JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Rational Design of a Non-canonical "Sticky-Ended" Collagen Triple Helix

Abhishek A. Jalan, Katherine A. Jochim, and Jeffrey D. Hartgerink*

Departments of Chemistry and Bioengineering, Rice University, 6100 Main Street, Houston, Texas 77005, United States

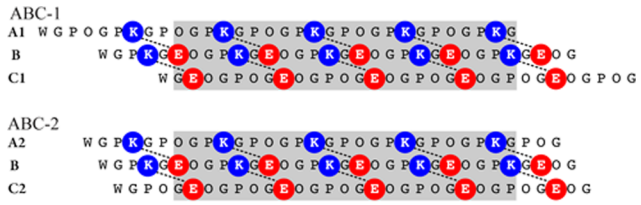
Supporting Information

ABSTRACT: In a canonical collagen triple helix, three peptides self-assemble into a supercoiled motif with a one-amino-acid offset between the peptide chains. Design of triple helices that contain more than one residue offset is lucrative, as it leaves the non-covalent interactions unsatisfied at the termini and renders the termini "sticky" to further self-assemble into collagen-like nanofibers. Here we use lysine−glutamate axial salt-bridges to design a heterotrimeric collagen triple helix, ABC-1, containing a non-canonical offset of four residues between the peptide chains. The four-residue offset is necessary to prevent aggregation, which would prevent characterization of the non-canonical chain arrangement at the molecular level by NMR spectroscopy. A second heterotrimer, ABC-2, also stabilized by axial salt-bridges, is designed containing a canonical one-amino-acid offset to facilitate comparison of structure and stability by CD and NMR. ABC-1 and ABC-2 demonstrate our ability to modulate chain offset in a collagen triple helix. This lays the groundwork to design longer, and therefore stickier, offsets allowing access to a new class of collagen-related nanostructures.

In DNA origami,[1] the adenine−thymine and guanine−cytosine nucleotide pairing is used to create sticky-ended motifs that self-assemble into complex nanostructures.[2−4] Similarly, in α-helical coiled-coil design,[5,6] electrostatic interactions between side chains of acidic and basic amino acids patterned at the *e* and *g* positions of the *abcdefg* heptad repeat introduce offset in an otherwise blunt-ended oligomer and nucleate fiber formation.[7,8] Therefore, design of protein folds containing unsatisfied polar or electrostatic contacts at the N- or C-termini is of interest from a supramolecular assembly perspective. Sticky-ended motifs can be routinely designed for DNA and α-helical coiled coils. Although a covalently linked sticky-ended collagen triple helix has been reported before,[9] proof-of-principle design of a non-covalently self-assembled sticky-ended collagen triple helix has remained elusive. Here we use axial salt-bridge interactions to place the three peptides of a collagen triple helix in a non-canonical four-amino-acid offset. The sticky ends are short and do not promote fiber formation, which allows us to confirm the intended molecular arrangement within the triple helix.

In a collagen triple helix, three peptides containing Xaa-Yaa-Gly triplet amino acid repeats supercoil, with a one-amino-acid offset between the peptide chains. The one-residue offset maximizes inter-peptide hydrogen bonds and optimizes helical packing.[10] Gly residues form the helical core, and therefore higher order assembly of the triple helices can only be mediated by patterning residues in the solvent-exposed Xaa and Yaa positions.[11] Using short peptides that mimic the triple-helical protein fold of native collagen, we recently elucidated sequence rules for interpeptide salt-bridge interactions between a Yaa-positioned Lys and an Xaa-positioned Asp or Glu residue.[12] We call these "axial" salt-bridges, as both side-chain conformations are approximately coaxial with the helical axis. Each Gly residue of the triple helix defines a cross-sectional plane that also contains the Xaa and Yaa residues of the other two chains. In

Scheme 1. Sequences of Non-canonical Offset ABC-1 and Canonical Offset ABC-2 Heterotrimers, Indicating Positions of Lys−Glu Axial Salt-Bridges[a]



[a]Gray highlighted area indicates the core region which is identical in both triple helices.

Table 1. Sequences, Abbreviations, and Melting Temperatures of Triple Helices Studied

triple helix sequence[a]	abbreviation	T_m (°C)
WG(POGPKG)$_2$PO**G**PKG(POGPKG)$_2$	A1	–
WG(PKGPOG)$_2$PKGPO**G**(PKGPOG)$_2$	A2	–
WG(PKGEOG)$_2$PK**G**EOG(PKGEOG)$_2$	B	32
WG(EOGPOG)$_2$EO**G**POG(EOGPOG)$_2$	C1	45
WG(POGEOG)$_2$PO**G**EOG(POGEOG)$_2$	C2	42
WG(POGPKG)$_5$/WG(PKGEOG)$_5$	A1B	36
WG(PKGPOG)$_5$/WG(PKGEOG)$_5$	A2B	39
WG(PKGEOG)$_5$/WG(EOGPOG)$_5$	BC1	44
WG(PKGEOG)$_5$/WG(POGEOG)$_5$	BC2	41
WG(POGPKG)$_5$/WG(EOGPOG)$_5$	A1C1	*m*[a]
WG(PKGPOG)$_5$/WG(POGEOG)$_5$	A2C2	*m*[b]
WG(POGPKG)$_5$·WG(PKGEOG)$_5$·WG(EOGPOG)$_5$	ABC-1	47
WG(PKGPOG)$_5$·WG(PKGEOG)$_5$·WG(POGEOG)$_5$	ABC-2	52

[a]**G** indicates the position of the ^{15}N label. [b]Broad melting transition.

Received: January 6, 2014
Published: May 14, 2014

   dx.doi.org/10.1021/ja5001246 | *J. Am. Chem. Soc.* 2014, 136, 7535−7538

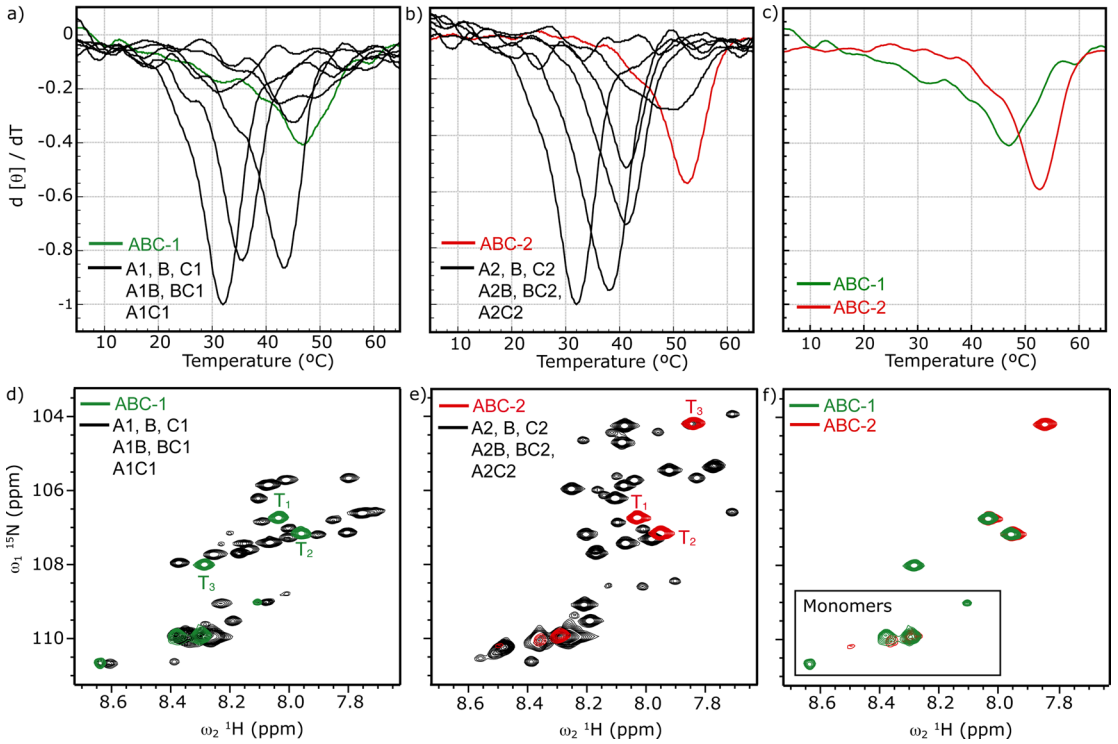


Figure 1. (Top) First-derivative plots of thermal unfolding curves of (a) ABC-1 and (b) ABC-2 overlaid with those of component peptides (shown in black) and (c) overlay of ABC-1 and ABC-2 spectra. (Bottom) ^{1}H,^{15}N-HSQC NMR spectra of (d) ABC-1 and (e) ABC-2 overlaid with those of component peptides (shown in black) and (f) overlay of ABC-1 and ABC-2 HSQC spectra, demonstrating the similarity of two of the three N−H peaks. Monomer peaks are indicated within the box.

terms of sequence position, an axial salt-bridge forms between the Lys in the *n*th and the Asp or Glu in the *n*+3 cross-sectional plane. The energetic stabilization derived from the interaction between Lys and Asp/Glu residues present within the same cross-sectional plane, which we call a "lateral" interaction, has been shown to be conformationally less robust than an axial salt-bridge.[13]

We have used axial salt-bridges to design AAA-,[14] AAB-,[15,16] and ABC-type[17−19] collagen triple helices. Of particular interest here is the peptide $(PKG)_4(POG)_4(DOG)_4$ that rapidly self-assembles into long fibers and ultimately a hydrogel in aqueous solution.[14] To explain the rapid self-assembly, it was proposed that the peptide forms a transient homotrimeric collagen triple helix with a 20-amino-acid offset between the peptide chains. The unusually long offset would contain numerous unsatisfied hydrogen-bonding partners as well as unpaired and electrostatically charged Lys and Asp residues, which could drive rapid aggregation. However, due to the fibrous morphology of the aggregates, the offset within the collagen triple helix could not be elucidated using any biophysical tool available. Control over peptide offset in a triple helix has the potential to provide unprecedented control over its supramolecular fate. Therefore, in this Communication we demonstrate a proof-of-principle design of a collagen triple helix, ABC-1, with a four-residue offset between the peptide chains. The short offset reduces the likelihood of aggregation, allowing us to study its structure and stability in molecular detail by NMR. A second collagen triple helix, ABC-2, with identical amino acid composition but containing a one-amino-acid offset, is also designed. These two systems demonstrate our ability to modulate chain offsets in a collagen triple helix and lay the groundwork for design of a new class of collagen-related nanostructures.

ABC-1 and ABC-2 are identical with respect to the triple-helical core (see Scheme 1, shaded area), which facilitates comparison of their stability and molecular structure. All peptides were synthesized with an N-terminal Trp to determine accurate concentration and one ^{15}N-isotopically enriched Gly (^{15}N-Gly) per peptide for NMR analysis (Table 1). Individual peptides and their 1:1 binary and 1:1:1 ternary mixtures were examined via circular dichroism (CD) and NMR at 1 and 3 mM total peptide concentration, respectively, in 10 mM Tris buffer at pH 7.2. Synthesis, purification, and sample preparation methods are provided in the Supporting Information. The abbreviations and melting temperatures (T_m) of all peptide systems examined in this study are listed in Table 1.

Assuming a canonical single-amino-acid offset between peptides, a mixture of three peptides, A, B, and C, can self-assemble into 27 unique triple helices. Axial Lys−Asp/Glu salt-bridges have been used to bias the ensemble of competing states to selectively populate one AAB-[16] or ABC-type[17] heterotrimer of desired chain composition and register. In the case of ABC-1, we placed Lys and Glu residues such that they only form axial salt-bridges if the peptide chains adopt a *four*-residue offset, therefore setting the stage for an offset or sticky-ended assembly.

The thermal melting curve of ABC-1 shows a broad transition with a major unfolding event at 47 °C (Figure 1a). In addition to a substantial decrease in van der Waals surface area, ABC-1 loses four inter-peptide hydrogen bonds at the termini due to the four-residue offset (Figure S5). The resulting conformational disorder is expected to result in a broader thermal transition, as terminal regions unfold before the more stable core of the triple helix. In comparison, ABC-2 shows a cleaner thermal transition, with a higher T_m value of 52 °C

dx.doi.org/10.1021/ja5001246 | J. Am. Chem. Soc. 2014, 136, 7535−7538

(Figure 1b). The overlay of their spectra in Figure 1c emphasizes the difference in unfolding of the two heterotrimers. The broad unfolding transition observed in ABC-1 could also be attributed to the presence of multiple competing triple helical states. To rule out this possibility, we performed NMR analysis.

Due to the difference in hydrogen bonds formed by the ^{15}N-Gly amide protons, each of the competing triple-helical states in a mixture of three peptides yields a fingerprint spectrum in a ^{1}H,^{15}N heteronuclear single-quantum coherence (HSQC) experiment. The HSQC spectra of ABC-1 shows three trimer cross-peaks labeled T_1, T_2, and T_3 (Figure 1d), which do not overlap with cross-peaks of individual peptides or their binary mixtures. This suggests that mixing peptides A1, B, and C1 results in only one triple helix of A1·B·C1 composition. Additionally, the presence of only three trimer cross-peaks suggests that the three peptides in ABC-1 are arranged in only one register. Similar conclusions can be drawn from the HSQC spectra of ABC-2 and its component peptides in Figure 1e.

As shown in Figure 1f, two trimer cross-peaks in the HSQC spectra of ABC-1 and ABC-2 have very similar chemical shifts. A comparison of their amino acid sequence in Figure S6 indicates that two of the three inter-peptide hydrogen bonds formed by ^{15}N-Gly(NH) are between identical sets of amino acid triplets. In contrast, the third hydrogen bond is formed between dissimilar triplets, which explains the observed overlap in only two of the three cross-peaks. Importantly, if the ABC-1 were to adopt a canonical one-residue offset, all three hydrogen bonds, and therefore HSQC peaks, would be chemically distinct from those of ABC-2, and no cross-peak overlap would be expected. This result provides the first molecular-level evidence for the four-residue chain offset in ABC-1.

Additional evidence for the intended chain offset comes from homonuclear correlation NMR experiments. A two-dimensional ^{1}H−^{1}H plane of the three-dimensional ^{1}H,^{15}N-NOESY-HSQC experiment (hereafter called edited-NOESY) shows nuclear Overhauser effect (NOE)-based cross-peaks between ^{15}N-Gly(NH) and protons within ~6 Å of it. Since the intensity of the cross-peaks varies as $1/r^6$, where r is the spatial distance between the correlating protons, generally no cross-peaks are observed at distances greater than 6 Å. Due to the four-residue offset in ABC-1, the ^{15}N-Gly amide protons of the A1, B, and C1 chains are more than 11.5 Å apart (Figure 2a). As a result, the edited-NOESY spectrum of ABC-1 shows no cross-peaks for correlations between the amide protons (Figure 2b, red dotted circles). Additionally, ^{15}N-Gly amide protons are also placed more than 10 Å apart from Hα of the residues preceding ^{15}N-Gly of another chain. As expected, these ^{15}NH-Hα correlations are also not observed (Figure 2b, blue dotted circles).

Analysis of the edited-NOESY spectra of ABC-2 confirms that the absence of NOE cross-peaks in ABC-1 is not an experimental artifact. As shown in Figure 3a, the ^{15}N-Gly's of A2, B, and C2 chains of ABC-2 heterotrimer are within the observable NOE distance of 6 Å. Therefore, both ^{15}NH−^{15}NH and ^{15}NH−Hα correlations are observed in its edited-NOESY spectra (Figure 3b). The accompanying homology model in Figure 3c, built from PDB 2KLW,[12] indicates the observed NOEs.

The peptides A2, B, and C2 can arrange in six different ways within the ABC-2 heterotrimer, but pairing of all Lys and Glu residues is only possible in one of them. Comparison of the

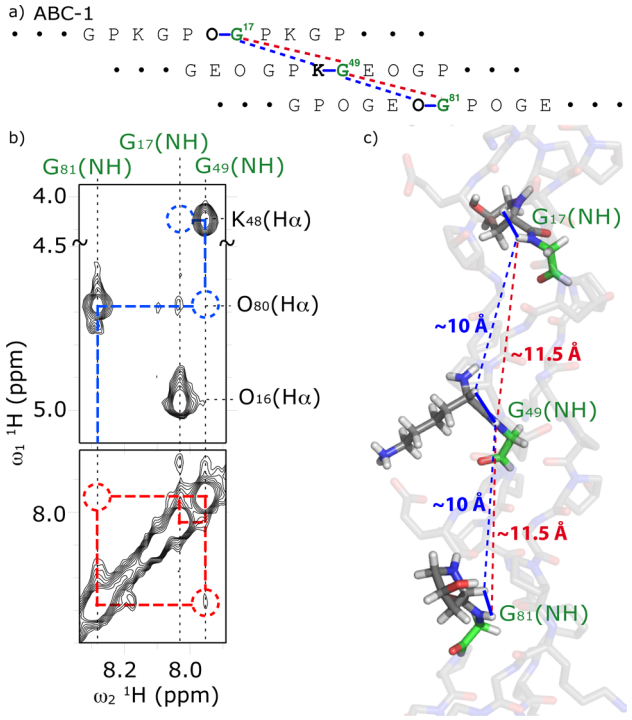


Figure 2. (a) Sequence, (b) ^{15}N-edited-NOESY spectrum, and (c) model of ABC-1 indicating absence of *inter-peptide* ^{15}NH−^{15}NH (red lines) and ^{15}NH−Hα (blue lines) NOE cross-peaks due to >6 Å separation between the protons.

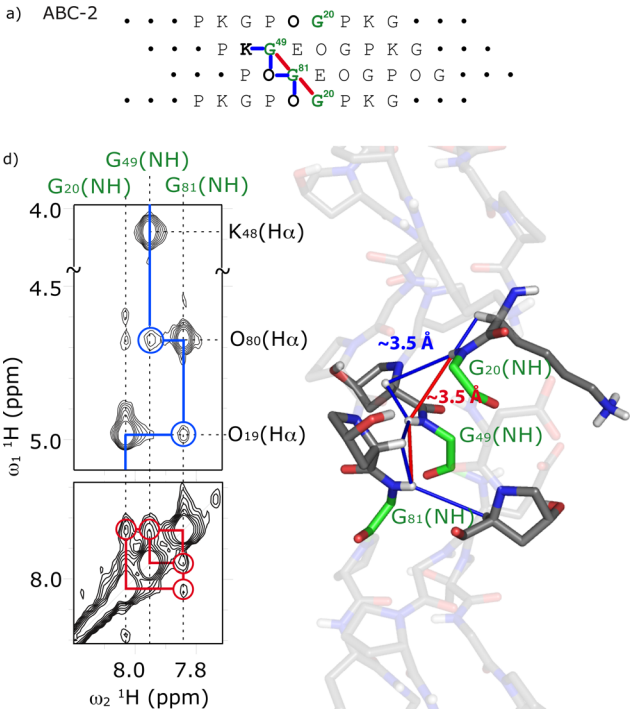


Figure 3. (a) Sequence, (b) ^{15}N-edited-NOESY spectrum, and (c) model of ABC-2 indicating absence of *inter-peptide* ^{15}NH−^{15}NH (red lines) and ^{15}NH−Hα (blue lines) NOE cross-peaks due to >6 Å separation between the protons.

observed ^{15}NH−^{15}NH and ^{15}NH−Hα correlations to those expected in the six possible peptide arrangements in Table S1 confirms that the register of A2, B, and C2 shown in Scheme 1 is correct, allowing all Lys and Glu residues to form salt-bridges.

dx.doi.org/10.1021/ja5001246 | J. Am. Chem. Soc. 2014, 136, 7535−7538

Similarly, the relative chain arrangement in ABC-1 depicted in Scheme 1 is critical for salt-bridge formation. To determine the peptide positions in ABC-1, we compared its two-dimensional ^{1}H,^{1}H-NOESY spectra to that of ABC-2 (Figure S9). Due to identical triple-helical core regions, their NOESY spectra are nearly identical, indicating similarity in chain arrrangement within the triple helix. Furthermore, a NOE cross-peak between E(NH) and K(Hε), characteristic of an axial salt-bridge,[13,19] is observed in the NOESY spectra of both heterotrimers (Figure S9). These results strongly suggest that the relative positions of peptide chains in both heterotrimers are as shown in Scheme 1. To further confirm the four-residue offset, we designed an alternative ABC system, discussed in Supporting Information section 1.5 and Figures S11−S13, that also supports this conclusion by moving the isotopic labels in the offset systems into close proximity to one another, resulting in the expected NOE correlations for a four-amino-acid offset.

The ABC-1 heterotrimer can putatively form 10 *lateral* salt-bridges while maintaining a canonical single-residue offset. Instead, our data suggest that the peptide chains in ABC-1 adopt a four-residue offset that sacrifices four hydrogen bonds at the termini in addition to significant van der Waals contact area to allow the formation of 10 *axial* salt-bridges. This demonstrates a clear energetic preference of axial over lateral salt-bridges. A similar preference for an axial salt-bridge has been demonstrated in systems where both contacts are feasible.[12] It has also been shown that Lys and Asp/Glu residues in lateral positions are conformationally frustrated due to favorable backbone polar contacts.[13] Although lateral interactions have been used in heterotrimeric design,[20,21] the results presented here add strong evidence of the dominance of axial interactions over lateral ones.

Characterization of supramolecular polymers is typically limited to relatively low resolution techniques such as AFM and TEM, leaving the exact mechanism of assembly vague and open to debate. This makes improvement and modification of these systems more difficult. Here we have created a model system of a sticky-ended collagen helix which, while having offset termini, does not continue assembling due to the relatively short, four-amino-acid stagger. This has allowed us to examine the structure by NMR and reveal the individual amino acid interactions responsible for the structural organization. Our study gives molecular-level evidence for axial salt-bridge-driven formation of a sticky-ended collagen triple helix for the first time. The design can be readily adapted to create collagen triple helices with longer offset ends that hybridize into higher order fibrillar assemblies. Failure to observe a lateral salt-bridge-driven canonical triple helix indicates that axial salt-bridges are better suited for controlling self-assembly of collagen triple helices.

The design of triple helices becomes an increasingly challenging topic as the number of possible alternative structures is considered. Moving from a homotrimer to a canonical ABC heterotrimer increases the number of reasonable structure from 1 to 27. An ABC system that makes non-canonical, sticky-ended structures feasible increases the number of competing structures to well over 100, even if only one triplet of overhang is considered. By increasing the length of the overhang, more competing states can be populated resulting in a rougher folding landscape, which in turn increases the time of folding. The effect of this is to roughen the folding landscape and therefore substantially increase the time of folding. From one perspective, the folding time may eventually make even well-designed systems impractical. From another perspective, in truly well-designed systems, this additional challenge will be taken to heart, and negative design strategies will be used to engineer the folding landscape.

■ ASSOCIATED CONTENT

Supporting Information

Peptide synthesis, mass-spectrometric analysis, experimental parameters, and additional CD and NMR data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

jdh@rice.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the NSF (DMR-1206899) and the Robert A. Welch Foundation (Grant No. C1557). The authors thank Prof. B. Pettitt and Dr. Tianzhi Wang at UTMB, Galveston, for help with triple-resonance NMR experiments.

■ REFERENCES

(1) Rothemund, P. W. K. *Nature* **2006**, *440*, 297.
(2) Shih, W. M.; Quispe, J. D.; Joyce, G. F. *Nature* **2004**, *427*, 618.
(3) Goodman, R. P.; Berry, R. M.; Turberfield, A. J. *Chem. Commun.* **2004**, 1372.
(4) He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. D. *Nature* **2008**, *452*, 198.
(5) Gradisar, H.; Bozic, S.; Doles, T.; Vengust, D.; Hafner-Bratkovic, I.; Mertelj, A.; Webb, B.; Sali, A.; Klavzar, S.; Jerala, R. *Nat. Chem. Biol.* **2013**, *9*, 362.
(6) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. *Science* **2013**, *340*, 595.
(7) Pandya, M. J.; Spooner, G. M.; Sunde, M.; Thorpe, J. R.; Rodger, A.; Woolfson, D. N. *Biochemistry* **2000**, *39*, 8728.
(8) Gribbon, C.; Channon, K. J.; Zhang, W. J.; Banwell, E. F.; Bromley, E. H. C.; Chaudhuri, J. B.; Oreffo, R. O. C.; Woolfson, D. N. *Biochemistry* **2008**, *47*, 10365.
(9) Kotch, F. W.; Raines, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 3028.
(10) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. *Science* **1994**, *266*, 75.
(11) Okuyama, K.; Miyama, K.; Mizuno, K.; Bachinger, H. P. *Biopolymers* **2012**, *97*, 607.
(12) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851.
(13) Fallas, J. A.; Dong, J. H.; Tao, Y. Z. J.; Hartgerink, J. D. *J. Biol. Chem.* **2012**, *287*, 8039.
(14) O'Leary, L. E. R.; Fallas, J. A.; Bakota, E. L.; Kang, M. K.; Hartgerink, J. D. *Nat. Chem.* **2011**, *3*, 821.
(15) O'Leary, L. E.; Fallas, J. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2011**, *133*, 5432.
(16) Jalan, A. A.; Hartgerink, J. D. *Biomacromolecules* **2013**, *14*, 179.
(17) Fallas, J. A.; Hartgerink, J. D. *Nat. Commun.* **2012**, *3*, No. 1087.
(18) Jalan, A. A.; Demeler, B.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2013**, *135*, 6014.
(19) Fallas, J. A.; Lee, M. A.; Jalan, A. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2012**, *134*, 1430.
(20) Xu, F.; Zahid, S.; Silva, T.; Nanda, V. *J. Am. Chem. Soc.* **2011**, *133*, 15260.
(21) Xu, F.; Zhang, L.; Koder, R. L.; Nanda, V. *Biochemistry* **2010**, *49*, 2307.