Electrostatic Interactions Involving Lysine Make Major Contributions to Collagen Triple-Helix Stability†

Anton V. Persikov,‡ John A. M. Ramshaw,§ Alan Kirkpatrick,§ and Barbara Brodsky*,‡

Department of Biochemistry, UMDNJ−Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854, and CSIRO, 343 Royal Parade, Parkville, Australia

Received August 17, 2004; Revised Manuscript Received November 12, 2004

ABSTRACT: Important stabilizing features for the collagen triple helix include the presence of Gly as every third residue, a high content of imino acids, and interchain hydrogen bonds. Host−guest peptides have been used previously to characterize triple-helix propensities of individual residues and Gly-X-Y triplets. Here, comparison of the thermal stabilities of host-guest peptides of the form $(Gly\text{-}Pro\text{-}Hyp)_3$-Gly-X-Y-Gly-X′-Y′-$(Gly\text{-}Pro\text{-}Hyp)_3$ extends the study to adjacent tripeptide sequences, to encompass the major classes of potential direct intramolecular interactions. Favorable hydrophobic interactions were observed, as well as stabilizing intrachain interactions between residues of opposite charge in the *i* and $i + 3$ positions. However, the greatest gain in triple-helix stability was achieved in the presence of Gly-Pro-Lys-Gly-Asp/Glu-Hyp sequences, leading to a T_m value equal to that seen for a Gly-Pro-Hyp-Gly-Pro-Hyp sequence. This stabilization is seen for Lys but not for Arg and can be assigned to interchain ion pairs, as shown by molecular modeling. Computational analysis shows that Lys-Gly-Asp/Glu sequences are present at a frequency much greater than expected in collagen, suggesting this interaction is biologically important. These results add significantly to the understanding of which surface ion pairs can contribute to protein stability.

Proteins require a stable native structure at the temperature at which they act, and in certain protein classes, the modulation of stability plays a pivotal role in biological function. For collagens, thermal stability of triple-helical structure is critical to biosynthesis, function, and degradation. Early studies revealed a close correlation between the thermal stability of collagen and the upper environmental temperature of the organism, ranging from 6 °C in ice fish to 39 °C in mammals and 46 °C in thermophilic worms (*1*−*3*). Recent *in vitro* studies suggest isolated mammalian collagen molecules show a gradual unfolding of the triple helix at 35 °C, indicating a very slow approach to equilibrium and a thermal stability less than body temperature (*4*). It is only the self-association of collagen molecules into fibrils that gives them the added stability to ensure a long lifetime for their *in vivo* structural function. In addition to global collagen stability, regions of local instability or flexibility have been implicated in biological processes (*5*). Microunfolding, a local dynamic loss of triple-helix structure, is detected in the triple helix below its melting temperature and is thought to be a prerequisite for fibril formation (*4*, *6*, *7*). Ligand-binding regions of collagen sometimes are less stable and more flexible (*5*, *8*, *9*), including an imino-acid-deficient region adjacent to the unique collagenase (matrix metalloproteinase) cleavage site (*10*). In addition, it has been suggested that stability of regions surrounding mutations may affect their clinical phenotype (*11*). Thus, clarification of the sequence dependence of triple-helix global and local stability is important to understanding collagen function.

The triple helix is a supercoiled form of three polyproline-II-like helices, and steric requirements for this structure dictate the unique features of the collagen amino acid sequence. The presence of glycine as every third residue allows for close packing of the three polypeptide chains, while the high content of the imino acids proline (Pro) and hydroxyproline (Hyp)[1] promote the polyproline-II-like conformation of individual chains (*12*−*14*). While imino acids themselves confer strong stabilization, the post-translational modification of Pro to Hyp in the position prior to Gly confers an additional advantage, which is likely to result from its promotion of the more favorable ring pucker for the Y position and its involvement in solvent-mediated hydrogen bonding (*15*−*17*). The sequence Gly-Pro-Hyp is the most common and stabilizing tripeptide unit in the $(Gly\text{-}Xaa\text{-}Yaa)_n$ pattern, while the identities of nonimino acid residues in the X and Y positions are important for biological recognition and specificity.

† This work was supported by NIH Grant GM60048 (to B.B.). A.P. was supported by the Michael Geismann Fellowship from the Osteogenesis Imperfecta Foundation, and B.B. was a recipient of the Ruth Kirchstein National Research Service Award (NIGMS).

* To whom correspondence should be addressed: Department of Biochemistry, UMDNJ−Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854. Telephone: (732) 235-4397. Fax: (732) 235-4783. E-mail: brodsky@umdnj.edu.

‡ UMDNJ−Robert Wood Johnson Medical School.

§ CSIRO.

[1] Abbreviations: Hydroxyproline is designated by "Hyp" in the 3-letter code and by "O" in the one-letter amino acid code. All peptides are denoted according to their guest hexapeptide sequence, using the one-letter amino acid code.

10.1021/bi048216r CCC: $30.25
Published on Web 01/11/2005

Peptides have been useful in defining factors involved in stabilization of the collagen triple helix (*16*, *18*, *19*). The propensity of all 20 amino acids for the X and Y positions were evaluated by comparing thermal stabilities for a host-guest peptide system, where guest residues are introduced in the middle of a Gly-Pro-Hyp repeating sequence (*20*). Except for Arg in the Y position, replacement of Pro or Hyp in the Gly-Pro-Hyp host triplet by any other residue decreased stability. For the X position, the most stabilizing residues are Pro and charged amino acids, while Hyp, Arg, and Met are most stabilizing in the Y position. In both positions, Gly, Cys, and aromatic amino acids are highly destabilizing. Thus, thermal stability will be varied according to the identity of the residue in the X or Y position, ranging from 32 to 47 °C for the X position and from 26 to 47 °C for the Y position in this system.

The contributions of individual amino acids to triple-helix stability can be altered by interactions between residues, which are determined by the geometry of the individual extended chains (2.9 Å rise per residue) and the proximity of the three chains staggered by one residue in the superhelix (*21*). The side chains of residues in the X and Y position are located on the outside of the triple helix, in contrast to the buried central Gly residues (*12*, *14*). The linear nature of the collagen triple helix limits possible intra- and interchain X and Y interactions to sequentially close neighboring residues. Molecular modeling indicates that even long side chains, e.g., Lys, Arg, or Glu, can interact within a homotrimeric triple helix only if they are not more than 3 residues apart (*22*−*24*). This limits the possibilities of pairwise interactions within a triple helix to XY, YX′, XX′, and YY′, where neighboring tripeptide units are designated as Gly-X-Y-Gly-X′-Y′. A previous report characterized XY pairwise interactions (*21*). About half of Gly-X-Y guests studied exhibited no evidence of interaction, showing a simple additivity in terms of destabilization from the residue in the X position relative to Pro and destabilization from the residue in the Y position relative to Hyp. The remaining Gly-X-Y guests showed small but significant degrees of stabilization or destabilization, with the most prominent being the 7 °C stabilization for GKD and GRD tripeptides (*21*).

The present study extends the host-guest system to include two adjacent tripeptide units, Gly-X-Y-Gly-X′-Y′, which now encompass all possible interactions within the collagen triple helix. Although there are many possible combinations of residues in adjacent tripeptide units, a selection was made of prototypes for likely electrostatic and hydrophobic interactions, using the residues found most frequently in collagens. Ala residues were incorporated in all positions in control peptides, because of the low interaction potential and high frequency of Ala in fibrillar collagens. Peptides were designed with Lys, Arg, Glu, and Asp residues, because of the high occurrence of these charged residues, and with Leu, because it is the most common hydrophobic residue in collagens. The results indicate intrachain XX′ hydrophobic and YY′ electrostatic stabilizing interactions, and a very dramatic interchain YX′ stabilization when Lys is in the Y position and acidic residues are in the following X′ position.

MATERIALS AND METHODS

Peptide Synthesis and Purification. All host-guest peptides studied consist of 8 triplets and had a uniform sequence of Ac-(Gly-Pro-Hyp)$_3$-Gly-X-Y-Gly-X′-Y′-(Gly-Pro-Hyp)$_3$-Gly-Gly-CONH$_2$. The N terminus was acetylated, and the C terminus was amidated to eliminate destabilizing charge repulsion. A C-terminal Gly-Gly sequence was included to eliminate the likelihood of diketopiperazine formation during synthesis. In some peptides, the C-terminal Gly was substituted by a Tyr for concentration determination, and this replacement was found to have no effect on folding and stability.

Peptides were synthesized by solid-phase chemistry by AusPep (Parkville Victoria, Australia), Tufts University Core Facility (Boston, MA) or in our laboratory as previously described (*21*). Peptides were purified to >95% purity on a C-18 column using a SHIMADZU reverse-phase HPLC system and were eluted in 0.1% trifluoroacetic acid with a binary gradient of 0−40% (v/v) water/acetonitrile. Laser desorption mass spectrometry (MALDI) confirmed peptide identity.

The peptide concentration was determined by dry weight, and for the peptides including Tyr at the C terminus, concentration was determined using the molar extinction coefficient 1400 M^{-1} cm^{-1} at 275 nm on a Beckmann model DU640 spectrophotometer.

Circular Dichroism (CD) Spectroscopy. CD measurements were made on an Aviv Model 62DS spectrometer. Peptide solutions of 1 mg/mL (0.43 mM) were characterized in phosphate-buffered saline (PBS) buffer (0.15 M NaCl and 10 mM sodium phosphate at pH 7.0) or in 0.1 M HAc (pH 3.0 ± 0.1). For thermal melting experiments, the ellipticity at 225 nm was monitored as the sample temperature was increased from 0 to 80 °C, with an average heating rate of 0.1 °C/min under standard conditions. Although these heating rates do not produce an equilibrium melting curve (*25*), the transition approximates a two-state model and the apparent melting temperature (T_m) values obtained under these standard conditions are useful for comparisons (*20*, *21*).

A trimer−monomer two-state model is applied for the analysis of thermal transitions

$$T_3 \underset{k_1}{\overset{k_2}{\rightleftharpoons}} 3M \qquad (1)$$

where T_3 is the triple-helical trimer state and M is the monomer state. Fraction folded was calculated from CD melting curves as a ratio

$$F(T) = \frac{\theta(T) - \theta_M(T)}{\theta_T(T) - \theta_M(T)} \qquad (2)$$

where θ is the observed ellipticity and θ_T and θ_M are the ellipticities in the native and monomer forms, respectively, which are linear functions of the temperature. T_m values were obtained as a midpoint of the transitions ($F = 0.5$) at these defined heating conditions. From experiments repeated on independently prepared samples, the error of determination of the T_m was ±0.4 °C.

Differential Scanning Calorimetry (DSC). DSC experiments were performed on a Nano-DSC II (Calorimetry

Sciences Corp., Model 6100) instrument at scanning rate of 1.0 °C/min. The peptide concentration was 1.0 mg/mL in PBS at pH 7.0. Calorimetric enthalpy values were obtained by temperature integration of the excess heat capacity. The experimental error of determination of the calorimetric enthalpy did not exceed 5%.

Thermal Stability Predictions. Experimental T_m values for peptides were compared with those predicted on the basis of independent, noninteracting tripeptide units, to evaluate if there were stabilizing or destabilizing interactions between adjacent triplets. The T_m at standard conditions was used as a measure of thermal stability. The predicted T_m value of a given host-guest peptide was calculated from the known stability contributions of the individual GXY tripeptides in its sequence. The decrease in stability that results when the most stabilizing GPO unit is replaced by a guest GXY sequence is taken from previously published data on host−guest peptides (*20*, *21*). The predicted stability of a host−guest peptide with a guest hexapeptide sequence GXYGX′Y′ is calculated as the decrease in stability caused by the first GXY triplet plus the decrease caused by the second GX′Y′ triplet subtracted from the T_m value for the $(GPO)_8$ host peptide (47.3 °C)

$$T_m^{GPOGPO} - T_m^{GXYGX'Y'} = [T_m^{GPO} - T_m^{GXY}] + [T_m^{GPO} - T_m^{GX'Y'}] \quad (3)$$

For example, the prediction for the peptide with a GPLGLO guest makes use of the experimental information that the T_m values for the host $(GPO)_8$, GPL, and GLO peptides are $T_m = 47.3$, 31.7, and 39.0 °C, respectively (*20*). The substitution of one GPO triplet by GPL would lead to a decrease in stability by 15.6 °C, and the GPO to GLO replacement would destabilize the peptide by 8.3 °C. If these effects are independent and there is no stabilizing/destabilizing interactions between two Leu residues, then the melting temperature for GPLGLO peptide would be 47.3 °C − (15.6 °C + 8.3 °C) = 23.4 °C. If the observed stability is greater/smaller than the predicted value, then a stabilizing/destabilizing interaction between the two adjacent tripeptide units in the triple helix is suggested.

Sequence Analysis. The search for the presence of nonrandom XY pair combinations in collagens was carried out using amino acid sequences obtained from SWISS-PROT annotated protein sequence database (http://au.expasy.org/sprot/). The sequences of chains of human fibril-forming collagens type I (α1 and α2), type II (α1), type III (α1), type V (α1, α2, and α3), and type XI (α1 and α2) were used, giving a total of 3046 Gly-X-Y triplets. The χ^2 test is carried out to test the null hypothesis that the distribution of amino acid pairs in collagen sequences is not statistically different from that expected on the basis of the observed frequencies of each of the two residues in a given pair. For instance, the observed frequency of KE pairs in YX′ positions was tested by finding all KGE sequences in fibrillar collagens (3.0%) and comparing the number found with that expected on the basis of the frequency of K residues in the Y position (9.0%) and the frequency of E residues in the X position (13.0%) ($e_i = 0.09 \times 0.13 = 0.12$ or 1.2%). This calculation is done for all pairs, and then the χ^2 value for the distribution of all pairs is calculated as

$$\chi^2 = \sum_{i=1}^{M} \frac{(n_i - e_i)^2}{e_i} \quad (4)$$

where n_i is the observed number of given amino acid pairs, e_i is the number expected on the basis of random X and Y distribution, and M is the total number of analyzed pairs (117, 117, 112, and 118 for XY, YX, XX′, and YY′ pairs, respectively). Because of statistical requirements, only pairs with expected frequencies of 5 or more were included in the analysis. If the calculated χ^2 value is larger than that obtained from statistical tables or eq 5, for a given degree of freedom, then the null hypothesis is rejected at a defined level of significance.

$$\chi_f^2 = f\left(1 - \frac{2}{9f} + z_p\sqrt{\frac{2}{9f}}\right)^3 \quad (5)$$

The number of degrees of freedom ($f = N - 1$) is determined from the total number N of possible pairs, and z_p is determined by the level of significance p chosen (for the level $p = 0.05$ used in this study, $z_p = 1.645$). There are 400 possible amino acid pair combinations. However, there is no Trp in the X position in all sequences analyzed; therefore $N = 380$ for XY and YX pairs and $N = 360$ for XX pairs. Because there is only one Cys, Trp, and Tyr reported in the Y position, it is impossible to get Cys-Cys, Trp-Trp, or Tyr-Tyr pairs and N is reduced to 397 for YY′ pairs. As a result, the number for degrees of freedom is 379, 359, and 396 respectively for XY (or YX′), XX′, and YY′ pairs.

Molecular Modeling. Molecular models for the peptides were based on ideal atomic coordinates of the peptide $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ (*26*) (http://www.rcsb.org/pdb, PDB ID 1K6F) and built using the Molecular Operating Environment 2002.03 (Chemical Computing Group Inc., Montreal, Canada). Every Pro in the Y position was substituted with Hyp, because protein chemistry indicates these Pro are almost completely hydroxylated *in vivo*. The structure was solvated with a standard MOE Water Soak procedure and refined by energy minimization using Kollman's all-atom force field (*27*). One Pro residue in the center of each polypeptide chain was replaced by Glu, and one Hyp was replaced by Lys to generate the required sequences: GEKGPO, GPKGEO, and GEOGPK. The accessible side-chain conformations of Lys and Glu were evaluated using the Stochastic Conformation Search in MOE (*28*) while holding fixed the sites of all other atoms. The Stochastic Conformation Search method generates new molecular conformations by randomly perturbing the position of each atom in the molecule by 1 Å and by rotating about single bonds. The conformers generated were minimized until their root-mean-square gradient was <0.01 kcal/mol. Two conformations were considered distinct if their energies differed by 0.01 kcal/mol and their atom positions exceeded the tolerance of 0.1 Å. The conformation with minimum energy for each structure was kept, and the distances between nitrogen of Lys side chain and closest oxygen of Glu side chain were measured.

RESULTS

Starting with the host peptide, $(Gly\text{-}Pro\text{-}Hyp)_8$, Pro, and/or Hyp residues in the two central tripeptide units were

Table 1: Experimentally Observed Melting Temperatures for Hexapeptide Host−Guest Peptides Compared with Those Predicted on the Basis of Individual Triplet Contributions

hexapeptide	T_m^{exp} (°C)[a]	T_m^{pred} (°C)[b]	ΔT_m (°C)
GPOGPO[c]	47.3		
GAAGPO[d]	32.9		
GPOGAA	33.3	32.9	+0.4
GAAGAA	20.0	18.5	+1.5
GAOGAO	36.9	36.1	+0.8
GPAGPA	35.8	34.5	+1.3
GPLGLO	28.2	24.4	+3.8
GLOGLO	38.1	30.7	+7.4
GEKGPO[c]	35.0		
GPKGEO	47.8	32.4	+15.4
GEOGPK	38.0	32.4	+5.6
GKOGEO	37.9	37.1	+0.8
GPKGPE	36.5	29.2	+7.3
GPKGDO	47.1	29.6	+17.5
GPRGDO	39.6	40.0	−0.4
GDOGPR	36.2	40.0	−3.8
GPRGEO	42.8	42.8	0.0

[a] T_m values were determined by CD spectroscopy under standard conditions as described. [b] Predicted T_m values were calculated by taking the destabilization predicted by two individual tripeptide units, according to eq 3. [c] See ref *20*. [d] See ref *21*.

replaced by Ala, Leu, Lys, Arg, Glu, or Asp residues. The peptides, of the form (Gly-Pro-Hyp)$_3$-Gly-X-Y-Gly-X′-Y′-(Gly-Pro-Hyp)$_4$, all showed a CD spectrum typical of a triple helix, with its characteristic maximum near 225 nm and a minimum near 198 nm. The melting profiles were determined by monitoring the decrease in the 225 nm CD maximum as the temperature is increased at a very slow rate (average 0.1 °C/min), and all peptides show a cooperative thermal transition (Table 1 and Figure 1). Previous NMR and ultracentrifugation studies indicate that the CD melting profile corresponds to the trimer−monomer transition (*29*, *30*). Thus, the hexapeptide host-guest peptides under study all form homotrimeric stable triple helices. The objective is to use the thermal transition data to evaluate whether there are favorable or unfavorable YX′, YY′, and YX′ interactions of an intra- or interchain nature. One established way of determining whether there are interactions between a defined pair of residues is to use a double mutant/thermodynamic cycle analysis (*31*−*33*). Although the very long equilibration times make it difficult to get equilibrium curves for the collagen-like host-guest peptides, standard type melting curves used in this study (average heating rates of about 0.1°C/min, and concentration of 1 mg/mL) are dominated by unfolding and have proven to be a useful practical measure of thermal stability (*25*, *34*). Analysis is based on empirical observations that the additivity of T_m values of double mutant sets using standard conditions can be used to ascertain whether there are favorable or unfavorable interactions.

The hypothesis of independent additivity of T_m values for individual tripeptide units in a Gly-X-Y-Gly-X′-Y′ guest sequence is tested for Ala, as a prototype of a noninteracting residue in the surface-exposed X and Y positions. A previous study used the 4th tripeptide unit from the N terminus as the GAA guest triplet and reported a T_m value of 32.9 °C (*21*). The current studies include the 4th and 5th tripeptide units as guest sequences. To confirm that these two unit positions are equivalent, the peptide GPOGAA was synthesized and is found to have the same stability as the previously studied GAAGPO peptide (Table 1). Studies of Ala residues in two adjacent X positions (GAOGAO), two adjacent Y positions (GPAGPA), and all 4 positions (GAAGAA) all give T_m values that are in good agreement with values predicted on the basis of simple additivity of the two tripeptide units

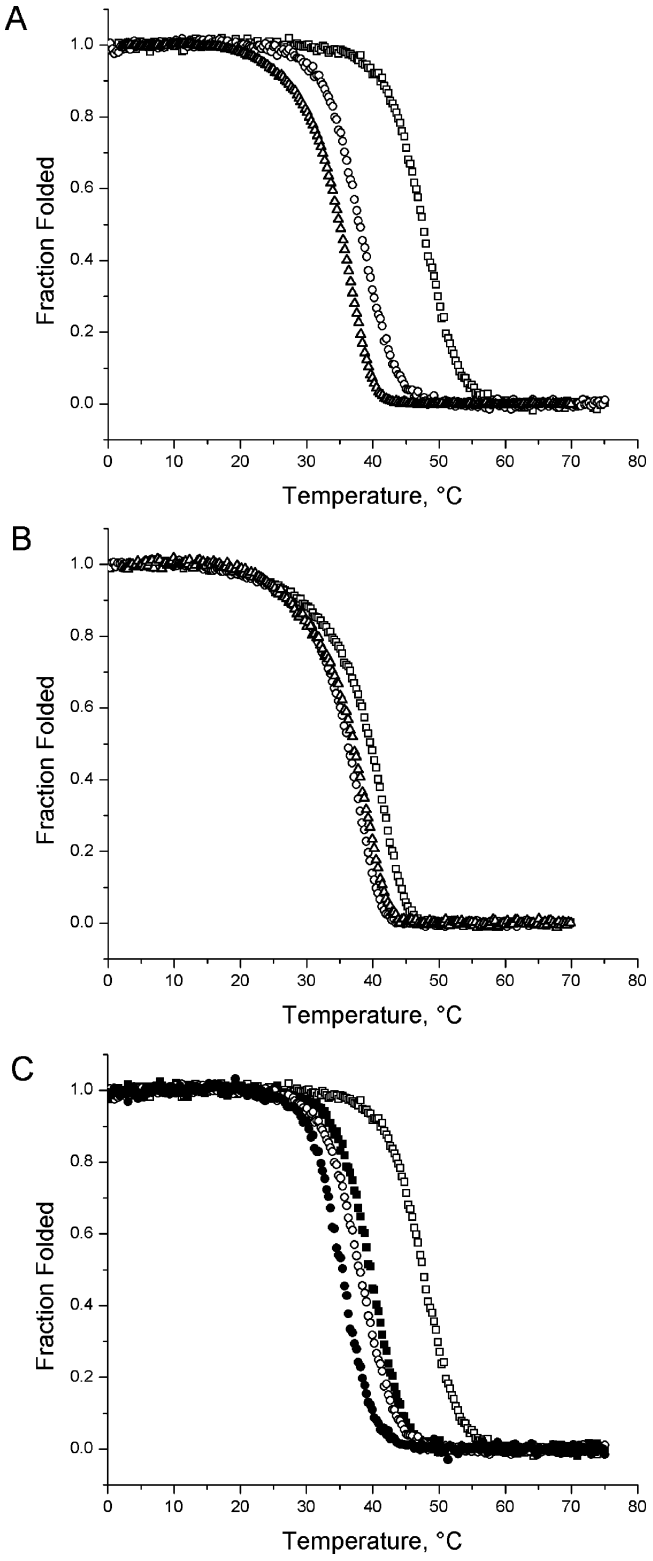

FIGURE 1: Temperature transitions monitored by CD spectroscopy for (A) GPKGEO (□), GEOGPK (○), and GEKGPO (△) peptides (1 mg/mL, PBS at pH 7). (B) GPRGDO (□), GDOGPR (○), and GDRGPO (△) peptides (1 mg/mL, PBS at pH 7). (C) GPKGEO (□ and ■) and GEOGPK (○ and ●) peptides at pH 7 (1 mg/mL, PBS, □ and ○) and pH 3 (1 mg/mL, HAc, ■ and ●).

(Table 1). This is consistent with the noninteracting nature of Ala in the triple helix and supports the use of T_m values in assessing the additivity of contributions of individual tripeptides.

This analysis was applied to hexapeptide guest sequences containing hydrophobic residues. Leu was introduced in adjacent tripeptide units in positions XX′ and YX′ (GLOGLO and GPLGLO peptides, see Table 1). The predicted stability for GPLGLO would be 24.4 °C (see the Materials and Methods), which is 3.8 °C smaller than the observed T_m = 28.2 °C. The Leu-Leu interaction was even more significant in an XX′ combination: the GLOGLO peptide was 7.4 °C more stable than that predicted on the basis of additivity (Table 1).

Pairs of oppositely charged residues were introduced into adjacent tripeptide units to investigate possible electrostatic effects. Guest sequences with a Lys or Arg residue in the Y position of the first triplet and an Asp or Glu residue in the X′ position of the following tripeptide unit were studied because of the strong preference of basic residues for Y positions and acidic residues for X positions in collagens. The GPKGEO and GPKGDO peptides both show dramatic stabilization compared to predictions (Table 1 and Figure 1). From a predictive view, the replacement of the first Hyp in GPOGPO by a Lys should result in a decrease in T_m by 10 °C, while replacing the second Pro by Asp would reduce the T_m by 7 °C, giving a net expected decrease of 17 °C for GPKGDO compared with GPOGPO. However, the observed T_m value for GPKGDO is 47.1 °C, indicating that some very favorable interaction has increased the T_m by ∼17 °C, bringing it to the same value as the T_m for GPOGPO.

A highly stabilizing interaction of 15.4 °C was also observed for GPKGEO (Table 1) and was further investigated by reversing the tripeptide order and by varying pH. A peptide with the same tripeptide composition but in a reverse order (GEOGPK) was studied, where the distance between the charged side chains is too great for interaction. The stability of GEOGPK was 5.6 °C greater than expected (38.0 °C versus 32.4 °C). However, this increase is much less than the 15 °C difference between GPKGEO expected and observed T_m values, suggesting the importance of direct interactions involving KGE. To further characterize the nature of these interactions, the stability of GPKGEO and GEOGPK peptides was measured at pH 3, which is below the p*K* of the Glu side chain. The T_m for GEOGPK decreased by 2.7 °C when the pH is lowered, consistent with pH studies on individual GEO and GPK peptides (*35*), while the GPKGEO peptide decreased by 8.2 °C, indicating a charge-mediated contribution because of ion-pair formation (Figure 1C).

In contrast to the striking stability gained in KGE/D sequences, peptides with RGD and RGE sequences (GPRGDO and GPRGEO) showed a stability close to that predicted (Table 1). However, a comparison with the reverse peptide, with the same tripeptide composition (GDOGPR versus GPRGDO) indicated that there may be a small stabilizing contribution for RGD, of about 3−4 °C (Figure 1B). The observed stability of the GDOGPR peptide was less than that predicted and is consistent with other observations suggesting that the GPR unit contributes less to stability when it is not surrounded by all GPO units (*20*, *21*, *36*).

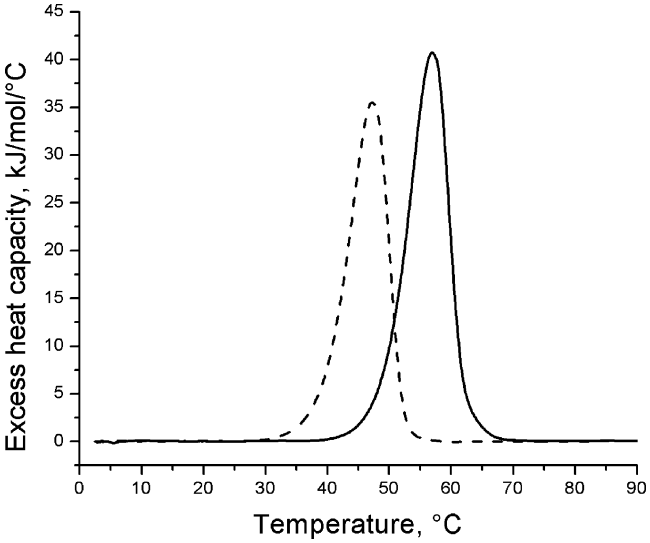

FIGURE 2: DSC scans showing the temperature dependence of excess heat capacity for peptides GPKGEO (—) and GEOGPK (- - -) at 1 mg/mL in PBS at pH 7.0.

Modeling and calculations have suggested the potential for intrachain interactions in the collagen triple helix for charged side chains of amino acids 3 residues apart (*22*). Lys and Glu residues were introduced into adjacent X positions (GKOGEO) and show good agreement with that predicted, indicating little interaction. When Lys and Glu were introduced into adjacent Y positions (GPKGPE), a significant stabilization was seen (Table 1).

Calorimetry was used to further investigate the nature of the very large stabilization seen for KGE/D sequences. A comparison of GPKGEO with GEOGPK confirmed the much greater thermal stability (∼10 °C) of the former peptide, although the T_m values are now higher because of the faster heating rate in these nonequilibrium conditions (Figure 2) (*25*). The higher enthalpy of GPKGEO (350 kJ/mol) compared to that seen for GEOGPK (295 kJ/mol) suggests a contribution of hydrogen bonding to the KGE stabilization.

If the pairwise interactions between adjacent tripeptide units found in these peptide studies are important for modulating the thermal stability of collagens, favorable amino acid pairs may be expected to be found preferentially in collagen sequences. The principle of equal probability is applied here: the frequency of any amino acid pair is compared to that expected based on the observed frequencies of the individual residues, assuming random distribution of each amino acid residue along the collagen sequence. It was previously reported that the distribution of residues within Gly-X-Y triplets (XY pairs) differs from that expected on the basis of frequencies of individual residues (*21*). Here, the observed versus expected frequency is calculated for the pairwise combinations that have the potential for close contact within a Gly-X-Y-Gly-X′-Y′ sequence context (YX′, XX′, and YY′). The resulting distributions were analyzed statistically using a χ^2 test, (Table 2), and the YX′ pairs distribution differed significantly from the expected distribution at a significance level of $p = 0.05$. Analysis indicated that KGD and KGE pairs made the highest χ^2 contributions, supporting the involvement of highly stabilizing electrostatic interactions involving Lys-Glu/Asp YX′ pairs in collagen (Figure 3). The XX′ and YY′ pairs and pairs of residues

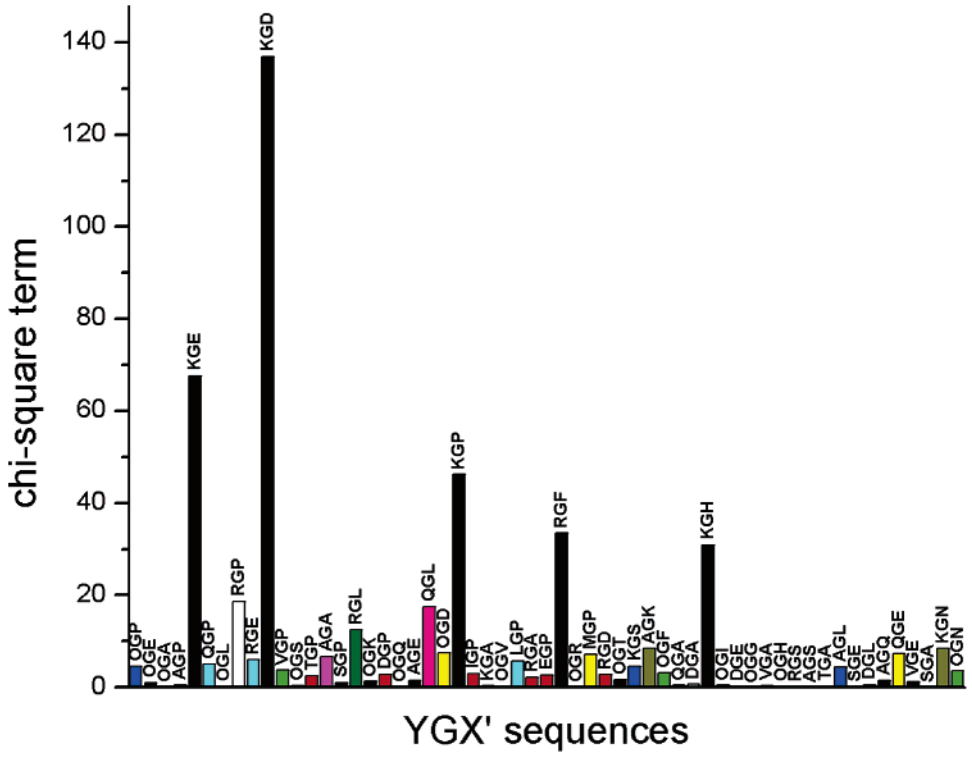

FIGURE 3: χ^2 terms for comparison of the occurrences of YX′ amino acid pairs in fibrillar collagens with those expected on the basis of individual residue frequencies. These are plotted in order of frequency of YX′ occurrences in human fibrillar collagens (from left, highest occurrence, to right, lowest occurrence). Five YX′ pairs are the major contributors to the high χ^2 value: KGD, KGE, KGP, RGF, and KGH (black bars).

Table 2: χ^2 Values for Comparison of the Occurrences of Different Amino Acid Pairs in Fibrillar Collagen Sequences GXYGX′Y′GX″Y″GX‴Y‴ with that Expected on the Basis of Individual Residue Frequencies[a]

amino acid pair	χ^2 values	χ^2 expected at $p = 0.05$
XY	**441**	425
XY′	172	425
XY″	164	425
YX′	**614**	425
YX″	173	425
YX‴	184	425
XX′	194	404
XX″	185	404
XX‴	170	404
YY′	229	443
YY″	194	443
YY‴	200	443

[a] χ^2 values exceeding the critical values at a $p = 0.05$ level of confidence are shown in bold, indicating rejection of the null hypothesis and a preferential occurrence of XY and YX′ pairs.

further apart did not differ statistically from those predicted on a random basis (Table 2).

DISCUSSION

Unraveling the relationship between amino acid sequence and protein stability is, in principle, more approachable for a protein such as collagen, where the supercoiled rodlike nature prevents interactions between distant amino acids and where repetitive patterns constrain the sequence. Sequences of the form Gly-Pro-Hyp, which are the most common tripeptide unit, are considered to confer maximal stability and rigidity to the collagen triple helix (*20*, *21*, *37*, *38*). Variations in the identities of the residues in the X and Y positions of the $(Gly\text{-}X\text{-}Y)_n$ pattern are considered to lower stability and rigidity, providing sites capable of specifying self-association and ligand binding. The peptide studies reported here complete the characterization of intramolecular triple-helix interactions and raise the possibility that sequences other than Gly-Pro-Hyp may be equally stabilizing for the triple helix.

In addition to the previously reported GPR sequence (*36*), the sequences GPKGEO and GPKGDO are just as favorable as GPOGPO in the host-guest peptide system. The mechanism of the very high stabilization seen from GPKGEO sequences contains an electrostatic component, as seen by the effect of pH, and also contains a hydrogen-bonding component as seen by its favorable enthalpy. It is likely that GPKGDO follows a similar mechanism. Molecular modeling shows that there are no possible intrachain interactions, because the side chains are pointing in directions 108° apart in each turn of the helix. However, three interchain ion pairs (K^C-E^A, K^B-E^C, and K^A-E^B, where superscript indicates chain A, B, or C in the triple helix) are possible between the three staggered chains, because their side chains come in close proximity in the GPKGEO peptide (Figure 4). The very high stability suggests that one or more of these ion pairs is very stabilizing. The GPRGE/DO oppositely charged side chains could also make similar close contact, yet these contribute little to triple-helix stability. The highly stabilizing effect of KGD and KGE sequences raises the possibility of pH-modulated triple-helix stability, which could be important during secretion, degradation, or ligand binding.

The possibility of intrachain ion pairs was raised previously, because of the potential for interactions between oppositely charged side chains of residues i and $i + 3$ in a

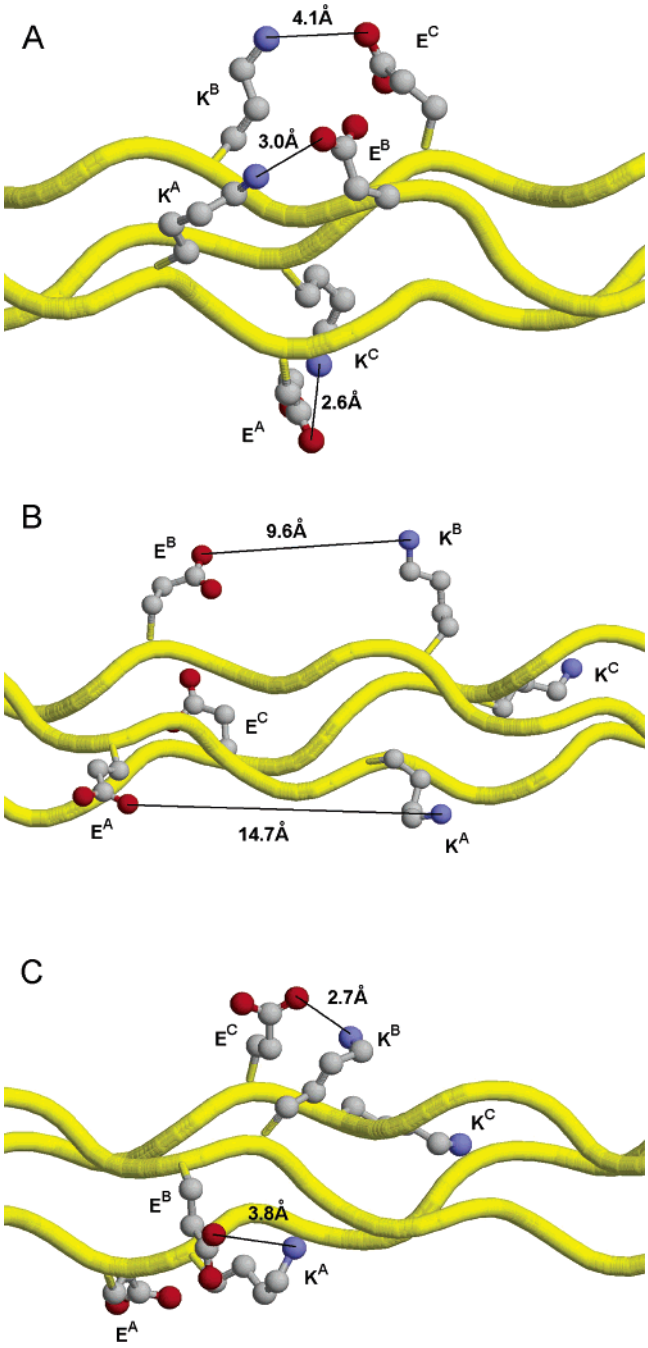

FIGURE 4: Computer models for pairwise interactions between Glu and Lys residues in different collagen sequence combinations. The distances are shown in angstroms. (A) GPKGEO. (B) GEOGPK. (C) GEKGPO.

polyproline II helix, which has 3 residues per turn (*22*, *39*). A recent investigation of ion pairs in a polyproline-II-like helix reported no stabilizing effect from salt bridges (*39*). Here, experimental data is reported for the first time on the effects of XX′ and YY′ amino acid pairs within one chain of a triple helix. A significant hydrophobic side-chain intrachain interaction was observed for two Leu residues in XX′ positions (GLOGLO) in the collagen triple helix, and significant electrostatic stabilizing interactions were observed for YY′ positions (GPKGPE). Surprisingly, no stabilization was seen for opposite charges in XX′ pairs (GKOGEO).

The data presented here, together with previous reports, show that close proximity of oppositely charge side chains within a triple helix can have diverse effects, depending upon the identity of the residue and its location: GPKGEO and GPKGDO sequences are highly stabilizing; GPKGPE, GEKGPO, and GEOGPK contribute significant but less stability; and GPRGEO, GPRGDO, GDOGPR, and GKOGEO contribute little or no stability. It has also been reported recently that there is no destabilizing effect of clusters of positive charge in GRKGRO and GKRGKO peptides (*40*). Significant advances have been made in understanding when ion pairs between surface-located oppositely charged side chains will lead to increased protein stability (*31*, *41*−*43*). Considerations include spatial features of the salt bridge, the interactions of the individual side chains with the rest of the protein, the entropic cost of restricting the side chains, and the degree of desolvation. All residues in the X and Y position of a triple helix are highly exposed to the solvent, and the variation in stabilizing effects cannot be explained solely on the basis of the difference in intrinsic strength based on models. For instance, both models of KGE/D and RGE/D show the potential for three close interactions, while only the former are highly stabilizing. Thus, the local context, involving interactions with the rest of the protein, may be critical for the triple helix. It is possible that RGE/D salt bridges contribute little to stability perhaps because Arg side-chain interactions with the backbone carbonyl are highly favorable, while similar Lys side-chain−backbone interactions are not as favorable (*20*, *21*, *44*, *45*). Both GKOGEO and GPKGPE have similar distances between oppositely charged side chains, yet only the latter lead to increased thermal stability. In high-resolution triple-helix structures containing the potential ion-pair sequence GEK or GER, the charged residues are seen to be involved in a range of hydrogen-bonding interactions, often to available peptide backbone carbonyls and often through water (*46*, *47*). This suggests charges may serve as a focal point for an ensemble of dynamic potential interactions involving the peptide backbone and water, and this local context may be the determining factor in whether salt bridges act to stabilize the triple helix.

This study completes all possible pairwise interactions geometrically possible within the collagen triple helix and provides a database for stability predictions. The interactions deduced here from homotrimeric peptides can be applied to collagens with three identical chains. However, for heterotrimeric collagens, the interactions between chains with different sequences would need to be considered. The experimental results on peptides have been correlated with statistical analysis of sequences, showing the preferential occurrence of the stabilizing KGD and KGE pairs in collagens and suggesting a functional role. In addition to being found in greater than expected amounts in the abundant human fibrillar collagens, the hyperstable KGE and KGD sequences are frequent in nonfibrillar collagens and in the collagen domains recently reported in bacteria and viruses (*48*, *49*). For instance, the hypothetical collagen from white spot syndrome virus (GenBank accession number AF440570) contains a 99-residue long $(KGE)_{33}$ repeating sequence. Because there is no Hyp in prokaryotes, the strong electrostatic interactions may provide an alternate evolutionary strategy for triple-helix stabilization.

ACKNOWLEDGMENT

We thank Tim Hardingham for advice on statistical analysis, Teresita Silva for technical assistance, and Drs. Angela Mohs and Jean Baum for helpful discussions.

REFERENCES

1. Rigby, B. J. (1968) Amino acid composition and thermal stability of the skin collagen of the Antarctic ice-fish, *Nature 219*, 166−167.
2. Mathews, M. B. (1975) *Connective Tissue: Macromolecular Structure and Evolution*, Springer-Verlag, New York.
3. Gaill, F., Mann, K., Wiedemann, H., Engel, J., and Timpl, R. (1995) Structural comparison of cuticle and interstitial collagens from annelids living in shallow sea-water and at deep-sea hydrothermal vents, *J. Mol. Biol. 246*, 284−294.
4. Leikina, E., Mertts, M. V., Kuznetsova, N., and Leikin, S. (2002) Type I collagen is thermally unstable at body temperature, *Proc. Natl. Acad. Sci. U.S.A. 99*, 1314−1318.
5. Sacca, B., Fiori, S., and Moroder, L. (2003) Studies of the local conformational properties of the cell-adhesion domain of collagen type IV in synthetic heterotrimeric peptides, *Biochemistry 42*, 3429−3436.
6. Ryhanen, L., Zaragoza, E. J., and Uitto, J. (1983) Conformational stability of type I collagen triple helix: Evidence for temporary and local relaxation of the protein conformation using a proteolytic probe, *Arch. Biochem. Biophys. 223*, 562−571.
7. Kadler, K. E., Hojima, Y., and Prockop, D. J. (1988) Assembly of type I collagen fibrils *de novo*. Between 37 and 41 °C the process is limited by micro-unfolding of monomers, *J. Biol. Chem. 263*, 10517−10523.
8. Deprez, P., Doss-Pepe, E., Brodsky, B., and Inestrosa, N. C. (2000) Interaction of the collagen-like tail of asymmetric acetylcholinesterase with heparin depends on triple-helical conformation, sequence, and stability, *Biochem. J. 350*, 283−290.
9. Shah, N. K., Sharma, M., Kirkpatrick, A., Ramshaw, J. A. M., and Brodsky, B. (1997) Gly-Gly-containing triplets of low stability adjacent to a type III collagen epitope, *Biochemistry 36*, 5878−5883.
10. Fields, G. B. (1991) A model for interstitial collagen catabolism by mammalian collagenases, *J. Theor. Biol. 153*, 585−602.
11. Bachinger, H. P., Morris, N. P., and Davis, J. M. (1993) Thermal stability and folding of the collagen triple helix and the effects of mutations in osteogenesis imperfecta on the triple helix of type I collagen, *Am. J. Med. Genet. 45*, 152−162.
12. Rich, A., and Crick, F. H. (1961) The molecular structure of collagen, *J. Mol. Biol. 3*, 483−506.
13. Ramachandran, G. N. (1964) in *Treatise on Collagen* (Ramachandran, G. N., Ed.) pp 103−183, Academic, New York.
14. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution, *Science 266*, 75−81.
15. Privalov, P. L. (1982) Stability of proteins. Proteins which do not present a single cooperative system, *Adv. Protein Chem. 35*, 1−104.
16. Jenkins, C. L., and Raines, R. T. (2002) Insights on the conformational stability of collagen, *Nat. Prod. Rep. 19*, 49−59.
17. Vitagliano, L., Berisio, R., Mastrangelo, A., Mazzarella, L., and Zagari, A. (2001) Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability, *Protein Sci. 10*, 2627−2632.
18. Baum, J., and Brodsky, B. (1999) Folding of peptide models of collagen and misfolding in disease, *Curr. Opin. Struct. Biol. 9*, 122−128.
19. Mechling, D. E., and Bachinger, H. P. (2000) The collagen-like peptide $(GER)^{15}GPCCG$ forms pH-dependent covalently linked triple helical trimers, *J. Biol. Chem. 275*, 14532−14536.
20. Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2000) Amino acid propensities for the collagen triple-helix, *Biochemistry 39*, 14960−14967.
21. Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2002) Peptide investigations of pairwise interactions in the collagen triple-helix, *J. Mol. Biol. 316*, 385−394.
22. Katz, E. P., and David, C. W. (1990) Energetics of intrachain salt-linkage formation in collagen, *Biopolymers 29*, 791−798.
23. Katz, E. P., and David, C. W. (1992) Unique side-chain conformation encoding for chirality and azimuthal orientation in the molecular packing of skin collagen, *J. Mol. Biol. 228*, 963−969.
24. Salem, G., and Traub, W. (1975) Conformational implications of amino acid sequence regularities in collagen, *FEBS Lett. 51*, 94−99.
25. Persikov, A. V., Xu, Y., and Brodsky, B. (2004) Equilibrium thermal transitions of collagen model peptides, *Protein Sci. 13*, 893−902.
26. Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2002) Crystal structure of the collagen triple helix model $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$, *Protein Sci. 11*, 262−270.
27. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986) An all atom force-field for simulations of proteins and nucleic-acids, *J. Comput. Chem. 7*, 230−252.
28. Ferguson, D. M., Glauser, W. A., and Raber, D. J. (1989) Molecular mechanics conformational-analysis of cyclononane using the rips method and comparison with quantum-mechanical calculations, *J. Comput. Chem. 10*, 903−910.
29. Li, M. H., Fan, P., Brodsky, B., and Baum, J. (1993) Two-dimensional NMR assignments and conformation of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and a designed collagen triple-helical peptide, *Biochemistry 32*, 7377−7387.
30. Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J., and Brodsky, B. (1993) Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence, *Biochemistry 32*, 11688−11695.
31. Ibarra-Molero, B., Zitzewitz, J. A., and Matthews, C. R. (2004) Salt-bridges can stabilize but do not accelerate the folding of the homodimeric coiled-coil peptide GCN4-p1, *J. Mol. Biol. 336*, 989−996.
32. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1990) Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles, *Biochemistry 29*, 9343−9352.
33. Horovitz, A. (1996) Double-mutant cycles: A powerful tool for analyzing protein structure and function, *Folding Des. 1*, R121−R126.
34. Miles, C. A., and Bailey, A. J. (2004) Studies of the collagen-like peptide $(Pro\text{-}Pro\text{-}Gly)_{10}$ confirm that the shape and position of the type I collagen denaturation endotherm is governed by the rate of helix unfolding, *J. Mol. Biol. 337*, 917−931.
35. Chan, V. C., Ramshaw, J. A. M., Kirkpatrick, A., Beck, K., and Brodsky, B. (1997) Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix, *J. Biol. Chem. 272*, 31441−31446.
36. Yang, W., Chan, V. C., Kirkpatrick, A., Ramshaw, J. A., and Brodsky, B. (1997) Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides, *J. Biol. Chem. 272*, 28837−28840.
37. Fan, P., Li, M. H., Brodsky, B., and Baum, J. (1993) Backbone dynamics of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and a designed collagen-like triple-helical peptide by N^{15} NMR relaxation and hydrogen-exchange measurements, *Biochemistry 32*, 13299−13309.
38. Xu, Y. J., Hyde, T., Wang, X., Bhate, M., Brodsky, B., and Baum, J. (2003) NMR and CD spectroscopy show that imino acid restriction of the unfolded state leads to efficient folding, *Biochemistry 42*, 8696−8703.
39. Whittington, S. J., and Creamer, T. P. (2003) Salt bridges do not stabilize polyproline II helices, *Biochemistry 42*, 14690−14695.
40. Doss-Pepe, E., Deprez, P., Silva, T., Inestrosa, N. C., Kirkpatrick, A., Ramshaw, J. A., and Brodsky, B. (2004) Stepwise construction of triple-helical heparin binding sites using peptide models, *Biochim. Biophys. Acta 1698*, 187−195.
41. Kumar, S., and Nussinov, R. (2002) Close-range electrostatic interactions in proteins, *ChemBioChem 3*, 604−617.
42. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: Guidelines for protein engineering, *J. Mol. Biol. 327*, 1135−1148.
43. Marti, D. N., and Bosshard, H. R. (2003) Electrostatic interactions in leucine zippers: Thermodynamic analysis of the contributions

of Glu and His residues and the effect of mutating salt bridges, *J. Mol. Biol. 330*, 621−637.
44. Vitagliano, L., Nemethy, G., Zagari, A., and Scheraga, H. A. (1993) Stabilization of the triple-helical structure of natural collagen by side-chain interactions, *Biochemistry 32*, 7354−7359.
45. Kramer, R. Z., Bella, J., Brodsky, B., and Berman, H. M. (2001) The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence, *J. Mol. Biol. 311*, 131−147.
46. Kramer, R. Z., Venugopal, M. G., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (2000) Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair, *J. Mol. Biol. 301*, 1191−1205.
47. Emsley, J., Knight, C. G., Farndale, R. W., and Barnes, M. J. (2004) Structure of the integrin $\alpha 2\beta 1$-binding collagen peptide, *J. Mol. Biol. 335*, 1019−1028.
48. Rasmussen, M., Jacobsson, M., and Bjorck, L. (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins, *J. Biol. Chem. 278*, 32313−32316.
49. Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices, *J. Biol. Chem. 277*, 27312−27318.

BI048216R