Conformational Features of a Natural Break in the Type IV Collagen Gly-X-Y Repeat*

Fibrillar collagens have an absolute requirement for Gly as every 3rd residue, whereas breaks in the Gly-X-Y repeating pattern are found normally in the triple helix domains of non-fibrillar collagens, such as type IV collagen in basement membranes. In this study, a model 30-mer peptide is designed to include the interruption GPOGAAVMGPOGPO found in the α5 chain of type IV collagen. The GAAVM peptide forms a stable triple helix, with $T_m = 29^\circ$C. When compared with a control peptide with Gly as every 3rd residue, the GAAVM peptide has a marked decrease in the 225 nm maximum of its CD spectrum and a 10 °C drop in stability. A 50% decrease in calorimetric enthalpy is observed, which may result from disruption of ordered water structure anchored by regularly placed backbone carbonyls. NMR studies on specific $^{15}$N-labeled residues within the GAAVM peptide indicate a normal triple helical structure for Gly-Pro-Hyp residues flanking the break. The sequence within the break is not disordered but shows altered hydrogen exchange rates and an abnormal Val chemical shift. It was previously reported that a peptide designed to model a similar kind of interruption in the peptide (Pro-Hyp-Gly)$_{10}$ (GPOG-POPOGPO), is unable to form a stable triple helix, and replacement of GAA by GPO or VM by PO within the GAAVM break decreases the stability. Thus, rigid imido acids are unfavorable within a break, despite their favorable stabilization of the triple helix itself. These results suggest some non-random structure typical of this category of breaks in the Gly-X-Y repeat of the triple helix.

The repeating (Gly-X-Y)$_n$ sequence of the collagen triple helix allows for easy identification of this structural motif from amino acid sequences. The collagen triple helix structure consists of three supercoiled polypeptide chains, each in a polyproline II-like conformation (1–4). The repeating pattern of Gly as every 3rd residue is generated by steric requirements of the close packing of the three supercoiled polypeptide chains near a central axis. The collagen amino acid sequence is also characterized by a high frequency of proline and hydroxyproline (Hyp)$^9$ in the $X$ and $Y$ positions, respectively, which serves to stabilize the polyproline II-like helix of the individual chains. The collagen triple helix is the defining motif of all extracellular matrix proteins classified as collagens and is also found as a domain in various other proteins, including C1q, macrophage scavenger receptor, collagenous tail of asymmetric acetylcholinesterase, and the bacterial proteins Streptococcus pyogenes sc1 and Bacillus anthracis bcl1 (5–9).

Although the (Gly-X-Y)$_n$ repeating sequence is a requirement to form a collagen triple helix, this pattern is perfectly maintained in some collagen domains but not in others (9). The most abundant collagens are those found in fibrils with a 670 Å axial period (types I, II, III, V, XI), and these all maintain a precise Gly-X-Y repeat throughout their ~1000-residue triple helix domain. The replacement of even one Gly by another residue as a result of a single base change results in a pathological condition (10–12). For instance, missense mutations leading to osteogenesis imperfecta have been identified at more than 200 Gly sites along the α1 and α2 chains of type I collagen (13). However, such a strict regularity of Gly as every 3rd residue throughout the triple helix domain is not required in non-fibrillar collagens, which may be found as membrane proteins, on the surface of periodic collagen fibrils, as networks, or in anchoring fibrils (9). For instance, type IV collagen, which forms a network-like array in basement membranes, contains more than 20 breaks in the (Gly-X-Y)$_n$ pattern of its ~1350-residue long triple helix (14–16). These normal interruptions found in non-fibrillar collagens may play a role in molecular structure, self-association within the network, or binding (17).

A number of possible structural perturbations have been proposed to occur at interruptions in the Gly-X-Y repeating pattern of a triple helix. A kink in the triple helix is known to occur in C1q at a site where the A chain has an extra Thr, the C chain has a Gly-to-Ala replacement, and the B chain has an uninterrupted Gly-X-Y sequence (18, 19). Alternatively, an interruption could lead to a flexible site within the triple helix (16, 20). Visualization of type IV collagen chains by rotary shadowing shows some flexible sites, and the positions of flexible sites within the molecule correlate with several sequence interruptions. Two crystal structures were solved for the homotrimeric model peptide (Pro-Hyp-Gly)$_{10}$ containing an interruption. Peptides containing either a Gly-to-Ala replacement (3) or a deletion of one Hyp residue$^5$ show straight rod-like triple helices but with a loss of angular or axial register at the interruption site.

In addition to affecting molecular and higher order structure, breaks in the (Gly-X-Y)$_n$ pattern can also serve as interaction sites of biological importance. For example, the sequence of the α1 chain of type IV collagen contains a break GAGKGPQGFFYFDLRKLKGDKGP, which was shown to promote the adhesion of melanoma, ovarian carcinoma, and Jurkat cells (21, 22). In the mannose-binding protein, the serine proteinases are shown to bind near a kink in the triple helix domain (23). A kink, flexible site, or loss of loss of helix registration would introduce a structurally distinct site that could be involved in recognition.

Structural and functional consequences could be influenced by the

* This work was supported in part by National Institutes of Health Grants GM60048 (to B. J.) and GM45302 (J. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 Both authors contributed equally to this work.

2 Supported by a Deutsche Forschungsgemeinschaft postdoctoral fellowship.

3 To whom correspondence should be addressed: Dept. of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, 675 Hoes La., Piscataway, NJ 08854. Tel.: 732-235-4397; Fax: 732-235-4783; E-mail: brodsky@umdnj.edu.

4 The abbreviations used are: Hyp, hydroxyproline; DSC, differential scanning calorimetry; HSQC, heteronuclear single quantum coherence; deg, degrees.

5 J. Bella, J. Liu, and H. M. Berman, personal communication.
nature of the break and by the sequence of surrounding residues (24). Breaks can be classified by the way in which they interrupt the Gly-X-Y pattern. In many cases, the presence of Gly as every 3rd residue is altered such that glycines are separated by 1 or 4 residues, rather than the normal 2 residues, whereas in other cases, there are larger stretches of non-repeating residues separating Gly-X-Y repeats. It has been suggested that the Gly-X-Y residues surrounding breaks have some characteristic features with respect to charges and imino acids (24).

Peptide models provide an approach to characterizing the effects of breaks on triple helix structure and stability (3, 25, 26). Here, studies are presented on a peptide that models a break in the triple helix found in a chain of type IV collagen. Type IV collagen is a family of six chains, α1(IV), α2(IV), α3(IV), α4(IV), α5(IV), and α6(IV), which can form three distinct heterotrimeric collagen molecules with tissue-specific locations: (α1(IV))₃, α2(IV) in all mammalian basement membranes; α3(IV)α4(IV)α5(IV) in renal glomerular basement membrane; and (α5(IV))₂α6(IV) in smooth muscle basement membrane (27). Each molecule has a long triple helix domain, with a globular C terminus involved in chain selection and dimerization of molecules. There are over 20 breaks in the Gly-X-Y repeat of each type IV chain (14). Despite the presence of these interruptions in native type IV collagen, an additional break due to the replacement of one Gly residue in the Gly-X-Y repeat is not tolerated. More than 150 mutations leading to Alport syndrome have been defined within the X-linked α5(IV) chain, and these are predominantly single base mutations replacing one Gly by another residue (15, 28, 29). Studies are presented here on a homotrimer peptide model for a natural break found in the α5(IV) chain of type IV collagen, GPOGAAVMGPOGPO (residues 386–399) and a set of homologous peptides with imino acids introduced within the break. A peptide in which the Gly as every 3rd residue is restored was designed as a control.

**Circular Dichroism Spectroscopy**—CD spectra were recorded on an Aviv model 62DS spectropolarimeter. Peptides were purified on a Shimadzu reversed-phase HPLC. 1H-15N heteronuclear single quantum coherence (HSQC) spectra of GAAVM peptides 1 and 2 were overlaid using a Silicon Graphics workstation using the FELIX 2004 software package (MSI, San Diego, CA). HSQC spectra of GAAVM peptides 1 and 2 were overlaid using NMRView (31).

**Hydrogen exchange experiments** were carried out at 10°C, pH 2.5–2.6, where the pH includes the correction for the glass electrode solvent isotope artifact (32). The sample was equilibrated in D₂O at 10°C for a minimum of 72 h to ensure that the monomer-trimer interconversion had reached equilibrium. The sample was then lyophilized and dissolved in D₂O. The HSQC spectra were acquired with a 5-min acquisition time for the first 14 spectra and a 44-min acquisition time for the following spectra to allow the measurements of both fast- and slow-exchanging amide protons. Peak volumes were measured as a function of time. The volumes were normalized by the integration of the non-exchangeable resonance of DSS in one-dimensional spectra collected before and after the HSQC experiments. The hydrogen/deuterium exchange rates were single exponential with a non-zero asymptote \( \lambda \) in the equation

\[
I(t) = (I_0 - \lambda) \exp(-k_{\lambda} t) + \lambda
\]

(Eq. 1)

The error in \( k_{\lambda} \) was the fitting error from the variance-covariance matrix. The protection factors (P) were calculated by the equation...
A Break in Triple Helix Structure

TABLE 1

| Peptide     | Sequence                      | $T_{m}$ | $[\Theta]_{225\ \text{nm}}$ | $\Delta H_{\text{cal}}$ |
|-------------|-------------------------------|---------|------------------------------|------------------------|
| GAAVM       | Ac-(GPO)$_2$GAAGVM(GPO)$_4$-CONH$_2$ | 29.1 ± 0.5 | 2360                        | 188 ± 4                |
| GAAGVM      | Ac-(GPO)$_2$GAAGVM(GPO)$_4$-CONH$_2$ | 39.7 ± 0.5 | 4950                        | 354 ± 8                |

$P = k_{\text{int}}/k_{\text{ex}}$ (Eq. 2)

where $k_{\text{int}}$ is the theoretical monomer exchange rate for the amide proton at a specific pH and temperature (33). The error in the intrinsic rate was estimated by assuming an experimental uncertainty of ±0.2 °C and ±0.1 pH units. The error in the protection factor $\sigma_p$ was then calculated using the equation

$$\sigma_p = P^* \sqrt{(\sigma_{k_{\text{ex}}}/k_{\text{ex}})^2 + (\sigma_{k_{\text{int}}}/k_{\text{int}})^2}$$ (Eq. 3)

where $\sigma_{k_{\text{ex}}}$ is the error in the measured exchange rate and $\sigma_{k_{\text{int}}}$ is the error in the intrinsic rate.

Statistical Analyses of Amino Acids within Breaks—The $\chi^2$ test of statistical significance at a level of $p < 0.001$ was applied to test whether amino acids within the Gly-AA$_1$-AA$_2$-AA$_3$-AA$_4$-Gly breaks differ from amino acids in the X- and Y-position of the triple helical region. Since the number of Gly-AA$_1$-AA$_2$-AA$_3$-AA$_4$-Gly breaks in all type IV chains is only 28, the amino acids were grouped in categories as follows: hydrophobic residues, Ile, Leu, Met, Val; small residues, Ala, Cys, Gly, Ser; charged residues, Asp, Glu, His, Lys, Arg; imino acid, Pro; aromatic residues, Phe, Trp, Tyr; and polar residues, Asn, Gln, Thr.

RESULTS

Peptide Design—In the α5 chain of human type IV collagen, there are 22 breaks in the Gly-X-Y repeating sequence distributed all along the triple helix (14) (Fig. 1). A set of peptides was synthesized to investigate one of these naturally occurring breaks, by incorporating the sequence GPOGAAVMGPOGPO (residues 386–399) into a peptide model. Additional Gly-Pro-Hyp sequences were added on each side to ensure stability and to create a 30-mer peptide that could be compared with earlier peptides of this length (3, 34–38). All peptides in this study are designated by the single-letter code for the central sequence, which is flanked by Gly-Pro-Hyp units, so this peptide is labeled the GAAVM peptide (Table 1). Two peptides were synthesized with this sequence, with $^{15}$N-Gly, $^{15}$N-Ala, or $^{15}$N-Val at various positions, so that residue-specific NMR assignments could be made.

CD Spectroscopy and Differential Scanning Calorimetry of the GAAVM and GAAGVM Peptides—The peptide including the native GPO-GAAVMGPOGPO sequence was found to form a structure with a substantial amount of triple helix. The GAAVM peptide has a CD spectrum with a maximum at 225 nm (MRE$_{225\ \text{nm}}$ = 2360 deg cm$^{-2}$/dmol) and a minimum at 198 nm (MRE$_{198\ \text{nm}}$ = −38366 deg cm$^{-2}$/dmol). This spectrum is characteristic of the collagen triple helix. Monitoring the ellipticity at 225 nm as a function of temperature showed a cooperative thermal transition with a $T_{m}$ value of 29.1 °C (Fig. 2). As shown below, NMR studies support a trimeric nature throughout this peptide at low temperature, suggesting that this is a trimer-to-monomer transition (see Fig. 5).

To investigate the consequences of the interruption, comparison was made with a homologous peptide, with the Gly inserted to recreate the Gly-X-Y pattern: GPOGAAVMGPOGPO. The magnitude of the 225 nm maximum and the 198 nm minimum are both greater for the GAAGVM peptide (MRE$_{225\ \text{nm}}$ = 4950 deg cm$^{-2}$/dmol, MRE$_{198\ \text{nm}}$ = −4112 deg cm$^{-2}$/dmol) when compared with the GAAVM peptide. The thermal stability of GAAGVM is 39.7 °C, 11 °C higher than that of GAAVM, indicating that the absence of the Gly also caused a substantial decrease in stability (Fig. 2). The observed $T_{m}$ value of 39.7 °C is high and close to that predicted ($T_{pred}$ = 36 °C) on the basis of host-guest peptides using the collagen stability calculator (37). Temperature-induced refolding profiles were very similar for both GAAVM and GAAGVM peptides at concentration of 1 mg/ml (data not shown), which may reflect a similar nucleation step.

Differential scanning calorimetry indicates that the calorimetric enthalpy for GAAGVM is 354 kJ/mol in phosphate-buffered saline (pH 7), whereas the value for GAAVM is 188 kJ/mol (Fig. 3, Table 1), suggesting a substantial loss of hydrogen bonding. The decrease in thermal stability is again observed, although the DSC $T_m$ values are higher than the CD $T_m$ values because of the faster heating rate. The $\Delta H_{\text{cal}}$ values for both peptides are considerably higher at lower ionic strength (50 mM sodium phosphate), at both neutral and acidic pH (428 kJ/mol for GAAGVM; 259 kJ/mol for GAAVM). Such large ionic strength effects were not observed for other peptides (30).

Consequence of Increasing Imino Acid Content in the Break—The effect of an artificial break with 4 adjacent non-Gly residues was previously studied in an all Gly-Pro-Hyp environment (25). The stable (Pro-Hyp-Gly)$_4$ peptide was compared with (Pro-Hyp-Gly)$_4$-Pro-Hyp-(Pro-Hyp-Gly)$_4$ which is denoted here as the GPOPO peptide (previously designated as the Gly deletion peptide) (25). The peptide (Pro-Hyp-Gly)$_4$ has a $T_{m}$ of 60 °C with a calorimetric enthalpy of 390
\textbf{A Break in Triple Helix Structure}

kJ/mol (30). The GPOPO peptide does not form a stable triple helix at \( c = 1 \text{ mg/ml} \) (Table 2), in contrast to the GAAVM peptide containing a natural break (25).

Given the striking observation that a GPOPO sequence cannot be accommodated in a triple helix, whereas GAAVM is allowed, further peptides were designed to investigate the effect of imino acids within the interruption. The sequence GAA was replaced by GPO to form a GPOVM peptide, and the sequence VM was replaced by PO to form the GAAP peptide. Although the two resultant peptides precipitate at neutral pH, both GAAP and GPOVM peptides form stable triple helices at pH 2.2. The ellipticity of the GPOVM peptide at 225 nm is similar to that of the GAAVM peptide at both low and high temperatures (Fig. 4), but its stability is decreased (from \( T_m = 23.6 \text{ °C} \) to 18.5 °C) for unblockered forms of both peptides, pH 2.2 (Table 2). The GAAP peptide has a similar ellipticity value at low temperature, but at high temperatures, the unfolded state appears to have a higher ellipticity value than seen for GAAVM and GPOVM. The thermal stability of GAAP is about 12.8 °C, significantly lower than seen for GAAVM and GPOVM (Fig. 4). The calorimetric enthalpies of GPOVM and GAAP did not appear to differ significantly from that obtained for GAAVM, but the low stability of these peptides made it problematic to obtain accurate baseline values (Table 2).

The decrease in stability in going from a Gly-\( X \)-\( Y \) repeating sequence to a sequence containing a break was calculated. To estimate the loss of thermal stability resulting from the absence of a Gly residue in a break, the observed \( T_m \) for each peptide with a break was subtracted from the \( T_m \) value predicted for the same sequence with Gly as every 3rd residue using the collagen stability calculator (GAAVM versus GAAGVM; GPOVM versus GPOGVM; GAAP versus GAAPG, and GPOPO versus GPOPO). As seen in Table 2, the extent of triple helix destabilization depends strongly on the imino acid content of the break. The decrease in stability \( \Delta T_m \) values ranges from \( \sim 8 \text{ °C} \) for the GAAVM peptide, which has no imino acids in the break to \( \sim 30 \text{ °C} \) for the peptides with two imino acids in the break (GAAP and GPOVM) to more than 60 °C for the GPOPO peptide where all 4 residues in the break are imino acids.

\textbf{NMR Spectroscopy of GAAVM Peptide—}The NMR spectrum of peptide GAAVM contains the typical features previously found in triple helical peptides (Fig. 5A). The \( ^{15}\text{N} \)-labeled resonances of the GAAVM peptide each show distinct monomer and trimer peaks, which can be differentiated by their \( R_2 \) and nuclear Overhauser effect values as well as their chemical shifts. The appearance of three trimer peaks as well as a single monomer resonance for each labeled residue supports a trimer form of the peptide in equilibrium with monomer. Residues Gly-24 and Gly-7 show only a single trimer resonance due to their repetitive Gly-Pro-Hyp environment. The proton and nitrogen chemical shifts for the trimer resonances of Gly-7 and Gly-24 are similar to those reported previously for Gly-Pro-Hyp triplets in other 30-mer triple helical peptides (35). Residues Gly-13, Ala-14, Ala-15, and Val-16 each have three trimer resonances, indicating that the three chains differ in their environ-

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Peptide} & \textbf{Sequence} & \textbf{\( T_m \)} & \textbf{\( \Delta T_m \)} & \textbf{\( \Delta H_{cal} \)} \\
\hline
GAAVM & NH\(_2\)-\( (GPO)4 \) GAAVM\( (GPO)4 \)-COOH & 23.6 ± 0.5 & 7.7 & (204 ± 5) \\
GAAP & NH\(_2\)-\( (GPO)4 \) GAAP\( (GPO)4 \)-COOH & 12.8 ± 0.5 & 31.6\(^a\) & (180 ± 40)\(^e\) \\
GPOVM & NH\(_2\)-\( (GPO)4 \) GPOVM\( (GPO)4 \)-COOH & 18.5 ± 0.5 & 27.2\(^a\) & (225 ± 45)\(^e\) \\
GPOPO & NH\(_2\)-\( (GPO)4 \) GPOPO\( (GPO)4 \)-COOH & <0 & >60 & NA\(^f\) \\
\hline
\end{tabular}
\caption{Effect of imino acid content within a break on thermal stability and calorimetric enthalpy. All peptides were unblocked at their termini and were characterized in 50 mM sodium phosphate, pH 2.2, since there were solubility problems at neutral pH.}
\end{table}

\(^a\) \( T_m \) values were obtained by CD monitoring of the 225 nm ellipticity, at a concentration of 5 mg/ml.
\(^b\) \( \Delta T_m \) is the difference in \( T_m \) values for peptides with a glycine and without a glycine.
\(^c\) Calorimetric enthalpies were determined by DSC at concentrations of ~5 mg/ml (GAAVMG, GAAPG) or 2.5 mg/ml (GPOVM).
\(^d\) \( T_m \) values of GAAPG and GPOVM were estimated using the stability calculator (37).
\(^e\) Since the denaturation temperatures were so low, the DSC baselines for the native state could not be determined accurately, so the values shown are approximations.
\(^f\) See Ref. 25.

\(^a\) NA, not applicable.
shielding from solvent. The three resonances assigned to Val-16 residue, with their atypical chemical shifts, show a protection factor of \( \sim 102 \), indicative of significant shielding from solvent or hydrogen bonding but somewhat lower than the Gly-13 value. The 2 Ala residues in the break, Ala-14 and Ala-15, have protection factors in the range of 2–18, indicating that these 2 residues are fully exposed to solvent and are not hydrogen bonded.

Analysis of Residues within Breaks—The break studied in the \( \alpha_5(IV) \) chain of type IV collagen, GAAVMG, is one of a category of breaks that has 4 non-Gly residues between two glycines Gly-AA1-AA2-AA3-AA4-Gly and is denoted as a G4G break. This category of breaks could also be considered as having a Gly missing in the periodic Gly-X-Y-Gly-X-Y pattern or a Gly deletion type (24). In the six chains of the type IV collagen family, there are 28 G4G breaks: five in \( \alpha_1(IV) \); two in \( \alpha_2(IV) \); seven in \( \alpha_3(IV) \); five in \( \alpha_4(IV) \); seven in \( \alpha_5(IV) \); and two in \( \alpha_6(IV) \). To test whether residues within the G4G breaks differ significantly from the residues found within the \( X \) and \( Y \) positions of the standard Gly-X-Y segments of these chains, a statistical analysis was carried out. Comparison of hydrophobic (Ile, Leu, Met, Val), small (Ala, Cys, Gly, Ser), imino acids (Pro), charged (Asp, Glu, His, Lys, Arg), aromatic (Phe, Tyr, Trp), and polar (Asn, Gln,Thr) residues indicated that the residues in the breaks were significantly different from \( X \) and \( Y \) residues in \( (\text{Gly}-X-Y) \) regions \( (p < 0.001) \). The major contributors to these differences are the high number of small amino acids and of hydrophobic residues in the break; the proline content is also well below that seen in the \( X \) and \( Y \) positions (Fig. 6).

Examination of the 28 G4G breaks suggests that the AA2 position is frequently occupied by Ala, Ser, or Asp, whereas the AA3 position is predominantly (\( \sim 70\% \)) occupied by a hydrophobic residue. In addition, positions AA1, AA2, and AA3 are particularly low in Pro residues. In terms of composition and the identity of residues in the AA1 and AA3 positions, the GAAVMG break studied here represents a typical G4G break in type IV collagen.

**DISCUSSION**

The characterization reported here of a triple helical peptide incorporating a type IV collagen break provides an opportunity to investigate conformational, energetic, and dynamic features at and surrounding an interrup-
A Break in Triple Helix Structure

tion in the Gly-X-Y repeating sequence. The GAAVM sequence chosen is hydrophobic and is surrounded in the α5(IV) chain by highly stabilizing Gly-Pro-Hyp units. It is not sterically possible to continue the triple helix through this break, where Gly is not present as every 3rd residue. This sequence with a break is compared with the artificially created control peptide, GAAGVVM, which forms a stable triple helix, with a $T_m$ of 40°C. Enthalpy is clearly an important stabilizing factor for the GAAVM control triple helical peptide, which has a calorimetric enthalpy even greater than seen for (Pro-Hyp-Gly)$_n$ at low ionic strength. Previous work suggests that exposed hydrophobic Val and Met residues are likely to be coated with a backbone anchored hydration network, which would contribute to its high enthalpy value (38).

Evidence for perturbation of the triple helix at the GAAVM break includes loss of the 225 nm CD signal, lower stability, and loss of calorimetric enthalpy when compared with the GAAGVVM peptide. The loss of almost half the enthalpy, going from GAAGVVM to GAAVM, is striking and is consistent with an interruption in the hydration network now that the backbone is not in a regular triple helix. It is likely that there is now some hydrophobic burying of the Val side chain and/or hydrogen bonding of its amide group, consistent with the NMR data showing slower exchange rates that are slower than a random chain for the Gly and Val within the GAAVM sequence. The presence of multiple imino acids in the break is found to be unfavorable for a stable triple helix. The decrease in $T_m$ of the triple helix peptide caused by the introduction of a break is smallest for GAAVM, followed by GAAPO, GPOVM, and finally the highly destabilizing GPOPO, which cannot be incorporated into a triple helix. This contrast with the NMR stabilization of the triple helix itself by imino acids. Although the polyproline II-like conformation of the individual chains in the triple helix is highly favored by a high content of imino acids, proline residues appear to be too rigid to be accommodated within a break. Such rigidity might prevent hydrophobic interactions between Val side chains or some hydrogen-bonding interaction. Their destabilizing nature is consistent with the observation of low contents of Pro residues within G4G breaks.

The formation of some ordered structure within the break is also consistent with NMR studies showing exchange rates that are slower than a random chain for the Gly and Val within the GAAVM sequence.

The peptides studied here form homotrimers, whereas type IV collagens are present as heterotrimers consisting of two or three distinct types of chains (27). The locations of interruption sites in the Gly-X-Y sequence tend to be conserved among type IV chains, whereas little conservation is seen for the number of amino acids and their sequence within corresponding breaks of different chains (14). Further studies are needed to investigate whether the features observed for homotrimers also hold for heterotrimers.

REFERENCES

1. Ramachandran, G. N., and Kartha, G. (1955) Nature 176, 593–595.
2. Rich, A., and Crick, F. H. (1961) J. Mol. Biol. 3, 483–506.
3. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) Science 266, 75–81.
4. Brodsky, B., and Persikov, A. V. (2005) Adv. Protein Chem. 70, 301–339.
5. Hoppe, H. J., and Reid, K. B. (1994) Protein Sci. 3, 1143–1158.
6. Deprez, D., Doss-Pepe, E., Brodsky, B., and Inestrosa, N. C. (2000) Biochem. J. 350, 283–290.
7. Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukomski, S. (2002) J. Biol. Chem. 277, 27312–27328.
8. Boydston, J. A., Chen, P., Steichen, C. T., and Turnbough, C. L., Jr. (2005) J. Bacterial. 187, 5310–5317.
9. Kielsey, C. M., and Grant, M. E. (2002) in Connective Tissue and Its Heritable Disorders Molecular, Genetic, and Medical Aspect (Royce, P. M., and Steinmann, B., eds) pp 159–222, Wiley-Liss, Inc., New York.
10. Myllyharju, J., and Kivirikko, K. I. (2001) Ann. Med. 33, 7–21.
11. Myllyharju, J., and Kivirikko, K. I. (2004) Trends Genet. 20, 33–43.
12. Kuivaniemi, H., Tromp, G., and Prockop, D. J. (1991) FASEB J. 5, 2052–2060.
13. Byers, P. H., and Cole, W. G. (2002) in Connective Tissue and Its Heritable Disorders Molecular, Genetic, and Medical Aspect (Royce, P. M., and Steinmann, B., eds) pp 385–430, Wiley-Liss, Inc., New York.
14. Zhou, J., Hertz, J. M., Leitonen, A., and Tryggvason, K. (1992) J. Biol. Chem. 267, 12475–12481.
15. Hudson, B. G., Tryggvason, K., Sundaramoorthy, M., and Neilson, E. G. (2003) N. Engl. J. Med. 348, 2543–2556.
16. Brazel, D., Oberbaumer, I., Dieringer, H., Babel, W., Glanville, R. W., Deutzmann, R., and Kuhn, K. (1987) Eur. J. Biochem. 168, 529–536.
17. Yurchenco, P. D., and Ruben, G. C. (1988) Am. J. Pathol. 132, 278–291.
18. Kilschier, E., Hofmann, H., Steigemann, W., and Engel, J. (1985) J. Biol. Chem. 260, 403–415.
19. Brodsky-Doyle, B., Leonard, K. R., and Reid, K. B. (1976) Biochem. J. 159, 279–286.
20. Hofmann, U., Voss, T., Kuhn, K., and Engel, J. (1984) J. Biol. Chem. 259, 325–343.
21. Miles, A. J., Skubit, A. P., Furcht, L. T., and Fields, G. B. (1994) J. Biol. Chem. 269, 30939–30945.
22. Furcht, L. T., Skubit, A. P., and Fields, G. B. (1994) Lab. Investig. 70, 781–783.
23. Wallis, R., Shaw, J. M., Unlebach, J., Chen, C. B., Toegersen, D., and Drickamer, K. (2004) J. Biol. Chem. 279, 14065–14073.
24. Long, C. G., Thomas, M., and Brodsky, B. (1995) Biopolymers 35, 621–628.
25. Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J., and Brodsky, B. (1993) Biochemistry 32, 11688–11695.
26. Baum, J., and Brodsky, B. (1999) Curr. Opin. Struct. Biol. 9, 122–128.
27. Borza, D. B., Bondar, O., Ninomiya, Y., Sado, Y., Naito, I., Todd, P., and Hudson, B. G. (2001) J. Biol. Chem. 276, 28532–28540.
28. Hudson, B. G., Reeder, S. T., and Tryggvason, K. (1993) J. Biol. Chem. 268, 26033–26036.
29. Lemmink, H. H., Schroder, C. H., Mooms, L. A., and Smeets, H. J. (1997) Hum. Mutat. 9, 477–499.
30. Persikov, A. V., Xu, Y., and Brodsky, B. (2004) Protein Sci. 13, 893–902.
31. Johnson, B. A., and Blevins, R. A. (1994) J. Biomol. NMR 4, 603–614.
32. Glasoe, P. K., and Long, F. A. (1960) J. Phys. Chem. 64, 188–189.
33. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Proteins 17, 75–86.
34. Fan, P., Li, M. H., Brodsky, B., and Baum, J. (1993) Biochemistry 32, 13299–13309.
35. Buevich, A. V., Silva, T., Brodsky, B., and Baum, J. (2004) J. Biol. Chem. 279, 46890–46895.
36. Li, M. H., Fan, P., Brodsky, B., and Baum, J. (1993) Biochemistry 32, 7377–7387.
37. Persikov, A. V., Ramshaw, J. A., and Brodsky, B. (2005) J. Biol. Chem. 280, 19343–19349.
38. Kramer, R. Z., Bella, J., Brodsky, B., and Berman, H. M. (2001) J. Mol. Biol. 311, 131–147.