

Published in final edited form as:

*J Am Chem Soc.* 2013 April 24; 135(16): 6014–6017. doi:10.1021/ja402187t.

## Hydroxyproline-free Single Composition ABC Collagen Heterotrimer

Abhishek A. Jalan<sup>§</sup>, Borries Demeler<sup>†</sup>, and Jeffrey D. Hartgerink<sup>§,\*</sup>

<sup>§</sup>Department of Chemistry, Rice University, 6100 Main St., Houston, TX 77005

<sup>†</sup>Department of Biochemistry, U. T Health Science Center, San Antonio, TX 78229

### Abstract

Hydroxyproline plays a major role in stabilizing collagenous domains in eukaryotic organisms. Lack of this modification is associated with significant lowering in thermal stability of the collagen triple helix and may also affect fibrillogenesis and folding of the peptide chains. In contrast, even though bacterial collagens lack hydroxyproline, their thermal stability is comparable to fibrillar collagen. This has been attributed to the high frequency of charged amino acids found in bacterial collagen. Here we report a thermally stable hydroxyproline-free ABC heterotrimeric collagen mimetic system composed of decapositive and decanegative peptides and a zwitterionic peptide. None of the peptides contain hydroxyproline and furthermore the zwitterionic peptide does not even contain proline. The heterotrimer is electrostatically stabilized via multiple interpeptide lysine-aspartate and lysine-glutamate salt-bridges and maintains good thermal stability with a melting temperature of 37 °C. The ternary peptide mixture also populates a single composition ABC heterotrimer as confirmed by circular dichroism (CD) and Nuclear Magnetic Resonance (NMR) spectroscopy. This system illustrates the power of axial salt-bridges to direct and stabilize the self-assembly of a triple helix and may be useful in analogous designs in expression systems where the incorporation of hydroxyproline is challenging.

Stability of the collagen triple helix in phylogenetically distant organisms such as mammals and bacteria is determined by a combination of enthalpic and entropic factors that depend on the amino acids occupying the X and Y positions of the canonical Xaa-Yaa-Gly triplet repeat.<sup>1</sup> For example, the cyclic amino acids proline (Pro, P) and hydroxyproline (Hyp, O) found predominantly in the collagenous domains of eukaryotic organisms,<sup>2</sup> restrict the backbone dihedral angles close to the native folded state<sup>3,4</sup> providing entropic stabilization. The stereoelectronic and inductive effect of the electronegative hydroxyl substituent provides additional stabilization by pre-organizing the Hyp ring pucker<sup>5,6</sup> and lowering the barrier to cis-trans isomerization of hydroxyprolyl amide bond.<sup>7,8</sup> Therefore, lowering the Hyp content is often associated with significant loss in thermal stability of the triple helix<sup>9,10</sup> that can also prevent fibrillogenesis in some native collagen types.<sup>11</sup>

Bacteria lack a crucial enzyme, prolyl hydroxylase, required for post-translational hydroxylation of proline, however, the consequent lack of hydroxyproline does not significantly affect the thermal stability of their collagenous domains.<sup>12,13</sup> For example, the

\*Corresponding Author: jdh@rice.edu.

### ASSOCIATED CONTENT

Synthesis and characterization of peptides, preparation of CD and NMR samples, experimental parameters and additional NMR and analytical ultracentrifugation experiments and analysis. This material is available free of charge via the Internet at <http://pubs.acs.org>.

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

melting temperature of the collagenous domains of Scl1 and Scl2 bacterial proteins in *Streptococcus Pyogenes* have been shown to be comparable to that of vertebrate fibrillar collagen.<sup>14</sup> These organisms have evolved alternate mechanisms to compensate for the lack of hydroxyproline.<sup>15</sup> *In silico* genomic analysis of a variety of bacteria and bacteriophages has indicated a high occurrence of threonine in the Y position where its hydroxyl group can form hydrogen bonds to the backbone carbonyl.<sup>16</sup> However, organisms that lack a predominance of threonine as well as proline residues contain an unusually high frequency of the charged amino acids lysine, arginine, glutamate and aspartate. Although, these residues generally destabilize the triple helix,<sup>17</sup> it has been demonstrated that the loss in stability is significantly compensated if the acidic and basic residues form specific charge pairs.<sup>18-20</sup>

The stability of previously reported collagen mimetic systems containing salt-bridge interactions had sizable contribution from proline and hydroxyproline.<sup>21,22</sup> In contrast, collagen constructs that contained no hydroxyproline were stabilized by terminally fused trimerization domains such as T4 fibrin foldon<sup>23</sup> or cysteine knots.<sup>24</sup>

Here we report a hydroxyproline free ABC heterotrimeric collagen triple helix composed of decacharged positive and negative peptides and a zwitterionic peptide that contains neither proline nor hydroxyproline. Multiple lysine-glutamate and lysine-aspartate axial charge pairs electrostatically stabilize the heterotrimer. To our knowledge, this is the first report of a collagen mimetic system that self-assembles into a stable triple helix despite the complete lack of cyclic amino acids in one of the peptide chains.

Using salt-bridges rationally engineered into a 3•(POG)<sub>10</sub> template, we have previously reported AAB and ABC heterotrimeric collagen mimetic systems that show controlled composition and/or register.<sup>25-27</sup> Of particular interest is the system composed of (PKG)<sub>10</sub>, (EOG)<sub>10</sub> and (POG)<sub>10</sub> peptides that self-assemble into an ABC heterotrimer with a melting temperature of 54 °C.<sup>21</sup> The primary sources of stabilization in this heterotrimer are the multiple lysine-glutamate charge pairs and the high content of cyclic amino acids. In order to design a hydroxyproline-free ABC heterotrimer, we made two modifications to this system: first the 10 hydroxyproline residues in (EOG)<sub>10</sub> were substituted with prolines and second, the neutral (POG)<sub>10</sub> chain was substituted with a zwitterionic peptide (DKG)<sub>10</sub>. This eliminates twenty hydroxyproline residues and ten proline residues replacing them with ten lysine, ten glutamate and ten aspartate residues, all of which are known to destabilize a triple helix unless they form axial charge pairs (Figure 1A).<sup>1,21</sup>

All peptides synthesized in this study include an N-terminal tyrosine for spectrophotometric determination of concentration. The Gly17 residue is <sup>15</sup>N-isotopically enriched for NMR spectroscopic study. YG labels are omitted in further discussion of the peptides for brevity. Amino acid residues are abbreviated by their single letter code followed by an XYZ subscript indicating their peptide chains.

The CD spectrograph of a mixture of (PKG)<sub>10</sub>, (EPG)<sub>10</sub> and (DKG)<sub>10</sub> peptides in Figure 1B shows a single cooperative unfolding transition at 37 °C. All monomers and possible binary mixture of the three peptides show a linear decrease in molar residue ellipticity (MRE) indicating the absence of well folded states. The energetic compensation from fewer charge pairs in the competing AAB heterotrimeric and homotrimeric states is not sufficient to offset the loss in stability due to lower cyclic amino acid content. The homologous ABC heterotrimer, containing (EOG)<sub>10</sub> instead of (EPG)<sub>10</sub>, unfolds at a higher melting temperature of 43 °C (Figure S2.C) while the two ABC heterotrimers have nearly identical CD spectra at 5 °C (Figure 2). Interestingly, the mixture of (PKG)<sub>10</sub> and (EOG)<sub>10</sub> unfolds at 39 °C (Figure S2.B) but a mixture of (PKG)<sub>10</sub> and (EPG)<sub>10</sub> does not form a triple helix. The

lack of triple helices formed by binary mixtures of peptides in the (EPG)<sub>10</sub> containing system means that its ABC system has no compositional competitors, unlike the (EOG)<sub>10</sub> containing system, and therefore the (PKG)<sub>10</sub>•(DKG)<sub>10</sub>•(EPG)<sub>10</sub> triple helix self-assembles more cleanly.

Although the 225 nm maximum observed in the wavelength scan of the ternary peptide mixture (Figure 2) is indicative of a polyproline type II (PPII) helix,<sup>28,29</sup> it provides little information on the association state of the heterotrimer. The highly cooperative nature of the thermal unfolding data (Figure 1b) is strongly suggestive of the formation of a triple helical structure. In order to further confirm the formation of a triple helix, we performed analytical ultracentrifugation (AUC) sedimentation velocity experiments over a large range of sample concentrations. These experiments showed that both (PKG)<sub>10</sub>•(DKG)<sub>10</sub>•(EPG)<sub>10</sub> and (PKG)<sub>10</sub>•(DKG)<sub>10</sub>•(EOG)<sub>10</sub> heterotrimers assembled into a major species with a molecular weight consistent only with the trimer. Furthermore the observed frictional ratios were fully consistent with an extended helical structure. Detailed description of the AUC data is presented in the supporting information.

After conclusively determining the oligomerization state to be three, as is expected for a collagen triple helix, establishing the triple helical topology of the assembly is particularly important in this case because of the lack of hydroxyproline and low percentage of proline residues. Further, one chain of this helix, (DKG)<sub>10</sub>, lacks cyclic amino acids all together. These characteristics might be expected to significantly perturb the backbone dihedral angles and therefore the fold of the triple helical assembly. To assess this we use NMR.

In theory, a mixture of three peptides can populate ten unique triple helices of different compositions corresponding to the A<sub>3</sub>, B<sub>3</sub> and C<sub>3</sub> homotrimers and A<sub>2</sub>B, AB<sub>2</sub>, A<sub>2</sub>C, AC<sub>2</sub>, B<sub>2</sub>C, BC<sub>2</sub> and ABC heterotrimers. Detection of these competing states is difficult in CD, especially if they represent minor components. <sup>1</sup>H, <sup>15</sup>N-Heteronuclear Single Quantum Coherence (HSQC) provides higher sensitivity and greater resolution of competing states. The HSQC spectra of a mixture of (PKG)<sub>10</sub>, (EPG)<sub>10</sub> and (DKG)<sub>10</sub> (Figure 3A) obtained at 20 °C shows a monomer cross peak and three additional cross peaks of equal intensity that correspond to the <sup>15</sup>N-isotopically enriched Gly present in each chain of the triple helix. The presence of only three triple helical cross peaks in the HSQC spectra and a single thermal transition observed in CD suggest that, within the limit of detection, the mixture of (PKG)<sub>10</sub>, (EPG)<sub>10</sub> and (DKG)<sub>10</sub> populates a single composition ABC heterotrimer.

The supramolecular topology of the triple helix is characterized by the PPII conformation of the peptide chains and their single residue stagger with respect to each other. The definitive proof for this structure comes from our NMR data which shows that the three chains, (PKG)<sub>10</sub>, (DKG)<sub>10</sub> and (EPG)<sub>10</sub>, form contacts that are within 5 Å of one another (Figure 3b,c and 4). Further, the NOEs observed are those expected for a canonical triple helix. The crystal structure of (PPG)<sub>10</sub><sup>30</sup> shows that nine backbone NH-NH NOE correlations should be observed, of these we can definitively assign eight while the ninth may exist but is obscured by spectral overlap (Figure 4). See supporting information for a detailed discussion of the NMR assignment.

The single residue stagger of peptide chains in the ABC heterotrimer leads to a unique situation where six different triple helices, called registers, are possible for the permutation of A, B and C chains in leading, middle and lagging positions. The number of potential axial charge pairs varies for each register and populating some registers may be more energetically favorable compared to others (Scheme S1). In our case, the six registers can be divided into two subsets of three registers each. Registers within a subset are symmetry related and will generated nearly identical NOE correlation map. Also, as shown in Figure

S4, registers from one subset form nearly twice as many charge pairs as those from the other.  $^1\text{H}$ ,  $^1\text{H}$ -NOESY data shows that registers that contain fewer charge pairs and are therefore energetically unfavorable, are not populated while one or more of the more favorable register are formed (see Supporting Information for the discussion on determination of registers).

Chemically synthesized short peptides are most commonly used to study the structure-function relationship of native collagen. While these systems have been very useful, as we move from refining our knowledge of this relationship to obtaining insights into fibrillogenesis, disease modeling, protein binding and the design of a robust and functional extracellular biomimetic scaffold, the need for longer polypeptides with multiple role-specific sub-domains becomes indispensable. Being effectively inaccessible via solid phase synthesis due to their size, these will need to be obtained by recombinant expression in various hosts. Although eukaryotic expression systems capable of various post-translational modifications have been developed,<sup>31,32</sup> due to the difficulty of isolation and purification, protein translation in bacterial expression systems remains lucrative. In this article, we have successfully demonstrated that hydroxyproline can be substituted with charge pairs with moderate loss in thermal stability of the resulting collagen triple helix. This strategy could be used to design hydroxyproline-free collagen constructs that can be expressed in bacterial hosts such as *E. Coli*. Furthermore, the specificity of salt-bridge interaction can provide a level of control in designing composition and register specific collagen heterotrimers that is inaccessible with hydroxyproline. This is reflected in the single composition ABC heterotrimer populated by the ternary peptide mixture. The designed hydroxyproline free heterotrimer might also be used as an autonomous trimerization domain for designing composition specific heterotrimeric collagenous or non-collagenous fusion proteins such as the coiled-coil collagenous motif found in type IX collagen.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

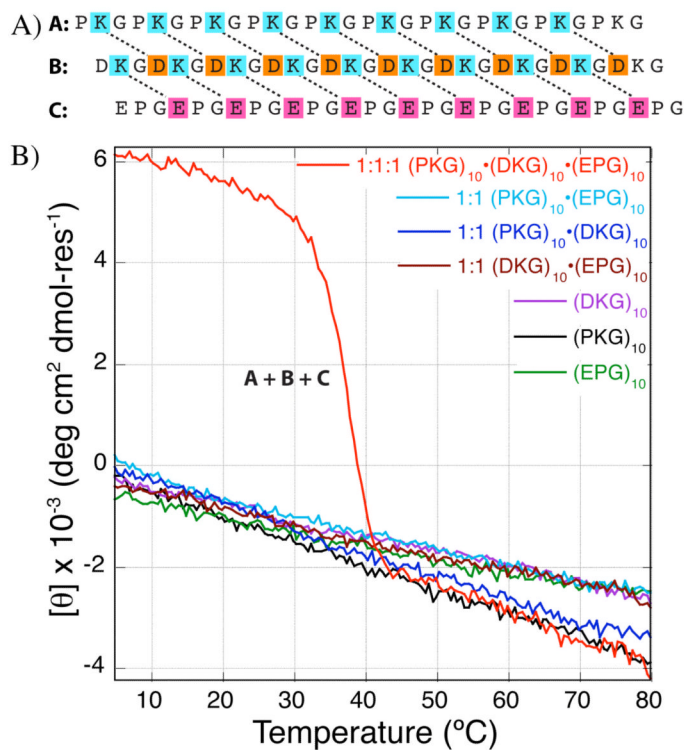
## Acknowledgments

This work was funded in part by National Science Foundation (DMR-0645474 and DMR-1206899) and the Robert A. Welch Foundation (Grant No. C1557). The development of the UltraScan software is supported by the National Institutes of Health through grant RR022200 (to BD). Supercomputer time allocations were provided through National Science Foundation grant TG-MCB070038 (to BD). BD acknowledges the support of the San Antonio Cancer Institute grant P30 CA054174 for the Center for Analytical Ultracentrifugation of Macromolecular Assemblies at the University of Texas Health Science Center at San Antonio.

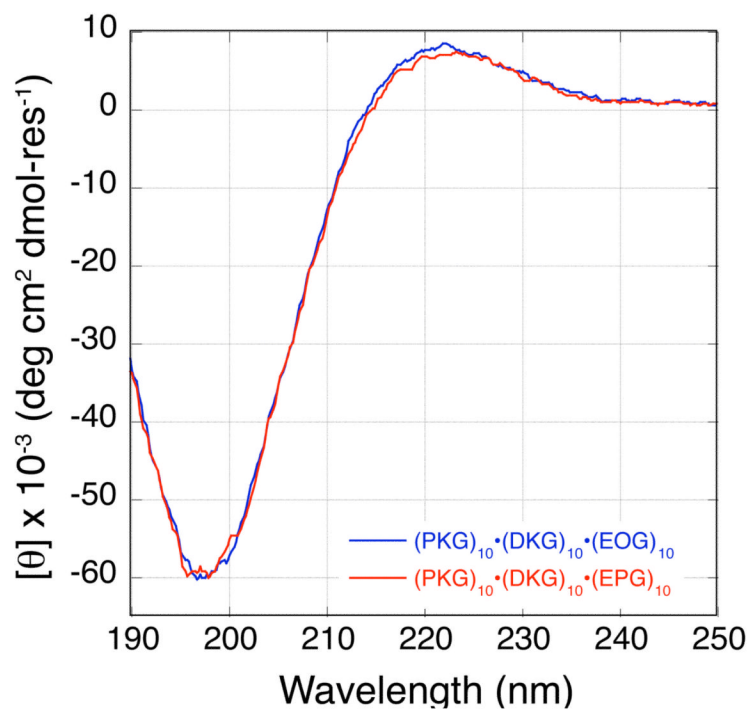
## References

- (1). Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Biochemistry. 2000; 39:14960. [PubMed: 11101312]
- (2). Ramshaw JA, Shah NK, Brodsky B. J. Struct. Biol. 1998; 122:86. [PubMed: 9724608]
- (3). Okuyama K. Conn. Tiss. Res. 2008; 49:299.
- (4). Bella J. J. Struct. Biol. 2010; 170:377. [PubMed: 20149876]
- (5). Shoulders MD, Satyshur KA, Forest KT, Raines RT. Proc. Nat. Acad. Sci. USA. 2010; 107:559. [PubMed: 20080719]
- (6). Holmgren SK, Taylor KM, Bretscher LE, Raines RT. Nature. 1998; 392:666. [PubMed: 9565027]
- (7). Eberhardt ES, Panisik N, Raines PR. J. Am. Chem. Soc. 1996; 118:12261. [PubMed: 21451735]
- (8). Reimer U, Fischer G. Biophys. Chem. 2002; 96:203. [PubMed: 12034441]
- (9). Berg RA, Prockop DJ. Biochem. Biophys. Res. Comm. 1973; 52:115. [PubMed: 4712181]

- (10). Jimenez S, Harsch M, Rosenblo J. *Biochem. Biophys. Res. Comm.* 1973; 52:106. [PubMed: 4576266]
- (11). Perret S, Merle C, Bernocco S, Berland P, Garrone R, Hulmes DJS, Theisen M, Ruggiero F. J. *Biol. Chem.* 2001; 276:43693. [PubMed: 11557756]
- (12). Boydston JA, Chen P, Steichen CT, Turnbough CL. *J. Bacter.* 2005; 187:5310.
- (13). Han RL, Zwiefka A, Caswell CC, Xu Y, Keene DR, Lukomska E, Zhao ZH, Hook M, Lukomski S. *Appl. Microbiol. Biotech.* 2006; 72:109.
- (14). Xu Y, Keene DR, Bujnicki JM, Hook M, Lukomski S. *J. Biol. Chem.* 2002; 277:27312. [PubMed: 11976327]
- (15). Mohs A, Silva T, Yoshida T, Amin R, Lukomski S, Inouye M, Brodsky B. *J. Biol. Chem.* 2007; 282:29757. [PubMed: 17693404]
- (16). Rasmussen M, Jacobsson M, Bjorck L. *J. Biol. Chem.* 2003; 278:32313. [PubMed: 12788919]
- (17). Chan VC, Ramshaw JA, Kirkpatrick A, Beck K, Brodsky B. *J. Biol. Chem.* 1997; 272:31441. [PubMed: 9395477]
- (18). Gurry T, Nerenberg PS, Stultz CM. *Biophys. J.* 2010; 98:2634. [PubMed: 20513408]
- (19). Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B. *J. Mol. Biol.* 2002; 316:385. [PubMed: 11851346]
- (20). Gauba V, Hartgerink JD. *J. Am. Chem. Soc.* 2007; 129:2683. [PubMed: 17295489]
- (21). Gauba V, Hartgerink JD. *J. Am. Chem. Soc.* 2007; 129:15034. [PubMed: 17988128]
- (22). Xu F, Zahid S, Silva T, Nanda V. *J. Am. Chem. Soc.* 2011; 133:15260. [PubMed: 21902217]
- (23). Frank S, Kammerer RA, Mechling D, Schulthess T, Landwehr R, Bann J, Guo Y, Lustig A, Bachinger HP, Engel J. *J. Mol. Biol.* 2001; 308:1081. [PubMed: 11352592]
- (24). Krishna OD, Kiick KL. *Biomacromolecules.* 2009; 10:2626. [PubMed: 19681603]
- (25). O'Leary LE, Fallas JA, Hartgerink JD. *J. Am. Chem. Soc.* 2011; 133:5432. [PubMed: 21428435]
- (26). Fallas JA, Lee MA, Jalan AA, Hartgerink JD. *J. Am. Chem. Soc.* 2012; 134:1430. [PubMed: 22239117]
- (27). Jalan AA, Hartgerink JD. *Biomacromolecules.* 2012; 14:179. [PubMed: 23210738]
- (28). Brown AM, Zondlo NJ. *Biochemistry.* 2012; 51:5041. [PubMed: 22667692]
- (29). Ramani R, Hanski S, Laiho A, Tuma R, Kilpelainen S, Tuomisto F, Ruokolainen J, Ikkala O. *Biomacromolecules.* 2008; 9:1390. [PubMed: 18419152]
- (30). Berisio R, Vitagliano L, Mazzarella L, Zagari A. *Protein Sci.* 2002; 11:262. [PubMed: 11790836]
- (31). Bulleid NJ, John DCA, Kadler KE. *Biochem. Soc. Trans.* 2000; 28:350. [PubMed: 10961917]
- (32). Pinkas DM, Ding S, Raines RT, Barron AE. *ACS Chem. Biol.* 2011; 6:320. [PubMed: 21210682]

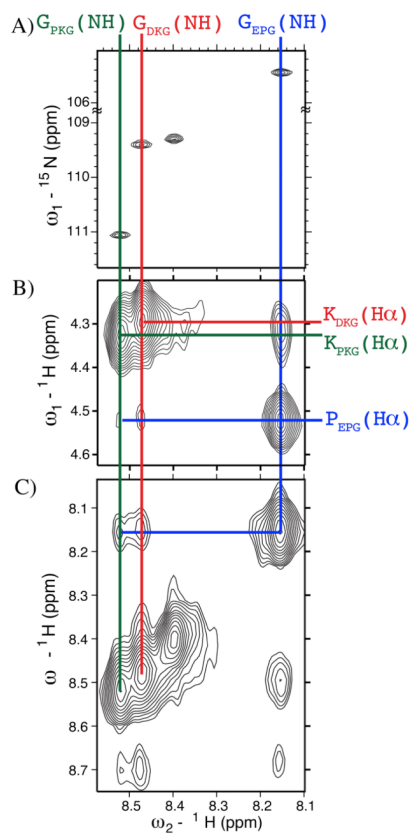
**Figure 1.**

(A) Schematic illustration of hydroxyproline-free ABC heterotrimer composed of (PKG)<sub>10</sub>, (DKG)<sub>10</sub> and (EPG)<sub>10</sub> peptides electrostatically stabilized through multiple lysine-aspartate and lysine-glutamate salt-bridges. All three peptides contain an N-terminal YG label to assist in determination of concentration. (B) Thermal melting profile of the individual, binary and ternary mixtures of (PKG)<sub>10</sub>, (DKG)<sub>10</sub> and (EPG)<sub>10</sub> peptide. The monomers as well as binary mixture were monitored at 225.0 nm and the ternary mixture was monitored at 222.4 nm.



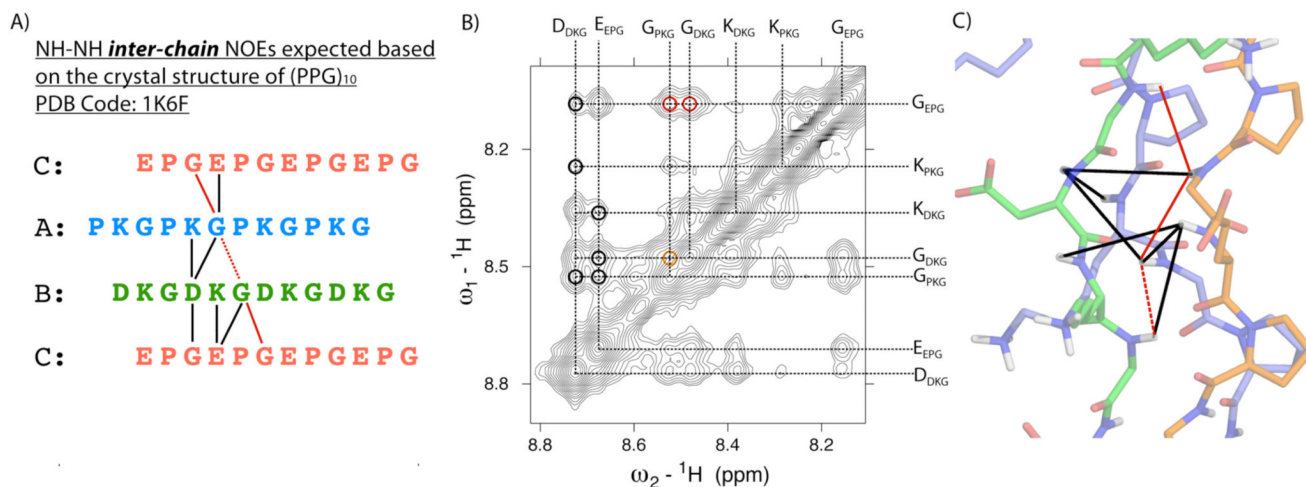
**Figure 2.**  
CD wavelength scan of  $(\text{PKG})_{10} \bullet (\text{DKG})_{10} \bullet (\text{EOG})_{10}$  in blue and  $(\text{PKG})_{10} \bullet (\text{DKG})_{10} \bullet (\text{EPG})_{10}$  in red recorded at 5 °C.





**Figure 3.**  ${}^1\text{H}$ ,  ${}^{15}\text{N}$ -HSQC spectrum (A) and two-dimensional plane of a three-dimensional  ${}^1\text{H}$ ,  ${}^{15}\text{N}$ -NOESY-HSQC (B and C) spectrum of a mixture of  $(\text{PKG})_{10}$ ,  $(\text{DKG})_{10}$  and  $(\text{EPG})_{10}$  peptides.





**Figure 4.**

(A) NH-NH *inter-chain* NOE expected based on the crystal structure of (PPG)<sub>10</sub> collagen triple helix<sup>30</sup> (PDB code: 1K6F). (B) <sup>1</sup>H, <sup>1</sup>H-NOESY spectrum of a mixture of (PKG)<sub>10</sub>, (EPG)<sub>10</sub> and (DKG)<sub>10</sub> showing the NH-NH NOE cross peaks. G(NH)-G(NH) correlations are indicated in black and red. The G<sub>DKG</sub>(NH)-G<sub>PKG</sub>(NH) correlation shown in orange/dashed could not be definitively assigned due to spectral overlap. (C) Model of the ABC heterotrimer composed of (PKG)<sub>10</sub>, (EPG)<sub>10</sub> and (DKG)<sub>10</sub> chains indicating observed NH-NH inter-chain correlations.