. Adler-Abramovich and E. Gazit, *Chem Soc Rev*, 2016, **45**, 3935-3953.
21. E. R. Draper and D. J. Adams, *Chem*, 2017, **3**, 390-410.
22. X. W. Du, J. Zhou, J. F. Shi and B. Xu, *Chem Rev*, 2015, **115**, 13165-13307.
23. R. R. Naik, S. J. Stringer, G. Agarwal, S. E. Jones and M. O. Stone, *Nat Mater*, 2002, **1**, 169-172.
24. R. Volkmer, V. Tapia and C. Landgraf, *FEBS Lett*, 2012, **586**, 2780-2786.
25. C. H. Gorbitz, *New J Chem*, 2003, **27**, 1789-1793.
26. C. Görbitz and E. Gundersen, *Acta Crystallographica Section C: Crystal Structure Communications*, 1996, **52**, 1764-1767.
27. H. Erdogan, E. Babur, M. Yilmaz, E. Candas, M. Gordesel, Y. Dede, E. E. Oren, G. B. Demirel, M. K. Ozturk, M. S. Yavuz and G. Demirel, *Langmuir*, 2015, **31**, 7337-7345.
28. C. H. Gorbitz, *Chem Commun (Camb)*, 2006, DOI: 10.1039/b603080g, 2332-2334.
29. I. Azuri, L. Adler-Abramovich, E. Gazit, O. Hod and L. Kronik, *Journal of the American Chemical Society*, 2014, **136**, 963-969.
30. C. Guo, Y. Luo, R. Zhou and G. Wei, *Acs Nano*, 2012, **6**, 3907-3918.
31. P. W. J. M. Frederix, R. V. Ulijn, N. T. Hunt and T. Tuttle, *Journal of Physical Chemistry Letters*, 2011, **2**, 2380-2384.
32. Z. Fan, L. Sun, Y. Huang, Y. Wang and M. Zhang, *Nature nanotechnology*, 2016, **11**, 388-394.

33. N. S. de Groot, T. Parella, F. X. Aviles, J. Vendrell and S. Ventura, *Biophysical journal*, 2007, **92**, 1732-1741.
34. M. Reches and E. Gazit, *Nano letters*, 2004, **4**, 581-585.
35. P. Tamamis, L. Adler-Abramovich, M. Reches, K. Marshall, P. Sikorski, L. Serpell, E. Gazit and G. Archontis, *Biophysical journal*, 2009, **96**, 5020-5029.
36. J. James and A. B. Mandal, *Journal of colloid and interface science*, 2011, **360**, 600-605.
37. P. Moitra, K. Kumar, P. Kondaiah and S. Bhattacharya, *Angewandte Chemie International Edition*, 2014, **53**, 1113-1117.
38. S. Zarzhitsky, T. P. Vinod, R. Jelinek and H. Rapaport, *Chem Commun*, 2015, **51**, 14409-14409.
39. S. Marchesan, C. D. Easton, F. Kushkaki, L. Waddington and P. G. Hartley, *Chem Commun*, 2012, **48**, 2195-2197.
40. S. Marchesan, C. Easton, K. Styan, L. Waddington, F. Kushkaki, L. Goodall, K. McLean, J. Forsythe and P. Hartley, *Nanoscale*, 2014, **6**, 5172-5180.
41. K. H. Chan, B. Xue, R. C. Robinson and C. A. E. Hauser, *Sci Rep-Uk*, 2017, **7**.
42. J. Smadbeck, K. H. Chan, G. A. Khoury, B. Xue, R. C. Robinson, C. A. E. Hauser and C. A. Floudas, *Plos Comput Biol*, 2014, **10**.
43. 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*, 2015, **7**, 30-37.
44. I. R. Sasselli, P. J. Halling, R. V. Ulijn and T. Tuttle, *ACS Nano*, 2016, **10**, 2661-2668.
45. G. G. Scott, P. J. McKnight, T. Tuttle and R. V. Ulijn, *Advanced Materials*, 2016, **28**, 1381-1386.
46. A. Lampel, S. A. McPhee, H.-A. Park, G. G. Scott, S. Humagain, D. R. Hekstra, B. Yoo, P. W. Frederix, R. R. Abzalimov and S. G. Greenbaum, *Science*, 2017, **356**, 1064-1068.
47. C. G. Pappas, R. Shafi, I. R. Sasselli, H. Siccardi, T. Wang, V. Narang, R. Abzalimov, N. Wijerathne and R. V. Ulijn, *Nat Nanotechnol*, 2016, **11**, 960-967.
48. H. Xu, Y. Wang, X. Ge, S. Han, S. Wang, P. Zhou, H. Shan, X. Zhao and J. R. Lu, *Chemistry of Materials*, 2010, **22**, 5165-5173.
49. S. Vauthey, S. Santoso, H. Gong, N. Watson and S. Zhang, *Proc Natl Acad Sci U S A*, 2002, **99**, 5355-5360.
50. S. Han, S. Cao, Y. Wang, J. Wang, D. Xia, H. Xu, X. Zhao and J. R. Lu, *Chemistry*, 2011, **17**, 13095-13102.
51. L. Tjernberg, W. Hosia, N. Bark, J. Thyberg and J. Johansson, *J Biol Chem*, 2002, **277**, 43243-43246.
52. J. B. Guilbaud, E. Vey, S. Boothroyd, A. M. Smith, R. V. Ulijn, A. Saiani and A. F. Miller, *Langmuir*, 2010, **26**, 11297-11303.
53. K. Tenidis, M. Waldner, J. Bernhagen, W. Fischle, M. Bergmann, M. Weber, M. L. Merkle, W. Voelter, H. Brunner and A. Kapurniotu, *J Mol Biol*, 2000, **295**, 1055-1071.
54. J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed and R. Tycko, *Biochemistry*, 2000, **39**, 13748-13759.
55. K. Lu, J. Jacob, P. Thiyagarajan, V. P. Conticello and D. G. Lynn, *J Am Chem Soc*, 2003, **125**, 6391-6393.
56. R. Azriel and E. Gazit, *J Biol Chem*, 2001, **276**, 34156-34161.
57. M. Reches, Y. Porat and E. Gazit, *J Biol Chem*, 2002, **277**, 35475-35480.
58. C. A. E. Hauser, R. S. Deng, A. Mishra, Y. H. Loo, U. Khoe, F. R. Zhuang, D. W. Cheong, A. Accardo, M. B. Sullivan, C. Riekel, J. Y. Ying and U. A. Hauser, *P Natl Acad Sci USA*, 2011, **108**, 1361-1366.
59. T. O. Omosun, M. C. Hsieh, W. S. Childers, D. Das, A. K. Mehta, N. R. Anthony, T. Pan, M. A. Grover, K. M. Berland and D. G. Lynn, *Nat Chem*, 2017, **9**, 805-809.
60. S. Fleming, S. Debnath, P. Frederix, T. Tuttle and R. V. Ulijn, *Chem Commun*, 2013, **49**, 10587-10589.
61. R. Vegners, I. Shestakova, I. Kalvinsh, R. M. Ezzell and P. A. Janmey, *Journal of Peptide Science*, 1995, **1**, 371-378.
62. Y. Zhang, H. Gu, Z. Yang and B. Xu, *Journal of the American Chemical Society*, 2003, **125**, 13680-13681.
63. Z. M. Yang, H. W. Gu, Y. Zhang, L. Wang and B. Xu, *Chem Commun*, 2004, DOI: 10.1039/b310574a, 208-209.
64. A. Mahler, M. Reches, M. Rechter, S. Cohen and E. Gazit, *Advanced Materials*, 2006, **18**, 1365-1370.
65. V. Jayawarna, M. Ali, T. A. Jowitt, A. F. Miller, A. Saiani, J. E. Gough and R. V. Ulijn, *Advanced Materials*, 2006, **18**, 611-614.
66. J. Raeburn, A. Zamith Cardoso and D. J. Adams, *Chem Soc Rev*, 2013, **42**, 5143-5156.
67. D. J. Adams, M. F. Butler, W. J. Frith, M. Kirkland, L. Mullen and P. Sanderson, *Soft Matter*, 2009, **5**, 1856-1862.
68. Z. Yang, G. Liang and B. Xu, *Acc Chem Res*, 2008, **41**, 315-326.
69. A. R. Hirst, S. Roy, M. Arora, A. K. Das, N. Hodson, P. Murray, S. Marshall, N. Javid, J. Sefcik, J. Boekhoven, J. H. van Esch, S. Santabarbara, N. T. Hunt and R. V. Ulijn, *Nat Chem*, 2010, **2**, 1089-1094.
70. M. Hughes, L. S. Birchall, K. Zuberi, L. A. Aitken, S. Debnath, N. Javid and R. V. Ulijn, *Soft Matter*, 2012, **8**, 11565-11574.
71. M. Hughes, S. Debnath, C. W. Knapp and R. V. Ulijn, *Biomater Sci-Uk*, 2013, **1**, 1138-1142.
72. R. J. Williams, A. M. Smith, R. Collins, N. Hodson, A. K. Das and R. V. Ulijn, *Nat Nanotechnol*, 2009, **4**, 19-24.
73. 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**, 5595-5602.
74. I. R. Sasselli, C. G. Pappas, E. Matthews, T. Wang, N. T. Hunt, R. V. Ulijn and T. Tuttle, *Soft Matter*, 2016, **12**, 8307-8315.
75. E. T. Pashuck, H. G. Cui and S. I. Stupp, *Journal of the American Chemical Society*, 2010, **132**, 6041-6046.
76. G. Cheng, V. Castelletto, C. M. Moulton, G. E. Newby and I. W. Hamley, *Langmuir*, 2010, **26**, 4990-4998.
77. C. Tang, R. V. Ulijn and A. Saiani, *Langmuir*, 2011, **27**, 14438-14449.
78. M. Tena-Solsona, J. F. Miravet and B. Escuder, *Chemistry*, 2014, **20**, 1023-1031.
79. H. G. Cui, A. G. Cheetham, E. T. Pashuck and S. I. Stupp, *Journal of the American Chemical Society*, 2014, **136**, 12461-12468.
80. L. Adler-Abramovich and E. Gazit, *Chem Soc Rev*, 2014, **43**, 6881-6893.
81. M. S. Ekiz, G. Cinar, M. A. Khalily and M. O. Guler, *Nanotechnology*, 2016, **27**.
82. C. R. L. Chapman, E. C. M. Ting, A. Kereszti and I. Paci, *Journal of Physical Chemistry C*, 2013, **117**, 19426-19435.
83. G. C. Schatz, *P Natl Acad Sci USA*, 2007, **104**, 6885-6892.
84. T. Tuttle, *Israel Journal of Chemistry*, 2015, **55**, 724-734.
85. M. Engels, D. Bashford and M. R. Ghadiri, *Journal of the American Chemical Society*, 1995, **117**, 9151-9158.

86. O.-S. Lee, S. I. Stupp and G. C. Schatz, *Journal of the American Chemical Society*, 2011, **133**, 3677-3683.
87. S. Esteban-Martin and J. Salgado, *Biophysical Journal*, 2007, **92**, 903-912.
88. E. Khurana, S. O. Nielsen, B. Ensing and M. L. Klein, *Journal of Physical Chemistry B*, 2006, **110**, 18965-18972.
89. J. Jeon, C. E. Mills and M. S. Shell, *Journal of Physical Chemistry B*, 2013, **117**, 3935-3943.
90. S. Tsonchev, A. Troisi, G. C. Schatz and M. A. Ratner, *Journal of Physical Chemistry B*, 2004, **108**, 15278-15284.
91. T. Yu, O.-S. Lee and G. C. Schatz, *Journal of Physical Chemistry A*, 2013, **117**, 7453-7460.
92. P. Tamamis, K. Terzaki, M. Kassinopoulos, L. Mastrogiannis, E. Mossou, V. T. Forsyth, E. P. Mitchell, A. Mitraki and G. Archontis, *Journal of Physical Chemistry B*, 2014, **118**, 1765-1774.
93. S. Boopathi and P. Kolandaivel, *Journal of Biomolecular Structure & Dynamics*, 2013, **31**, 158-173.
94. A. J. van Hell, A. Klymchenko, P. P. Burgers, E. E. Moret, W. Jiskoot, W. E. Hennink, D. J. A. Crommelin and E. Mastrobattista, *Journal of Physical Chemistry B*, 2010, **114**, 11046-11052.
95. C. Menard-Moyon, V. Venkatesh, K. V. Krishna, F. Bonachera, S. Verma and A. Bianco, *Chem-Eur J*, 2015, **21**, 11681-11686.
96. J. Jeon and M. S. Shell, *Journal of Physical Chemistry B*, 2014, **118**, 6644-6652.
97. I. R. Sasselli, R. V. Ulijn and T. Tuttle, *Physical Chemistry Chemical Physics*, 2016, **18**, 4659-4667.
98. E. Moroni, G. Scarabelli and G. Colombo, *Frontiers in Bioscience*, 2009, **14**, 523-U525.
99. C.-T. Lai, N. L. Rosi and G. C. Schatz, *Journal of Physical Chemistry Letters*, 2017, **8**, 2170-2174.
100. B. Banerji, M. Chatterjee, U. Pal and N. C. Maiti, *Journal of Physical Chemistry B*, 2017, **121**, 6367-6379.
101. N. Nishikawa, P. H. Nguyen, P. Derreumaux and Y. Okamoto, *Molecular Simulation*, 2015, **41**, 1041-1044.
102. S. Park, M. Lee and S. Shin, *Chemistry-an Asian Journal*, 2015, **10**, 1684-1689.
103. M. Y. Yeh, C. T. Huang, T. S. Lai, F. Y. Chen, N. T. Chu, D. T. H. Tseng, S. C. Hung and H. C. Lin, *Langmuir*, 2016, **32**, 7630-7638.
104. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S.-J. Marrink, *Journal of Chemical Theory and Computation*, 2008, **4**, 819-834.
105. S. O. Nielsen, C. F. Lopez, G. Srinivas and M. L. Klein, *Journal of Physics-Condensed Matter*, 2004, **16**, R481-R512.
106. M. L. Klein and W. Shinoda, *Science*, 2008, **321**, 798-800.
107. A. Villa, C. Peter and N. F. A. van der Vegt, *Physical Chemistry Chemical Physics*, 2009, **11**, 2077-2086.
108. G. Bellesia and J.-E. Shea, *Journal of Chemical Physics*, 2007, **126**.
109. S. P. Carmichael and M. S. Shell, *Journal of Physical Chemistry B*, 2012, **116**, 8383-8393.
110. M. Seo, S. Rauscher, R. Pomes and D. P. Tieleman, *Journal of Chemical Theory and Computation*, 2012, **8**, 1774-1785.
111. A. Villa, N. F. A. van der Vegt and C. Peter, *Physical Chemistry Chemical Physics*, 2009, **11**, 2068-2076.
112. R. B. Pandey, Z. Kuang and B. L. Farmer, *Plos One*, 2013, **8**.
113. D. H. de Jong, G. Singh, W. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman and S. J. Marrink, *Journal of Chemical Theory and Computation*, 2012, **9**, 687-697.
114. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. De Vries, *The journal of physical chemistry B*, 2007, **111**, 7812-7824.
115. I. W. Fu, C. B. Markegard, B. K. Chu and H. D. Nguyen, *Advanced Healthcare Materials*, 2013, **2**, 1388-1400.
116. I. W. Fu, C. B. Markegard, B. K. Chu and H. D. Nguyen, *Langmuir*, 2014, **30**, 7745-7754.
117. I. W. Fu, C. B. Markegard and H. D. Nguyen, *Langmuir*, 2015, **31**, 315-324.
118. E. Khurana, R. H. DeVane, A. Kohlmeyer and M. L. Klein, *Nano Letters*, 2008, **8**, 3626-3630.
119. R. A. Mansbach and A. L. Ferguson, *Journal of Physical Chemistry B*, 2017, **121**, 1684-1706.
120. L. Ruiz and S. Keten, *Soft Matter*, 2014, **10**, 851-861.
121. T. Yu and G. C. Schatz, *Journal of Physical Chemistry B*, 2013, **117**, 14059-14064.
122. C. Guo, Y. Luo, R. Zhou and G. Wei, *Nanoscale*, 2014, **6**, 2800-2811.
123. I. R. Sasselli, I. P. Moreira, R. V. Ulijn and T. Tuttle, *Organic & Biomolecular Chemistry*, 2017, **15**, 6541-6547.
124. D. Bochicchio and G. M. Pavan, *Acs Nano*, 2017, **11**, 1000-1011.
125. A. Raspa, G. A. A. Saracino, R. Pugliese, D. Silva, D. Cigognini, A. Vescovi and F. Gelain, *Advanced Functional Materials*, 2014, **24**, 6317-6328.
126. P. Ung and D. A. Winkler, *Journal of Medicinal Chemistry*, 2011, **54**, 1111-1125.
127. F.-X. Maquart, L. Pickart, M. Laurent, P. Gillery, J.-C. Monboisse and J.-P. Borel, *FEBS letters*, 1988, **238**, 343-346.
128. Y. M. Abul-Haija, G. G. Scott, J. K. Sahoo, T. Tuttle and R. V. Ulijn, *Chem Commun*, 2017, **53**, 9562-9565.
129. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *Journal of Chemical Theory and Computation*, 2008, **4**, 435-447.
130. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781-1802.
131. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *Journal of Physical Chemistry B*, 1998, **102**, 3586-3616.
132. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *Journal of the American Chemical Society*, 1995, **117**, 5179-5197.
133. I. P. Moreira, I. R. Sasselli, D. A. Cannon, M. Hughes, D. A. Lamprou, T. Tuttle and R. V. Ulijn, *Soft Matter*, 2016, **12**, 2623-2631.
134. M. Hughes, H. 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**, 10032-10038.
135. L. Liang, L.-W. Wang and J.-W. Shen, *Rsc Advances*, 2016, **6**, 100072-100078.
136. C. Levinthal, *Journal de chimie physique*, 1968, **65**, 44-45.
137. R. Zwanzig, A. Szabo and B. Bagchi, *Proceedings of the National Academy of Sciences*, 1992, **89**, 20-22.

138. R. J. Williams, A. M. Smith, R. Collins, N. Hodson, A. K. Das and R. V. Ulijn, *Nature nanotechnology*, 2009, **4**, 19-24.
139. J. W. Sadownik and R. V. Ulijn, *Current opinion in biotechnology*, 2010, **21**, 401-411.
140. A. K. Das, A. R. Hirst and R. V. Ulijn, *Faraday discussions*, 2009, **143**, 293-303.
141. S. Toledano, R. J. Williams, V. Jayawarna and R. V. Ulijn, *Journal of the American Chemical Society*, 2006, **128**, 1070-1071.
142. C. G. Pappas, R. Shafi, I. R. Sasselli, H. Siccardi, T. Wang, V. Narang, R. Abzalimov, N. Wijerathne and R. V. Ulijn, *Nature nanotechnology*, 2016, **11**, 960-967.
143. D. B. Rasale and A. K. Das, *International journal of molecular sciences*, 2015, **16**, 10797-10820.
144. K. Tsuchiya and K. Numata, *Macromolecular Bioscience*, 2017, **17**.
145. J. M. Lehn, *Chemistry-A European Journal*, 1999, **5**, 2455-2463.
146. E. Moulin, G. Cormos and N. Giuseppone, *Chemical Society Reviews*, 2012, **41**, 1031-1049.
147. A. Herrmann, *Chemical Society Reviews*, 2014, **43**, 1899-1933.
148. H. Y. Au-Yeung, G. D. Pantos and J. K. Sanders, *The Journal of organic chemistry*, 2011, **76**, 1257-1268.
149. E. H. Bromley, R. B. Sessions, A. R. Thomson and D. N. Woolfson, *Journal of the American Chemical Society*, 2008, **131**, 928-930.
150. Y. Krishnan-Ghosh and S. Balasubramanian, *Angewandte Chemie*, 2003, **115**, 2221-2223.
151. J. M. Carnall, C. A. Waudby, A. M. Belenguer, M. C. Stuart, J. J.-P. Peyralans and S. Otto, *Science*, 2010, **327**, 1502-1506.
152. M. Colomb-Delsuc, E. Mattia, J. W. Sadownik and S. Otto, *Nature communications*, 2015, **6**.
153. Y. Altay, M. Tezcan and S. Otto, *Journal of the American Chemical Society*, 2017, **139**, 13612-13615.
154. M. A. Case and G. L. McLendon, *Journal of the American Chemical Society*, 2000, **122**, 8089-8090.
155. H. J. Cooper, M. A. Case, G. L. McLendon and A. G. Marshall, *Journal of the American Chemical Society*, 2003, **125**, 5331-5339.
156. P. G. Swann, R. A. Casanova, A. Desai, M. M. Frauenhoff, M. Urbancic, U. Slomczynska, A. J. Hopfinger, G. C. Le Breton and D. L. Venton, *Peptide Science*, 1996, **40**, 617-625.
157. A. K. Das, A. R. Hirsth and R. V. Ulijn, *Faraday Discuss*, 2009, **143**, 293-303; discussion 359-272.
158. J.-B. Guilbaud, E. Vey, S. Boothroyd, A. M. Smith, R. V. Ulijn, A. Saiani and A. F. Miller, *Langmuir*, 2010, **26**, 11297-11303.
159. L. A. Castillo Diaz, M. Elsawy, A. Saiani, J. E. Gough and A. F. Miller, *Journal of tissue engineering*, 2016, **7**, 2041731416649789.
160. Y. M. Abul-Haija and R. V. Ulijn, *Biomacromolecules*, 2015, **16**, 3473-3479.
161. R. J. Williams, T. E. Hall, V. Glattauer, J. White, P. J. Pasic, A. B. Sorensen, L. Waddington, K. M. McLean, P. D. Currie and P. G. Hartley, *Biomaterials*, 2011, **32**, 5304-5310.
162. C. Berdugo, S. K. M. Nalluri, N. Javid, B. Escuder, J. F. Miravet and R. V. Ulijn, *ACS applied materials & interfaces*, 2015, **7**, 25946-25954.
163. S. K. M. Nalluri and R. V. Ulijn, *Chem Sci*, 2013, **4**, 3699-3705.
164. S. Debnath, S. Roy and R. V. Ulijn, *J Am Chem Soc*, 2013, **135**, 16789-16792.
165. C. G. Pappas, I. R. Sasselli and R. V. Ulijn, *Angew Chem Int Ed Engl*, 2015, **54**, 8119-8123.
166. M. P. Hughes, M. R. Sawaya, D. R. Boyer, L. Goldschmidt, J. A. Rodriguez, D. Cascio, L. Chong, T. Gonen and D. S. Eisenberg, *Science*, 2018, **359**, 698-701.

1