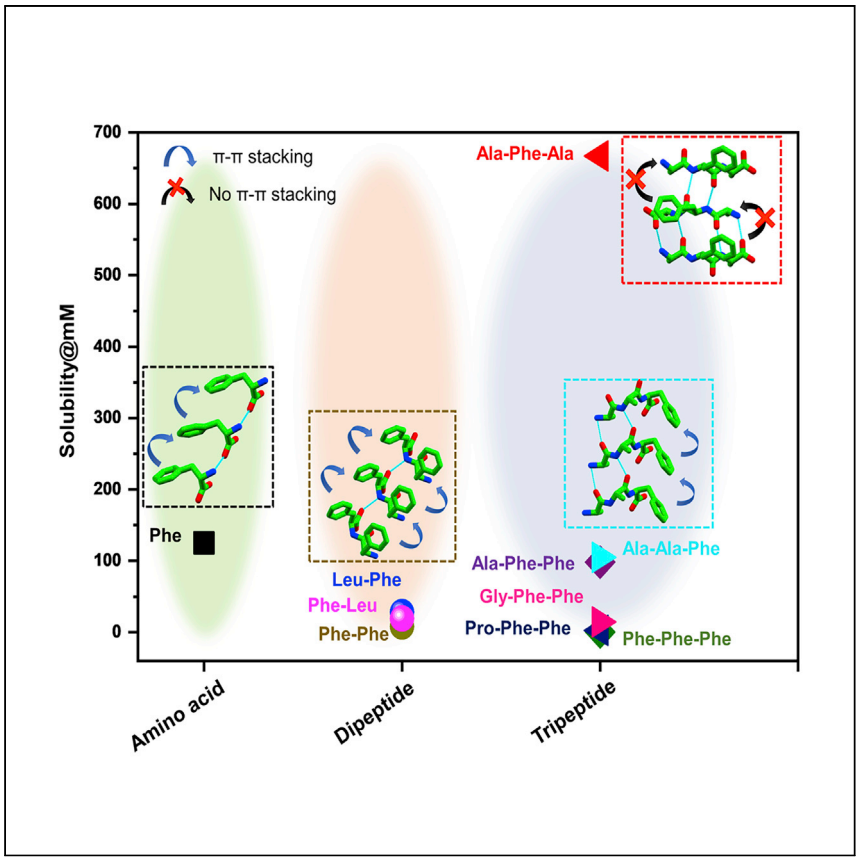
Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Solid-state packing dictates the unexpected solubility of aromatic peptides



The solubility of biomolecules in aqueous solutions is a central issue with regard to their physiological and pathological activities. Bera et al. provide insight into the solid-state packing-solubility relationship of biomolecules. This work may further extend the importance of solid-state biology for basic and applied science.

Santu Bera, Xuewei Dong, Bankala Krishnarjuna, ..., David E. Clemmer, Guanghong Wei, Ehud Gazit

ehudga@tauex.tau.ac.il

Highlights

Balance between solubility versus aggregation of aromatic peptides is examined

Significant differences in solubility are observed for analogous tripeptides

Solid-state packing is found to be underlying molecular basis for high solubility

Revealing the importance of biology at the solid state in physiology and pathology

Bera et al., Cell Reports Physical Science 2, 100391
April 21, 2021
https://doi.org/10.1016/j.xcrp.2021.100391

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Solid-state packing dictates the unexpected solubility of aromatic peptides

Santu Bera,[1] Xuewei Dong,[2] Bankala Krishnarjuna,[3] Shannon A. Raab,[4] David A. Hales,[4,5] Wei Ji,[1] Yiming Tang,[2] Linda J.W. Shimon,[6] Ayyalusamy Ramamoorthy,[3] David E. Clemmer,[4] Guanghong Wei,[2] and Ehud Gazit[1,7,*]

SUMMARY

The understanding and prediction of the solubility of biomolecules, even of the simplest ones, reflect an open question and unmet need. Short aromatic tripeptides are among the most highly aggregative biomolecules. However, in marked contrast, Ala-Phe-Ala (AFA) was surprisingly found to be non-aggregative and could be solubilized at millimolar concentrations. Here, aiming to uncover the underlying molecular basis of its high solubility, we explore in detail the solubility, aggregation propensity, and atomic-level structure of the tripeptide. We demonstrate an unexpectedly high water solubility of AFA reaching 672 mM, two orders of magnitude higher than reported previously. The single crystal structure reveals an anti-parallel β sheet conformation devoid of any aromatic interactions. This study provides clear mechanistic insight into the structural basis of solubility and suggests a simple and feasible tool for its estimation, bearing implications for design of peptide drugs, peptides materials, and advancement of peptide nanotechnology.

INTRODUCTION

Although extremely important, the problem of insolubility of peptides and proteins in aqueous solution is still far from being completely understood.[1–3] In nature, autonomous organization of disordered building blocks into ordered stable assemblies is a fundamental process.[4] Many major degenerative disorders share the common etiology of aggregation into amyloid fibrils by proteins and peptides.[5] Although, in more recent times, these nanostructures have found interesting applications as advanced materials for innovative medicine solutions, tissue engineering, renewable energy, environmental science, nanotechnology, and materials science,[6,7] the high propensity to assemble into insoluble aggregates is very common but poses a serious limitation to the production and use of many peptides in a wide range of biotechnological and pharmaceutical applications.[1,8] Increasing the aqueous solubility of poorly water-soluble drugs is the most frequent requirement in the pharmaceutical analysis and formulation fields. In principle, for a drug to exert its biological effect, it must be soluble in and transported by body fluids.[9] Many peptide-based drugs with great therapeutic potential are rendered ineffective during preclinical or clinical trials[1] simply because of the problems related to low water solubility or high propensity for irreversible self-assembly.

The self-assembly process is driven by a wide variety of weak non-covalent interactions,[10–12] including hydrogen bonding, van der Waals forces, Coulombic interactions, and hydrophobic effects. The rate of aggregation of biomolecules has been

[1]Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel

[2]Department of Physics, State Key Laboratory of Surface Physics, Key Laboratory for Computational Physical Sciences (MOE), Fudan University, Shanghai, 200433, People's Republic of China

[3]Biophysics Program, Department of Chemistry, Macromolecular Science and Engineering, The University of Michigan, Ann Arbor, MI 48109-1055, USA

[4]Department of Chemistry, Indiana University, 800 Kirkwood Avenue, Bloomington, IN 47401, USA

[5]Department of Chemistry, Hendrix College, Conway, AR 72032, USA

[6]Department of Chemical Research Support, Weizmann Institute of Science, Rehovot 7610001, Israel

[7]Lead contact

*Correspondence: ehudga@tauex.tau.ac.il

https://doi.org/10.1016/j.xcrp.2021.100391

shown to be altered significantly by the presence of metal ions.[13–16] Although aggregation does not occur under normal physiological conditions, it can be induced in the presence of certain metal ions, such as copper(II), zinc(II), and iron(II). However, among these interactions, π-π stacking between aromatic residues of peptide side chains has been postulated to be the major driving force for early recognition and acceleration of peptide self-assembly.[17–19] A recent *in silico* analysis aimed to predict formation of nanostructures by *de-novo*-designed short peptides has suggested aromatic tripeptides to be the most aggregation-prone biomolecules.[20] In accordance, the aromatic-rich small peptide segment ^{16}KLVFFAE^{22} was designated as the nucleating core of the β-amyloid (Aβ) peptide because of high aggregation propensity in producing insoluble fibrillar deposits with cross-β architecture, the hallmark of the full-length sequence.[21,22] In addition, the Phe residues play an important role in formation of key Aβ intermediates with a disordered 3_{10} helical structure.[23] The importance of aromatic moieties is also evident by the abundance of aromatic residues in amyloid-forming short peptide sequences as well as by the proposed design rule of amyloid inhibitors to include building units that specifically target the aromatic residues.[24–26] Moreover, applying a reductionist approach, FF, FFF, and their derivatives have been shown to be strong promoters of peptide aggregation into a variety of nanoarchitechtures.[27–29] Replacement of the aromatic residues with nonaromatic surrogates has been found to completely abrogate or attenuate aggregation, further signifying the role of aromatic moieties in peptide self-asssembly.[30–32]

The solubility of a building unit and the rate of its association process are inversely related to each other. Physiochemically, the degree of solubility depends on the power of peptide-solvent cross talk, the types and intensities of electrostatic and hydrogen bonding, and van der Waals interactions between the peptides and water molecules at the expense of peptide-peptide interactions. Protein and peptide aggregation is also strongly dependent on the concentration of the aggregating biomolecules because of their solubility limits. At the initial stage of the aggregation process, the biomolecular aggregates are completely soluble but gradually become insoluble while enhancing protein association as they exceed solubility limits, resulting in protein/peptide precipitation.[33,34] However, significantly different solubility of analogous peptide sequences with similar polarities and numbers of probable H-bonding sites remains a mystery. In this context, uncovering the underlying features that determine the ultimate solubility of biomolecules is of the utmost importance for basic science as well as for design of efficient peptide drugs.

In contrast to the commonly observed self-assembly propensity of aromatic short peptides, manifested in fast and efficient aggregation into ordered nanostructures in aqueous solution, AFA has been found, surprisingly, to be non-aggregative and quite highly soluble, with a reported critical concentration of up to 10 mM.[35–37] Primary study of its solution structure via solution NMR showed a combination of inverse γ turn and β strand conformations.[36] Follow-up studies revealed that AFA exists as an ensemble of rapidly interconverting conformers in aqueous solution,[37] but its unexpected solubility remains unexplained.

Here we extensively explore the solubility, aggregation propensity, and solid-state structure relationship of AFA and compare them with those of analogous peptides. We uncover that AFA can be solubilized up to a concentration of 672 mM at room temperature and 720 mM at 90°C. These values are unexpectedly high, considering the solubility of aromatic short peptides,[2,38] and two orders of magnitude higher than the previously reported solubility of AFA. The solid-state structure reveals

formation of an antiparallel β sheet organization surprisingly devoid of any aromatic π-π stacking between molecular units. Simple sequence shuffling results in decreased solubility by an order of magnitude for AAF (105 mM) and FAA (42 mM). The crystal structure of AAF demonstrates transformation of the molecular orientation into a helical structure with strong face-to-face π-π stacking interactions. In addition, screening of a focused peptide library at different length scales further confirms the aromatic stacking-dependent solubility of aromatic peptides. Our work clearly establishes that the solubility of analogous short peptides with similar polarity and H-bonding sites is determined by solid-state packing. These results could further advance our understanding of peptide solubility and pave the way for modular design of peptide drugs with high efficiency.

RESULTS AND DISCUSSION

Critical soluble concentration of tripeptide

We first established the threshold concentration of the AFA tripeptide that could be solubilized in pure water at room temperature and physiological pH of 6.8 (Figure 1A). Under these conditions, the maximum solubility of AFA was found to be 672 mM, unexpectedly high compared with other well-known short aromatic peptides.[2,38] The ultraviolet-visible (UV-vis) spectra at different concentrations indicated no major shift of the peak position, emphasizing the appearance of a yellowish color because of the extremely high concentrated solution of peptides (Figure S1). Next, to understand the thermodynamics of solubility, we investigated the dependence of critical concentration on temperature in the range of 25°C–100°C with a minimum interval of 10°C (Figure 1A). The solubility increased gradually with temperature and showed a steep climb to 702 mM at 70°C. Near the boiling point of water, the solubility reached 710 mM. This variation of solubility with temperature indicated the existence of heat capacity, and the corresponding enthalpy and entropy change could be realized from a van't Hoff relation.[39] The favorable free energy could arise from various interactions, such as hydrogen bonding and the presence of salt bridges. The aromatic side chain could also be engaged in an extended π-stacked structure that would also cause favorable free energy.[40] However, these types of interactions have only a moderate temperature dependence, and the temperature dependence of solubility or aggregation has been reported to most likely stem from the contribution of the hydrophobic effect of the aromatic side chains,[41] suggesting that such an effect takes place in AFA. Next, the role of electrostatic interactions was explored by measuring the solubility at different pH values and at room temperature (Figure 1B). The minimal critical concentration was found around pH 6–7 (668 mM), which is the corresponding pH of AFA lyophilized powder in pure water. At this pH, the predominant species in solution is the globally neutral, zwitterionic AFA monomer. As the pH of the solution approached the pK_a value of either terminus resulted rapidly increasing concentration of charged monomers; however, this increased the solubility only very slightly (670 mM and 672 mM). The relatively strong effect of temperature and weak effect of pH on the critical concentration of AFA can be rationalized on the basis of relatively weak absolute values of electrostatic driving forces, which are diminished by the strong hydrophobic effect of the aromatic side chains. At the same time, because the peptide molecules carry one positive and one negative charge, electrostatic interactions could also be involved in salt bridge formation.[40] To verify the importance of these electrostatic interactions, the solubility was investigated at different ionic strengths, adjusted by increasing NaCl concentrations from 1 mM to 2 M (Figure 1C). The results showed that the solubility remained constant for a large range of NaCl concentrations and started to change only at a very high NaCl concentration of 500 mM or above.

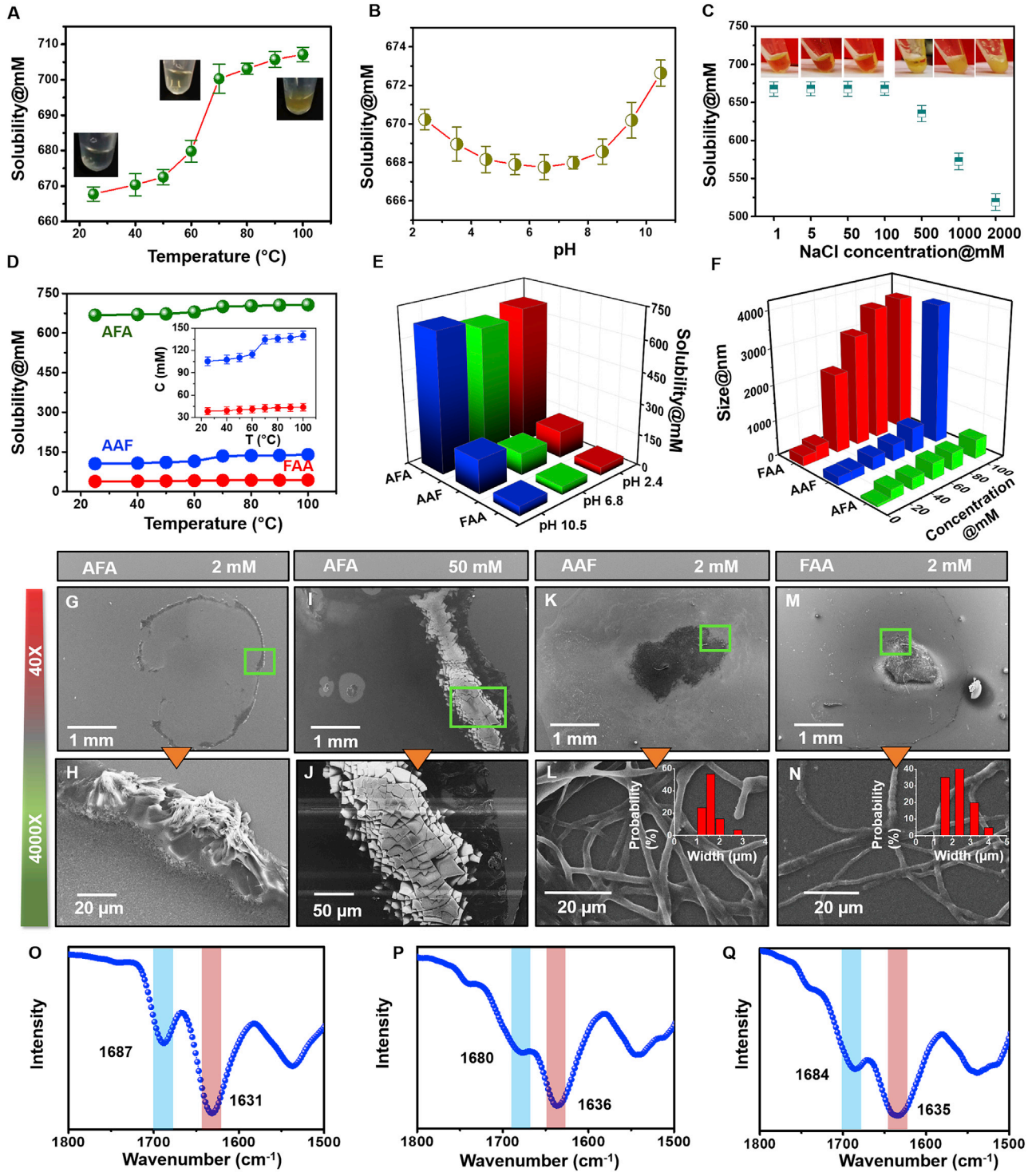


Figure 1. Solubility versus self-assembly of tripeptides

(A and B) Water solubility of AFA relative to (A) temperature and (B) pH of the solution. The insets in (A) show images of the solution at three different temperatures. The data represent the mean ± SEM for 3 independent experiments.

(C) Solubility of AFA in the presence of different concentrations of NaCl. The data represent the mean ± SEM for 3 independent experiments.

(D and E) Comparison of the critical concentrations of AFA, AAF, and FAA at different (D) temperature and (E) pH values. The inset in (D) shows a magnified graph for AAF and FAA.

This observation clearly indicates that the effect of NaCl is related to the well-established "salting out" effect of hydrophobic molecules, further supporting no significant role of the electrostatic effect in solubility of AFA.[42,43] Moreover, using different types of salts having altered cations or anions also had a similar effect, suggesting that simple electrostatic screening was not much responsible for the solubility of AFA (Figure S2). Therefore, the hydrophobic and aromatic interactions play the most striking role in maintaining the favorable energy landscape for the exceptional high solubility of AFA.

To further understand the role of the hydrophobic effect, we shuffled the sequence of the native peptide and designed two other possible combinations, AAF and FAA. Surprisingly, although displaying similar hydrophobicity, the two modified tripeptides exhibited solubility of nearly one order of magnitude lower compared with the parent peptide (105 mM for AAF and 42 mM for FAA) under similar conditions (Figure 1D). The solubility of both modified peptides revealed a similar dependency on temperature, although, even at the highest measured temperature, the critical concertation was found to be nearly 6-fold and 15-fold lower for AAF and FAA, respectively. The critical soluble concentration was explored at different pH values to understand the effect of electrostatic interactions (Figure 1E). However, similar to AFA, the temperature dependence stemming from hydrophobic or aromatic effects was found to be more significant than the driving forces imparted by electrostatic interactions. To obtain the size of existing clusters in the solution, dynamic light scattering (DLS) experiments were performed with a range of sub- and supercritical concentrations in water (Figures 1F and S3). Below 40 mM, for all three peptides, the scattering signal was dominated by monomeric or small oligomeric species with a diffusion coefficient corresponding to an equivalent size of approximately 200–400 nm. Above 40 mM, the cluster size of AFA and AAF remained constant, whereas FAA showed a presence of larger insoluble subcritical structures approximately 2000 nm in size. These significantly larger clusters dominated the scattering intensity by the Rayleigh scattering relation.[41] When the concentration reached 100 mM, AAF showed formation of insoluble large prenucleation clusters or micelles approximately 4,000 nm in size, indicating insolubility and an aggregation process. However, the size of clusters formed by AFA persisted, signifying the presence of non-aggregative monomeric species even at this high concentration or above.

Solubility and self-assembly/aggregation are opposite processes; the higher the solubility, the lower the aggregation. Support for this notion was obtained from the DLS results because AFA failed to form any higher-ordered structures even at very high concentrations. Aggregation was followed by investigating the nanostructures produced by the three peptides in solution. Scanning electron microscopy images showed AFA to mainly exist as monomers or small oligomers even after allowed to self-assemble for 2 weeks. Large-scale images capturing the entire area of solution drop clearly indicated the absence of any kind of ordered nanostructure (Figures 1G, 1H, and S4). During the drying process, the soluble AFA monomers were mainly deposited in a ring-like fashion along the original drop edge because of a coffee ring effect. Magnification of the edge of these drops showed deposition of unordered particulate materials resulting from coalescence of monomers. Upon increasing

(F) Calculated Stokes hydrodynamic diameter distribution of three tripeptide clusters in solution at different concentrations at pH 6.8 and room temperature.
(G–N) Scanning electron microscopy images of the resultant architectures after 2 weeks of self-assembly of (G–J) AFA, (K and L) AAF, and (M and N) FAA. The top panels show the entire area of the solution drop, and the corresponding magnified areas marked by green squares are shown in the bottom panels. The insets in (L) and (N) show the probability distribution of fiber width.
(O–Q) FTIR spectra of (O) AFA, (P) AAF, and (Q) FAA.

the concentration, larger crystalline clusters were found to deposit; however, no indication of formation of ordered assembled structures was observed (Figures 1I and 1J). On the other hand, AAF and FAA exhibited formation of well-ordered and closely entangled networks of fibers consisting of filaments that bundled into larger twisted fibrils because of self-assembly of small nucleating sites even at lower concentrations (Figures 1K–1N and S5). The average diameter of the filaments was approximately 2–3 μm with only a minor variation between the tripeptides, as calculated using a histogram. Therefore, the significant differences in solubility and aggregation of similar hydrophobic analogous peptides, along with an observed minimal effect of electrostatic interactions, most likely indicate that molecular structure, atomic-level packing, and aromatic stacking might play a pivotal role in solubility.

To characterize the secondary structure of the tripeptides, we used Fourier transform infrared (FTIR) and circular dichroism (CD) spectroscopy techniques. For AFA, the FTIR spectra at solid state manifested a sharp amide I peak at 1,631 cm^{-1} with a shoulder at 1,687 cm^{-1}, indicating the predominant β sheet structure (Figure 1O).[44] However, the amide I peak shifted to 1,636 cm^{-1} and 1,635 cm^{-1} for AAF and FAA, respectively, indicating structural transition toward a helical conformation (Figures 1P and 1Q). The CD spectrum of AFA at room temperature showed a clear positive ellipticity in the far-UV region with distinct double maxima at 197 and 217 nm, in accordance with previous studies (Figure S6A),[36,37] and provide support for a dominant β strand conformation. The CD spectrum of AAF showed similar double maxima with positive peaks at the same positions but a significantly different intensity ratio compared with the CD spectra of AFA (Figure S6D). Apparently, the corresponding differences in intensity ratios for AFA and AAF indicate different dihedral angles.[35] FAA exists preferably in a polyproline II (PPII)-type turn conformation because the CD spectra exhibited a negative maximum at 197 nm and a positive peak at 218 nm (Figure S6G).[35] To gain knowledge regarding the effect of temperature and pH on molecular conformation, CD spectra were acquired at different pH values (2.4–10.5) and temperatures (10°C–90°C) (Figure S6). Identical spectra were obtained at all the studied pH values, indicating conformation rigidity in the studied pH range. Upon increasing the temperature, the intensity of the CD signal decreased very slightly, likely because of partial unfolding. Moreover, our earlier study showed that ROESY spectra acquired in the presence of 10 mM to 1 M NaCl and at pH 1.2 and 12.1 did not significantly alter the NOE pattern,[36] further confirming the limited influence of the charged terminal groups on molecular conformation and solubility.

Solid-state packing of tripeptides

Next, to identify the most stable molecular conformation and preferred packing of the aromatic moieties, we grew diffraction-quality single crystals of AFA and AAF under similar conditions as used for other studies and analyzed their structure. Unfortunately, we were unable to produce diffractable single-crystal of FAA. The fundamental mechanisms underlying the oriented organization process into hierarchical self-assembly and macroscopic crystallization still remain unclear.[12] Different kinds of nanostructures and morphologies are stabilized according to the balance between noncovalent acceptor (hydrogen-bond acceptor, HBA) ability, polarity, surface tension, and dielectric constant, which fuse or dissociate to produce macroscopic crystals. Thus, we speculate that a correlation of the differences in rate of aggregation at nanoscale within specific timescale and macroscopic crystallization between FAA, AFA, and AAF is not feasible. The single-crystal structure of AFA revealed that two peptide molecules crystallized with water, acetic acid, and trifluoroacetic acid, one molecule of each, in the asymmetric unit of the monoclinic

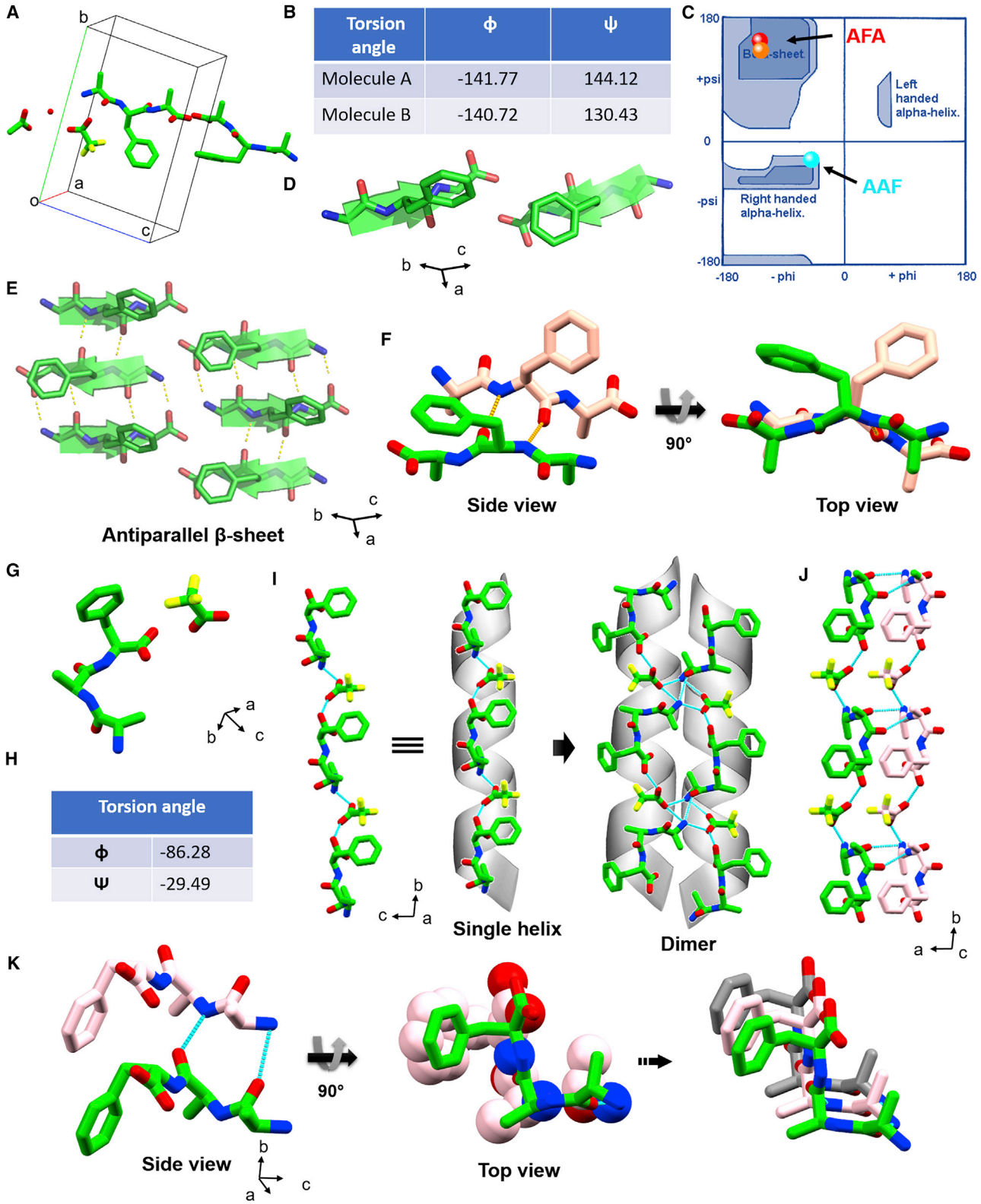


Figure 2. Solid-state structures of AFA and AAF

(A–F) AFA (CCDC:1834550):[28] shown are (A) the asymmetric unit; (B) the torsion angle of Phe_2 for the two different molecules present in the asymmetric unit; (C) a Ramachandran plot of the allowed dihedral angles of Phe_2 residues in red and orange, showing preference for the β sheet conformation; (D)

$P2_1$ space group (Figure 2A).[28] In the solid-state structure, all counterions were present with the peptide at a 1:1 ratio as part of their solid-phase synthesis and included in the crystals because of their role in packing; therefore, the solution should also contain equimolar concentrations of the tripeptide and its counterion. The two peptide molecules adopted a similar conformation with no intramolecular H-bond in the peptide molecules, which refutes the possibility of a γ turn structure. The torsion angles around the Phe_2 residue appeared to play a pivotal role in determining the overall structural features. The allowed torsion angles of the Phe_2 residue were constrained within the β sheet region of the Ramachandran plot,[45] with φ_2 and ψ_2 values of −141.77° and 144.12° for molecule A and −140.72° and 130.43° for molecule B, respectively (Figures 2B–2D). The tripeptide molecules exhibited two conventional $NH_3^{+\cdots}{}^{-}OOC$ head-to-tail hydrogen bonds, with the third amine hydrogen involving the acetic acid or trifluoroacetic acid moiety. In the crystallographic *a* direction, β strands were stacked in an antiparallel manner and interacted with the adjacent strands through intermolecular H-bonds incorporating terminal polar groups and internal amide groups, producing an antiparallel β sheet conformation (Figures 2D and 2E). The aromatic side of the Phe residue of one β strand resided in the opposite side of the next interacting strand, and, surprisingly, no aromatic π-π interactions were formed in the β sheet conformation (Figure 2F). Moreover, nearby β sheets were connected through head-to-tail H-bonds between polar head groups in the crystallographic *b* direction, generating a 1D layer where all aromatic residues were positioned on the same face of the layer (Figure S7). The individual layers stacked to afford a layer-by-layer structural organization of AFA stabilized through van der Waals interactions. However, π-π interactions between Phe residues were also not observed within cross-sheets along the *b* direction. The closest distance between two Phe residues was 9.5 Å (Figure S7), considerably higher than the defined distance for π-π interaction (4.5–7 Å).[46] Therefore, the solid-state structure of AFA is a clear exception from the structures of other aromatic short peptides, which are mainly stabilized by strong aromatic π-π interactions.[46] This special arrangement of the aromatic ring and the absence of π-π stacking could be the molecular basis of the exceptional solubility of AFA.

The modified tripeptide AAF crystallized with one peptide molecule and one trifluoroacetic acid molecule in the asymmetric unit of the orthorhombic $P2_12_12_1$ space group (Figures 2G–2K, Table S1). Shuffling the Phe moiety position from the central residue to the C terminus completely altered the molecular conformation of AAF compared with the native peptide. The allowed torsion angles of the Ala_2 were constrained in the right handed α-helical region of the Ramachandran plot, with φ_2 and ψ_2 values of –86.28° and −29.49°, respectively (Figure 2H). In the crystallographic *b* direction, the AAF molecules were connected via head-to-tail H-bonds through trifluoroacetic acid molecules and produced a single helical structure (Figure 2I). Two nearby helices connected by H-bonds in an antiparallel fashion in the *c* direction fabricated the helical dimer (Figure 2I). Adjacent dimers were mated through a hydrophobic zipper-like structural interaction exerted by Phe residues and the trifluoroacetic acid side chain, creating a helical sheet-like organization (Figure S8).[47,48] Most importantly, in the perpendicular direction (along the *a* axis), individual helices were stacked in parallel orientation to produce a single helical strand (Figure 2J). Together with H-bonds, strong aromatic face-to-face π-π interactions played a

the β sheet structure of AFA; (E) H-bonded antiparallel β sheet arrangement in higher-order packing; and (F) the absence of any π-π stacking interactions between two adjacent molecules.

(G–K) AAF: shown are (G) the asymmetric unit, (H) the allowed torsion angle of Ala_2, (I) helical structure formation by head-to-tail H-bond connection and the helical dimer, (J) stacking of adjacent helices via H-bond and aromatic interactions, and (K) strong face-to-face π-π stacking in the structure.

pivotal role to stabilize the helical strands (Figures 2K and S8). The centroid-to-centroid distance of Phe residues between two nearby AAF molecules was 5.32 Å in the lower region of the proposed π-π stacking distance.[46] Therefore, in accordance with other reported aromatic short peptide structures, the presence of aromatic π-π interaction in AAF solid-state packing is probably the major influential factor that determines the lower solubility of AAF compared with AFA. Although we were unable to produce a diffraction-quality single crystal of FAA, the powder X-ray pattern of FAA exhibited peaks at similar 2θ value positions with AAF (Figure S9). This indicates a similar atomic-level orientation, equatorial and longitudinal distances between molecular units, and higher-order organization of FAA with structural arrangement of AAF, which also resulted in a similar lower solubility.

Effect of aromatic amyloid inhibitors on solubility

Polyphenols such as epigallocatechin-3-gallate (EGCG), tannic acid (TA), and rosmarinic acid (RA) have been well studied as protein aggregation modulators that inhibit formation of amyloid aggregations as well as dissociate preformed fibrils by preferentially and reversibly binding to the structures, breaking the initial ordered pattern of two polymers and greatly reducing their β sheet content and enlarging their conformational space.[3,49–52] The aromatic rings of these polyphenols appear to play a key role by interfering in amyloid aggregation and partially replacing the original amino acids in π-π stacking. To understand the effect of inhibitors in structural disruption-mediated modulation of tripeptide solubility, critical concentration was measured in the presence of different ratios of EGCG, TA, and RA. The presence of inhibitors at molar ratios ranging from 1:0.1–1:0.3 was unable to significantly alter the solubility of AFA, probably because of the inability of the inhibitors to disrupt interpeptide π-π stacking (Figure 3A). However, we observed a significant increase in critical soluble concentration of AAF and FAA in a dose-dependent manner in the presence of polyphenols (Figures 3B and 3C). The AAF solubility of 105 mM without inhibitor increased sharply to 130 mM and 210 mM in the presence of increasing amounts of EGCG (1:0.25–1: 0.5). The lower water solubility of EGCG hindered a further increase of its molar ratio. The effect of TA was found to be extremely potent because the solubility of AAF could be increased by 3-fold (340 mM), whereas the ratio reached 1:2 in favor of TA. Addition of RA was also caused substantial improvement of AAF solubility to 263 mM by the occurrence of a same ratio of RA. A similar sharp increase in critical soluble concentration of FAA was also observed by increasing the concentration of all three inhibitors interference with peptide stacking. The results clearly demonstrate the dependence of solubility of the low-soluble aromatic peptides, but not soluble AFA, on the π-π stacking interactions because increasing the stacking-disrupting agent enables a gradual increase in solubility by intercalating peptide aggregation and partial replacement of the peptide molecules in higher-order π-π stacking.

Aggregation characterization using ESI-IMS-MS

In recent years, important advances in mass spectrometry (MS) have allowed the phenomenon of biological macromolecule aggregation to be uniquely characterized.[53–55] Combination of MS with orthogonal techniques such as ion mobility spectrometry (IMS) manifests a new dimension that also yields insight into the gas-phase structural composition of the corresponding parent ions. When used in conjunction with electrospray ionization (ESI), ESI-IMS-MS has provided interesting structural information of building units in the gas phase.[56–59] To understand aggregation of higher-order oligomers and non-aggregative characteristics, the tripeptide samples were run on a homebuilt linear 4-m drift tube IMS paired with a time-of-flight (ToF) mass spectrometer (Figure S10). The ESI mass spectrum of a 10-mM water solution of AAF at pH 6.8 manifested a predominant presence of higher-order aggregates of different charge

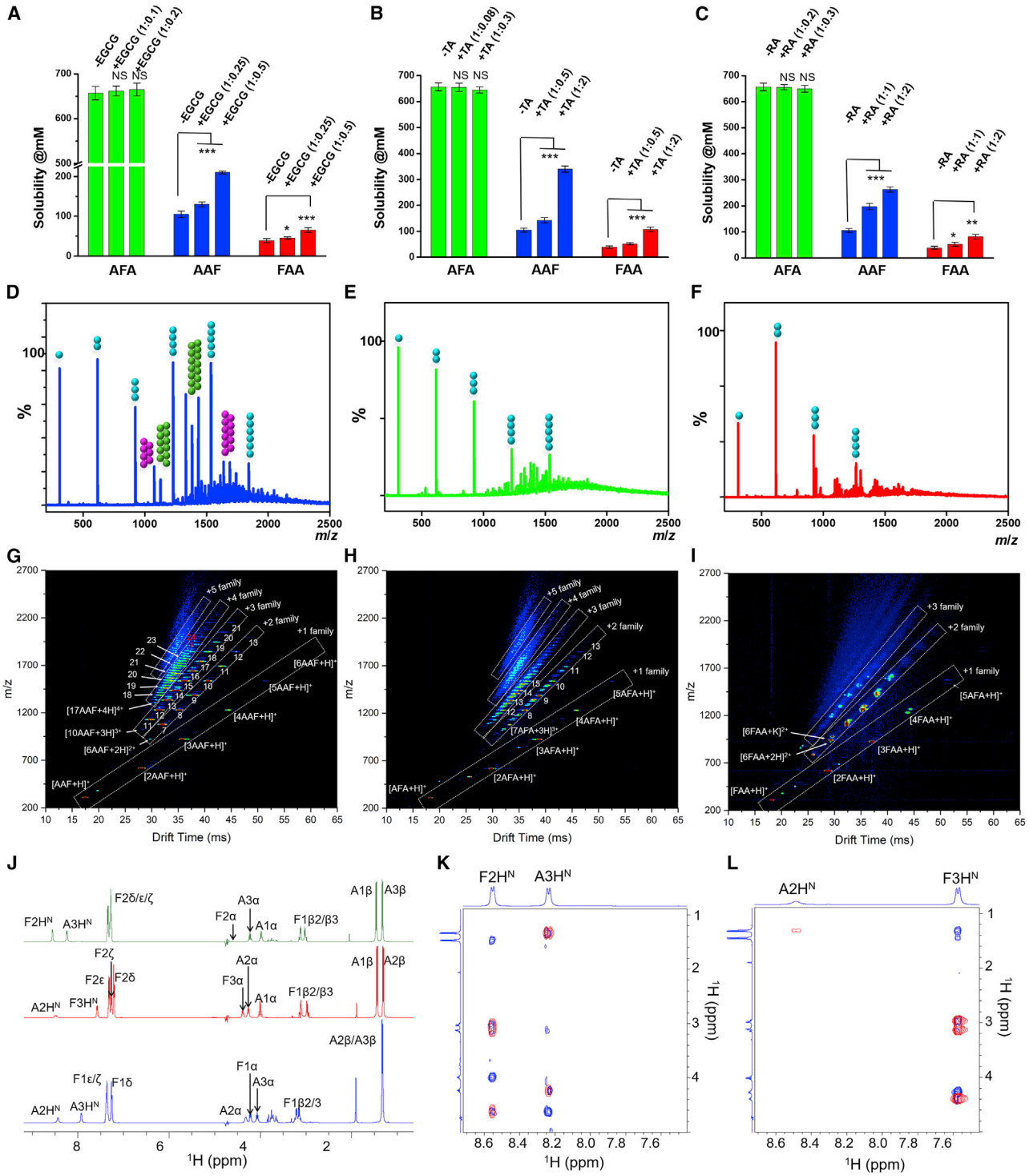


Figure 3. Structure-dependent solubility of the tripeptides

(A–C) The effect of the polyphenols EGCG (A), TA (B), and RA (C) on modulation of solubility of the tripeptides because of interference with structural organization. Statistical distribution and p values of tripeptides solubility by the inhibitors compared with their absence is shown. The data represent the mean ± SEM for 3 independent experiments. N.S., not significant; *p < 0.05, **p < 0.005, ***p < 0.001.

(D–F) ESI-MS mass spectra of (D) AAF, (E) AFA, and (F) FAA (10 mM in water, pH 6.8, 25°C). Circles above the peaks denote the oligomer order. The number of circles represents the number of monomers in the aggregate. +1, +2, and +3 charge states are represented as cyan, pink, and green circles, respectively.

families, as shown in Figures 3D and S11. For the +1 charge state, monomeric and oligomeric species were the most abundant, followed by dimers through pentamers, with traces of hexamer ions. High-intensity peaks observed for +2 (heptamer through undecamer), +3 (decamer through hexadecamer), and +4 (heptadecamer and nonadecamer) charge states indicated strong aggregation. A typical 2D plot of the data (a nested drift-[flight] time distribution), recorded upon ESI of the AAF solution, is shown in Figure 3G and S12. As reported in earlier studies, for amino acids and peptides, peaks fall linearly into families according to their charge states.[56,57] Under the experimental conditions employed here, AAF showed a multitude of higher-order species comprising five distributions of charge states: $[n(AAF)+H]^{+}$ (n = 1–6, where n corresponds to the number of monomer units), $[n(AAF)+2H]^{2+}$ (n = 6–13), $[n(AAF)+3H]^{3+}$ (n = 10–21), $[n(AAF)+4H]^{4+}$ (n = 17 to 24), and $[n(AAF)+5H]^{5+}$ (Table S2). The extensive population of the +4 and higher charge families indicated a strong aggregation propensity to form higher-order oligomers that finally assembled into fibrillary structures, as observed in scanning electron microscopy images (Figure 1L). In sharp contrast, the ESI mass spectrum of AFA under similar experimental conditions exhibited a much lower affinity of monomeric conformers for aggregation, as shown in Figures 3E and S11. The spectrum showed a predominant presence of the +1 charge state, with a lesser intensity of dimers and trimers along with only a tiny amount of tetramers and pentamers. Although AFA was found to form clusters up to the +5 charge state, prominent intensity was observed only up to oligomers of +3: $[n(AFA)+H]^{+}$ (n = 1–5), $[n(AFA)+2H]^{2+}$ (n = 7–13), and $[n(AFA)+3H]^{3+}$ (n = 12–15) (Figures 3H and S12; Table S3). The ESI-IMS-MS data presented clear differences between the spectral "fingerprints" of AAF undergoing strong aggregation into higher-order clusters, whereas AFA epitomized significantly lower aggregation propensity. This further clarified the behavior of AFA, which existed mainly as soluble monomeric conformers in solution and showed reluctance to form any ordered nanostructures (Figure 1G). Although FAA was expected to show a spectral pattern and ESI-IMS-MS plot similar to aggregated AAF, the traces of higher-order oligomers could not be detected clearly (Figures 3F, 3I, S11, and S12; Table S4). Under the applied experimental concentration, FAA converted into a gel-like, semi-solid matrix, which might affect the experimental results and the exact proportions of higher-order oligomeric structures compared with analogous tripeptides.

Solution structure and aggregation analysis by NMR spectroscopy

To probe the molecular conformation of the tripeptides in solution and ensure a similar molecular organization with the solid-state structure, comprehensive NMR spectroscopy studies were carried out in 9:1 $H_2O/^2H_2O$ solution at 10 mM concentration and pH ~7. 1D ^1H spectra showed a single set of resonances for all three tripeptides (Figure 3J). The NMR assignments were inferred from 2D [^1H-^1H]-TOCSY, 2D [^1H-^1H]-NOESY, [^{13}C-^1H]-HSQC, and [^{15}N-^1H]-HSQC spectra (Figures 3K, 3L, S13, and S14). The amide temperature coefficients were calculated from NMR spectra recorded at different temperatures ranging from 5°C–30°C (Figure S15). The amide temperature coefficients for Ala_2 (FAA), Phe_2 (AFA), Ala_2 (AAF), and Phe_3 (AAF) were less than 5 ppb/K (Table S5), suggesting their involvement in hydrogen bonding. The chemical shifts and $^3JH^N\alpha$ values were measured from 1D

(G–I) ESI-IMS-MS 2D plot of the (G) AAF, (H) AFA, and (I) FAA monomers through to oligomers. ESI-IMS-MS 2D plots show the IMS drift time versus mass/charge (m/z) versus intensity (z, square-root scale). The different charge families are marked with white boxes. The numbers indicate the corresponding oligomeric states (for details, see Supplemental information).
(J) One-dimensional ^1H NMR spectra of tripeptides (FAA, blue; AAF, red; AFA, green) dissolved in 90% H_2O/10% ^2H_2O, with assignments labeled.
(K and L) 2D [^1H-^1H]-NOESY (blue) and [^1H-^1H]-TOCSY (red) NMR spectra of AFA (K) and AAF (L). For clarity, only the fingerprint region of TOCSY/NOESY spectra is shown, and the assignments are labeled.

and 2D NMR spectra to obtain evidence regarding the nature of the secondary structures (Table S5). For AAF, the measured $^{3}JH^{N}\alpha$ values for Ala_2 (5.8 Hz) and Phe_3 (8 Hz) designate a helical-turn-like and an extended conformation for Ala_2 and Phe_3, respectively, showing the existence of a molecular organization in solution similar to that observed in the solid state. The $^{3}JH^{N}\alpha$ values for the residues of AFA were between 6 and less than 8 Hz, indicating that the peptide has a disordered/extended conformation, which is in accordance with the solid-state structural arrangement. FAA shows $^{3}JH^{N}\alpha$ values in a range similar to that of AFA. Additionally, for AAF, a large H^{N} chemical shift dispersion was observed compared with FAA and AFA, supporting the conformations mentioned above for these peptides (Figures 3J–3L and S13). Further, the self-association of tripeptides under these conditions was monitored by measuring translational diffusion coefficient measurements using DOSY NMR experiments (Figure S16). However, the measured diffusion coefficients were found to be quite similar for all three peptides, with hydrodynamic radii between 1.18–1.28 nm (Table S6), suggesting the absence of higher-order aggregates in solution. DOSY spectra were also recorded for FAA and AAF at a higher concentration of 40 mM. Although AAF remained entirely in solution at this concentration, FAA produced larger subcritical aggregates that came out of the solution, leaving monomers and small oligomers in solution, resulting in similar diffusion coefficient values observed at a lower concentration. Also, identical chemical shifts were observed at different concentrations (10 and 40 mM) and at different time intervals (day 1 and day 28) (Figure S17), further showing the existence of mainly non-aggregative units in solution in an isotropic environment.

Molecular dynamics simulations

Molecular dynamics (MD) simulations are used widely to dissect the mechanisms of peptide aggregation and interactions.[60–62] To understand the distinct solubility of AAF, FAA, and AFA at the molecular level, we performed all-atom MD simulations on systems consisting of 60 AAF, FAA, or AFA molecules in aqueous solutions (Figures 4, S18, and S19). All simulations started from randomly disordered states (Figures 4A–4C). Within 300 ns, AAF and FAA molecules aggregated into a large and relatively compact cluster with only a few molecules dispersed in solution and the compact aggregate maintained until the end of the simulation (Figures 4A and 4B). In contrast, AFA molecules first aggregated into a large but loosely packed cluster that split into two separate clusters at t = 500 ns (Figure 4C).

To characterize the shape of the aggregates, we calculated the probability density function (PDF) of RgY/RgX and RgZ/RgX of the largest assemblies, where RgX, RgY, and RgZ are the radii of gyration (Rgs) of a nanoassembly with respect to the three principal axes x, y, and z. RgY and RgZ of AAF were almost the same but larger than RgX, which corresponds to a prolate aggregate, suggesting elongation potential of AAF during fibril formation (Figure 4D). As for FAA and AFA, RgY and RgZ of the aggregates were much closer to RgX, indicating formation of sphere-like nanoaggregates (Figures 4E and 4F). Comparison of the cluster number and the largest cluster size in the three systems showed that AAF and FAA formed much larger aggregates than AFA, indicating their stronger aggregation propensity and lower solubility (Figure 4G). The smaller solvent-accessible surface area (SASA) of AAF and FAA reveals that more AAF and FAA molecules assemble into aggregates, consistent with their stronger aggregation propensity compared with AFA. More importantly, the side chains of the Phe residues in AAF and FAA showed a smaller solvent-exposed surface than those in AFA, indicating that the aromatic rings of AAF and FAA play a major role in aggregate formation, in contrast to AFA (Figure 4H). As shown in Figure 4I, AAF and FAA form more inter-peptide H-bonds

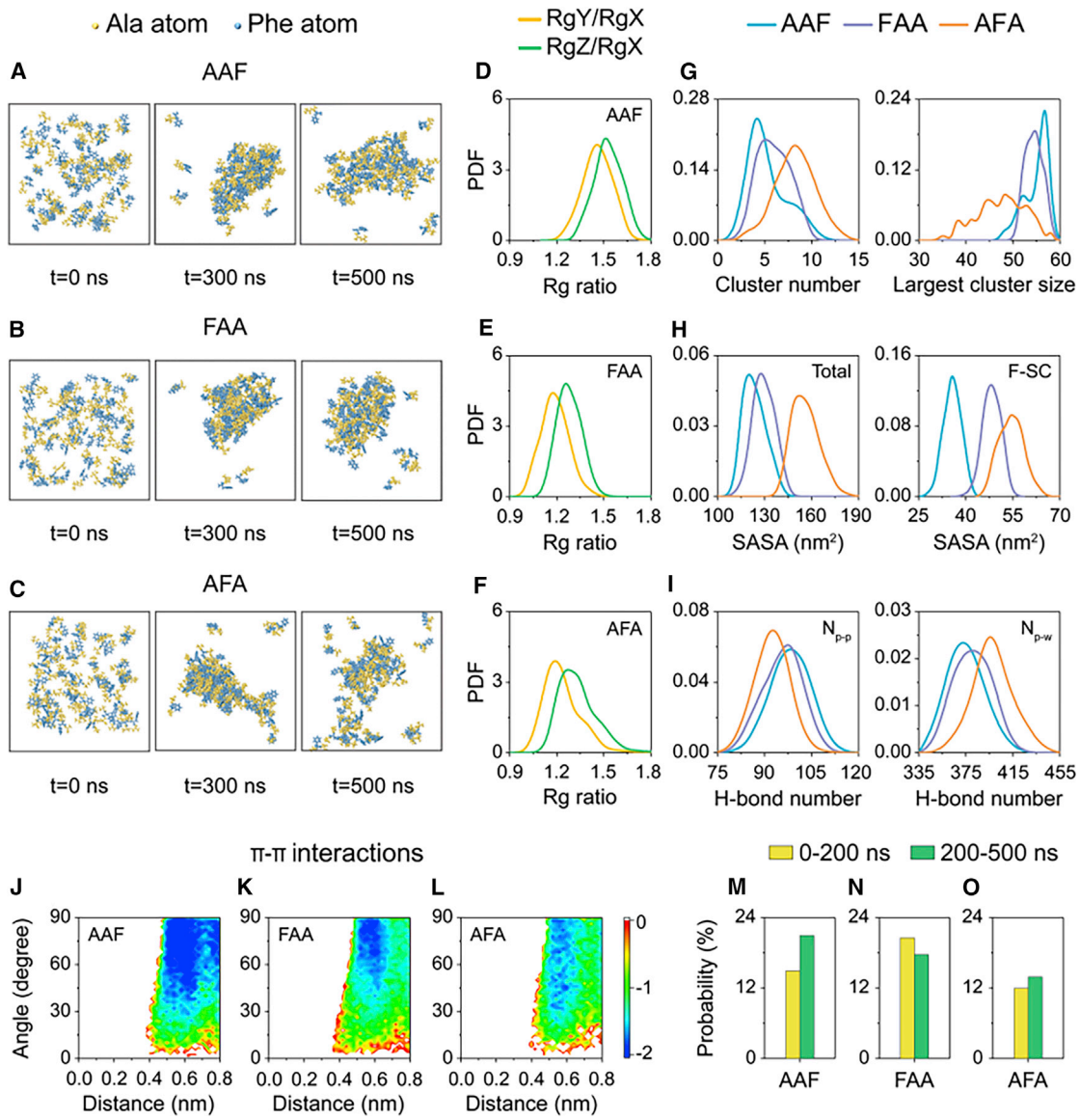


Figure 4. Distinct solubility of AAF, FAA, and AFA probed by all-atom MD simulations

(A–C) Snapshots of (A) AAF, (B) FAA, and (C) AFA aggregates at three time points. Tripeptide molecules are shown in van der Waals (VDW) representation, with Ala in yellow and Phe in blue.

(D–F) Probability density function (PDF) of RgY/RgX and RgZ/RgX of the largest aggregates, where RgX, RgY, and RgZ are the radii of gyration (Rg) of a nanoassembly with respect to the three principle axes x, y, and z.

(G–I) PDF of the number of clusters and the size of the largest cluster, SASA of tripeptide molecules and Phe residues, and number of H-bonds formed between tripeptides (Np-p) and between tripeptide and water molecules (Np-w).

(J–L) The free energy landscape (FEL) of π-π stacking plotted as a function of the centroid distance and the angle between two aromatic rings.

(M–O) Probability of the parallel π-π stacking pattern within 0–200 ns and 200–500 ns.

and fewer H-bonds with water molecules compared with AFA, revealing that AAF and FAA possess stronger aggregation propensity and lower solubility.

To reveal the contribution of each residue to the different aggregation propensities of the three peptides, we calculated the time evolution of their fraction of SASA (Figures S19A–S19C). The SASA fraction of the Phe side chains in AAF dropped rapidly from 0.38 to 0.29 within the first 100 ns and then fluctuated around this

value until the end of the simulation, whereas the SASA fraction of the N-terminal Ala increased quickly to 0.22, indicating that the aromatic ring of AAF is mostly buried in the aggregates, whereas the N-terminal Ala is solvent exposed. As for the second Ala in the peptide sequence, its SASA fraction did not change significantly throughout the entire simulation (Figure S19A). Although FAA exhibited strong aggregation propensity similar to AAF, the SASA fraction of Phe side chains increased slightly to 0.41 before t = 129 ns and then decreased to 0.38 with fluctuation, revealing that the aromatic rings in FAA are not buried in the aggregates like those in AAF. Instead, the C-terminal Ala tended to be buried, as evidenced by the slight decrease of the SASA fraction (Figure S19B). In AFA systems, the SASA fraction of each residue exhibited no significant change throughout the entire simulation (Figure S19C), indicating that the aggregation propensity of AFA is very weak. The Phe and Ala residues make sequence-dependent contributions to the solvent exposure of the peptide, leading to different solubilities for the three peptides.

We also calculated the contact probability for each residue pair to uncover interactions crucial for the different self-assembly capabilities of AAF, FAA, and AFA (Figures S19D–S19F). The Phe-Phe pair in AAF and AFA showed the highest contact probability (0.732 and 0.449, respectively), indicating a dominant role of the Phe-Phe interaction in self-assembly of AAF and AFA. In contrast, the highest contact probability in FAA was detected between Phe and the C-terminal Ala, and the contact probability between Phe residues ranked second, with a value of 0.421. The contact probability between Phe residues in AAF is much higher than that in FAA and AFA, indicating that the Phe-Phe interaction is crucial for aggregate formation by AAF. The interaction strength among different residue pairs in FAA is quite similar, revealing strong interaction competition and lack of a dominant interaction, which may lead to imperfect molecular packing and explain the inability of FAA to grow good-quality single crystals.

To explore the role of aromatic stacking in self-assembly of the three tripeptides, we calculated the free energy landscape (FEL) of π-π stacking as a function of the centroid distance and the angle between two aromatic rings of AAF, FAA, and AFA (Figures 4J–4L). The stacking patterns are classified into three categories according to the angles: parallel (0°–30°), herringbone (30°–60°), and perpendicular (T-shaped, 60°–90°). The FEL shows that the aromatic stacking of AAF and FAA is much stronger than that of AFA. Perpendicular and herringbone stacking patterns, rather than the parallel pattern, are dominant, in contrast to the aromatic stacking pattern observed in tripeptide crystal structures. This is probably because of the fact that our MD simulations at a nanosecond timescale can only explore the early stage of tripeptide self-assembly, which is far from the timescale required for crystallization. The early-formed aggregates still undergo structural reorganization toward thermodynamically stable crystal structures. Furthermore, statistical analysis of the parallel stacking within 0–200 ns and 200–500 ns shows that the probability of the parallel pattern in AAF increases significantly over the simulation time (Figure 4M), showing a strong tendency toward the stacking pattern observed in the crystal structure. However, such a tendency is not observed in the FAA and AFA systems (Figures 4N and 4O).

To investigate the aggregation propensity of AAF, FAA, and AFA at a larger size scale, we performed 3-ms coarse-grained (CG) MD simulations on systems consisting of 720 AAF, FAA, or AFA molecules in aqueous solution. All three tripeptides self-assembled into large aggregates from the dispersed states (Figures 5A–5C).

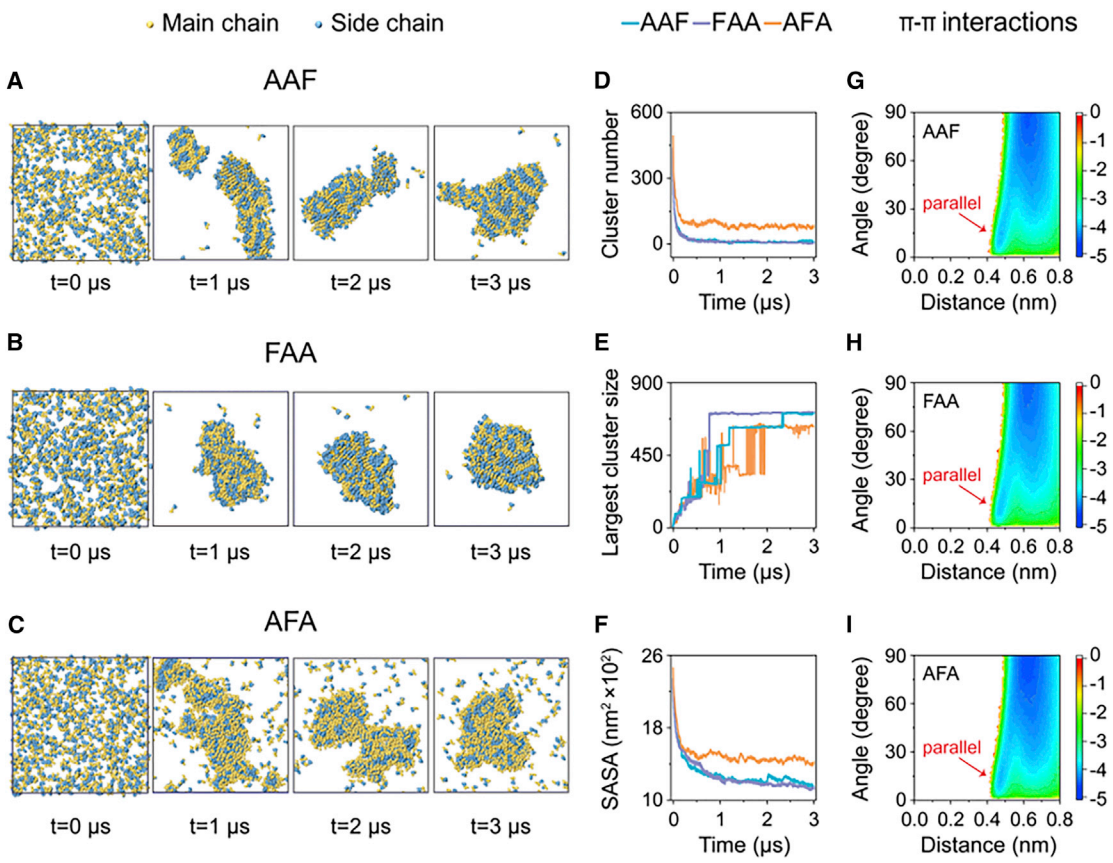


Figure 5. Large size scale analysis of the different aggregation propensities of AAF, FAA, and AFA using microsecond-long CG MD simulations

(A–C) Snapshots of AAF, FAA, and AFA aggregates at four time points. The CG beads of the main chain and side chain are colored yellow and blue, respectively.

(D–F) Time evolution of (D) the number of clusters, (E) the size of the largest cluster, and (F) the SASA of the aggregates.

(G–I) FEL of intermolecular π-π stacking in the largest cluster, plotted as a function of the centroid distance and the angle between two aromatic rings. The basins of the parallel stacking pattern are marked by arrows.

The cluster number of AAF and FAA dropped rapidly within the first 300 ns and reached a plateau of 12 and 5, respectively, at t = 500 ns, whereas the cluster number of AFA decreased slowly to a larger value of 86 (Figure 5D). The size of the largest cluster of FAA increased and reached a value of 706 at t = 770 ns, corresponding to formation of one large aggregate within 1 μs (purple line in Figure 5E and snapshot at t = 1 μs in Figure 5B). Interestingly, the size of the largest cluster of AAF was smaller than that of FAA after t = 770 ns, followed by an increase to a value only slightly smaller than that of FAA at t = 2.3 μs (blue line in Figure 5E). AAF self-assembled into two aggregates around t = 1 μs (snapshot at t = 1 μs in Figure 5A), which then started to fuse into a large prolate aggregate (snapshots at t = 2 μs and 3 μs in Figure 5A). AFA showed the smallest size of the largest cluster after t = 1 μs, and the largest cluster size only reached a value of 624 at t = 3 μs (orange line in Figure 5E), leaving almost 100 AFA molecules dispersed in solution (snapshot at t = 3 μs in Figure 5C). More importantly, the largest cluster size of AFA fluctuated the most, indicating that AFA molecules exchange dynamically into and out of the aggregates. Thus, AFA molecules display the highest mobility, in accordance with its highest solubility. The rapid decrease of the SASA also reveals that all three tripeptide molecules aggregate into large aggregates from dispersed states (Figure 5F). These data indicate that the aggregation propensities of the

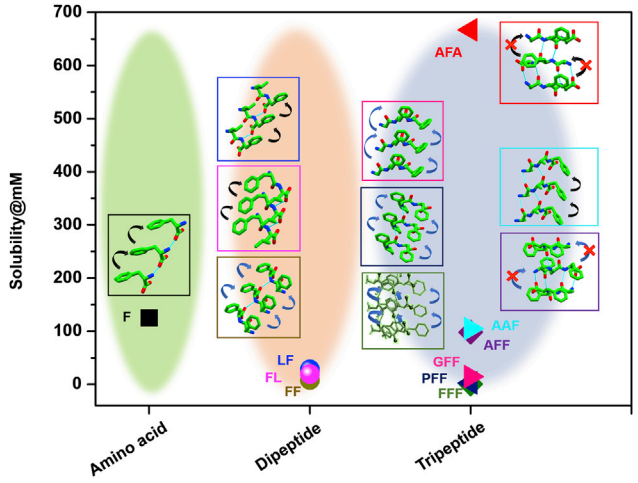


Figure 6. Solid-state packing underlies the solubility of Phe-containing peptides at different length scales

The solubility of single amino acids (NH_3^+-F-COO^-), dipeptides (NH_3^+-FF-COO^-.H_2O, NH_3^+-FL-COO^-.H_2O, and NH_3^+-LF-COO^-.H_2O), and tripeptides (NH_3^+-FFF-COO^-, NH_2^+-PFF-COO^-, NH_3^+-GFF-COO^-.H_2O, NH_3^+-AFF-COO^-.TFA.H_2O, NH_3^+-AAF-COO^-.TFA, and NH_3^+-AFA-COO^-.TFA.AcOH.H_2O) is represented by squares, circles, and triangles, respectively. The solid-state structure of the corresponding peptide is presented in a box of the same color as the symbol. Black and blue arrows indicate the presence of one and more than one type of aromatic-aromatic stacking, respectively, in the structure.

tripeptide molecules are in the order of FAA > AAF > AFA. Aggregates formed in CG MD simulations were indeed larger and more structurally ordered than those in all-atom MD simulations. We also calculated the FEL of π-π stacking in the largest cluster of AAF, FAA, and AFA as a function of the centroid distance and the angle between two aromatic rings (Figures 5G–5I). A minimum-energy basin located at 15° and 0.48 nm was observed in the free energy surface, corresponding to formation of the parallel stacking pattern in the aggregates of AAF, FAA, and AFA. These data reveal that the probability of parallel aromatic stacking patterns increases with system size and simulation timescale, but the perpendicular and herringbone stacking patterns are still dominant.

Screening of peptide library

To further investigate whether the notion of solid-state packing-dependent solubility is applicable only to the studied tripeptide series or whether it is a generic phenomenon that is lawful for all peptides, we prepared a focused library of Phe-based peptides with similar compositions at different length scales (Figure 6; Table S7). For proof of principle, a number of building units staring from the single amino acid (F)[63] through dipeptides (FF, LF, and FL)[64] to tripeptides (FFF, PFF, GFF, and AFF),[27,28,65] all having knowledge of structural information, were selected as queries to investigate water solubility at room temperature and pH 6.8 and compared with the three peptides studied here. The single amino acid F, with one π-π interaction per molecule, was soluble up to 125 mM. Upon increasing the chain length, the solubility of aromatic peptides decreased gradually. FF, comprising two types of π-π interactions per molecule, could be solubilized at a maximal concentration of 8.1 mM. However, replacing one F with an aliphatic amino acid raised the solubility to some extent because of the decrease in the number of aromatic-aromatic interactions. Thus, FL and LF with one aromatic stacking per molecule showed a higher solubility of 28.3 mM and

19.4 mM, respectively, than FF. The tripeptide FFF was found to exhibit significantly lower water solubility (0.5 mM) because of the increased number of π-π interaction (three types per molecule). However, lowering the number of aromatic side chains gradually increased the solubility; PFF and GFF, each comprising two types of aromatic-aromatic interactions per molecule, showed a threshold concentration of 2.6 mM and 14.5 mM, respectively. Surprisingly, despite the similar composition and hydrophobicity, AFF displayed a 6-fold higher solubility of 98.3 mM compared with GFF. The solid-state structure of AFF revealed absence of any π-π interaction in the packing because of antiparallel β-sheet structure formation (Figure S20). The observed lack of aromatic interactions in the molecular organization and the tremendous increase in AFF solubility is in accordance with the discovered 6-fold and 15-fold higher solubility of AFA without aromatic-aromatic stacking compared with the analogous peptides AAF and FAA, respectively, exerting strong aromatic interactions in the solid state. The solubility of a solid in water mainly depends on two factors: the stability of the solute crystal lattice and the interaction of the solute with water. For the molecule to dissolve in the aqueous solvent, it must be able to dissociate from its crystal lattice. Therefore, the general solubility equation (GSE) used to predict aqueous solubility indicates solubility at room temperature as a simple function of its melting temperature (T_m).[66,67] However, many suggested structure-property associations in the literature are valid only for the immediate system described and lack full transferability to a wider context. To investigate the relation between solubility and melting point, the T_m of all of the above mentioned amino acids and peptides was measured at room temperature (Table S7). Correlation between solubility and melting point was found only within isostructural groups having a similar length of chains, but anomalies appear for different lengths of peptides, further indicating our motivation for extended elucidation of the process of solubility. On the other hand, screening of a peptide library and the observed solid-state packing-dependent water solubility perfectly validates our proposed principle of the molecular origin of peptide solubility based on solid-state structure.

In summary, understanding and prediction of peptide solubility in water is of relevance in diverse areas such as basic chemistry, materials science, and molecular medicine. However, the puzzle still remains unresolved, and there is an extreme need for in-depth exploration. The experimental results presented here show that solid-state packing guides the water solubility of aromatic peptides. The analogous tripeptides AFA, AAF, and FAA, obtained only by sequence alteration, exhibit significantly different solubility stemming from their different molecular packing at the atomic level. Lacking any π-π interactions, AFA has 6-fold and 15-fold higher solubility than AAF and FAA, respectively, which comprise aromatic stacking in the solid-state structure. The extremely poor aggregation of AFA further supports its predominant presence as a soluble monomeric conformer in solution. Comprehensive NMR analysis reveals the molecular conformation in solution to be similar to the solid-state structure. All-atom and CG MD simulations demonstrate their differences in aggregation propensity and solubility and provide explanations at the molecular level for the unexpected solubility of AFA. Additionally, screening of a focused peptide library based on Phe-containing peptides at different length scales establishes this notion more firmly. Solubility and the number of aromatic side chains are inversely related as long as all aromatic side chains are involved in the π-π interaction. However, a major discrepancy of this general trend is found in peptides lacking π-π stacking in the solid-state structure, which exhibits a remarkably high critical soluble concentration. The results illustrate the molecular origin of high solubility of certain aromatic peptides and are significant steps

forward in understanding peptide solubility, potentially paving the way for modular design of future peptide drugs based on highly soluble aromatic peptides.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Ehud Gazit (ehudga@tauex.tau.ac.il).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The data supporting the findings of this study are available in the article and the Supplemental information. X-ray crystallography data for AAF (cif) have been deposited in the Cambridge Crystallographic Data Centre (CCDC). The accession number is CCDC:1942277. These data can be obtained free of charge from the CCDC. All other data are available from the lead contact upon reasonable request.

Preparation of peptide solution and concentration measurements

The peptides were synthesized by DGpeptides (Hangzhou, Zhejiang, China) and purified to more than 95% followed by MS confirmation of their identity (Figures S21–S23). All peptides were stored at −20°C. Solutions of peptides were prepared by suspension in distilled water, followed by vigorous vortexing for 2 min at room temperature. The samples were then centrifuged for 15 min at 6,000 rpm. The critical concentrations of the peptide solutions were determined by UV spectrophotometry using a Cary 100 spectrophotometer. Spectra were recorded between 300 and 200 nm with a reading every 0.5 nm and a scan rate of 400 nm/min. Glass cuvettes of 10-mm path length were used for the experiments. Concentration was calculated based on absorption at 257 nm. All other experimental details are described in the Supplemental experimental procedures.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrp.2021.100391.

ACKNOWLEDGMENTS

S.B. thanks Tel Aviv University for a post-doctoral fellowship. This project received funding from ERC under the European Union Horizon 2020 Research and Innovation Program (grant agreement BISON-694426 to E.G and 5R01GM121751-03 to D.E.C). The authors thank Dr. Sigal Rencus-Lazar for help with scientific and language editing.

AUTHOR CONTRIBUTIONS

S.B. and E.G. designed the experiments. S.B. performed the experiments and crystallization of peptides. X.D., Y.T., and G.W. performed the all-atoms MD and CG MD analysis. B.K. and A.R. performed the NMR experiments. S.A.R., D.A.H., and D.E.C. performed the ESI-IMS-MS analysis. W.J. performed DLS of the peptides. L.J.W.S. solved the crystal structures. S.B. coordinated all of the work and analyzed the results. S.B., X.D., G.W., and E.G, wrote the paper. All authors commented on the manuscript.

DECLARATIONS OF INTEREST

The authors declare no competing interests.

Received: December 3, 2020
Revised: February 11, 2021
Accepted: March 12, 2021
Published: April 2, 2021

REFERENCES

1. Malavolta, L., Pinto, M.R.S., Cuvero, J.H., and Nakaie, C.R. (2006). Interpretation of the dissolution of insoluble peptide sequences based on the acid-base properties of the solvent. Protein Sci. *15*, 1476–1488.

2. Perrett, S., Buell, A., and Knowles, T. (2019). Biological and Bio-inspired Nanomaterials (Springer).

3. Ji, W., Xue, B., Arnon, Z.A., Yuan, H., Bera, S., Li, Q., Zaguri, D., Reynolds, N.P., Li, H., Chen, Y., et al. (2019). Rigid Tightly Packed Amino Acid Crystals as Functional Supramolecular Materials. ACS Nano *13*, 14477–14485.

4. Tadepalli, S., Slocik, J.M., Gupta, M.K., Naik, R.R., and Singamaneni, S. (2017). Bio-Optics and Bio-Inspired Optical Materials. Chem. Rev. *117*, 12705–12763.

5. Knowles, T.P.J., and Buehler, M.J. (2011). Nanomechanics of functional and pathological amyloid materials. Nat. Nanotechnol. *6*, 469–479.

6. Abdelrahman, S., Alghrably, M., Lachowicz, J.I., Emwas, A.-H., Hauser, C.A.E., and Jaremko, M. (2020). "What Doesn't Kill You Makes You Stronger": Future Applications of Amyloid Aggregates in Biomedicine. Molecules *25*, 5245.

7. Wei, G., Su, Z., Reynolds, N.P., Arosio, P., Hamley, I.W., Gazit, E., and Mezzenga, R. (2017). Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. Chem. Soc. Rev. *46*, 4661–4708.

8. Bak, A., Leung, D., Barrett, S.E., Forster, S., Minnihan, E.C., Leithead, A.W., Cunningham, J., Toussaint, N., and Crocker, L.S. (2015). Physicochemical and formulation developability assessment for therapeutic peptide delivery–a primer. AAPS J. *17*, 144–155.

9. Sherje, A.P., and Desai, K.J. (2011). Spectrophotometric Determination of Poorly Water Soluble Drug Rosiglitazone Using Hydrotropic Solubilization technique. Indian J. Pharm. Sci. *73*, 579–582.

10. Zhang, S. (2003). Fabrication of novel biomaterials through molecular self-assembly. Nat. Biotechnol. *21*, 1171–1178.

11. Whitesides, G.M., Mathias, J.P., and Seto, C.T. (1991). Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. Science *254*, 1312–1319.

12. Yuan, C., Ji, W., Xing, R., Li, J., Gazit, E., and Yan, X. (2019). Hierarchically Oriented Organization in Supramolecular Peptide Crystals. Nat. Rev. Chem. *3*, 567–588.

13. Alghrably, M., Dudek, D., Emwas, A.-H., Jaremko, Ł., Jaremko, M., and Rowińska-Żyrek, M. (2020). Copper(II) and Amylin Analogues: A Complicated Relationship. Inorg. Chem. *59*, 2527–2535.

14. Łoboda, D., and Rowińska-Żyrek, M. (2017). Zn(II) - pramlintide: Stability, binding sites and unexpected aggregation. J. Inorg. Biochem. *174*, 150–155.

15. Alghrably, M., Czaban, I., Jaremko, Ł., and Jaremko, M. (2019). Interaction of amylin species with transition metals and membranes. J. Inorg. Biochem. *191*, 69–76.

16. Brender, J.R., Hartman, K., Nanga, R.P.R., Popovych, N., de la Salud Bea, R., Vivekanandan, S., Marsh, E.N.G., and Ramamoorthy, A. (2010). Role of zinc in human islet amyloid polypeptide aggregation. J. Am. Chem. Soc. *132*, 8973–8983.

17. Fleming, S., and Ulijn, R.V. (2014). Design of nanostructures based on aromatic peptide amphiphiles. Chem. Soc. Rev. *43*, 8150–8177.

18. Gazit, E. (2002). A possible role for pi-stacking in the self-assembly of amyloid fibrils. FASEB J. *16*, 77–83.

19. Lampel, A., McPhee, S.A., Park, H.-A., Scott, G.G., Humagain, S., Hekstra, D.R., Yoo, B., Frederix, P.W.J.M., Li, T.-D., Abzalimov, R.R., et al. (2017). Polymeric peptide pigments with sequence-encoded properties. Science *356*, 1064–1068.

20. Frederix, P.W.J.M., Scott, G.G., Abul-Haija, Y.M., Kalafatovic, D., Pappas, C.G., Javid, N., Hunt, N.T., Ulijn, R.V., and Tuttle, T. (2015). Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels. Nat. Chem. *7*, 30–37.

21. Hsieh, M.C., Liang, C., Mehta, A.K., Lynn, D.G., and Grover, M.A. (2017). Multistep Conformation Selection in Amyloid Assembly. J. Am. Chem. Soc. *139*, 17007–17010.

22. Omosun, T.O., Hsieh, M.C., Childers, W.S., Das, D., Mehta, A.K., Anthony, N.R., Pan, T., Grover, M.A., Berland, K.M., and Lynn, D.G. (2017). Catalytic diversity in self-propagating peptide assemblies. Nat. Chem. *9*, 805–809.

23. Vivekanandan, S., Brender, J.R., Lee, S.Y., and Ramamoorthy, A. (2011). A partially folded structure of amyloid-beta(1-40) in an aqueous environment. Biochem. Biophys. Res. Commun. *411*, 312–316.

24. Doran, T.M., Kamens, A.J., Byrnes, N.K., and Nilsson, B.L. (2012). Role of amino acid hydrophobicity, aromaticity, and molecular volume on IAPP(20-29) amyloid self-assembly. Proteins *80*, 1053–1065.

25. Madhusudan Makwana, K., and Mahalakshmi, R. (2015). Implications of aromatic-aromatic interactions: From protein structures to peptide models. Protein Sci. *24*, 1920–1933.

26. Ma, M., Kuang, Y., Gao, Y., Zhang, Y., Gao, P., and Xu, B. (2010). Aromatic-aromatic interactions induce the self-assembly of pentapeptidic derivatives in water to form nanofibers and supramolecular hydrogels. J. Am. Chem. Soc. *132*, 2719–2728.

27. Reches, M., and Gazit, E. (2003). Casting metal nanowires within discrete self-assembled peptide nanotubes. Science *300*, 625–627.

28. Bera, S., Mondal, S., Xue, B., Shimon, L.J.W., Cao, Y., and Gazit, E. (2019). Rigid helical-like assemblies from a self-aggregating tripeptide. Nat. Mater. *18*, 503–509.

29. Mayans, E., Casanovas, J., Gil, A.M., Jiménez, A.I., Cativiela, C., Puiggalí, J., and Alemán, C. (2017). Diversity and Hierarchy in Supramolecular Assemblies of Triphenylalanine: From Laminated Helical Ribbons to Toroids. Langmuir *33*, 4036–4048.

30. Marshall, K.E., Morris, K.L., Charlton, D., O'Reilly, N., Lewis, L., Walden, H., and Serpell, L.C. (2011). Hydrophobic, aromatic, and electrostatic interactions play a central role in amyloid fibril formation and stability. Biochemistry *50*, 2061–2071.

31. Bowerman, C.J., Liyanage, W., Federation, A.J., and Nilsson, B.L. (2011). Tuning β-sheet peptide self-assembly and hydrogelation behavior by modification of sequence hydrophobicity and aromaticity. Biomacromolecules *12*, 2735–2745.

32. Azriel, R., and Gazit, E. (2001). Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation. J. Biol. Chem. *276*, 34156–34161.

33. Kodaka, M. (2004). Interpretation of concentration-dependence in aggregation kinetics. Biophys. Chem. *109*, 325–332.

34. Poulson, B.G., Szczepski, K., Lachowicz, J.I., Jaremko, L., Emwas, A.-H., and Jaremko, M. (2020). Aggregation of Biologically Important Peptides and Proteins: Inhibition or Acceleration Depending on Protein and Metal Ion Concentrations. RSC Advances *10*, 215–227.

35. Eker, F., Griebenow, K., Cao, X., Nafie, L.A., and Schweitzer-Stenner, R. (2004). Preferred peptide backbone conformations in the unfolded state revealed by the structure analysis of alanine-based (AXA) tripeptides in aqueous solution. Proc. Natl. Acad. Sci. USA *101*, 10054–10059.

36. Motta, A., Reches, M., Pappalardo, L., Andreotti, G., and Gazit, E. (2005). The preferred conformation of the tripeptide Ala-Phe-Ala in water is an inverse γ-turn: implications for protein folding and drug design. Biochemistry *44*, 14170–14178.

37. Pizzanelli, S., Forte, C., Monti, S., Zandomeneghi, G., Hagarman, A., Measey, T.J., and Schweitzer-Stenner, R. (2010). Conformations of phenylalanine in the tripeptides AFA and GFG probed by combining MD simulations with NMR, FTIR, polarized Raman, and VCD spectroscopy. J. Phys. Chem. B *114*, 3965–3978.

38. Sarma, R., Wong, K.Y., Lynch, G.C., and Pettitt, B.M. (2018). Peptide Solubility Limits: Backbone and Side-Chain Interactions. J. Phys. Chem. B *122*, 3528–3539.

39. Mason, T.O., Michaels, T.C.T., Levin, A., Dobson, C.M., Gazit, E., Knowles, T.P.J., and Buell, A.K. (2017). Thermodynamics of Polypeptide Supramolecular Assembly in the Short-Chain Limit. J. Am. Chem. Soc. *139*, 16134–16142.

40. Mason, T.O., Chirgadze, D.Y., Levin, A., Adler-Abramovich, L., Gazit, E., Knowles, T.P.J., and Buell, A.K. (2014). Expanding the solvent chemical space for self-assembly of dipeptide nanostructures. ACS Nano *8*, 1243–1253.

41. Mason, T.O., Michaels, T.C.T., Levin, A., Gazit, E., Dobson, C.M., Buell, A.K., and Knowles, T.P.J. (2016). Synthesis of Nonequilibrium Supramolecular Peptide Polymers on a Microfluidic Platform. J. Am. Chem. Soc. *138*, 9589–9596.

42. Graziano, G. (2009). On the Salting out of Benzene by Alkali Chlorides. J. Chem. Eng. Data *54*, 464–467.

43. Arakawa, T., and Timasheff, S.N. (1984). Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. Biochemistry *23*, 5912–5923.

44. Kong, J., and Yu, S. (2007). Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta Biochim. Biophys. Sin. (Shanghai) *39*, 549–559.

45. Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J. Mol. Biol. *7*, 95–99.

46. Anjana, R., Vaishnavi, M.K., Sherlin, D., Kumar, S.P., Naveen, K., Kanth, P.S., and Sekar, K. (2012). Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. Bioinformation *8*, 1220–1224.

47. Tayeb-Fligelman, E., Tabachnikov, O., Moshe, A., Goldshmidt-Tran, O., Sawaya, M.R., Coquelle, N., Colletier, J.P., and Landau, M. (2017). The cytotoxic *Staphylococcus aureus* PSMα3 reveals a cross-α amyloid-like fibril. Science *355*, 831–833.

48. Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.Ø., Riekel, C., Grothe, R., and Eisenberg, D. (2005). Structure of the cross-β spine of amyloid-like fibrils. Nature *435*, 773–778.

49. Maity, S., Kumar, R., Maity, S.K., Jana, P., Bera, S., and Haldar, D. (2013). Synthesis and Study of 2-Acetyl Amino-3-[4-(2-Amino-5-Sulfo-Phenylazo)- Phenyl]-Propionic Acid: A New Class of Inhibitor for Hen Egg White Lysozyme Amyloidogenesis. MedChemComm *4*, 530–536.

50. Hora, M., Carballo-Pacheco, M., Weber, B., Morris, V.K., Wittkopf, A., Buchner, J., Strodel, B., and Reif, B. (2017). Epigallocatechin-3-gallate preferentially induces aggregation of amyloidogenic immunoglobulin light chains. Sci. Rep. *7*, 41515.

51. Lopez del Amo, J.M., Fink, U., Dasari, M., Grelle, G., Wanker, E.E., Bieschke, J., and Reif, B. (2012). Structural properties of EGCG-induced, nontoxic Alzheimer's disease Aβ oligomers. J. Mol. Biol. *421*, 517–524.

52. Wang, Q., Guo, J., Jiao, P., Liu, H., and Yao, X. (2014). Exploring the influence of EGCG on the β-sheet-rich oligomers of human islet amyloid polypeptide (hIAPP1-37) and identifying its possible binding sites from molecular dynamics simulation. PLoS ONE *9*, e94796.

53. Cukalevski, R., Yang, X., Meisl, G., Weininger, U., Bernfur, K., Frohm, B., Knowles, T.P.J., and Linse, S. (2015). The Aβ40 and Aβ42 peptides self-assemble into separate homomolecular fibrils in binary mixtures but cross-react during primary nucleation. Chem. Sci. (Camb.) *6*, 4215–4233.

54. Bera, S., Mondal, S., Tang, Y., Jacoby, G., Arad, E., Guterman, T., Jelinek, R., Beck, R., Wei, G., and Gazit, E. (2019). Deciphering the Rules for Amino Acid Co-Assembly Based on Interlayer Distances. ACS Nano *13*, 1703–1712.

55. Do, T.D., de Almeida, N.E.C., LaPointe, N.E., Chamas, A., Feinstein, S.C., and Bowers, M.T. (2016). Amino Acid Metaclusters: Implications of Growth Trends on Peptide Self-Assembly and Structure. Anal. Chem. *88*, 868–876.

56. Counterman, A.E., and Clemmer, D.E. (2001). Magic Number Clusters of Serine in the Gas Phase. J. Phys. Chem. B *105*, 8092–8096.

57. Chen, X., Raab, S.A., Poe, T., Clemmer, D.E., and Larriba-Andaluz, C. (2019). Determination of Gas-Phase Ion Structures of Locally Polar Homopolymers Through High-Resolution Ion Mobility Spectrometry-Mass Spectrometry. J. Am. Soc. Mass Spectrom. *30*, 905–918.

58. Young, L.M., Saunders, J.C., Mahood, R.A., Revill, C.H., Foster, R.J., Tu, L.H., Raleigh, D.P., Radford, S.E., and Ashcroft, A.E. (2015). Screening and classifying small-molecule inhibitors of amyloid formation using ion mobility spectrometry-mass spectrometry. Nat. Chem. *7*, 73–81.

59. Trimpin, S., Plasencia, M., Isailovic, D., and Clemmer, D.E. (2007). Resolving oligomers from fully grown polymers with IMS-MS. Anal. Chem. *79*, 7965–7974.

60. Yuan, C., Li, S., Zou, Q., Ren, Y., and Yan, X. (2017). Multiscale simulations for understanding the evolution and mechanism of hierarchical peptide self-assembly. Phys. Chem. Chem. Phys. *19*, 23614–23631.

61. Tang, Y., Yao, Y., and Wei, G. (2020). Expanding the structural diversity of peptide assemblies by coassembling dipeptides with diphenylalanine. Nanoscale *12*, 3038–3049.

62. Guo, C., Luo, Y., Zhou, R., and Wei, G. (2012). Probing the self-assembly mechanism of diphenylalanine-based peptide nanovesicles and nanotubes. ACS Nano *6*, 3907–3918.

63. Ihlefeldt, F.S., Pettersen, F.B., von Bonin, A., Zawadzka, M., and Görbitz, C.H. (2014). The polymorphs of L-phenylalanine. Angew. Chem. Int. Ed. Engl. *53*, 13600–13604.

64. Görbitz, C.H. (2001). Nanotube formation by hydrophobic dipeptides. Chemistry *7*, 5153–5159.

65. Parthasarathy, R., Chaturvedi, S., and Go, K. (1990). Design of crystalline helices of short oligopeptides as a possible model for nucleation of α-helix: role of water molecules in stabilizing helices. Proc. Natl. Acad. Sci. USA *87*, 871–875.

66. Ran, Y., and Yalkowsky, S.H. (2001). Prediction of drug solubility by the general solubility equation (GSE). J. Chem. Inf. Comput. Sci. *41*, 354–357.

67. Bergström, C.A.S., and Larsson, P. (2018). Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting. Int. J. Pharm. *540*, 185–193.

Cell Reports Physical Science, Volume *2*

Supplemental information

Solid-state packing dictates the unexpected solubility of aromatic peptides

Santu Bera, Xuewei Dong, Bankala Krishnarjuna, Shannon A. Raab, David A. Hales, Wei Ji, Yiming Tang, Linda J.W. Shimon, Ayyalusamy Ramamoorthy, David E. Clemmer, Guanghong Wei, and Ehud Gazit

1. Supplemental Experimental Procedures

Preparation of peptide assemblies. For assembly, peptides in the required concentration were dissolved in double distilled water by vigorous vortexing for 2 min. The peptide solutions were then incubated at 18°C for two weeks with frequent shaking before examination.

Dynamic light scattering (DLS). Eight hundred µL of the sample solution at required concentration was introduced into a DTS1070 folded capillary cell (Malvern, Worcestershire, U.K.), and the size was measured using a Zetasizer Nano ZS analyzer (Malvern Instruments, Malvern, UK) at 25.0 °C and a backscatter detector (173°). Three measurements were performed and averaged for accuracy.

Scanning electron microscopy (SEM). A 5 µL aliquot was allowed to dry on a microscope glass cover slip at ambient conditions over night and coated with Au. SEM images were recorded using a JSM-6700F FE-SEM (JEOL, Tokyo, Japan) operating at 10 kV.

Fourier-transform infrared (FTIR) spectroscopy. A 30 µL aliquot of the peptide solution was deposited onto disposable KBr infrared sample cards (Sigma-Aldrich, Rehovot, Israel), which were then allowed to dry under vacuum. The samples were saturated twice with 30 µL of D_2O and vacuum dried. FTIR spectra were collected using a nitrogen purged Nicolet Nexus 470 FTIR spectrometer (Nicolet, Offenbach, Germany) equipped with a deuterated triglycine sulfate (DTGS) detector. Measurements were performed using a 4 cm^{-1} resolution and by averaging 64 scans. The absorbance maxima

values were determined using an OMNIC analysis program (Nicolet). The background was subtracted using a control spectrum.

Circular dichroism (CD) spectroscopy. CD spectra were collected using a Chirascan spectrometer (Applied Photophysics, Leatherhead, UK) fitted with a Peltier temperature controller set to desired temperature, using quartz cuvettes with an optical path length of 0.1 mm (Hellma Analytics, Müllheim, Germany). Absorbance of the sample was kept within the linear range of the instrument during measurement. Data acquisition was performed in steps of 1 nm at a wavelength range of 190 to 260 nm with a spectral bandwidth of 1.0 nm and an averaging time of 3 s. The spectrum of each sample was collected three times and averaged. Baseline was similarly recorded for phosphate buffer and subtracted from the samples spectra. Data processing was performed using Pro-Data Viewer software (Applied Photophysics, Leatherhead, UK).

Powder X-ray diffraction (XRD). The lyophilized peptide powder was dissolved in double distilled water and allowed to self-assemble by incubation at 18°C for four weeks. The sample was then centrifuged for 10 min at 6000 rpm and the solution was decanted to remove non-assembled peptide molecules. The assembled structures were lyophilized and poured inside a glass capillary 0.5 mm in diameter. X-ray diffraction was collected using a Bruker D8 Discover theta/theta diffractometer with liquid-nitrogen-cooled intrinsic Ge solid-state linear position detector.

Crystal preparation and data collection. Crystals used for data collection were grown using the vapor diffusion method. The dry peptide was first dissolved in water, at a concentration of 5 mg/ml. Then, 50 µL was deposited into a series of 8x40 mm vessels.

Each tube was sealed with Parafilm®, in which a single small hole was pricked using a needle. The samples were placed inside a larger vessel filled with 2 mL of acetonitrile. The systems were ultimately capped and incubated at 18 °C for several days. Needle-like crystals grew within 7-8 days. For data collection, crystals were coated in paratone oil (Hampton Research), mounted on a MiTeGen cryo-loop and flash frozen in liquid nitrogen. Diffraction data were collected at 100 K on a Rigaku XtaLabPro with a Dectris PilatusR 200K-A detector using CuKα radiation λ= 1.54184Å.

Processing and structural refinement of crystal data. The diffraction data were processed using CrysAlisPro 1.171.39.22a. Structure was solved by direct methods in SHELXT-2016/4.[1] The refinements were performed with SHELXL-2016/4 and weighted full-matrix least-squares against $|F^2|$ using all data. Atoms were refined independently and anisotropically, with the exception of hydrogen atoms, which were placed in calculated positions and refined in a riding mode. Crystal data collection and refinement parameters are shown in Supplementary Table 1 and the complete data can be found in the cif file as supplementary information. The crystallographic data have been deposited in the CCDC with no. 1942277 for AAF.

Ion mobility spectrometry-mass spectrometry (IMS-MS) experiments. To fabricate electrospray emitters, uncoated borosilicate glass capillaries (ID: 1.2 mm OD: 1.5 mm) were purchased from Sutter Instrument Co. (Novato, CA). The 10 cm long capillaries were pulled with a Sutter p-97 micropipette puller to produce electrospray emitters with 1 µm tip sizes. The tripeptide solutions (each prepared at 10 mM in Milli-Q water at pH 6.8) were inserted into the back of a pulled emitter and a 0.25 mm platinum wire was

inserted into the solution. An ESI potential between 1-2 kV was connected to the platinum wire to generate electrospray ions. A custom-made 4-meter drift tube coupled to a time-of-flight (ToF) mass spectrometer was used for IMS-MS analysis. The instrument is shown in **Figure S10**. A detailed description of IMS theory and instrumentation have been detailed previously.[2,3] Briefly, ions produced by ESI enter the IMS-MS instrument through a narrow capillary and are stored in an hourglass-shaped ion funnel (F1) until being pulsed into the drift region by an electrostatic gate (G1). The ion packet then traverses the drift region which is filled with a neutral buffer gas (~3 Torr He) and has a constant electric field drop (~12 $V \cdot cm^{-1}$). After every meter of separation, the diffuse ion packet is radially focused by ion funnels with applied RF potentials (F2/F3/F4/F5). Ions then exit the drift tube through a differentially pumped region and are pulsed into an orthogonal reflection-geometry ToF-MS where they are separated by *m*/*z*.

NMR spectroscopy. Samples were prepared by dissolving lyophilized peptides in 90% H_2O/10% ^2H_2O at 10 mM concentration. AAF and FAA were also prepared at 40 mM concentration. All the NMR experiments were performed on a Bruker 500 MHz NMR spectrometer equipped with a triple-resonance TXI (5 mm with xyz gradient) probe operated at 25 °C.[4] One-dimensional (1D) ^1H and two-dimensional (2D) ^1H-^1H TOCSY and 2D ^1H-^1H NOESY were recorded for ^1H chemical shift assignments. [^{13}C-^1H]-HSQC and [^{15}N-^1H]-HSQC spectra were recorded for ^{13}C and ^{15}N chemical shift assignments, respectively. 1D ^1H NMR spectra were also recorded at different temperatures ranging from 5 to 30 °C. All NMR data were processed in Bruker TopSpin (4.0.6) and analyzed using the Bruker TopSpin/CcpNmr analysis software. DOSY experiments were recorded using a simulated echo sequence with a 3-9-19 pulse sequence for water suppression. The

DOSY data were processed/analyzed in phased mode using T_1/T_2 analysis module in Bruker TopSpin.

All-atom molecular dynamic simulations (All-atom MD)

All-atom MD simulations of 60 AAF/FAA/AFA systems were carried out in an isothermal-isobaric (NPT) ensemble using the GROMACS-2016.4 software package[5] in combination with OPLS-AA force field.[6] 60 AAF/FAA/AFA molecules were randomly placed in an 8×8×8 nm^3 cubic box filled with 15796 TIP4P water molecules,[7] as the initial state of each simulated system. Na^+ and Cl^- ions were also added to the simulation boxes. The bond lengths of peptides and water molecules were constrained using the LINCS[8] and SETTLE[9] algorithms, respectively, allowing an integration time step of 2 fs. The peptide and non-peptide (water molecules and counterions) groups were separately coupled to an external heat bath using a velocity rescaling coupling method,[10] maintaining the temperature at 310 K. The pressure was kept at 1 bar using the Parrinello-Rahman method.[11] Electrostatic interactions were calculated using the particle mesh Ewald (PME) method with a real space cutoff of 1.4 nm.[12] The van der Waals interactions were calculated using the same cutoff of 1.4 nm. An important consideration in choosing a cutoff for the calculation of electrostatic and van der Waals interactions is a tradeoff between accuracy and computational cost, larger cutoff will improve accuracy, albeit an increased computational cost. In order to improve accuracy, we choose 1.4 nm as the cutoff for both electrostatic and van der Waals interactions in our MD simulations. In fact, a cutoff of 1.4 nm for electrostatic and van der Waals interactions has been used in extensive computational studies on peptides, polymers and chemical compounds.[13-16] Verlet cutoff-scheme was used for neighbor searching. To equilibrate the systems after

energy minimization, simulations were performed in the 100 ps NVT MD run first, followed by the 100 ps NPT MD run. Subsequently, three individual 500 ns MD simulations were carried out for the AAF/FAA/AFA systems.

Coarse-grained molecular dynamic simulations (CG-MD). CG-MD simulations on 720 AAF/FAA/AFA systems were performed using the GROMACS-2018.3 software package[5] in combination with the MARTINI coarse-grained model (version 2.1).[17,18] The mapping from the all-atom model of the AAF/FAA/AFA molecules to the CG model and the interaction types of the CG beads are shown in Figure S18. The AAF/FAA/AFA molecule can be divided into two groups: main chain and side chain (the aromatic ring), which are colored in yellow and blue, respectively (Figure S18). Each AAF/FAA/AFA molecule was represented by six CG beads: three beads for the main chain and three beads for the aromatic side chain. Water molecules were represented using P4 interaction types of beads. In the initial state of three simulated systems, 720 AAF/FAA/AFA molecules were randomly placed in a solution containing 40000 water beads, yielding a peptide concentration of ~65 mg/mL. Electrostatic interactions were calculated using the PME method with a real space cutoff of 1.2 nm,[12] and the same cutoff was used for van der Waals interactions. The solute and solvent were separately coupled to an external heat bath using a velocity rescaling coupling method,[10] and a pressure bath using the Parrinello-Rahman method.[11] After 200 ps NVT MD run and the 800 ps NPT MD run, three microsecond-long (3 μs) MD simulations were performed on 720 AAF/FAA/AFA systems.

Analysis methods for MD simulations

Data analyses were performed using in-house-developed codes and tools implemented in the GROMACS package.

For all-atom MD simulation data, a hydrogen bond (H-bond) was considered to be formed when (i) the distance between N and O was smaller than 0.35 nm and (ii) the angle of N − H ⋯ O (or O − H ⋯ N) was larger than 150°. The SASA fraction of Ala residues and Phe side chains was defined as the percentage of the SASA of Ala or Phe side chain relative to the SASA of all AAF/FAA/AFA molecules at each time point. Residue-pair contact probabilities were used to estimate the inter-peptide interactions. The angle between two aromatic rings refers to the angle between the normal vectors of the two rings. If the angle was larger than 90°, the supplementary angle was used as the angle between the two aromatic rings. Two aromatic rings were considered to form π-π stackings when their centroid distance is within 0.7 nm.[19] The 2D free energy landscape was constructed using the formula –RTln[P(angle, centroid distance)], where P (angle, centroid distance) is the probability of a stacking pattern to have a certain value of angle and centroid distance. Probability of the parallel aromatic stacking pattern was calculated using a distance cutoff of 0.5 nm.

The data in the last 300 ns of all-atom MD trajectories were used to calculate the probability density function, the free energy landscape and the probability of the parallel stacking pattern. Trajectory visualization and graphical structure analysis were performed using the VMD[20] and PyMOL[21] software suite.

2. Critical soluble concentration of tripeptides

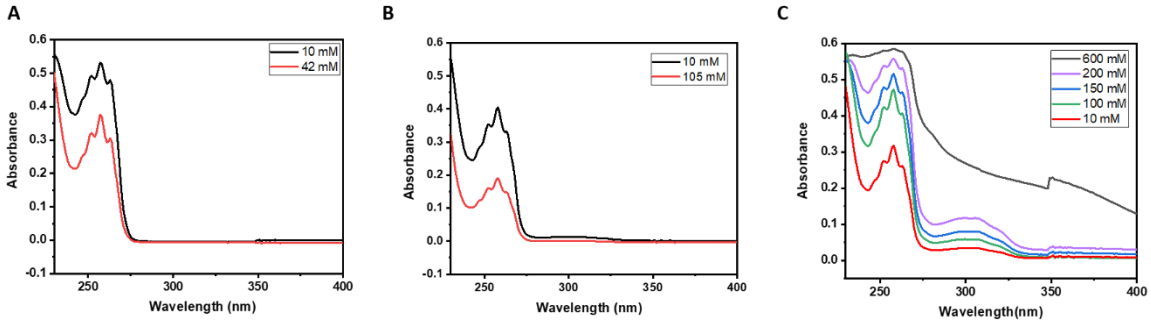


Figure S1: UV/Vis characterization of tripeptides. UV/Vis spectra of (**A**) FAA, (**B**) AAF and (**C**) AFA at different concentrations showing increasing absorbance intensity with rise of concentration without any shift of peak position.

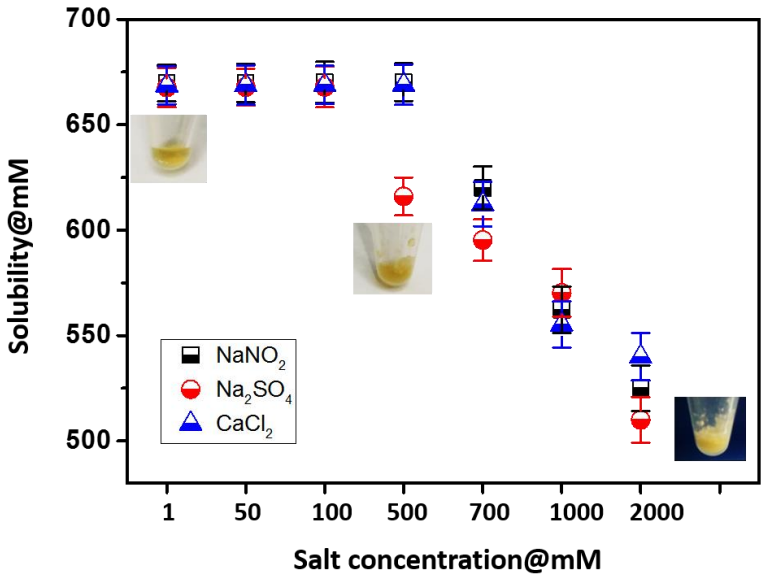


Figure S2: Critical soluble concentration of AFA in the presence salts. Solubility was measured in presence of different concentrations of various salts such as $NaNO_2$, Na_2SO_4 and $CaCl_2$. The effect of salts was quite similar even though changing the cation as well as anion. Solubility decreased only at very high concentration of salt (nearly 500 mM or above) due to commonly known "salting out" effect.

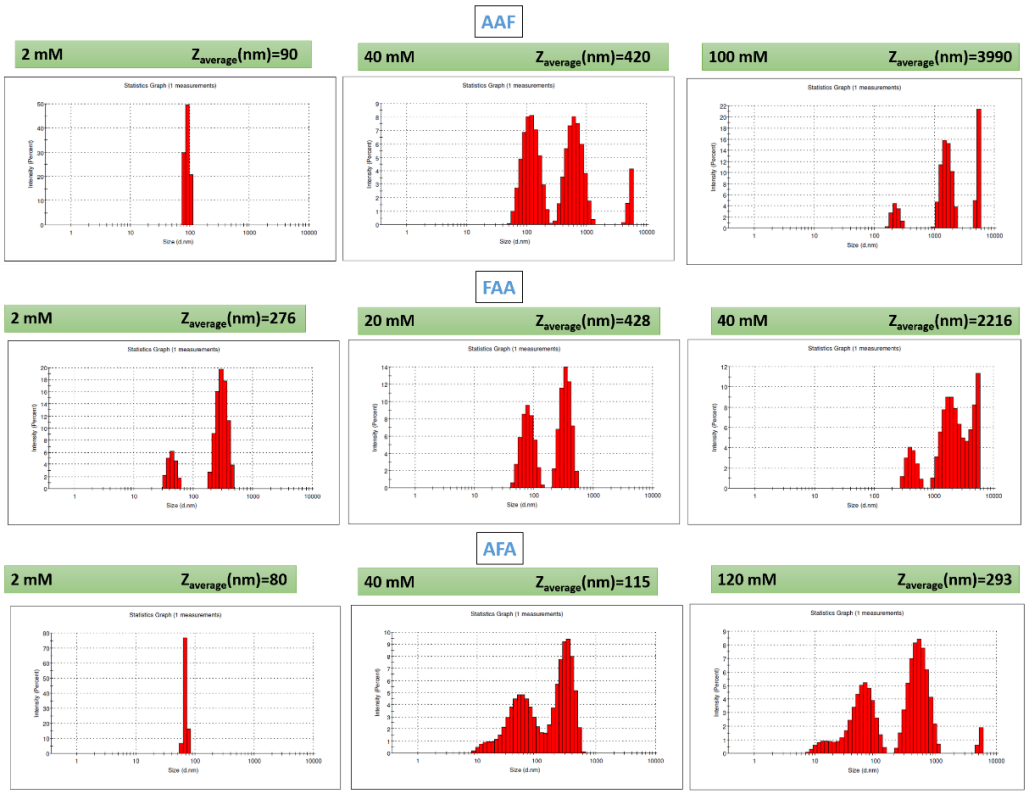


Figure S3: DLS characterization of tripeptides. The results showing the average hydrodynamic diameter of assembled structures at different concentrations.

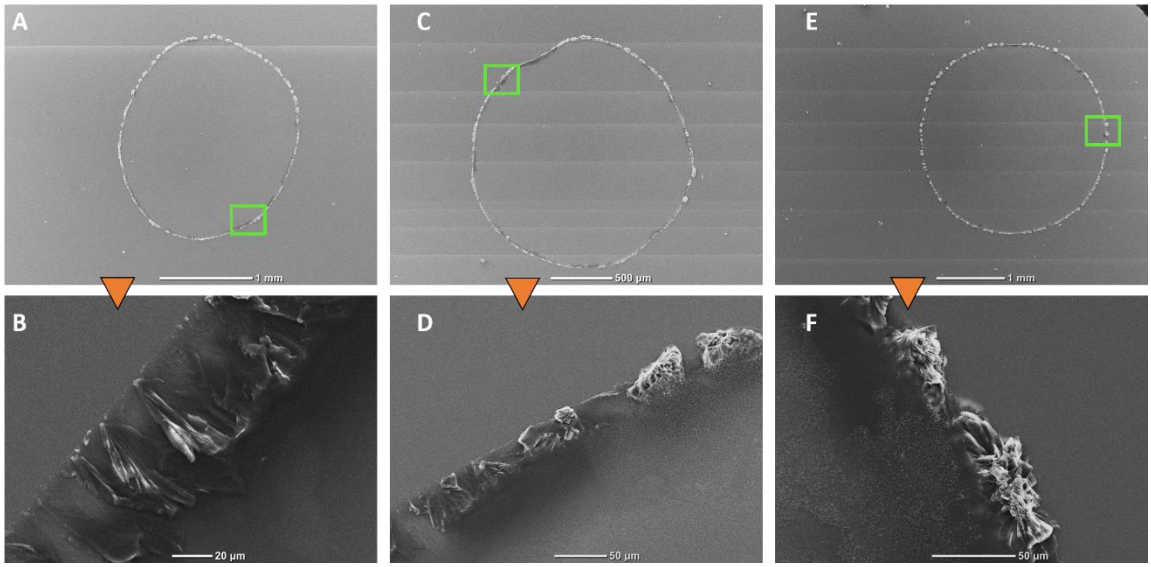


Figure S4: SEM images of AFA. The micrographs showing the resultant architectures formed by 2 mM AFA at pH 6.8 after two weeks of self-assembly. The upper panel represents the entire area of the solution drop and the corresponding zoom-in area marked by a green square is shown in the lower panel.

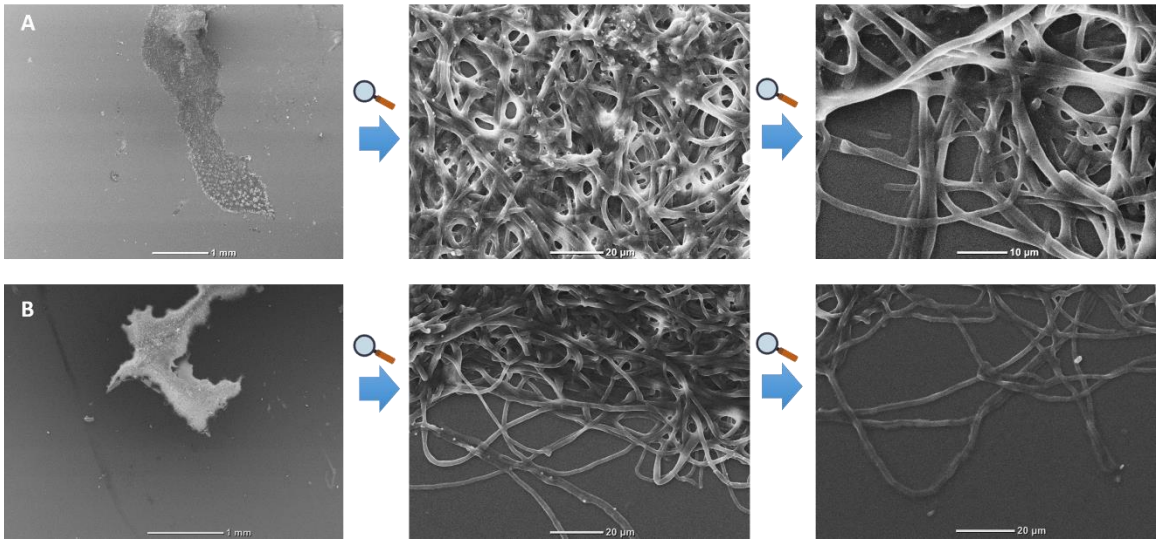


Figure S5: SEM images of AAF and FAA. The micrographs of (**A**) AAF and (**B**) FAA were recorded at 2 mM concentration and pH=6.8.

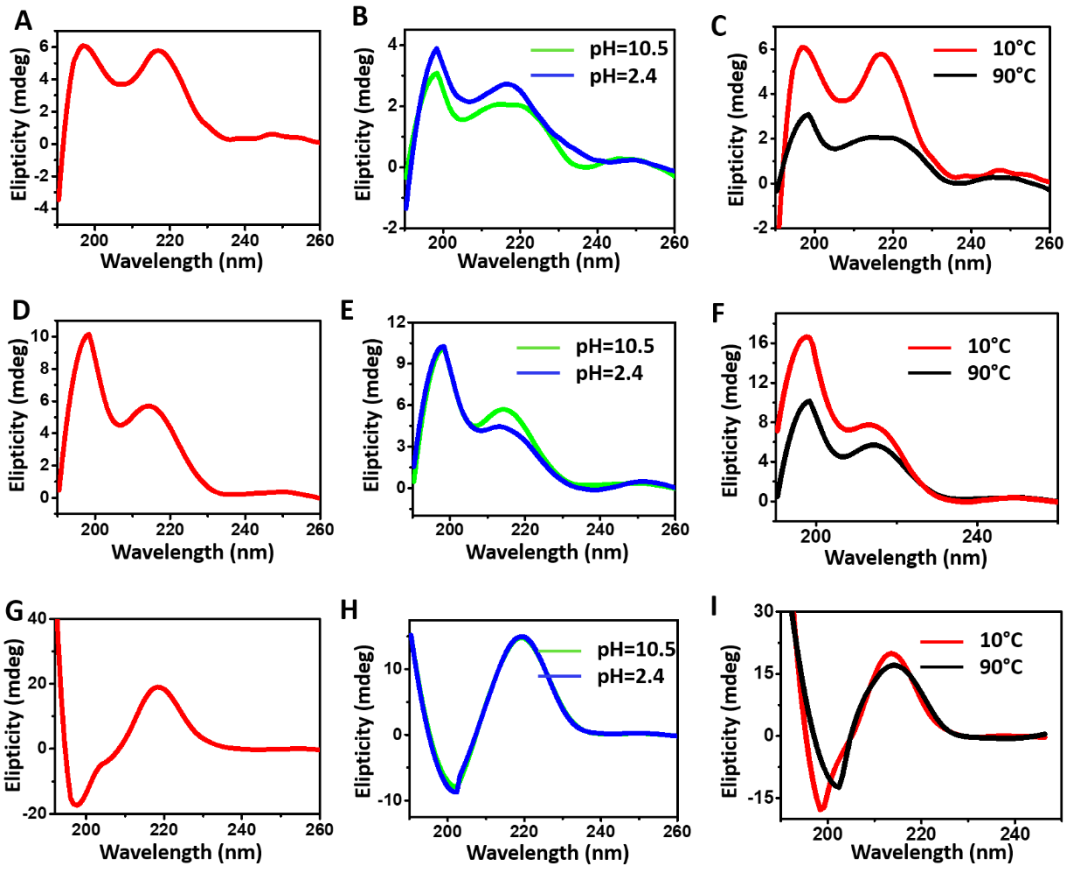


Figure S6: CD spectroscopy of the tripeptides. The spectra of (**A**-**C**) AFA, (**D**-**F**) AAF and (**G**-**I**) FAA were measured under different conditions. (**A**,**D**,**G**) The solution secondary structure at 2 mM concentration and pH 6.8. (**B**,**E**,**H**) At high (pH=10.5) and low (pH=2.4) pH values, the CD spectra did not exhibit any marked change, indicating preservation of the secondary structure. (**C**,**F**,**I**) Upon raising the temperature from 10°C to 90°C, all three peptides showed partial unfolding as the intensity of the major peaks decreased slightly.

3. X-ray diffraction of tripeptides

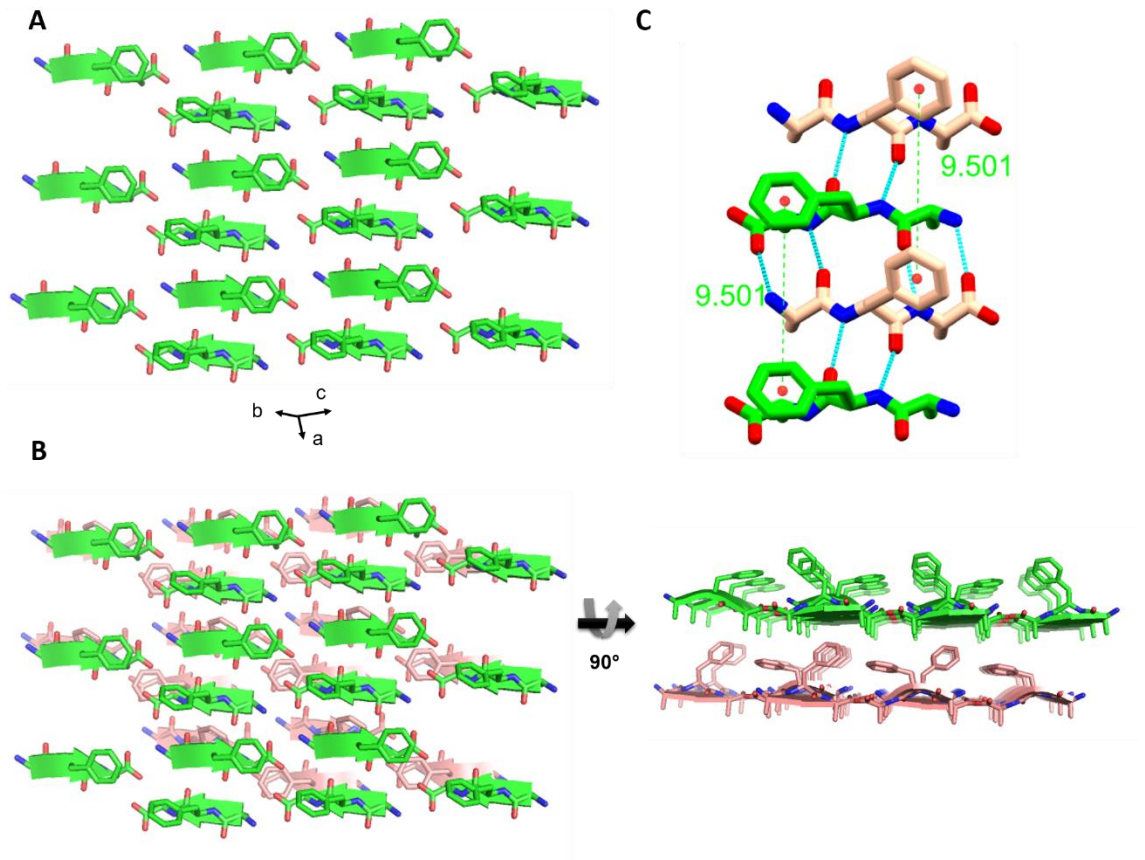


Figure S7: Single crystal structure of AFA. (**A**) Single sheet. (**B**) Hydrophobic interaction between two adjacent sheets. (C) The closest possible interacting distance between two aromatic rings of F is 9.5 Å.

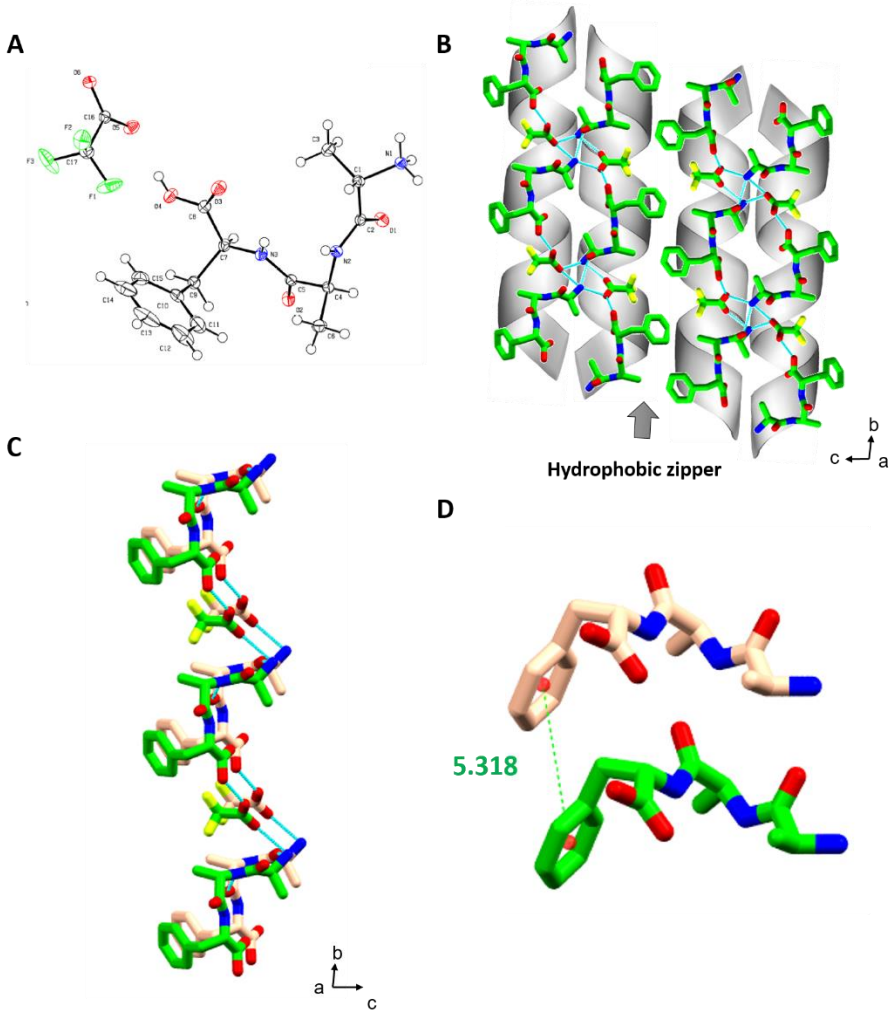


Figure S8: Single crystal structure of AAF. (**A**) The ORTEP diagram of the asymmetric unit in 50% probability displacement ellipsoids. (**B**) Stabilization of nearby helical dimers through hydrophobic zipper-like interactions of aromatic rings. (**C**) Stacking of adjacent helixes in the crystallographic *a*-direction. (**D**) The closest distance of π-π stacking between aromatic rings.

Table S1: Data collection and refinement statistics

Experimental details:

Crystal data	AAF
Chemical formula	C17 H22 F3 N3 O6
Mr	421.37
Crystal system	Orthorhombic
Space group	$P2_12_12_1$
a (Å)	5.31786(5)
b (Å)	12.0314(1)
c (Å)	31.7122(3)
α (°)	90
β (°)	90
γ (°)	90
V ($Å_3$)	2028.99(3)
Z , *Z*'	4
μ (mm^{-1})	1.061
Temperature (K)	100 (2)
Data collection	
Diffractometer	Rigaku XtaLAB AFC12 (RINC): Kappa dual offset/far
Wavelength (Å)	1.54184
Crystal size (mm)	0.262x0.032x0.025
*T*min , *T*max	0.960,0.974
$N_{measured}$(unique)	17797(4384)
$N_{observed}$ [$I > 2\sigma(I)$]	4237
R_{int}	0.0406
θ_{max} (°)	80.155
Refinement	
R[F2 > 2σ(F2)]	0.0349
wR	0.0361
wR[*F*2 > 2σ(*F*2)]	0.0879
wR(*F*2)	0.0892
Goodness-of-fit	1.037
No. of reflections	4384
No. of parameters	266
No. of restraints	0
H-atom treatment	H-atom parameters constrained

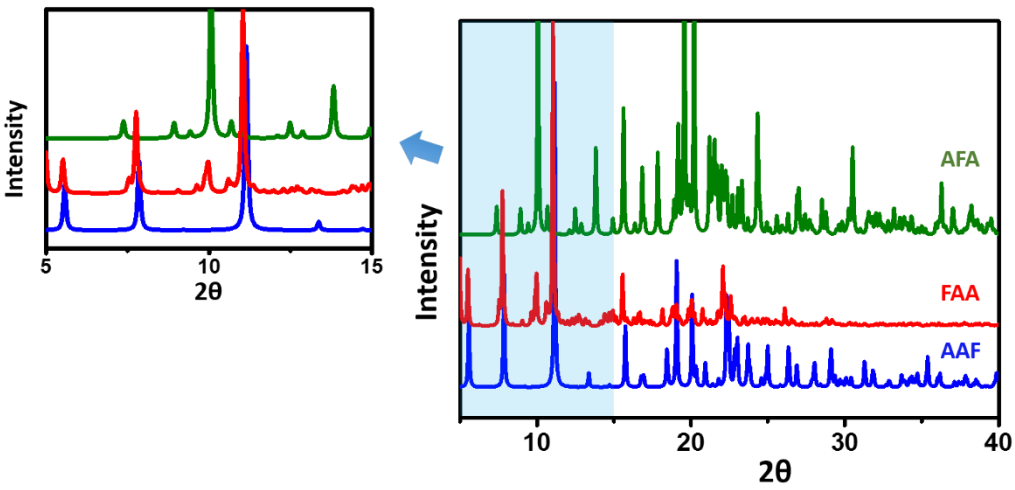


Figure S9: Powder X-ray diffraction of the tripeptides. The region in the graph highlighted by cyan has been enlarged in the left side. The AAF and FAA show quite similar diffraction pattern indicating their similar molecular arrangement in the atomic level and higher order packing.

4. Aggregation characterization using ESI–IMS–MS

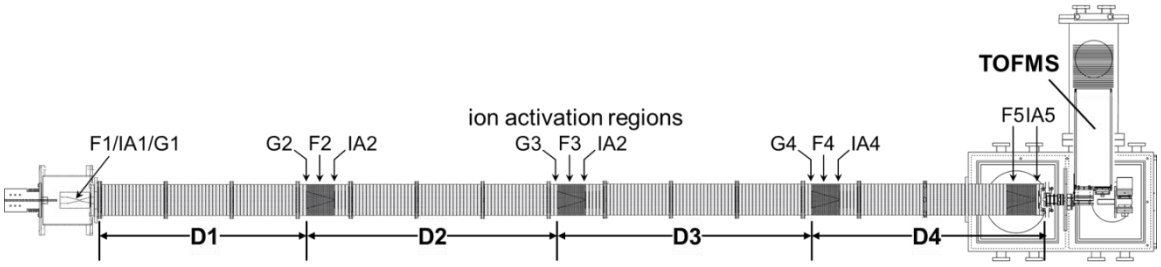


Figure S10: Schematic illustration of the 4m IMS-MS instrument used for tripeptide experiments.

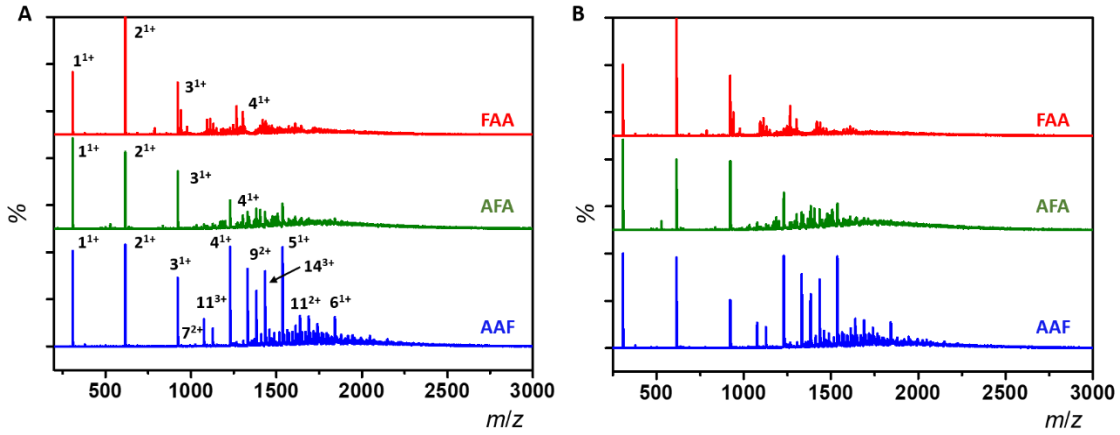


Figure S11: ESI-MS mass spectrum of the tripeptides. The spectrum were recorded at 10 mM peptide concentration and at two different time point after preparation of samples, after 1 day (**A**) and after two weeks (**B**). The numbers above the peaks denote the oligomer order, with the positive-charge state of ions in superscript. The two spectra are showing quite similar peak position and intensity, indicating presence of similar oligomers.

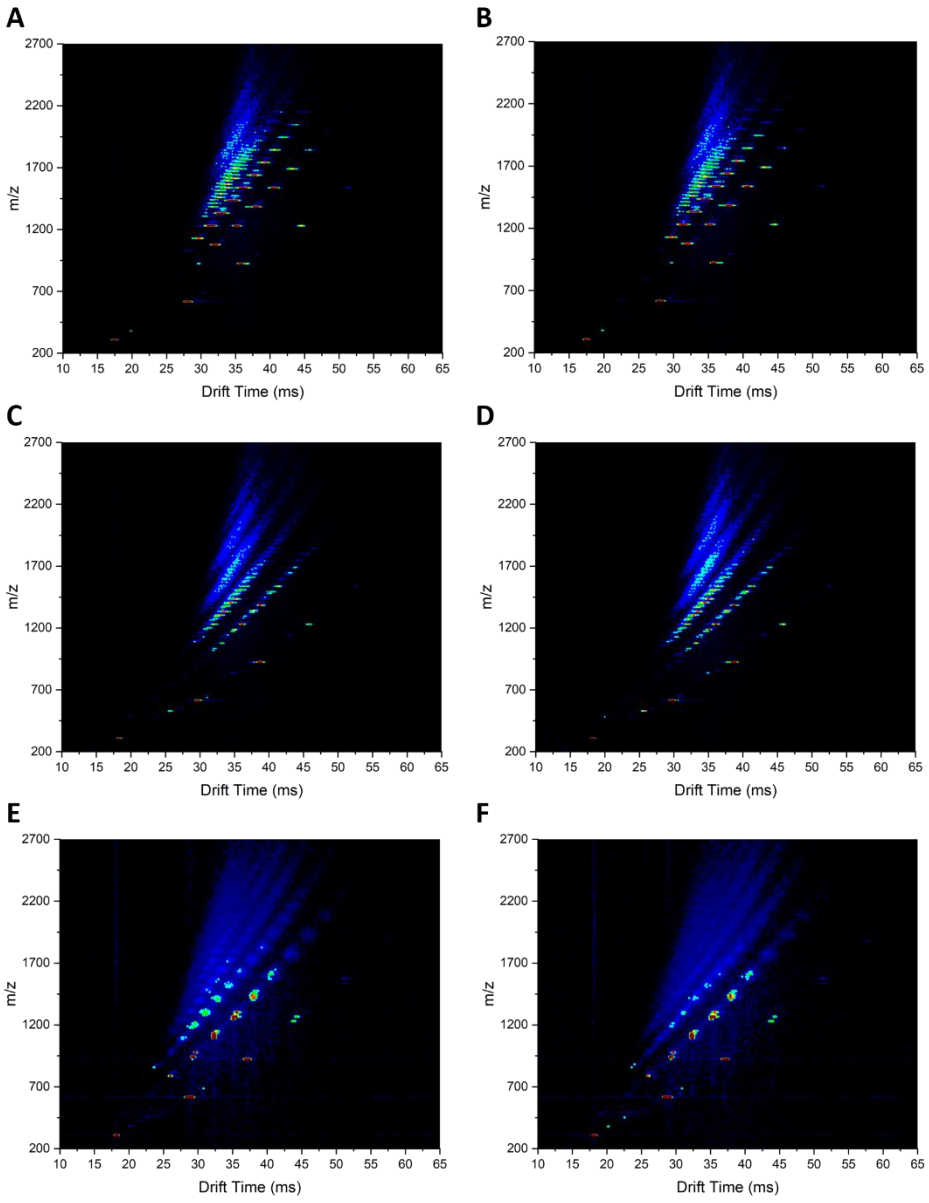


Figure S12: ESI–IMS–MS 2-dimensional plot of the tripeptides. The images of (**A**,**B**) AAF, (**C**,**D**) AFA and (**E**,**F**) FAA monomers through to oligomers. The experiments were carried out at 10 mM peptide concentration and at two different time points after preparation of samples, after 1 days (**A**,**C**,**E**) and after two weeks (**B**,**D**,**E**). However, the 2-dimensional plot showed quite a similar pattern during this time scale.

Table S2: Self-assembled clusters of AAF obtained from ESI–IMS–MS experiment.

Family	symbol	m/z	N	Z	Assembly
+1	1	308	1	1	$[1(AAF)+H]^{+}$
	2	615	2	1	$[2(AAF)+H]^{+}$
	3	922	3	1	$[3(AAF)+H]^{+}$
	4	1229	4	1	$[4(AAF)+H]^{+}$
	5	1536	5	1	$[5(AAF)+H]^{+}$
	6	1843	6	1	$[6(AAF)+H]^{+}$
+2	6	922	6	2	$[6(AAF)+2H]^{2+}$
	7	1076	7	2	$[7(AAF)+2H]^{2+}$
	8	1229	8	2	$[8(AAF)+2H]^{2+}$
	9	1383	9	2	$[9(AAF)+2H]^{2+}$
	10	1536	10	2	$[10(AAF)+2H]^{2+}$
	11	1690	11	2	$[11(AAF)+2H]^{2+}$
	12	1843	12	2	$[12(AAF)+2H]^{2+}$
	13	1997	13	2	$[13(AAF)+2H]^{2+}$
+3	10	1024	10	3	$[10(AAF)+3H]^{3+}$
	11	1127	11	3	$[11(AAF)+3H]^{3+}$
	12	1229	12	3	$[12(AAF)+3H]^{3+}$
	13	1331	13	3	$[13(AAF)+3H]^{3+}$
	14	1434	14	3	$[14(AAF)+3H]^{3+}$
	15	1536	15	3	$[15(AAF)+3H]^{3+}$
	16	1638	16	3	$[16(AAF)+3H]^{3+}$
	17	1741	17	3	$[17(AAF)+3H]^{3+}$
	18	1843	18	3	$[18(AAF)+3H]^{3+}$
	19	1945	19	3	$[19(AAF)+3H]^{3+}$
	20	2048	20	3	$[20(AAF)+3H]^{3+}$
	21	2150	21	3	$[21(AAF)+3H]^{3+}$
	17	1306	17	4	$[17(AAF)+4H]^{4+}$

+4	18	1383	18	4	$[18(AAF)+4H]^{4+}$
	19	1459	19	4	$[19(AAF)+4H]^{4+}$
	20	1536	20	4	$[20(AAF)+4H]^{4+}$
	21	1613	21	4	$[21(AAF)+4H]^{4+}$
	22	1690	22	4	$[22(AAF)+4H]^{4+}$
	23	1766	23	4	$[23(AAF)+4H]^{4+}$
	24	1843	24	4	$[24(AAF)+4H]^{4+}$

Table S3: Self-assembled higher order clusters of AFA obtained from ESI–IMS–MS experiment.

Family	symbol	m/z	N	Z	Assembly
+1	1	308	1	1	$[1(AFA)+H]^{+}$
	2	615	2	1	$[2(AFA)+H]^{+}$
	3	922	3	1	$[3(AFA)+H]^{+}$
	4	1229	4	1	$[4(AFA)+H]^{+}$
	5	1536	5	1	$[5(AFA)+H]^{+}$
+2	7	1076	7	2	$[7(AFA)+2H]^{2+}$
	8	1229	8	2	$[8(AFA)+2H]^{2+}$
	9	1383	9	2	$[9(AFA)+2H]^{2+}$
	10	1536	10	2	$[10(AFA)+2H]^{2+}$
	11	1690	11	2	$[11(AFA)+2H]^{2+}$
	12	1843	12	2	$[12(AFA)+2H]^{2+}$
	13	1997	13	2	$[13(AFA)+2H]^{2+}$
+3	12	1229	12	3	$[12(AFA)+3H]^{3+}$
	13	1331	13	3	$[13(AFA)+3H]^{3+}$
	14	1434	14	3	$[14(AFA)+3H]^{3+}$
	15	1536	15	3	$[15(AFA)+3H]^{3+}$

Table S4: Self-assembled higher order clusters of FAA obtained from ESI–IMS–MS experiment.

Family	symbol	m/z	N	Z	Assembly
+1	1	308	1	1	$[1(FAA)+H]^{+}$
	2	615	2	1	$[2(FAA)+H]^{+}$
	3	922	3	1	$[3(FAA)+H]^{+}$
	4	1229	4	1	$[4(FAA)+H]^{+}$
	5	1536	5	1	$[5(FAA)+H]^{+}$
+2	6	922	6	2	$[6(FAA)+2H]^{2+}$
	6	940	6	2	$[6(FAA)+K]^{2+}$

5. Aggregation analysis by NMR spectroscopy

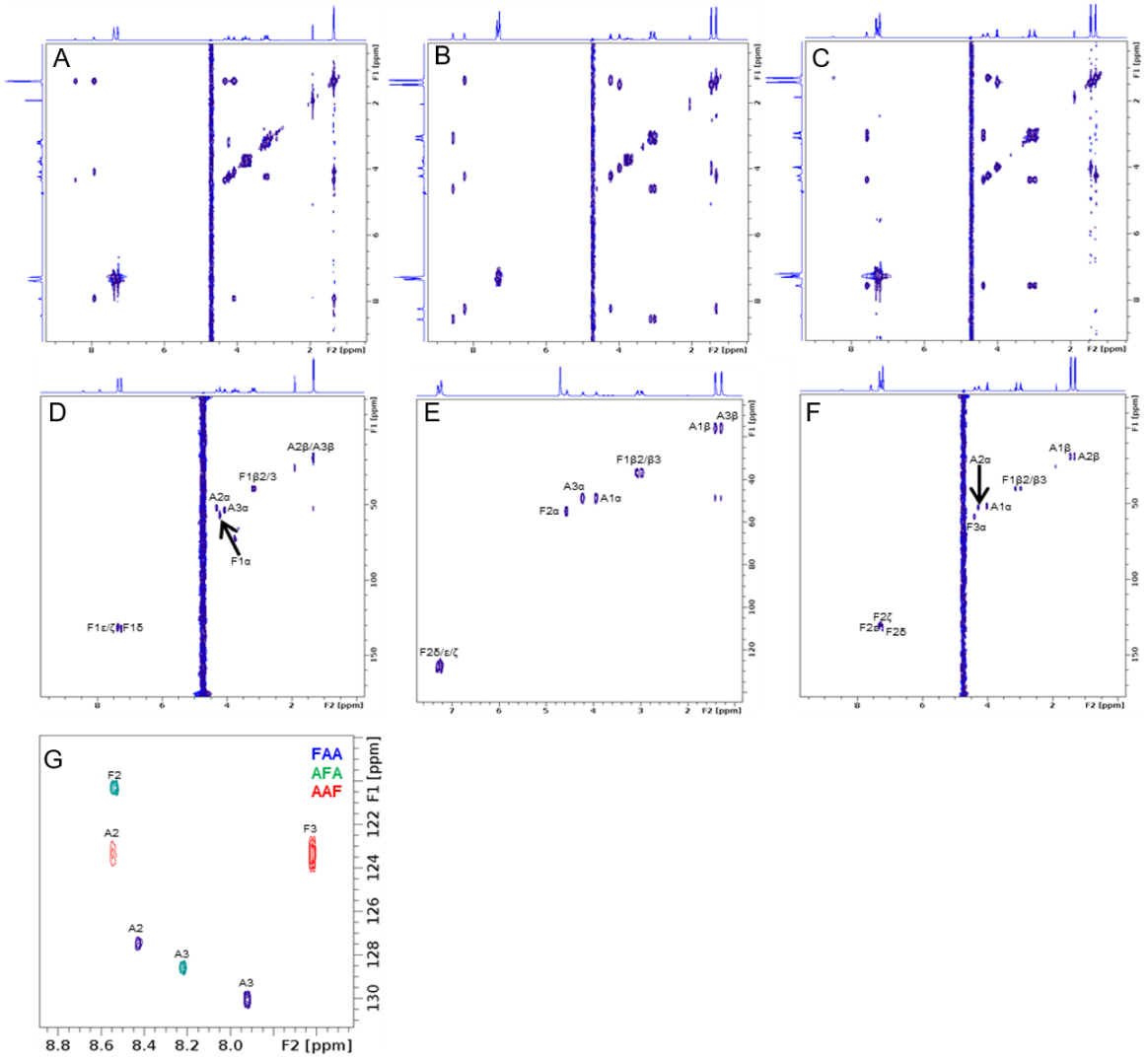


Figure S13. Two-dimensional NMR spectra of tripeptides. The [^{1}H-^{1}H]-TOCSY (**A-C**) and [^{13}C-^{1}H]-HSQC (**D-F**) NMR spectra of FAA (**A,D**), AFA (**B,E**) and AAF (**C,F**) recorded in 100 % $^{2}H_2O$ or 90 % H_2O/ 10 % $^{2}H_2O$. (**G**) An overlay of [^{15}N-^{1}H]-HSQC spectra obtained from all three tripeptides.

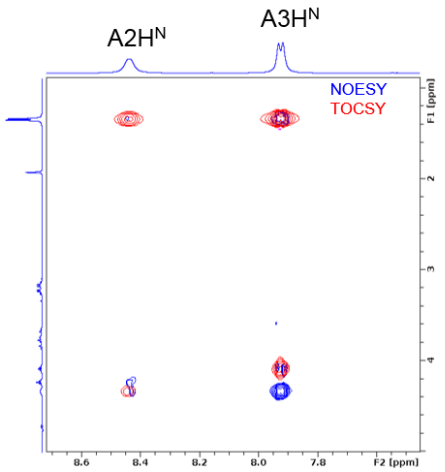


Figure S14. Two-dimensional NMR spectra of FAA. The [^{1}H-^{1}H]-NOESY and [^{1}H-^{1}H]-TOCSY NMR spectra of FAA recorded in 90 % H_2O/ 10 % ^2H_2O.

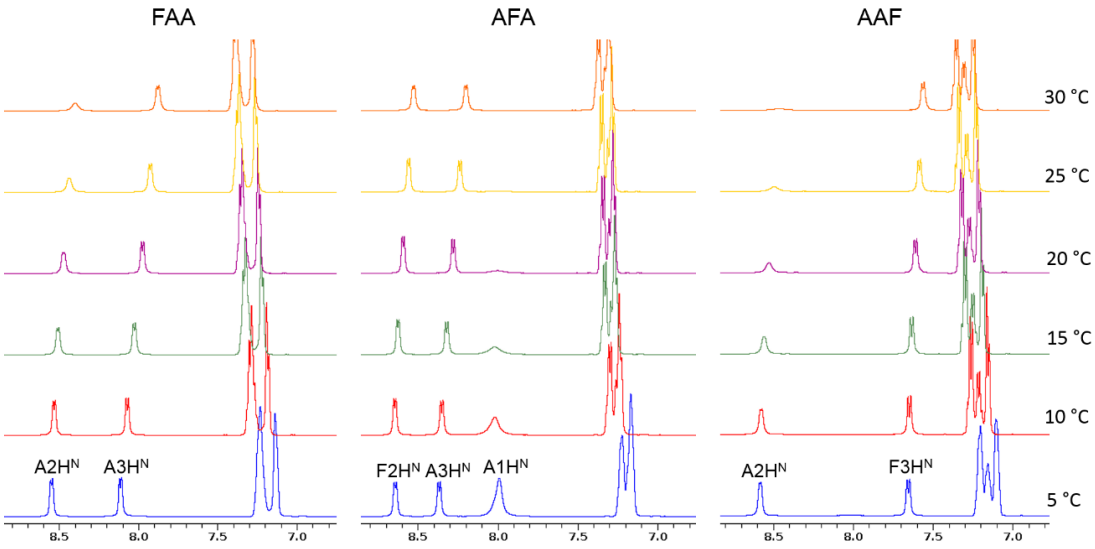


Figure S15. Temperature dependent ^{1}H NMR spectra of the tripeptides. The spectra were recorded in 90 % H_2O/ 10 % ^2H_2O at the indicated temperatures.

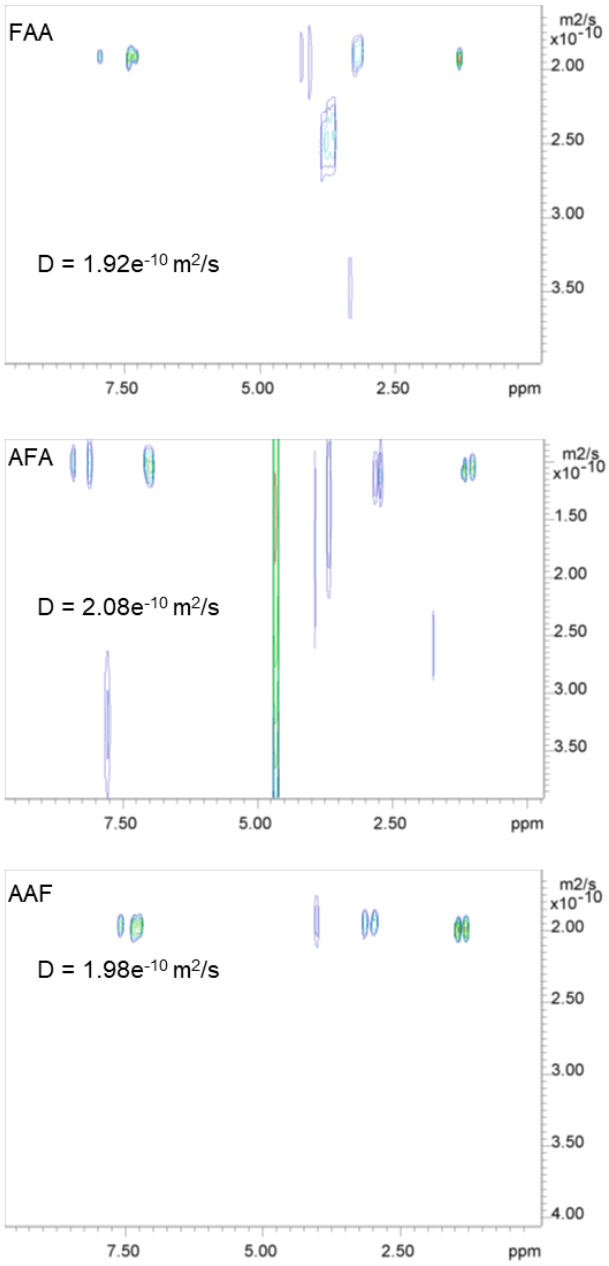


Figure S16. DOSY NMR spectra of tripeptides. The spectra were recorded in 90 % H_2O/ 10 % ^2H_2O. The experimentally determined diffusion coefficients are indicated in the corresponding spectrum for each peptide.

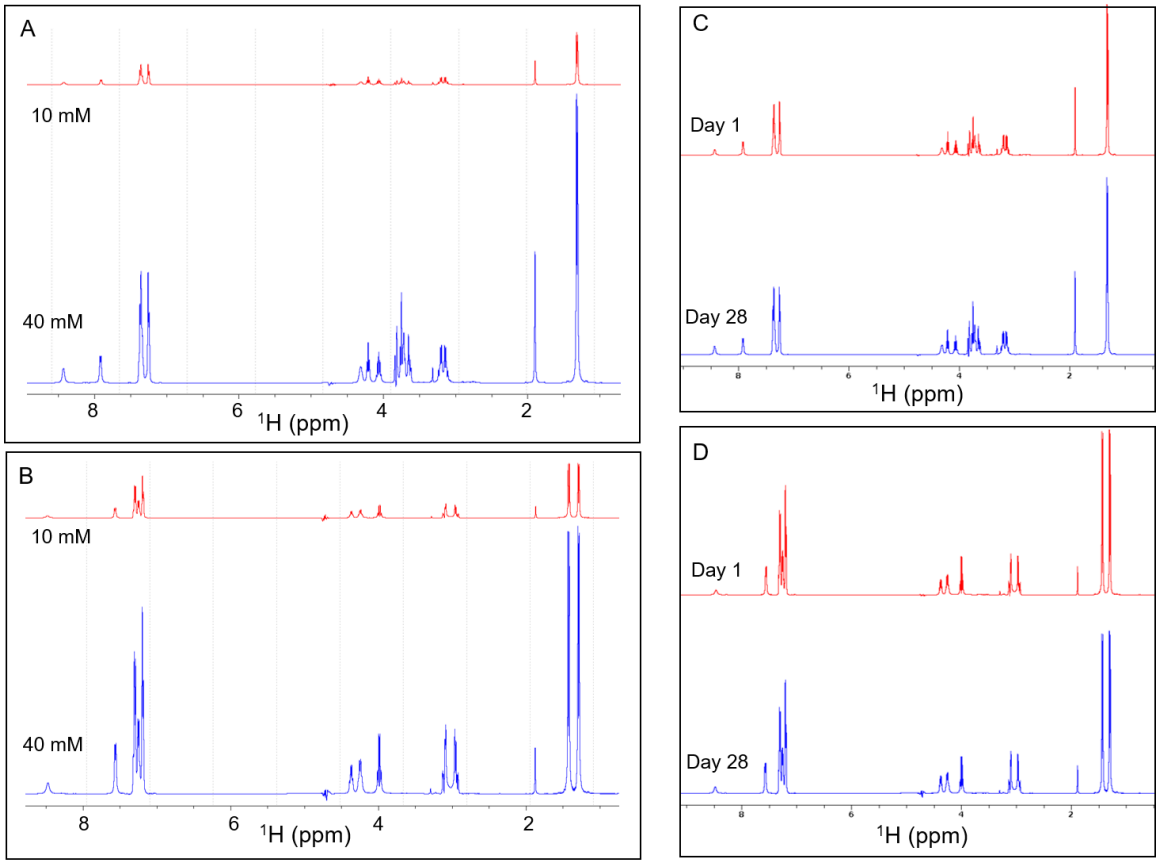


Figure S17. Concentration and time dependent ^{1}H NMR experimental results. (**A,B**) ^{1}H NMR spectra of FAA (**A**) and AAF (**B**) recorded in 90 % H_2O/ 10 % ^2H_2O at 10 mM (red) and 40 mM (blue) concentrations. (**C,D**) Time-dependent ^{1}H NMR spectra of FAA (**C**) and AAF (**D**) recorded in 90 % H_2O/ 10 % D2O at 40 mM concentration.

Table S5. $^{3}JH^{N}\alpha$ values (in Hz) and temperature coefficients for the tripeptides investigates in this study

FAA	**F1**	**A2**	**A3**
$^{3}JH^{N}\alpha$ at 5 °C	-	6.4	6.9
$^{3}JH^{N}\alpha$ at 25 °C	-	-	7.2
Temp. Coefft. (ppb/K)	-	2.8	9.2
AFA	**A1**	**F2**	**A3**
$^{3}JH^{N}\alpha$ at 5 °C	-	6.9	7.4
$^{3}JH^{N}\alpha$ at 25 °C	-	6.7	7.3
Temp. Coefft. (ppb/K)	-	4.8	6.8
AAF	**A1**	**A2**	**F3**
$^{3}JH^{N}\alpha$ at 5 °C	-	5.8	8.6
$^{3}JH^{N}\alpha$ at 25 °C	-	-	8
Temp. Coefft. (ppb/K)	-	4.8	3.6

Table S6. Translational diffusion coefficients and hydrodynamic radii for the tripeptides investigated in this study.

Peptide	**Diffusion coefficient ($e^{-10}m^2/s$)**	**Hydrodynamic radius(nm)**
FAA	1.92	1.28
AFA	2.08	1.18
AAF	1.98	1.24

6. Molecular dynamic simulations

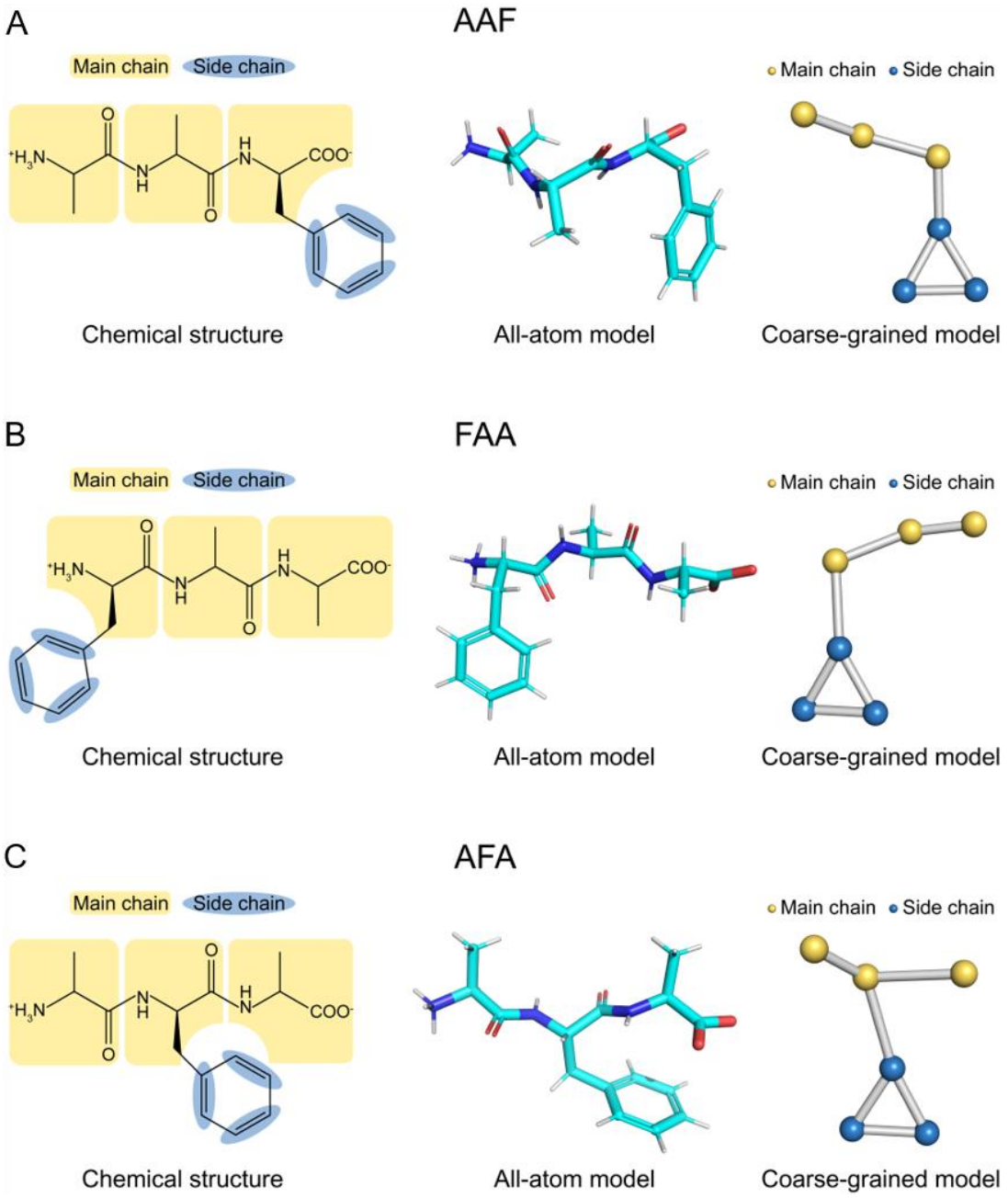


Figure S18: Chemical structure, all-atom model and coarse-grained model of tripeptides. (**A**) AAF, (**B**) FAA and (**C**) AFA molecules.

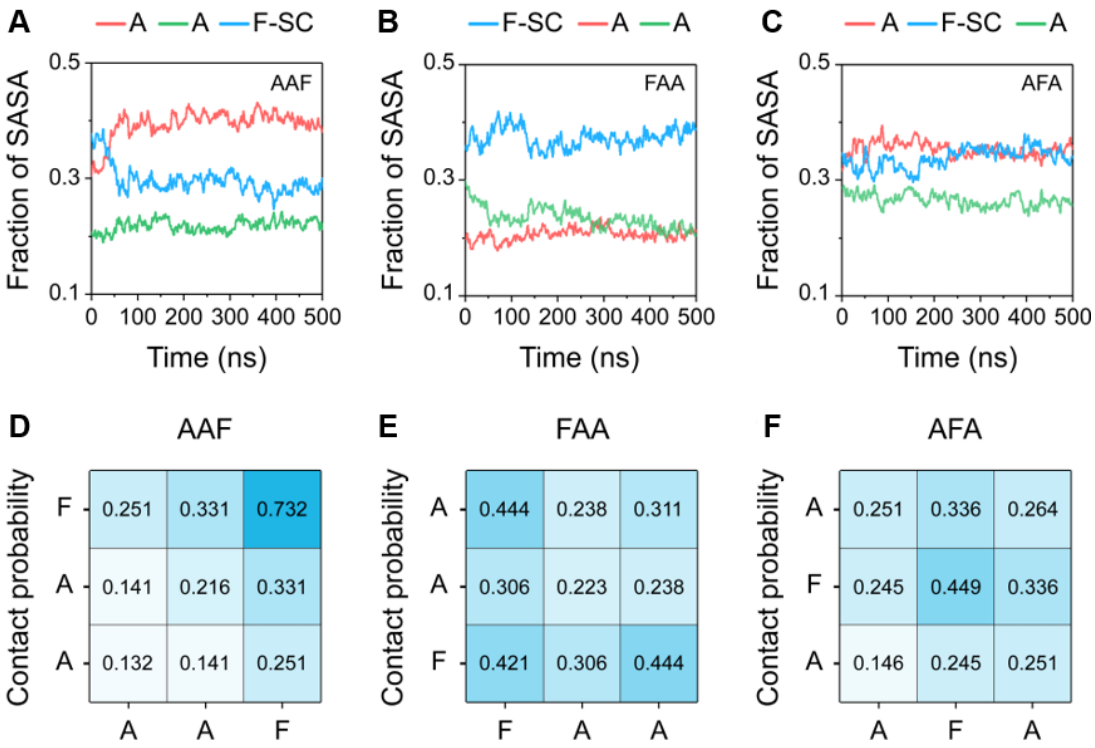


Figure S19: 6. Molecular dynamic simulations of tripeptides. (A-C) Time evolution of the SASA fraction of Ala residues and Phe side chain in AAF, FAA and AFA. (**D-F**) Contact probability for each residue pair of AAF, FAA and AFA.

7. X-ray structure analysis of AFF

A

Torsion angle	ϕ	ψ
Molecule A	-144.20	150.81
Molecule B	-127.95	135.10

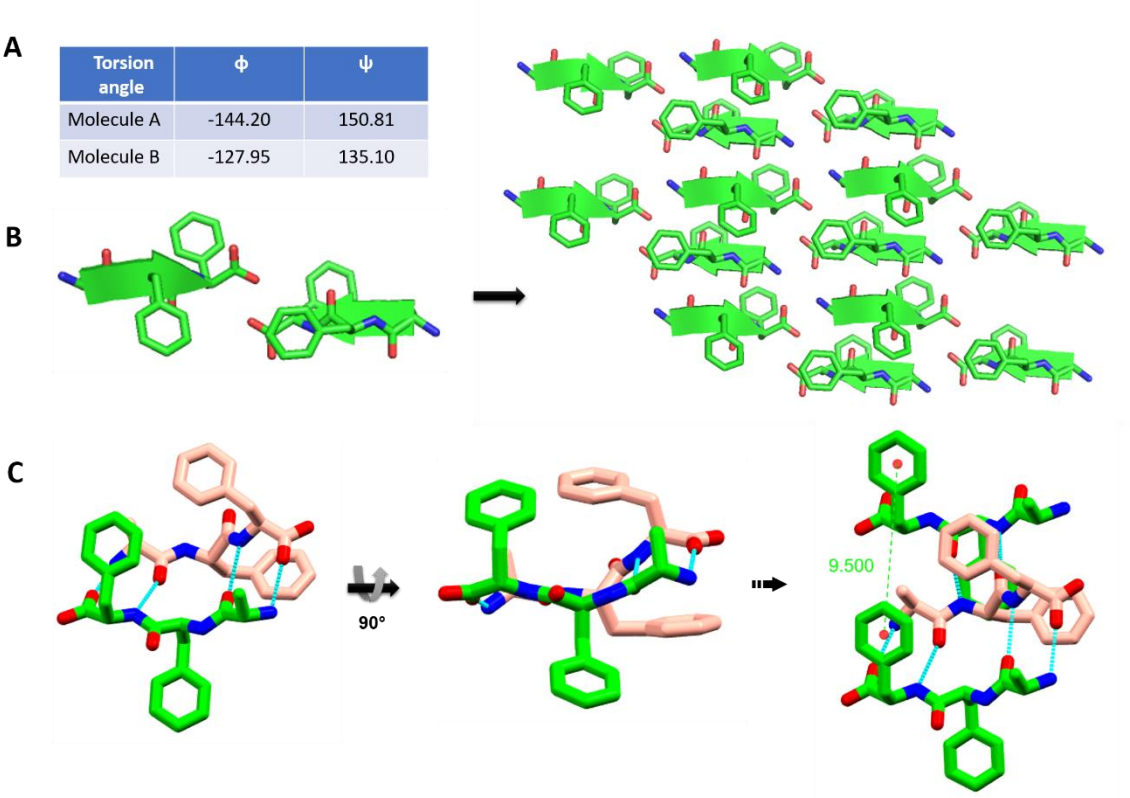


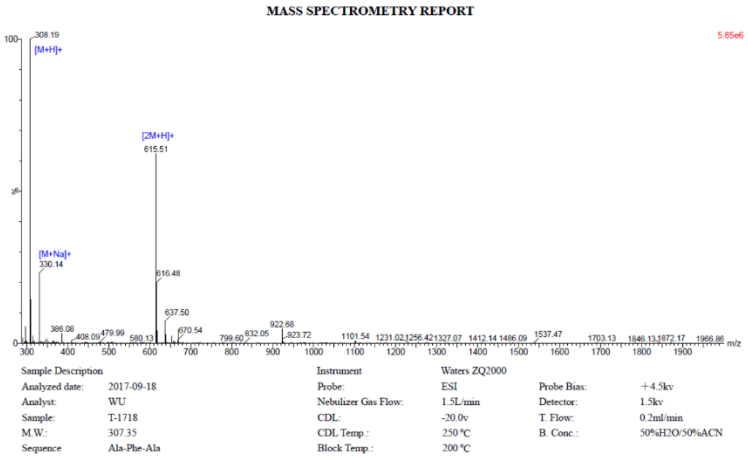
Figure S20: Crystal structure of AFF (CCDC ref no. 1862583[22]). (**A**) Torsion angle of Phe_2 for the two different molecules present in the asymmetric unit. (**B**) The antiparallel β-sheet structure. (**C**) Absence of any π-π interactions in the solid state packing as observed from different directions.

Table S7: Solid-state packing dependent solubility of peptides obtained through screening of peptide library at different length-scale.

Chain length	Compound	Aromatic interaction exits?	No of aromatic group involve in stacking	Solubility (mM)	Melting point(°C)
Amino Acid	NH_3^+-F-COO^-	Yes	1	125	283
Dipeptide	NH_3^+-FF-COO^-.H_2O	Yes	2	8.1	360
	NH_3^+-FL-COO^-.H_2O	Yes	1	28.3	258
	NH_3^+-LF-COO^-.H_2O	Yes	1	19.4	218
Tripeptide	NH_3^+-FFF-COO^-	Yes	3	0.5	219
	NH_2^+-PFF-COO^-	Yes	2	2.6	297
	NH_3^+-GFF-COO^-.H_2O	Yes	2	14.5	225
	NH_3^+-AFF-COO^- .TFA.H_2O	**No**	-	**98.3 ◄**	243
	NH_3^+-AAF-COO^- .TFA	Yes	1	105	292
	NH_3^+-AFA-COO^- .TFA.AcOH.H_2O	**No**	-	**672 ◄**	257

8. Characterization of tripeptides

A



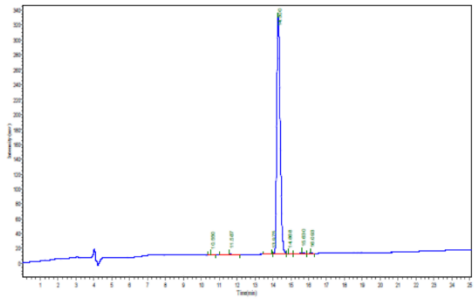
B

HPLC

Sample: T-1718 Analyzed date: 2017-09-15
Sequence Ala-Phe-Ala
Lot No DG-64378
Column: 4.6×250mm, Kromasil 100-5-C18
Solvent A A: 0.1% Trifluoroacetic Acid in 100% Acetonitrile
Solvent B B: 0.1% Trifluoroacetic Acid in 100% Water
Gradient:

	A	B
0.0min	12%	88%
25.0min	43%	57%
25.1min	100%	0%
30.0min	Stop	

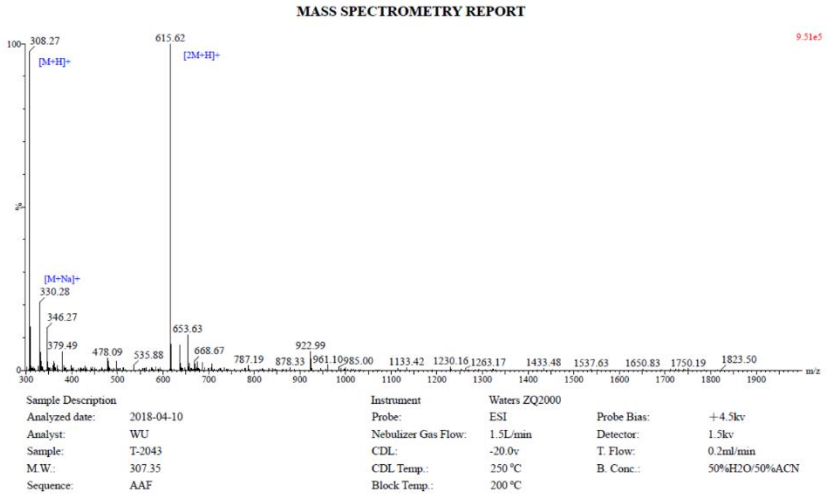
Volume: 5μl
Wavelength: 214nm
Flow rate: 1.0ml/min



Peak	Time	Height	Area	Conc.
1	10.550	676.331	6652.780	0.1523
2	11.587	562.015	17937.611	0.4106
3	13.975	396.370	5749.300	0.1316
4	14.300	318425.719	4306767.500	98.5845
5	14.868	666.012	8328.713	0.1906
6	15.630	1667.626	18634.463	0.4266
7	16.093	402.934	4536.835	0.1039
Total				100.000

Figure S21: Characterization of AFA. (**A**) Mass Spectra, (**B**) HPLC trace.

A



B

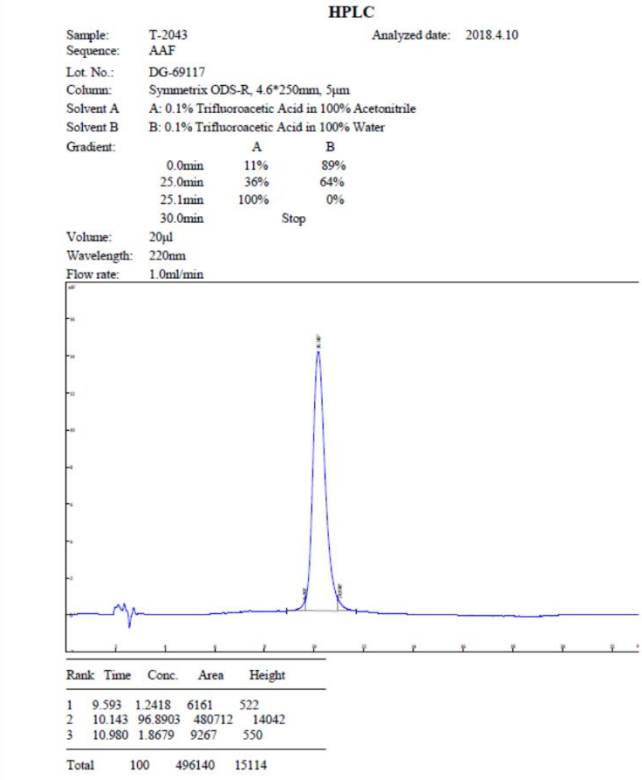
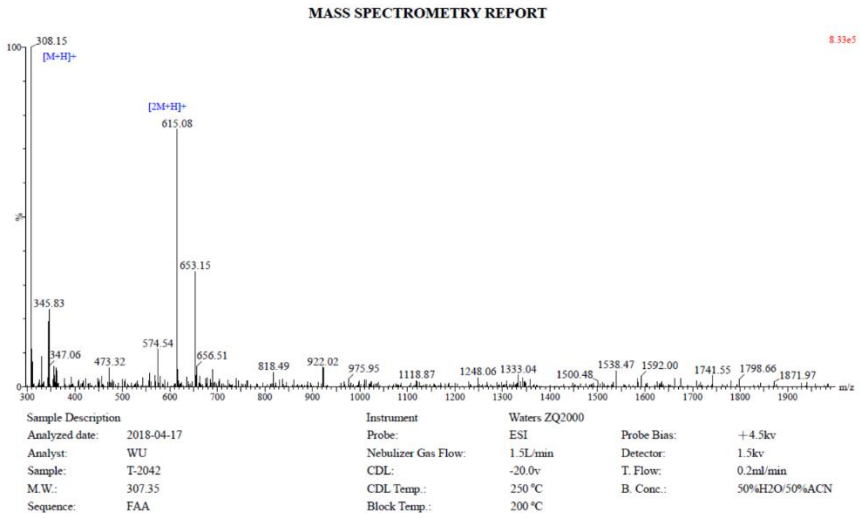


Figure S22: Characterization of AAF. (**A**) Mass Spectra, (**B**) HPLC trace.

A



B

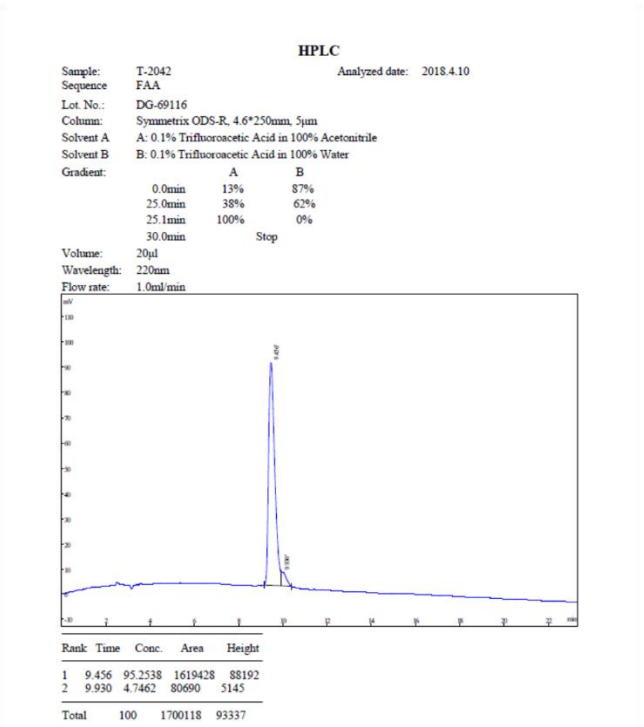


Figure S23: Characterization of FAA. (**A**) Mass Spectra, (**B**) HPLC trace.

Supplemental References

1. Sheldrick, G. (2013). SHELXL-2013. Univ. Göttingen, Göttingen, Ger.

2. Koeniger, S. L. *et al.* (2006). An IMS-IMS analogue of MS-MS. Anal. Chem. *78*, 4161–4174.

3. Merenbloom, S. I., Koeniger, S. L., Valentine, S. J., Plasencia, M. D., Clemmer, D. E. (2006). IMS-IMS and IMS-IMS-IMS/MS for separating peptide and protein fragment ions. Anal. Chem. *78*, 2802–2809.

4. Yao, S., Howlett, G. J., Norton, R. S. (2000). Peptide self-association in aqueous trifluoroethanol monitored by pulsed field gradient NMR diffusion measurements J. Biomol. NMR *16*, 109–119.

5. Abraham, M. J. *et al.* (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25.

6. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., Jorgensen, W. L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J. Phys. Chem. B *105*, 6474–6487.

7. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

8. Hess, B., Bekker, H., Berendsen, H. J. C., Fraaije, J. G. E. M. (1997). LINCS: A Linear Constraint Solver for molecular simulations. J. Comput. Chem. *18*, 1463–

1472.

9. Miyamoto, S., Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. *13*, 952–962.

10. Bussi, G., Donadio, D., Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*, 014101.

11. Parrinello, M., Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. *52*, 7182–7190.

12. Darden, T., York, D., Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. *98*, 10089–10092.

13. Dutta, M. S., Basu, S. (2021). Identifying the Key Residues Instrumental in Imparting Stability to Amyloid Beta Protofibrils – a Comparative Study Using MD Simulations of 17–42 Residues. J. Biomol. Struct. Dyn., 39, 431-456.

14. Lee, H., Ostadhassan, M., Sun, Z., Pu, H., Liu, B., Varma, R. S., Jang, H. W., Shokouhimher, M. (2020). RSC Adv. *10*, 37938.

15. Liu, J., Xu, Q., Jiang, J. (2019). A Molecular Simulation Protocol for Swelling and Organic Solvent Nanofiltration of Polymer Membranes. J. Membr. Sci., *573*, 639-646.

16. Kamat, S., Lin, R., Chiew, Y. C. (2019). Structure and Properties of Dicarboxylic Acids at Hexane/Water Interface: A Molecular Dynamics Study. Colloids Surf. A, *580*, 123725.

17. Monticelli, L. *et al.* (2008). The MARTINI coarse-grained force field: Extension

to proteins. J. Chem. Theory Comput. *4*, 819–834.

18. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., De Vries, A. H. (2007). The MARTINI force field: Coarse grained model for biomolecular simulations. J. Phys. Chem. B *111*, 7812–7824.

19. Burley, S. K., Petsko, G. A. (1985). Aromatic-aromatic interaction: A mechanism of protein structure stabilization. Science *229*, 23–28.

20. Humphrey, W., Dalke, A., Schulten, K. (1996). Sartorius products. J. Mol. Graph. *14*, 33–38.

21. Schrödinger, L. (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

22. Bera, S., Mondal, S., Xue, B., Shimon, L. J. W., Cao, Y., Gazit, E. (2019). Rigid Helical-like Assemblies from a Self-Aggregating Tripeptide. Nat. Mater. *18*, 503–509.