pubs.acs.org/Biomac

Article

Cation−π Interactions and Their Role in Assembling Collagen Triple Helices

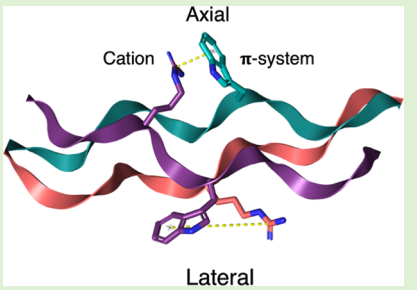
Carson C. Cole, Mikita Misiura, Sarah A. H. Hulgan, Caroline M. Peterson, Joseph W. Williams, III, Anatoly B. Kolomeisky, and Jeffrey D. Hartgerink*

Cite This: *Biomacromolecules* 2022, 23, 4645−4654

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Cation−π interactions play a significant role in the stabilization of globular proteins. However, their role in collagen triple helices is less well understood and they have rarely been used in de novo designed collagen mimetic systems. In this study, we analyze the stabilizing and destabilizing effects in pairwise amino acid interactions between cationic and aromatic residues in both axial and lateral sequential relationships. Thermal unfolding experiments demonstrated that only axial pairs are stabilizing, while the lateral pairs are uniformly destabilizing. Molecular dynamics simulations show that pairs with an axial relationship can achieve a near-ideal interaction distance, but pairs in a lateral relationship do not. Arginine−π systems were found to be more stabilizing than lysine−π and histidine−π. Arginine−π interactions were then studied in more chemically diverse ABC-type heterotrimeric helices, where arginine−tyrosine pairs were found to form the best helix. This work helps elucidate the role of cation−π interactions in triple helices and illustrates their utility in designing collagen mimetic peptides.



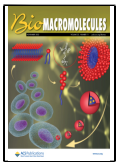
INTRODUCTION

Cation−π interactions have been shown to contribute important stabilization to globular proteins and are of interest due to their high frequency as a pairwise interaction in the proteome.[1−5] A cation−π interaction may occur between a cationic residue of lysine (K), arginine (R), or histidine (H) with the aromatic residues phenylalanine (F), tyrosine (Y), and tryptophan (W).[6] Gallivan and Dougherty initially noted that cation−π interactions occur at a high incidence and began to elucidate their greater than expected stabilizing properties in protein folding.[7] The quantification of these interactions has been determined experimentally and is also supported by the quantum mechanical theory.[8] Previously, studies of cation−π interactions in homotrimers quantified their expression in natural collagens and discussed their application in designing collagen mimetic triple helices.[9−12] In our current work, further clarification as to how the geometry of a triple helix is involved in each cation−π interaction is detailed and supported with both experimental and molecular dynamics (MD) approaches.

Collagen and collagen-like proteins are found in all domains of life, from prokaryotes to eukaryotes and even viruses.[1,13−15] Within humans, there are 28 traditional collagens that give rise to a variety of structural tissues, signaling molecules, and extracellular components.[16−20] In addition to these, there are a variety of collagen-like proteins, such as the family of "defense collagens" that contain triple-helical motifs, which epitomize all collagen proteins.[21−23] The triple helix is composed of three left-handed polyproline type II helices that wind around one another to form a right-handed superhelix.[24,25] The primary sequence of each peptide strand in the helix is comprised of Xaa−Yaa−Gly repeats, where Xaa and Yaa are frequently proline (P) and 4-hydroxyproline (O), respectively.[26] Glycine is required every third amino acid both for its small size and to provide an intrahelical hydrogen bond that helps drive folding. This hydrogen bond also results in a single amino acid offset between each polypeptide monomer differentiating a leading, middle, and trailing strand.[27,28] The relative alignment of each peptide strand differentiates specific registers where, for example, a heterotrimer with an ABC register is not equivalent to a BAC register.[29−31]

To differentially stabilize one registration, pairwise interactions (both stabilizing and destabilizing) are needed.[32−34] It has been shown that collagen's triple helix allows for two major types of pairwise interaction geometries, lateral and axial. In a lateral pairwise interaction, an amino acid in the Yaa position of one peptide strand interacts with the Xaa amino acid of the next strand (e.g., leading to middle, middle to trailing, or trailing to leading). In an axial pairwise interaction, the Yaa position amino acid interacts with a Xaa amino acid in a similar fashion to the lateral interaction but offset by one full triplet

Received: July 12, 2022
Revised: September 26, 2022
Published: October 14, 2022



https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654

toward the C-termini. When the Xaa or Yaa position is changed from a proline (P) or hydroxyproline (O), respectively, there is a loss of stability to the triple helix, but if there is a favorable, pairwise interaction between the two new amino acids, then some or all of the lost stability can be regained. Of the pairwise interactions studied so far, axial lysine−aspartate pairs offer the highest degree of stabilization and this lysine−aspartate axial interaction has been utilized in an de novo design to yield a variety of supramolecular collagen mimetic peptides (CMPs).[35−38]

Arginine and lysine are predominantly found in the Yaa position of triple-helical sequences. Aromatic residues phenylalanine and tyrosine have low frequency of expression and tryptophan has no expression in the proteome of natural collagens. When aromatic residues are found in the proteome they occur more frequently in the Xaa position.[39] Based on this natural bias of amino acid position, we set out to clearly define the stabilizing or destabilizing interactions between the cationic amino acids lysine, arginine, and histidine in the Yaa position with the aromatic amino acids phenylalanine, tyrosine, and tryptophan in the Xaa position and to differentiate these interaction based on their presentation in a lateral or axial arrangement. Our results clarify the ambiguity in previous literature studies by showing that the axial interaction is indeed favored over the lateral. Then using MD, a measure of the interaction distance and relative strain to the polyproline-II backbone between solvated monomers is calculated. Finally, a comparison of cation−π containing ABC-type heterotrimers is used to demonstrate how axial cation−π interactions can be used to tailor CMP designs even in complex environments.

EXPERIMENTAL METHODS

Peptide Synthesis. Fmoc-protected amino acids and resin were purchased from EMD Chemicals, and 2-(1*H*-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uranium hexafluorophosphate methanaminium (HATU) was purchased from P3Bio. All chemicals not otherwise specified were purchased from Sigma-Aldrich.

Peptides were synthesized using standard Fmoc-protected amino acids using a low loading rink amide MBHA (0.35 loading) resin to yield C-terminal amidation. A mixture of 25% v/v piperidine in dimethylformamide (DMF) was used for deprotecting steps. Coupling was performed using HATU and diisopropylethylamine (DiEA) in DMF in a ratio of 1:4:4:6 (resin/amino acid/HATU/DiEA, respectively). Acetylation of the N-terminus was performed twice with an excess of acetic anhydride and DiEA in dichloromethane. Cleavage was performed with 7.5% (v/v) scavengers (triisopropylsilane, H_2O, and anisole) in trifluoroacetic acid (TFA). Excess TFA was removed after cleavage from resins by a nitrogen flow. The crude peptide was triturated twice with cold diethyl ether.

Crude peptides were dissolved in H_2O to a concentration of 15 mg/mL. This was sonicated and then filtered with 0.3 μm syringe filter before purification by reverse-phase high-pressure liquid chromatography with water and acetonitrile with 0.05% TFA at a gradient of 0.85% per minute on a Waters XC18 column. Samples were roto-evaporated to remove acetonitrile and then lyophilized. Matrix-assisted laser desorption ionization time of flight mass spectrometry (Bruker Instruments) was used to confirm the peptide mass. All synthesized peptides and their characterization can be found in the Supporting Information.

Circular Dichroism. Circular dichroism (CD) spectroscopy was performed on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature-controlled stage. The samples were diluted with Milli-Q water to 0.03 mM for spectra and 0.3 mM for melts. Any tryptophan (W) containing peptide was diluted to 0.015 mM due to their innate absorption in CD. After being prepared, 200 μL of sample was transferred to a quartz cuvette of 0.1 cm path length. Wavelength scans were performed between 180 and 250 nm to confirm triple helical secondary structures. The maximum, which falls near 225 nm, is then followed as the temperature is increased from 5 to 65 °C at 10 °C/h. Using a first degree Savitzky−Golay smoothing algorithm, the first derivative curve was calculated for melting curves. The minimum of the first derivative is defined as the melting temperature (T_m). The MRE was calculated as previously reported.[40]

NMR Analysis. For NMR, all samples were prepared by mixing 500 μL of 3 mM peptide solution and 50 μL D_2O. Peptide YB for the heterotrimer was prepared with 500 μL of 2.6 mM solution and 50 μL of D_2O. Around 0.5 mg of trimethylsilylpropionic acid, purchased from Sigma, was used as a reference. Tyrosine and tryptophan containing peptides' concentrations were confirmed with a Thermo NanoDrop.

Two-dimensional (2D) $^1H-^{15}N$ heteronuclear single quantum coherence (HSQC) and three-dimensional (3D) $^1H-^1H-^{15}N$ nuclear Overhauser effect spectroscopy (NOESY) HSQC experiments were performed on an 600 MHz Bruker spectrometer containing a cryogenic probe. The nitrogen carrier frequency was set at 108 ppm for the 3D experiment and at 112 ppm for the 2D experiments. The proton carrier frequency was set to match that of the water signal. All experiments were performed at 25 °C. The $^1H-^{15}N$ HSQC experiments were collected using 1922 × 256 complex points for each of 8 scans with 32 preceding dummy scans. A sweep width of 12 ppm was used in the ^1H dimension and 25 ppm in the ^{15}N dimension. The 3D $^1H-^1H-^{15}N$ NOESY HSQC experiment was collected with a 90 ms mixing time, 4 scans, and 2048 × 32 × 256 complex points with 64 dummy scans. The sweep width was 12 ppm in the direct dimension, 10 ppm in the ^{15}N dimension, and 8 ppm in the indirect dimension. Raw NMR data were processed using TopSpin 4.0 software. A baseline correction (qpol for 2D experiments and sfil for the 3D experiment) and a window multiplication function were applied (sine squared function for the direct dimension of the 2D experiments and all three dimensions of the 3D experiment and sine function for the indirect dimension of the 2D experiments). Data were zero-filled to process 2048 × 1024 complex points in the HSQC experiments and 2018 × 128 × 1024 complex points in the NOESY HSQC experiment. Each dimension for each experiment was Fourier-transformed and phase-corrected manually using the TopSpin 4.0 software. NOE correlation was rendered using DLPacker and then visualized in PyMol.[41,42]

MD Simulations. All-atom MD simulations were utilized to understand side-chain interaction effects on the stability of the OGXY′ and YGXO peptides, where X was substituted with Phe (F) and Trp (W), and Y was substituted with either Lys (K) or Arg (R). The initial structure of the homotrimer collagen was obtained from the crystal structure of $(GPO)_9$ with protein data bank (PDB) id: 3B0S.[43] We then focused our analysis around the central portion of the triple helix and applied the mutations of interest to each structure. To build YGXO, peptides, we replaced the 12th and 14th residues of the $(GPO)_9$ structure with cations and π-contributors, respectively. OGXY′ homotrimer structures were constructed by changing the 14th and 15th residues of $(GPO)_9$ collagen to π contributors and cation amino acids, respectively. We placed each homotrimer in the center of a cubic box and solvated with TIP3P water molecules in Gromacs. Resulting systems were comprised roughly 167,000 atoms in a cubic box with sides of 12 nm. GROMACS 5.1 was used to carry out MD simulations using the AMBER14-ffSB force field.[44,45] According to the experimental conditions, pressure and temperature were kept at 1 bar and 300 K using Langevin dynamics in the *NPT* ensemble. Periodic boundary conditions were applied in all directions, and the particle mesh Ewald method was utilized to evaluate the electrostatic potentials.[44] First, energy minimization was run for 5000 steps. After that, 10 ns runs using first *NVT* and then *NPT* ensemble were performed to equilibrate the system. Lastly, 100 ns production runs were performed. A time step of 2 fs was chosen in all simulations and snapshots of the systems were saved every 20 ps. Inter-residue distance was measured with the center of mass of each aromatic ring to the cation. For the angle (ω) calculations, this inter-residue

https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654

distance was then correlated to the normal vector from the center of the aromatic ring as previously used by Steiner.[46]

RESULTS AND DISCUSSION

Determining Axial and Lateral Interaction Strength. We initially sought to understand the relative stabilities of cation−π interactions between the cationic amino acids arginine, lysine, and histidine, and the π-systems of phenylalanine, tyrosine, and tryptophan. These pairwise interactions can in principle be presented in one of two sequential relationships: axial and lateral. This leads to a total of nine possible pairwise interactions presented in one of two different geometries. To examine all 18 of these relationships, a series of single- and double-substitution peptides were prepared and their thermal unfolding temperature (T_m) was determined by CD polarimetry. Each peptide was generated from a general substitution scheme of a peptide Ac-$(POG)_3$PYGXY′G-$(POG)_3$-Am, where the OGXY′ substitution results in two lateral interactions, and the YGXO substitution has one axial and one lateral interactions. Table 1 shows the complete set of

Table 1. Substitutions to a Ac-$(POG)_3$PYGXY′G$(POG)_3$-Am (YGXY′) Collagen Mimetic Peptide Allows for Relative Destabilization of Single Amino Acids[a]

	T_m (°C)		T_m (°C)		T_m (°C)
		Parent Ac-$(POG)_8$-Am			
OGPO	50.0[47]				
		Xaa Substitutions			
OGFO	37.0[47]	OGYO	34.0[47]	OGWO	33.5[47]
		Yaa Substitutions			
HGPO	38.0[47]	KGPO	40.5[47]	RGPO	48.5[47]
		YGXO Substitutions			
HGFO	26.5	HGYO	24.0	HGWO	22.5[47]
KGFO	29.5[47]	KGYO	28.5	KGWO	27.0
RGFO	41.5[47]	RGYO	41.5	RGWO	40.0
		OGXY′ Substitutions			
OGFH	20.0	OGYH	17.5	OGWH	14.5[47]
OGFK	22.0[47]	OGYK	17.5	OGWK	13.5
OGFR	29.5[47]	OGYR	26.0	OGWR	21.0

[a]The OGXY′ and YGXO substitutions help in determining the axial and lateral interaction's stabilization to the triple helix. Peptides previously determined are referenced.[47]

melting temperatures and Figure 1 shows the first derivative of the CD melting data. This data can then be used to deconvolve the 18 different types of pairwise interactions being investigated.

As an example of how this deconvolution works, consider the axial interaction between arginine and tryptophan. We can see that the parent peptide $(POG)_8$ has a melting temperature of 50 °C, while the single substitution of arginine in RGPO has a melting temperature of 48.5 °C and the single substitution of tryptophan in OGWO of 33.5 °C. This suggests that a single arginine in the Yaa position destabilized a triple helix by 0.5 °C [(50.0 − 48.5)/3] and a single tryptophan in the Xaa position destabilizes a triple helix by 5.5 °C [(50 − 33.5)/3]. Based on this approach, if there are no pairwise interactions, the double substitution peptides RGWO and OGWR should have an identical melting temperature of 32 °C for their homotrimeric helices. Instead, we see that they are quite different with RGWO being 40.0 °C and OGWR being 21.0 °C. The main difference between these two peptides is that RGWO can form two axial interactions and one lateral interaction between arginine and tryptophan while OGWR can only form two lateral interactions and no axial interactions. The strength of the lateral interactions can be estimated from the stability of OGWR with a lateral interaction = (21 − 32)/2 = −5.5 °C. We can then solve for the strength of an axial interaction as follows = (40 − 32 − 5.5)/2 = 6.75 °C. Using this approach, the results of the deconvolution of all interactions are presented in Table 2. The most obvious conclusion from these data is that cation−π interactions, when presented in an axial sequential geometry, are uniformly stabilizing while cation−π interactions presented in a lateral sequential geometry are uniformly destabilizing. Of these interactions, the most stabilizing are those formed between arginine−tryptophan (6.8 °C), arginine−tyrosine (6.2 °C), arginine−phenylalanine (4.5 °C), and lysine−tryptophan (4.2 °C). The arginine−tryptophan and arginine−tyrosine interactions are approximately equal to the best-known pairwise interaction within the collagen triple helix, the axial interaction of lysine−aspartate. Combined with the fact that each of these stabilizing axial cation−π interactions are paired with a significantly destabilizing lateral interaction, the use of cation−π interactions for both positive and negative designs of collagen triple helices is strongly supported. In contrast, interactions between histidine and π-systems were found to be only just slightly above our estimated experimental error and are probably only useful as secondary interactions in the de novo design. To provide additional proof that the stabilizing effects we observe are in fact from cation−π interactions, we performed two additional experiments: the use of physiological salt concentrations and the substitution of 5F-tryptophan previously demonstrated to reduce tryptophan's ability to form cation−π interactions.[48] Increased salt concentration mildly destabilized the triple helix while the electronegative fluorine derivative dramatically destabilized the triple helix (see Supporting Information, Figure S17). Both results support the stabilizing forces in these systems as being canonical cation−π interactions.

MD Illustrate Axial Versus Lateral Differences. In order to gain a better understanding of the cation−π interactions in the context of a triple helix and especially to help understand the stark difference between the axial and lateral sequential geometry, we performed MD simulations.

We started with the published crystal structure of $(GPO)_9$ (PDB: 3B0S) and prepared double substitution mutants with arginine, lysine, histidine, phenylalanine, and tryptophan. We define an interaction of two residues' side chains as anything within 6 Å.[6] As was discussed above with the algebraic deconvolution, in an YGXO substitution scheme, there are two axial interactions between the leading to middle (L to M) and middle to trailing (M to T) peptides, respectively. Additionally, the trailing to leading (T to L) peptides can form a lateral interaction. Detailed methods for MDs are presented in the Experimental Methods, but we monitored the distance between cationic and π-containing amino acids over time to generate a population map versus distance plot, as presented in Figure 2.

Amino acids in an axial sequential geometry in the RGFO homotrimer revealed two major populations of distances (labeled 1 and 2), the first of which is approximately 4 Å and allows a near-ideal cation−π interaction between the two amino acids. The second population is over 6 Å and is not likely to contribute to the stabilization of the helix. These

https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654

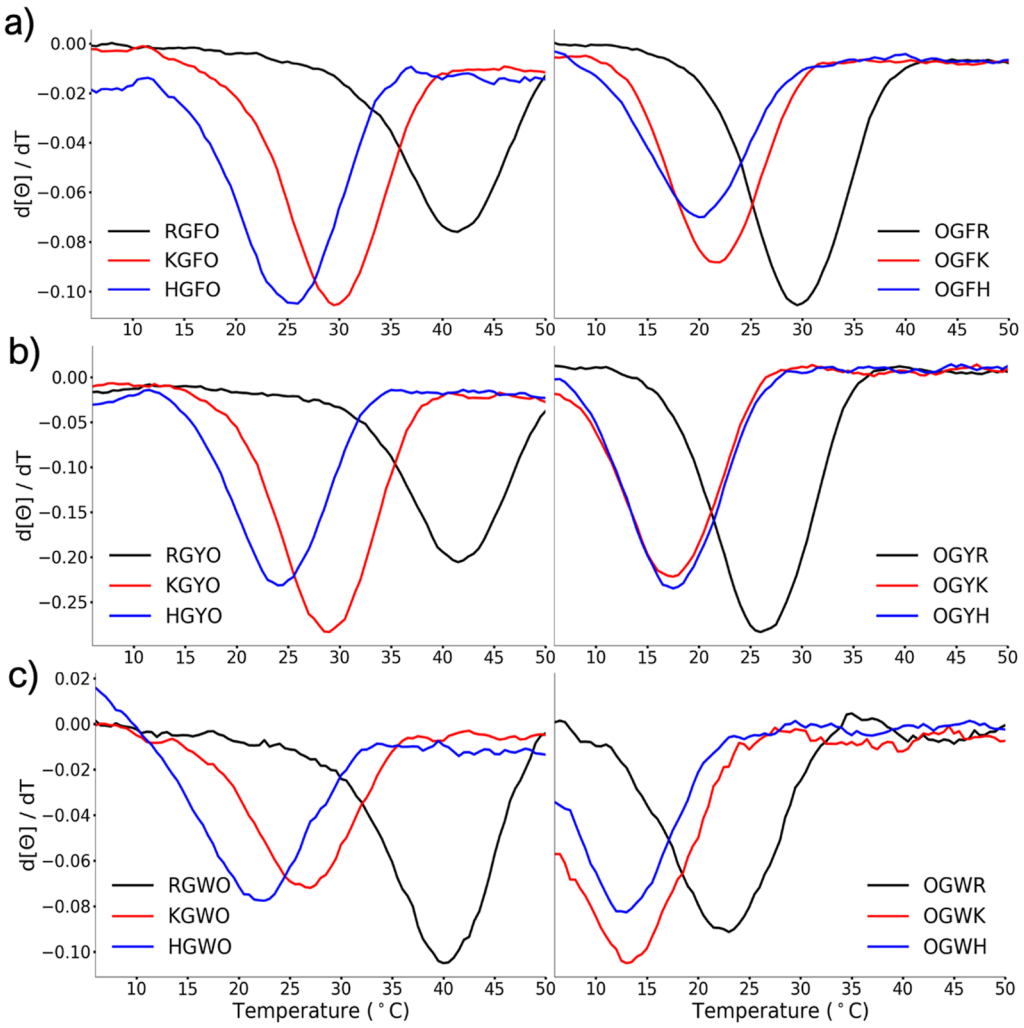


Figure 1. CD melt derivatives used to deconvolute homotrimers for both axial and lateral interactions. The YGXO substitutions are aligned on the left and the OGXY′ substitutions are on the right. (a) Phenylalanine (F) containing triple helices, (b) tyrosine (Y) containing triple helices, and (c) tryptophan (W) containing triple helices.

Table 2. Relative Interaction Strengths for All Deconvoluted Cation−π Interactions

	axial (°C)	lateral (°C)
RW	6.8	−5.5
RY	6.2	−3.3
RF	4.5	−3.0
KW	4.2	−5.3
KY	3.8	−3.5
KF	2.4	−2.3
HW	2.3	−3.5
HY	2.2	−2.3
HF	2.0	−2.5

interactions can be observed in both the L to M and M to T peptides as expected. The T to L peptide interaction, which has the lateral geometry, shows a very different arrangement with the closest contact being in excess of 5 Å and a longer contact near 7 Å. The MD simulations then are clearly showing the differences between axial and lateral sequential geometries, in which only the axial relationship allows for amino acids to come into close contact. The OGFR peptide, in contrast to RGFO, is never able to achieve close contacts between arginine and phenylalanine and show distance population similar to the lateral interactions observed in RGFO. Similar results were found for lysine−phenylalanine, arginine−phenylalanine, histidine−phenylalanine, arginine−tryptophan, and histidine−tryptophan (see the Supporting Information). The relative orientation of the cation to the aromatic ring also plays a role in its ability to stabilize the helix. By measuring the angle between the normal vector of the aromatic ring and the inter-residue distance further support for axial being favored over lateral and arginine over lysine was apparent. From Figure S22C, it is observed that the RW axial interaction is constrained near an ideal interaction angle of 45°. While the RW lateral interaction is disfavored in almost all cases. In Figure S22B, it is observed that the KW axial interaction has a closer inter-residue distance than the KW lateral but does not have as concentrated of a population density for ω as arginine containing peptides.[46,49] These MD simulations reinforce the conclusions drawn from CD thermal melts: axial interactions, not lateral, promote triple helix stability, and can play a useful role in the design of heterotrimers.

We also probed the strain imposed on the backbone surrounding a cation−tryptophan containing peptide. Tryptophan was of interest because it had the largest loss of stability upon substitution, and it has the most sterically demanding R-

https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654

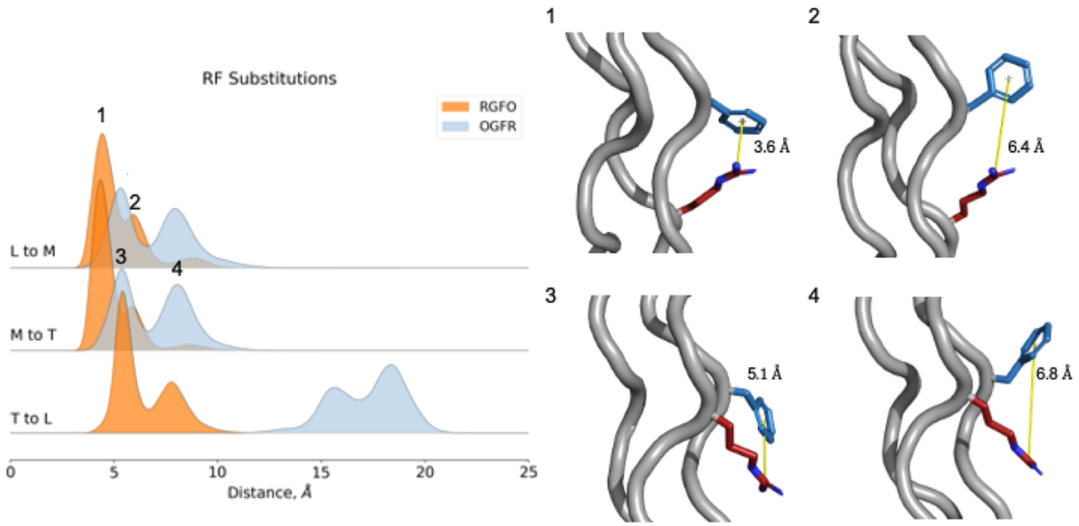


Figure 2. MD simulations for homotrimers containing arginine and phenylalanine. Left: Ridge-plots overlaid for RGFO (orange) and OGFR (blue). Right: Snapshots from the MD simulations shows a representation of the most commonly observed orientations. (1) From RGFO, the closest proximity and largest population for the axial orientation. (2) Less-ideal population from an axial interaction. (3) Closest interaction for a lateral OGFR interaction. (4) Non-ideal lateral orientation results in no favorable cation−π interaction.

group. Ramachandran plots were able to suggest that lateral interactions did induce a greater strain on the triple helix as ϕ/ψ angles show a greater deviation from the ideal PPII coordinates (Figure S23) when compared with axial interactions. In contrast, when the tryptophan was substituted for phenylalanine, the ϕ/ψ angles showed little deviation (Figure S24). This demonstrates the importance of steric considerations in de novo design.

Cation−π Interactions in ABC-Type Heterotrimers. Of the natural collagens, many exist as heterotrimers, where not all strands are equivalent. When there are three different strands forming a ternary solution of monomers A, B, and C, 27 different compositions and registers can arise. This folding problem has been tackled with both chemical intuition and computational models.[30,31,47,50,51] In particular, a previously published ABC-type heterotrimer utilized lysine−aspartate axial interactions from the leading to middle strands, lysine−glutamate axial interactions from the middle to trailing, and axial arginine−phenylalanine, cation−π interactions from the trailing to leading strands. Each of these peptides is listed in Table 3.

Using this successful design as a template, we examined different A peptides that contained as their π component either phenylalanine (A_F), tyrosine (A_Y), or tryptophan (A_W) while C peptides were prepared with arginine (C_R). As a control, an additional C peptide was prepared in which the cationic amino acids were replaced with alanine (C_A). The B peptide was unchanged in all experiments. This allowed us to study how cation−π interactions impact stability in a much more complicated and chemically diverse heterotrimeric systems (Figure 3).

Table 3. Sequences of Peptides Used to Fold an ABC-Type Heterotrimer[a]

monomer	sequence
A_F[47]	PKGPKGFOGPOGFKGFKGPKGPOGFKGPOG
A_Y	PKGPKGYOGPOGYKGYKGPKGPOGYKGPOG
A_W	PKGPKGWOGPOGWKGWKGPKGPOGWKGPOG
B[47]	PKGDOGDKGPOGPOGDKGDOGDKGPKGDOG
C_R[47]	PRGEPGPRGERGPPGPPGERGPPGEPGEPG
C_A[52]	PAGEPGPAGEAGPPGPPGEAGPPGEPGEPG

[a]Peptides previously determined are referenced.[47]

Thermal melts of heterotrimers (A_FBC_R), (A_YBC_R), or (A_WBC_R) showed similar melting temperatures of 39.5, 40.0, and 37.0 °C, respectively. While these differences are modest, the lowest of the three, (A_WBC_R), also has features that suggest its self-assembly is not as clean as that of (A_FBC_R) and (A_YBC_R). There appears to be a significant tail on the low-temperature side of the unfolding curve, which suggests additional species in solution or non-uniform unfolding of the helix. By replacing all of the arginines with alanines in the C-monomer, we were able to observe the importance of the cation−π interaction in the heterotrimer assembly. As indicated in Table 4 and Figure S17, thermal melts of (A_FBC_A), (A_YBC_A), and (A_WBC_A) showed a decrease of thermal stability of 15.5, 19.0, and 17.5°, respectively. The (A_YBC_A) and (A_WBC_A) lose so much stability that their assembly appears to compete with other heterotrimer species, as a result forming different compositions and registers.

NMR was used to gain a molecular-level understanding of these assemblies. First, while the composition of each triple helix can be determined from CD alone, the registration between peptides requires NOESY and HSQC NMR. The determination of the registration of triple helices has been accomplished in the past and we followed a similar approach here.[29] Strategically placed ^{15}N-labeled glycines (at position 21) were introduced during synthesis. First, the HSQCs of all monomers, binary solutions, and ternary mixtures were determined. This allows clear assignments of possible assemblies (Figure 4). The A_YBC_R data in Figure 4a,b show clearly that only one new triple-helical signal is observed. This is in good agreement with A_FBC_R, previously published, which also showed only one triple-helical signal. In contrast to this, but within agreement with CD melting data, A_WBC_R shows some additional peaks, which suggest an additional register may be forming. It is also noteworthy that HSQC spectra of (A_FBC_A), (A_YBC_A), and (A_WBC_A) mixtures resulted in no

https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654

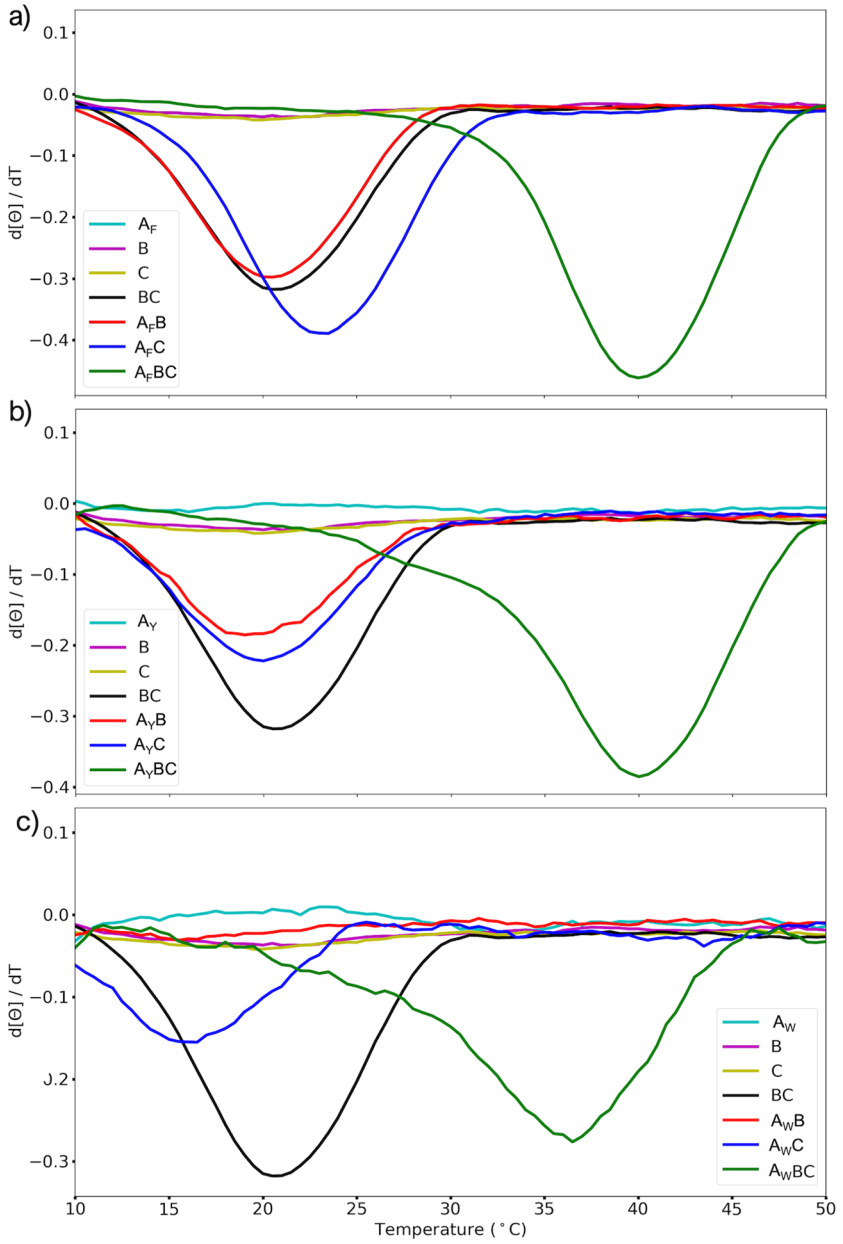


Figure 3. CD melt derivatives of each ABC-type heterotrimer. (a) Phenylalanine containing triple helices, (b) tyrosine containing triple helices, and (c) tryptophan containing triple helices.

Table 4. T_m Data for Different Mixtures of Cation−π Containing ABC-Type Heterotrimers[a]

mixture	T_m (°C)	mixture	T_m (°C)	mixture	T_m (°C)
A_FBC_R[47]	39.5	A_YBC_R	40.0	A_WBC_R	37.0
A_FB[47]	21.0	A_YB	20.0	A_WB	DNF
A_FC_R[47]	23.5	A_YC_R	19.5	A_WC_R	15.5
A_FBC_A[47]	24.0	A_YBC_A	21.0	A_WBC_A	19.5

[a]Peptides previously determined are referenced.[47]

heterotrimer peaks, confirming that cation−π plays an integral role in the assembly process.

3D NOESY-HSQC studies, as depicted in Figure 5a, allow for the deconvolution of interstrand interactions and determination of the register of the triple helix, from its 27 possibilities. By replicating the 3D NOESY-HSQC experiment for both A_YBC_R and A_WBC_R samples, we can prove a specific register of the leading, middle, and trailing strands. Three slices from the 3D spectra were found to correlate the strands of A_YBC_R at 105.19, 109.56, and 110.50 ppm, and the results are presented in Figure 5b. To correlate which monomer is the leading, middle, and trailing strands, the amide region shows a cross-peak from the leading strand A_Y (109.56 ppm) and the middle strand (105.19 ppm). These amide NOE peaks are then shown to shift an upstream and then correlate to the trailing strand in the final slice of Figure 5b. While these do confirm that the heterotrimer is comprised of three different peptides, the Cα region is necessary to confirm the registration of each strand. The NOE environments for each monomer, as illustrated in Figure 5a, are distinct.

The hydroxyproline (O) of the middle strand yields a larger downfield shift when compared to the lysine (K) and arginine (R) protons. To confirm that the register is indeed A_YBC_R, we cross-correlated all of the interactions that can take place in

https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654

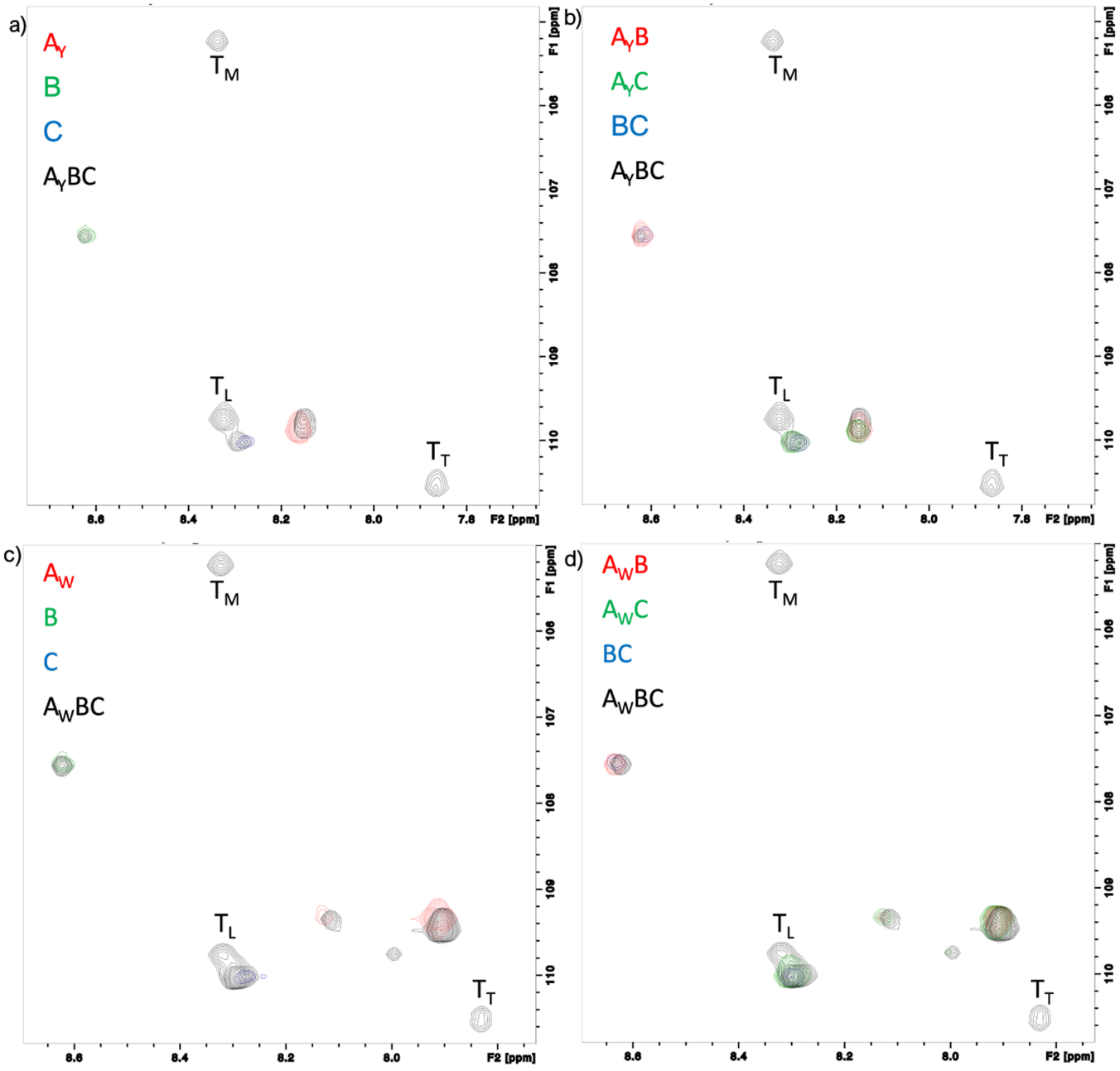


Figure 4. NMR ^{1}H, ^{15}N HSQC for both A_YBC_R and A_WBC_R mixtures. (a) Unary samples of A_YBC, (b) binary samples of A_YBC_R, (c) unary samples of A_WBC_R, and (d) binary samples of A_WBC_R. Overlapping spectra points are indicative of a monomer species in solution. All samples labeled with a "T" indicate a heterotrimer folding with a distinct composition. Each subscript indicates the leading (L), middle (M), and trailing (T) strands. All experiments were conducted at 25 °C.

each register within 5 Å. The B-monomer had the O because of its distinct shift, as seen in Figure 5b's slice two. The O−Cα is then able to have a cross-peak with the monomer in slice one of Figure 5b. To determine that slice one is A_Y, the total registers were compared against each other to find that the O−Cα cross-peak can only occur in the A_YBC_R or C_RBA_Y registers. The NOE of the O−Cα in Figure 5c shows the most likely overlap. However, if the C_RBA_Y was the observed register then we would expect to see a cross-peak of the R−Cα in the spectrum. In addition, the lack of glutamine (E) Cα cross-peak further suggests that the leading strand is A_Y, confirming that there is an A_YBC_R register. The registration of A_WBC_R was confirmed similarly (see the Supporting Information). With these data in hand, we have demonstrated that A_FBC_R, A_YBC_R, and A_WBC_R all form controlled composition triple helices and that A_FBC_R and A_YBC_R have controlled registrations while A_WBC_R has primarily one registration, some signs of an alternate registration are suggested both by CD and NMR. This degradation of specificity in the case of tryptophan, despite the very high stabilization observed in simple cation−π systems, may be due to the general difficulty of the large tryptophan indole ring to adopt a triple-helical conformation, which may lead to alternate conformation, especially when used too many times in a single monomer.

By confirming the assembled ABC_R-type heterotrimer series, and also the lack of assembly of the $A_\pi BC_A$, we have demonstrated in the importance of cation−π in CMP design.

https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654

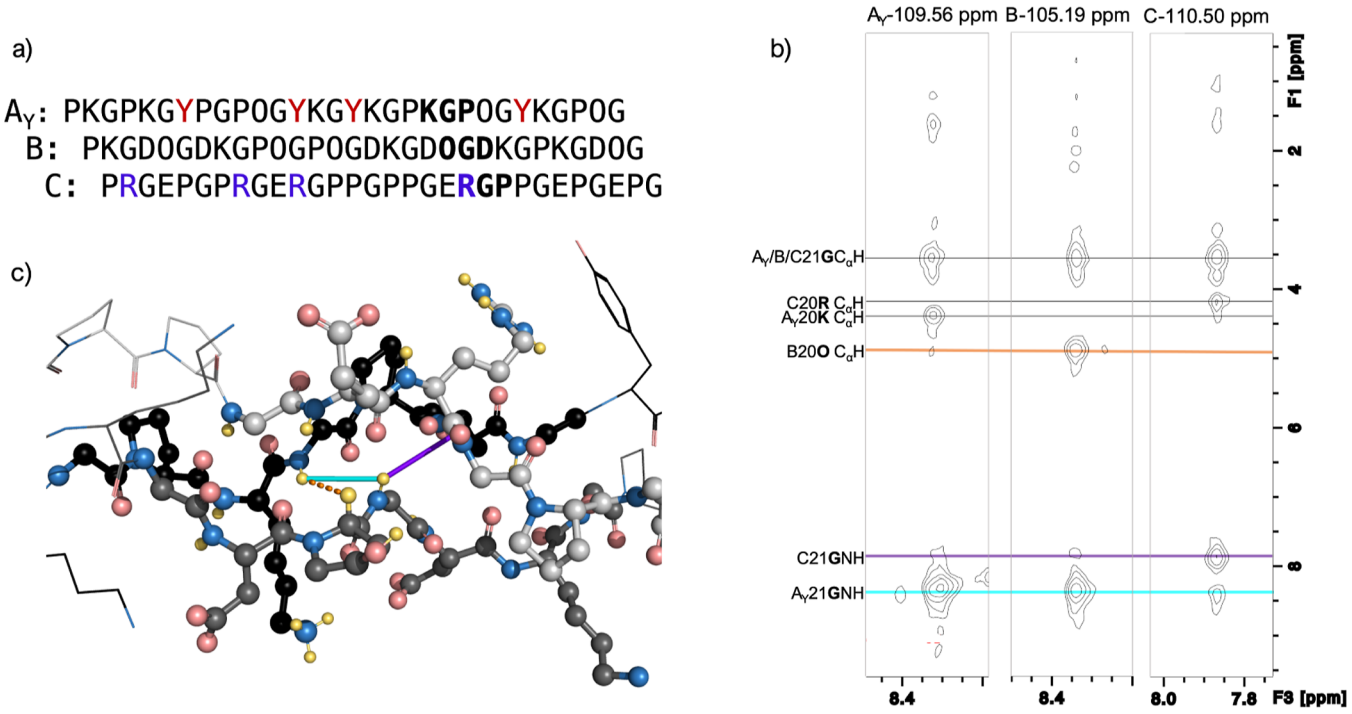


Figure 5. 3D NOESY-HSQC ^{1}H, ^{15}N, and ^{1}H NMR was used to determine the register for A_YBC, all cation−π interactions are highlighted. (a) Sequence of each leading, middle, and trailing strand. The bolded parts of each strand indicate the NOE region that is of interest. (b) Slices of each 3D NOESY-HSQC ^{1}H, ^{15}N, and ^{1}H NMR NOE's, which indicate the A_Y, B, and C strands. The downfield portion of the spectrum in the F1 dimension has the highlighted amide proton (NH) shifts. Upfield, the α-carbons and various side group protons are also labeled. (c) 3D rendering of the NOE region shows the B-monomer's NH-proton interacting with both the leading (blue) and trailing (purple) strands. Also the NOE for the B monomer's hydroxyproline (O) to the leading strand's NOE (orange).

Both the thermal melting temperature deconvolutions and MD support that only axial interactions can be favored in assembling triple helices. We systematically quantified that lateral interactions are not favorable for triple helix formation in contrast to the good stabilization realized in axial interactions, an issue which previously remained conflicted in collagen literature studies. Furthermore, by doing a complete deconvolution in a homotrimer system, we were able to quantitatively demonstrate trends in cations (R > K > H) and π-systems (W > Y > F). These trends do align with relative interaction action strengths as observed in globular proteins.[7] NMR studies of heterotrimers confirmed the expected registration of peptides allowing us to clearly define the specific pairwise interactions in each triple helical system and their contributions to stability.

CONCLUSIONS

From thermal unfolding experiments, we have demonstrated that cation−π interactions with arginine, lysine, or histidine, making the cation, to phenylalanine, tyrosine, or tryptophan, making the π-system, are uniformly stabilizing when presented in an axial sequential geometry while uniformly destabilizing when presented in a lateral sequential geometry. MD indicate that the lateral sequential arrangement does not lead to suitable close contacts while the axial sequential arrangement provides near-ideal separation between cation and the partner π-system. Among the nine possible axial pairwise relationships, arginine−tryptophan provides the highest degree of stabilization followed by arginine−tyrosine, arginine−phenylalanine, and lysine tryptophan also having substantial stabilizing interactions. In contrast, cation−π interactions between histidine and any other aromatic amino acid have a rather weak stabilizing effect, only just above experimental error. Heterotrimeric helices were prepared to compare the utility of arginine−phenylalanine, arginine−tyrosine, and arginine−tryptophan in systems with significantly greater chemical diversity. Each of these systems had four occurrences of a cation−π interaction. Of these, the arginine−tyrosine containing helix (A_YBC) and arginine−phenylalanine containing helix (A_FBC) had the highest melting temperatures and cleanest assembly. The arginine−tryptophan containing helix (A_WBC) was slightly less stable and was slightly less clean (competing triple helices were observed). This slightly poorer result despite having the best pairwise interaction may be due to the inherent destabilizing effect of tryptophan on collagen helices independent of interactions with other amino acids that become more problematic with larger numbers of tryptophan. This study helps to clarify the role of cation−π interactions in the context of collagen triple helices and illustrates methods in which they can be incorporated into designed heterotrimeric systems.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biomac.2c00856.

Full peptide characterization including UPLC, mass spectrometer data, CD melt curves; MD simulations; and further NMR characterization for other cation−π pairs (PDF)

https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654

AUTHOR INFORMATION

Corresponding Author

Jeffrey D. Hartgerink − *Department of Chemistry and Department of Bioengineering, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-3186-5395; Phone: +1-713-348-4142; Email: jdh@rice.edu

Authors

Carson C. Cole − *Department of Chemistry, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-7573-1882

Mikita Misiura − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Sarah A. H. Hulgan − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Caroline M. Peterson − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Joseph W. Williams, III − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Anatoly B. Kolomeisky − *Department of Chemistry, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biomac.2c00856

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded in part by the National Science Foundation (CHE 2203937) and the Robert A. Welch Foundation (C1559). This work was also supported in part by the Big-Data Private-Cloud Research Cyberinfrastructure MRI-award funded by the NSF under grant CNS-1338099 and by Rice University's Center for Research Computing (CRC). The authors would like to thank KMW and BP for their assistance in reviewing the manuscript.

REFERENCES

(1) Infield, D. T.; Rasouli, A.; Galles, G. D.; Chipot, C.; Tajkhorshid, E.; Ahern, C. A. Cation-π interactions and their functional roles in membrane proteins. *J. Mol. Biol.* **2021**, *433*, 167035.

(2) Dougherty, D. A. The cation-π interaction. *Acc. Chem. Res.* **2013**, *46*, 885−893.

(3) Zhang, H.; Li, C.; Yang, F.; Su, J.; Tan, J.; Zhang, X.; Wang, C. Cation-pi interactions at non-redundant protein-RNA interfaces. *Biochemistry* **2014**, *79*, 643−652.

(4) Crowley, P. B.; Golovin, A. Cation−π interactions in protein−protein interfaces. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 231−239.

(5) Mecozzi, S.; West, A. P.; Dougherty, D. A. Cation-pi interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10566−10571.

(6) Dougherty, D. A. Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp. *Science* **1996**, *271*, 163−168.

(7) Gallivan, J. P.; Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459−9464.

(8) Minoux, H.; Chipot, C. Cation−π Interactions in Proteins: Can Simple Models Provide an Accurate Description? *J. Am. Chem. Soc.* **1999**, *121*, 10366−10372.

(9) Chen, C.-C.; Hsu, W.; Hwang, K.-C.; Hwu, J. R.; Lin, C.-C.; Horng, J.-C. Contributions of cation−π interactions to the collagen triple helix stability. *Arch. Biochem. Biophys.* **2011**, *508*, 46−53.

(10) Zheng, H.; Liu, H.; Hu, J.; Xu, F. Using a collagen heterotrimer to screen for cation- interactions to stabilize triple helices. *Chem. Phys. Lett.* **2019**, *715*, 77−83.

(11) Chiang, C.-H.; Fu, Y.-H.; Horng, J.-C. Formation of AAB-Type Collagen Heterotrimers from Designed Cationic and Aromatic Collagen-Mimetic Peptides: Evaluation of the C-Terminal Cation Interactions. *Biomacromolecules* **2017**, *18*, 985−993.

(12) Chiang, C.-H.; Horng, J.-C. Cation−π Interaction Induced Folding of AAB-Type Collagen Heterotrimers. *J. Phys. Chem. B* **2016**, *120*, 1205−1211.

(13) Brodsky, B.; Ramshaw, J. A. M. The collagen triple-helix structure. *Matrix Biol.* **1997**, *15*, 545−554.

(14) Fratzl, P.; Weinkamer, R. Nature's hierarchical materials. *Prog. Mater. Sci.* **2007**, *52*, 1263−1334.

(15) Arora, S.; Gordon, J.; Hook, M. Collagen binding proteins of gram-positive pathogens. *Front. Microbiol.* **2021**, *12*, 628798.

(16) Ricard-Blum, S. The Collagen Family. *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, a004978.

(17) Canty, E. G.; Kadler, K. E. Collagen fibril biosynthesis in tendon: a review and recent insights. *Comp. Biochem. Physiol., Part A: Mol. Integr. Physiol.* **2002**, *133*, 979−985.

(18) Theocharis, A. D.; Skandalis, S. S.; Gialeli, C.; Karamanos, N. K. Extracellular matrix structure. *Adv. Drug Delivery Rev.* **2016**, *97*, 4−27.

(19) Birk, D. E.; Trelstad, R. L. Extracellular compartments in tendon morphogenesis: collagen fibril, bundle, and macroaggregate formation. *J. Cell Biol.* **1986**, *103*, 231−240.

(20) Hynes, R. O. The Extracellular Matrix: Not Just Pretty Fibrils. *Science* **2009**, *326*, 1216−1219.

(21) Reid, K. B. Complement component C1q: historical perspective of a functionally versatile, and structurally unusual, serum protein. *Front. Immunol.* **2018**, *9*, 764.

(22) Fouët, G.; Bally, I.; Signor, L.; Häußermann, K.; Thielens, N. M.; Rossi, V.; Gaboriaud, C. Headless C1q: a new molecular tool to decipher its collagen-like functions. *FEBS J.* **2021**, *288*, 2030−2041.

(23) Thielens, N. M.; Tedesco, F.; Bohlson, S. S.; Gaboriaud, C.; Tenner, A. J. C1q: A fresh look upon an old molecule. *Mol. Immunol.* **2017**, *89*, 73−83.

(24) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78*, 929−958.

(25) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. Code for collagen's stability deciphered. *Nature* **1998**, *392*, 666−667.

(26) Egli, J.; Schnitzer, T.; Dietschreit, J. C. B.; Ochsenfeld, C.; Wennemers, H. Why Proline? Influence of Ring-Size on the Collagen Triple Helix. *Org. Lett.* **2020**, *22*, 348−351.

(27) Improta, R.; Berisio, R.; Vitagliano, L. Contribution of dipole−dipole interactions to the stability of the collagen triple helix. *Protein Sci.* **2008**, *17*, 955−961.

(28) Gordon, M. K.; Hahn, R. A. Collagens. *Cell Tissue Res.* **2009**, *339*, 247.

(29) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. Solution Structure of an ABC Collagen Heterotrimer Reveals a Single-register Helix Stabilized by Electrostatic Interactions. *J. Biol. Chem.* **2009**, *284*, 26851−26859.

(30) Fallas, J. A.; Lee, M. A.; Jalan, A. A.; Hartgerink, J. D. Rational Design of Single-Composition ABC Collagen Heterotrimers. *J. Am. Chem. Soc.* **2012**, *134*, 1430−1433.

(31) Zheng, H.; Lu, C.; Lan, J.; Fan, S.; Nanda, V.; Xu, F. How electrostatic networks modulate specificity and stability of collagen. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 6207−6212.

(32) Jalan, A. A.; Hartgerink, J. D. Pairwise interactions in collagen and the design of heterotrimeric helices. *Curr. Opin. Chem. Biol.* **2013**, *17*, 960−967.

(33) Hentzen, N. B.; Islami, V.; Köhler, M.; Zenobi, R.; Wennemers, H. A Lateral Salt Bridge for the Specific Assembly of an ABC-Type Collagen Heterotrimer. *J. Am. Chem. Soc.* **2020**, *142*, 2208−2212.

(34) Xu, F.; Zahid, S.; Silva, T.; Nanda, V. Computational design of a collagen A: B: C-type heterotrimer. *J. Am. Chem. Soc.* **2011**, *133*, 15260−15263.

(35) Sun, T.; Qiang, S.; Lu, C.; Xu, F. Composition-dependent energetic contribution of complex salt bridges to collagen stability. *Biophys. J.* **2021**, *120*, 3429−3436.
(36) Tanrikulu, I. C.; Forticaux, A.; Jin, S.; Raines, R. T. Peptide tessellation yields micrometre-scale collagen triple helices. *Nat. Chem.* **2016**, *8*, 1008−1014.
(37) O'Leary, L. E. R.; Fallas, J. A.; Bakota, E. L.; Kang, M. K.; Hartgerink, J. D. Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nat. Chem.* **2011**, *3*, 821−828.
(38) Persikov, A. V.; Ramshaw, J. A. M.; Brodsky, B. Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **2005**, *280*, 19343−19349.
(39) Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. Amino Acid Propensities for the Collagen Triple-Helix. *Biochemistry* **2000**, *39*, 14960−14967.
(40) Gauba, V.; Hartgerink, J. D. Self-Assembled Heterotrimeric Collagen Triple Helices Directed through Electrostatic Interactions. *J. Am. Chem. Soc.* **2007**, *129*, 2683−2690.
(41) Schrödinger, L.; DeLano, W. PyMOL. http://www.pymol.org/pymol (accessed 2022).
(42) Misiura, M.; Shroff, R.; Thyer, R.; Kolomeisky, A. B. DLPacker: deep learning for prediction of amino acid side chain conformations in proteins. *Proteins: Struct., Funct., Bioinf.* **2022**, *90*, 1278−1290.
(43) Okuyama, K.; Miyama, K.; Mizuno, K.; Bächinger, H. P. Crystal structure of (Gly-Pro-Hyp)9: Implications for the collagen molecular model. *Biopolymers* **2012**, *97*, 607−616.
(44) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(45) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.
(46) Steiner, T.; Koellner, G. Hydrogen bonds with π-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **2001**, *305*, 535−557.
(47) Walker, D. R.; Hulgan, S. A. H.; Peterson, C. M.; Li, I.-C.; Gonzalez, K. J.; Hartgerink, J. D. Predicting the stability of homotrimeric and heterotrimeric collagen helices. *Nat. Chem.* **2021**, *13*, 260−269.
(48) Shao, J.; Kuiper, B. P.; Thunnissen, A.-M. W.; Cool, R. H.; Zhou, L.; Huang, C.; Dijkstra, B. W.; Broos, J. The Role of Tryptophan in π Interactions in Proteins: An Experimental Approach. *J. Am. Chem. Soc.* **2022**, *144*, 13815−13822.
(49) Makareeva, E.; Sun, G.; Mirigian, L. S.; Mertz, E. L.; Vera, J. C.; Espinoza, N. A.; Yang, K.; Chen, D.; Klein, T. E.; Byers, P. H.; et al. Substitutions for arginine at position 780 in triple helical domain of the α1 (I) chain alter folding of the type I procollagen molecule and cause osteogenesis imperfecta. *PLoS One* **2018**, *13*, No. e0200264.
(50) Hilderbrand, A. M.; Taylor, P. A.; Stanzione, F.; LaRue, M.; Guo, C.; Jayaraman, A.; Kloxin, A. M. Combining simulations and experiments for the molecular engineering of multifunctional collagen mimetic peptide-based materials. *Soft Matter* **2021**, *17*, 1985−1998.
(51) Brodsky, B.; Persikov, A. V. *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers*; Advances in Protein Chemistry; Academic Press, 2005; Vol. *70*, pp 301−339.
(52) Walker, D. R.; Alizadehmojarad, A. A.; Kolomeisky, A. B.; Hartgerink, J. D. Charge-Free, Stabilizing Amide−π Interactions Can Be Used to Control Collagen Triple-Helix Self-Assembly. *Biomacromolecules* **2021**, *22*, 2137−2147.

https://doi.org/10.1021/acs.biomac.2c00856
Biomacromolecules 2022, 23, 4645−4654