NMR Shows Hydrophobic Interactions Replace Glycine Packing in the Triple Helix at a Natural Break in the (Gly-X-Y)$_n$ Repeat*

Yingjie Li, Barbara Brodsky, and Jean Baum

From the Department of Chemistry and Chemical Biology, BIOMAPS Institute, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854 and the Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

Little is known about the structural consequences of the more than 20 breaks in the (Gly-X-Y)$_n$, repeating sequence found in the long triple helix domain of basement membrane type IV collagen. NMR triple resonance studies of doubly labeled residues within a set of collagen model peptides provide distance and dihedral angle restraints that allow determination of model structures of both a standard triple helix and of a triple helix with a break in solution. Although the standard triple helix cannot continue when Gly is not every third residue, the NMR data support rod-like molecules that have standard triple-helical structures on both sides of a well defined and highly localized perturbation. The GAAVM break region may be described as a “pseudo triple helix,” because it preserves the standard one-residue stagger of the triple helix but introduces hydrophobic interactions at the position normally occupied by the much smaller and hydrogen-bonded Gly residue of the repeating (Gly-X-Y)$_n$ sequence. This structure provides a rationale for the consensus presence of hydrophobic residues in breaks of similar length and defines a novel variant of a triple helix that could be involved in recognition.

Although the collagen triple helix can be considered as one of the most well defined protein motifs, there are still surprises and undefined molecular features. All non-fibrillar collagens contain sites where the repeating Gly-X-Y pattern is interrupted, and evidence suggests that these breaks are of functional importance, playing a role in molecular or higher order structure, or serving as recognition sites for interactions (1, 2). The structural consequences of a break for the triple helix are not understood.

The unique triple helix protein motif found in collagen has three extended polyproline II-like chains supercoiled about a common axis. The typical triple helix is characterized by a (Gly-X-Y)$_n$ repeating sequence. The three chains are staggered by one residue, and Gly residues must be present as every third residue in each chain so that the three chains can pack very tightly, burying the Gly residues and forming Gly N-H...C=O (X) hydrogen bonds (3–5). The abundant fibrillar collagens, e.g. types I, II, and III, maintain this precise Gly-X-Y repeat throughout their ~1000-residue triple helix domain, and the replacement of even one Gly by another residue results in a pathological condition (1). But many non-fibrillar collagens have been identified, and these all contain one or more breaks in the repeating Gly-X-Y pattern (1). The non-fibrillar type IV collagen is found in basement membranes of all multicellular animals, and its network architecture contributes to the essential function of basement membranes in providing mechanical support to cells, serving as semi-permeable barriers between tissues, and providing signals for differentiation (6–9). In type IV collagen, there are more than 20 breaks in the ~1350-residue-long triple helix of the form Gly-X-Y-(AA)$_n$-Gly-X-Y, and the lengths of the breaks range from common short breaks where $n = 1$, to long breaks where $n = 26$ (10, 11). To begin to define the effects that small breaks have on triple helix structure, NMR studies are presented here on a model peptide of a type IV collagen break where $n = 4$.

Peptide models can provide an approach to characterizing the effects of breaks on triple helix structure and stability (12, 13). Our laboratories recently reported that a 30-mer homotrimeric peptide model for a natural break found in the α5(IV) chain of type IV collagen, GPOGAAVMYPGPGPO (residues 386–399, where O is used as a single letter code for hydroxyproline (Hyp)) forms a stable trimeric structure, with a relatively small decrease in stability compared with the triple helical structure of a homologous peptide with Gly as every third residue, GPOGAGVMPGPO (13). The classic triple helix structure cannot continue when Gly is not present as every third residue, and biophysical studies indicate that the break decreases the triple helix content, destabilizes the triple helix by 10 °C, and reduces the enthalpy substantially. Although NMR experiments reveal a non-random conformation within the break, it was not clear whether there is a way to continue a modified triple helix through this GAAVM region, or whether a new local structure, such as a β-bend, is introduced.

Here NMR triple resonance experiments are presented on a set of doubly labeled peptides with and without the GAAVM

---

*This work was supported by National Institutes of Health Grants GM45302 (to J. B.) and GM60048 (to B. 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 To whom correspondence should be addressed: Dept. of Chemistry and Chemical Biology, BIOMAPS Institute, Rutgers, the State University of New Jersey, 610 Taylor Rd., Piscataway, NJ 08854. Tel.: 732-445-5666; Fax: 732-445-5312; E-mail: jean.baum@rutgers.edu.

2 The abbreviations used are: Hyp, hydroxyproline (three-letter code); O, hydroxyproline (single-letter code); HSQC, heteronuclear single quantum coherence; GFT, G-matrix Fourier Transform; TOCSY, total correlation spectroscopy; PPII, polyproline II; NOE, nuclear Overhauser effect.
NMR Studies on a Collagen Break

break (Table 1), allowing the assignment and tracking of individual chains, the identification of intra- and intermolecular NOEs, and the determination of $\phi$ dihedral angles through $^{3}J_{\text{HNHa}}$ coupling constants. These NMR parameters are used as restraints in energy minimization of molecular models derived from high resolution x-ray structures of several homologous peptides, to allow the first visualization of the solution molecular conformation of a standard triple helix and of a triple helix with a naturally occurring type IV collagen break. The break in the type IV collagen sequence leads to a highly localized distortion of the triple helix. The GAAVM region has a well defined conformation, which maintains the one-residue stagger but introduces hydrophobic packing of three Val residues at the center of the molecule, which would be normally occupied by Gly residues.

**EXPERIMENTAL PROCEDURES**

**Sample Preparation**—Peptides Ac-(GPO)$_4$GAAGVM(GPO)$_4$GY-CONH$_2$, designated as the GAAGVM peptide, and Ac-(GPO)$_4$GAAVM(GPO)$_4$GY-CONH$_2$, designated as the GAAVM peptide, were synthesized by Tufts University Core Facility (Boston, MA), as previously described (13). The peptides were made with selectively $^{13}$C/$^{15}$N doubly labeled residues (underlined below): residues Gly$^{13}$, Ala$^{14}$, Ala$^{15}$, Gly$^{16}$, Val$^{17}$, and Gly$^{25}$ were labeled in peptide GAAGVM; residues Gly$^{13}$, Ala$^{14}$, Ala$^{15}$, Val$^{16}$, and Gly$^{24}$ were labeled in the peptide GAAVM. Peptides were purified using a Waters XTerra Prep C18 column on an Amersham Biosciences fast-protein liquid chromatography system, and the identity of the peptides was confirmed by matrix-assisted laser desorption ionization mass spectrometry. Samples for both peptides were prepared in 10% D$_2$O/90% H$_2$O at pH 2 with concentrations of 3.0 mM for GAAGVM and 1.8 mM for GAAVM.

**NMR Experiments**—NMR experiments were performed on a Varian Inova 600-MHz spectrometer equipped with a cryo-probe, and the G-matrix Fourier Transform (GFT) (5,3)D HACACONHN/HACA, CONHN experiments (14, 15) were recorded on a Bruker 900 spectrometer equipped with a cryo-probe at the New York Structural Biology Center to improve dispersion and sensitivity. For sequential assignments, $^{1}$H–$^{15}$N heteronuclear single quantum coherence (HSQC) and a series of triple resonance experiments were obtained at 25 °C, including HNCO (16)/(HN(CA)CO (17, 18), HNCA (16), H(CA)NH/ HA(CACO)NH (19), HNCACB (20)/(CBA(CO)NH (21), and GFT (5,3)D HACACONHN/HACA, CONHN (14, 15). Three-dimensional $^{15}$N edited TOCSY-HSQC (22, 23) with a mixing time of 45 ms and three-dimensional $^{15}$N edited NOESY-HSQC (22–24) with mixing times of 30–80 ms were performed at different temperatures from 15–25 °C. Short mixing times (30 ms) in NOESY-HSQC were used to eliminate spin diffusion, and the data at various temperatures were used to help resolve overlapped resonances. Three-dimensional HNHA experiments (25) were performed to measure homonuclear $^{3}J_{\text{HNHa}}$ coupling constants at 25 °C, with an H–H coupling period of 25 ms. The correction factor for the $^{3}J_{\text{HNHa}}$ coupling constants was obtained by performing T$_{zz}$ measurements (26) of amide protons on a $^{15}$N-labeled GAAVM sample at 25 °C.

All data were processed using the FELIX 2004 software package (MSI, San Diego, CA) and/or NMRPipe (39) and analyzed with FELIX 2004 or NMRView (27). In the acquisition dimensions for all experiments, a solvent suppression filter was applied to the data prior to apodization with a 90° sine-bell window function. The data were subsequently zero-filled to 1024 complex points and Fourier transformed. The $t_1$ and $t_2$ dimensions in all three-dimensional experiments were increased 1.5 times by forward-backward linear prediction (28), multiplied by a sine-bell function, zero-filled to 256 complex points, and Fourier transformed. The final three-dimensional data for each experiment included 256 × 256 × 512 real points.

In the $^{3}J_{\text{HNHa}}$ experiments, the cross peak to diagonal peak volume ratios were taken to calculate the apparent coupling constants (25). Experimental error was calculated based on the experimental uncertainty in volume measurement. The experimental uncertainty was estimated as the standard deviation of the volume integration in the regions free of signals. The average value and the measurement error of the apparent $^{3}J_{\text{HNHa}}$ coupling constants was then calculated based on the maximum and minimum J coupling constants from the maximum and minimum volume ratio of cross/diagonal peaks with experimental uncertainty included. From the separate T$_{zz}$ measurement (26) an average $T_{zz}$ value of 84 ms was found for the amide protons of the $^{15}$N-labeled GAAVM peptide. Equation 1, below (29), was solved to determine accurate values of $^{3}J_{\text{HNHa}}$, and the correction factor of 1.16 was determined. The corrected J coupling constants of doubly labeled GAAGVM and GAAVM peptides were obtained by multiplying the apparent coupling by 1.16.

$$\frac{S_{\text{cross}}}{S_{\text{diagonal}}} = \left[\frac{\tan [(1 - \chi)^{1/2} \zeta]}{(1 - \chi)^{1/2} + \chi \tan [(1 - \chi)^{1/2} \zeta]}\right]^2$$

where $\chi = 1/(2\pi T_{\text{zz}}^{\text{sel}})^{3}J_{\text{HNHa}}$, $\zeta = \pi \sigma^{3}J_{\text{HNHa}}$, and $\delta$ is the H–H coupling period.

**Generation of NOE Contact Map and Molecular Modeling**—A computer model structure of GAAGVM was obtained based on the crystal structure of T3–785 (PDB ID: 1BKV) (30) and built using the Molecular Operating Environment 2005.06 (Chemical Computing Group Inc., Montreal, Canada). The GIT-GAR-GLA residues in T3–785 were replaced with the residues GPO-GAA-GVM. The structure was solvated with a standard MOE Water Soak procedure and energy-minimized from residue 6 to residue 21 using the Amber99 all-atom force field (31) to a room mean square gradient of 0.05. This structure was used to generate the background of the NOE map of a standard triple helix motif. Hydrogen atoms were added to the model structure of GAAGVM using the REDUCE program (32). The predicted background map was generated by calculating NH–H distances equal to or smaller than 5 Å and classifying these as NH-NH, NH–H$^\alpha$, and NH-side chains (H$^\beta$, H$^\gamma$, and H$^\delta$). NOE contact maps for peptides GAAGVM and GAAVM were made from observed NH–H NOEs in the three-dimensional $^{1}$H–$^{15}$N NOESY-HSQC experiment and classified as NH-NH, NH–H$^\alpha$, and NH side chains (H$^\beta$, H$^\gamma$, and H$^\delta$).
A refined model of GAAGVM was generated by implementing the $\phi$ angle restraints from the $^3J_{HNHa}$ values with varying restraint ranges to the energy minimization. Back calculation of $^3J_{HNHa}$ values and back calculation of an NOE map were used to select the best fit model with the experimental data. A model of GAAVM was generated based on the x-ray crystal structure of Hyp (PDB ID: 1EI8) (33). Residues GPO-GP were substituted to GAA-VM. The resulting model was energy-minimized with $^3J_{HNHa}$ coupling restraints similarly to the GAAGVM peptide minimization. Energy minimization is very sensitive to the initial starting structure and to the ranges of the restraints; therefore, a number of different input structures with varying restraint ranges were used. Back calculation of $^3J_{HNHa}$ values and NOEs indicated that most structures were not consistent with all of the experimental data. A representative structure that was consistent with experimental $^3J_{HNHa}$ values and all $^1$H-$^1$H NOEs was selected for GAAGVM and GAAVM.

RESULTS AND DISCUSSION

NMR Chain Assignments and Chemical Shift Differences between GAAVM and GAAGVM Peptides—For both peptides, all trimer resonances could be assigned to specific chains of the triple helix as indicated in the heteronuclear single quantum coherence (HSQC) spectrum by the superscripted number (Fig. 1). The sequential assignment is derived from the triple resonance experiments, and the chain stagger is derived from NOE experiments that define interchain interactions.

To obtain the sequence-specific assignments, a series of triple resonance experiments were performed on the control GAAGVM peptide, where each residue in the GAAGV segment of the peptide is double labeled (Table 1). For the GAAGVM peptide, traditional triple resonance experiments, including HNCO (16)/HN(CA)CO (17, 18), HNCA (16), and HNCACB (20)/CBCA(CO)NH (21) experiments were performed (Table 2). Unambiguous correlations were obtained for residues Gly$^{13}$-Ala$^{14}$, Ala$^{15}$-Gly$^{16}$, and Gly$^{16}$-Val$^{17}$. The resonances arising from the Ala$^{14}$ residue were highly overlapped in the proton and carbon dimensions, which is likely to be a result of the similarity in environment of the two Ala residues in the GAAG sequence. To resolve the three Ala$^{14}$-Ala$^{15}$ connectivities, an additional (5,3)D HACACONHN/HACA, CO NHN GFT experiment (14, 15) was used and successfully sequentially correlated the chemical shifts of the $^1$C$^\alpha$-$^1$C$^\alpha$-$^1$H moieties of residue $i$ and the NH group of residue $i$ for Ala$^{14}$-Ala$^{15}$. For the GAAVM peptide complete sequential assignments were obtained for the doubly labeled GAAV segment in the GAAVM peptide from a single pair of HA(CA)NH/HA(CACO)NH (19) experiments (Table 2B), because there was no chemical shift overlap in the NH or H$^\alpha$ dimensions. Assignments were confirmed by the HNCA (16) experiment (Table 2B).

NOE experiments were used to obtain interchain distances, which led to the identification of the three chains in terms of their 1-residue stagger. For example, NOEs observed between chain 3$^1$Gly$^{13}$NH and chain 1$^1$Ala$^{15}$ H$^\alpha$ in GAAGVM and between $^1$A$^{14}$NH and $^2$G$^{13}$H$^\alpha$ in GAAVM indicated the relationship between chains and allowed identification of the leading (chain 1), trailing chains (chain 3), and middle chains (chain 2) of the peptides. Chain-specific assignments of peaks set the stage for using distance and dihedral angle constraints in structure determination.

The HSQC spectrum of the GAAGVM peptide gives a typical

![FIGURE 1. $^1$H-$^1$5N HSQC spectra of peptides GAAGVM and GAAVM at 20 °C. A, peptide GAAGVM with a standard Gly-X-Y pattern throughout. B, peptide GAAVM containing a type IV collagen break. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T, respectively. Assignment of each resonance to chains 1, 2, or 3 is indicated as a number preceding the superscript T.](image-url)

| Peptide sequences of GAAGVM and GAAVM |
|----------------------------------------|
| Peptide name | Peptide sequence$^a$ |
| GAAGVM | Ac-GPO-GPO-GPO-GPO-GAAGVM-GPO-GPO-GPO-GPO-GY-CONH$_2$ |
| GAAVM | Ac-GPO-GPO-GPO-GPO-GAAVM-GPO-GPO-GPO-GPO-GY-CONH$_2$ |

$^a$ $^{15}$N- and $^{13}$C-labeled residues are underlined and bolded.
pattern expected for the triple helix. Each labeled residue shows at least one monomer peak and one or more trimer peaks, consistent with the presence of the triple helix structure. Gly25 shows only a single trimer resonance due to the repetitive Gly-Pro-Hyp environment (13), whereas residues Gly13, Ala14, Ala15, Gly16, and Val17 all have multiple trimer peaks (two peaks for Gly16 and three peaks for all the other residues) due to the non-repeating sequence environment (13, 34, 35). Ala14 and Ala15 have similar chemical shifts to each other, which may reflect their similar environments in the GAAG sequence. The HSQC spectrum of the GAAVM peptide again shows three trimer peaks as well as monomer peaks for each labeled residue. This indicates that a well defined structure with three non-equivalent chains is present in the region where the Gly-X-Y pattern is broken. Gly24, within the Gly-Pro-Hyp repeating region at the C terminus, shows only a single trimer resonance at the typical triple helix position similar to Gly25 above in GAAGVM, whereas three trimer peaks are seen for residues Gly13, Ala14, Ala15, and Val16. Comparison of the chemical shifts of the trimer resonances shows changes in chemical shift for most resonances in GAAVM compared with GAAGVM indicating that there is a difference in the local electronic environment of the central region. Most notably, the Val resonances in the GAAVM peptide have shifted downfield very significantly in both the \( ^1H \) and \( ^15N \) dimensions with \( ^3\text{Val} \) shifting the most from 7.8 down to 9.4 ppm. The large downfield shift may reflect a change in the hydrogen bonding behavior of Val in GAAVM relative to GAAGVM.

**Prediction of Short Range Distances in a Classic Triple Helical Conformation from the GAAGVM Model Structure**—A predicted contact map in a classic triple helical conformation is generated from a GAAGVM model structure. The high resolution structures of a number of collagen-like peptides have been solved by x-ray crystallography (30, 36–38), and these structures have confirmed and provided molecular coordinates for the triple-helical structure. A model structure of the GAAGVM peptide was generated by molecular modeling using the crystal structure of the peptide T3–785 ((POG)3-ITGARGLAG-(POG)4 sequence) (30), because it contains a central sequence with no imino acids surrounded by Gly-Pro-Hyp (GPO) triplets. The central GIT-GAR-GLA residues were replaced with the residues GPO-GAA-GVM, and the structure was energy-minimized (see “Experimental Procedures” for details). The cross-section of this minimized GAAGVM model structure shows the NH groups of the Gly residues are located near the center of the triple helix and form hydrogen bonds with the CO of the X residue of a neighboring chain (Fig. 2). The NH vectors of the Y residues point outward toward the solvent, while the NH vectors of the X residues point more tangentially toward a neighboring chain within the molecule making a water-bridged hydrogen bond with the C=O of the Gly of the neighboring chain.

A predicted contact map is generated from the computer model structure for the GAAGVM peptide and allows calculation of distances that are within 5 Å for the backbone NH to backbone NH, H|Xa|H and side-chain protons in this structure (Fig. 3A; shaded squares indicate that there is one or more predicted contact within 5 Å, and circles indicate the individual predicted NOE). A number of salient features can be seen from the contact map and include: (a) Intrachain sequential distances (indicated in gray highlighted boxes) are not symmetric. The NH atoms of residue \( i \) (indicated by \( ^1\text{NH}(i) \), where \( j \) refers to the
AUGUST 3, 2007• VOLUME 282 • NUMBER 31

JOURNAL OF BIOLOGICAL CHEMISTRY

NMR Studies on a Collagen Break

FIGURE 2. Representation of the peptide sequence GAAGVM and the cross-section view of the one residue stagger. A, sequence diagram of peptide GAAGVM shows the characteristic 1-residue stagger. The corresponding isotope-labeled residues are red. B, cross-section view from the model structure of GAAGVM showing the GAA segment. GAA triplets are red, blue, and green for Gly13, Ala14, and Ala15, respectively. NH vectors of GAA are shown in ball-and-stick model and labeled. One example of the Rich and Crick II interchain hydrogen bonding is shown between the Gly13 NH of chain 3 and the Ala14 CO of chain 1.
Hz, shifted from the 5- to 7-Hz values seen in GAAGVM. There are two solutions to the Karplus equation for values of $\gamma$ that include $\gamma$ angles of $100^\circ$ and $140^\circ$, respectively. The comparison of the values of the coupling constants between GAAGVM and GAAVM indicate that the distortion of the $\gamma$ angles is very local and that Ala15 and Val16 adopt conformational angles outside the PPII region.

Molecular Modeling of the GAAGVM and GAAVM Peptides—The $\gamma$ angle restraints from $^3J_{HNHa}$ measurements were used to refine the model of GAAGVM derived from an x-ray crystal structure as described above. The refined model did not differ significantly from the original model, indicating that the original model derived from x-ray structure is in good agreement with NMR (Fig. 5, A and B). In the GAAGVM model, all $\phi$, $\psi$ angles fall into the PPII region of the Ramachandran plot, consistent with a well-behaved triple helix (Fig. 5C).

The agreement between the model obtained from an x-ray crystal structure and the NMR solution data for the GAAGVM peptide established a strategy to use NMR data to define the conformation in solution for a GAAVM break. For this peptide, a model was generated based on a recent x-ray structure with a break in the Gly-$X$-$Y$ repeating sequence (33). Our peptide differs from the peptide solved by x-ray in having a 4-residue break compared with a 1-residue break and in having a natural type IV collagen sequence compared with the repeating (POG)$n$ sequence. The central GPOGP sequence of the crystal structure peptide (POG)$2$-POGPOGPG-(POG)$5$ was replaced by GAAVM, and the structure was energy-minimized incorporating $\gamma$ angle restraints from experimental GAAVM NMR $^3J_{HNHa}$ that one can see that they come from both intra- and intermolecular contacts in a complex manner because of the supercoiled trimeric structure of the triple helix. B, experimental contact map of peptide GAAGVM (represented by circles), overlaid on the shaded contacts of the GAAGVM model, indicating that some of the predicted interchain as well as intrachain NOEs are observed. The fact that many predicted NOEs are not observed is largely attributable to overlap. Half circles represent overlapped NOEs consistent with the GAAGVM model. C, experimental contact map for GAAVM. The observed GAAVM NOEs (circles) are overlaid on the background of the predicted intra- and intermolecular contacts in the GAAGVM model, showing the expected intrachain and interchain NOEs and some new interchain NOEs consistent with the packing of the Val residues at the break and the 1-residue stagger between chains. In C, thick half circles and thinner half circles represent overlapping NOEs.
couplings. The structure obtained gives good agreement with NOE NMR data.

The principal features of the model preserves the 1-residue stagger between the 3 chains and introduces close packing of the 3 chains and differs only in the dihedral angles of a small number of residues. The absence of one residue in GAAVM versus GAAGVM creates a loss of axial register of the triple helix on both sides of the break, as seen in the crystal structure for peptides with breaks (5, 33).

The Ramachandran plot of the GAAVM model indicates a local disruption of the PPII dihedral angles relative to the model of GAAGVM. The Val16 dihedral angles are the most perturbed with all three chains falling outside of the PPII range (Fig. 5F). In addition, the dihedral angles of 2Ala15 and 3Ala15 (just N-terminal to Val16), 2Ala14, 1Met17, and Gly18 all are outside the PPII region. The highly localized nature of the perturbation is recognized by the observation that only 8 residues in the 3 chains are altered from standard collagen dihedral angles, suggesting that the well-defined deformation in structure is primarily absorbed by a subset of residues in the break so that one can optimize the extent of normal triple helix structure on both sides of the break.

A number of lines of argument support this GAAVM model obtained by molecular modeling with NMR J-coupling restrictions. Back calculation of the $^{3}$JHNHa and NMR distances from the model shown in Fig. 5D and comparison to the experimentally observed $^{3}$JHNHa and distances indicate that the proposed model has a good fit to the experimental data (data not shown). Other models tested did not show good agreement with the experimental data. Therefore, although the

FIGURE 5. Model structures of peptide GAAGVM with a standard Gly-X-Y pattern and peptide GAAVM containing a type IV collagen break. A and D, ribbon diagram of GAAGVM (A) and GAAVM (D) represented with the N-terminal at the top and chain numbers. Leading, middle, and trailing chains are represented as 1, 2, and 3. For ease of comparison both peptides are shown with comparable lengths by selecting residues 4–30 for GAAGVM and residues 1–26 for GAAVM. The GAAGVMG segment in peptide GAAGVM and the GAAVMG segment in peptide GAAVM are green and red, respectively. Enlarged views of those two segments are shown on the right of their ribbon diagram. For GAAGVM, residues are color-coded as Gly13 (orange), Ala14 (green), Ala15 (dark blue), Gly16 (yellow), Val17 (red), Met18 (light blue), and Gly19 (orange). For GAAVM, residues are color-coded as Gly13 (orange), Ala14 (green), Ala15 (dark blue), Val16 (red), Met17 (light blue), and Gly18 (orange). B and E, space filling model of the cross-section view from the N terminus to the C terminus of GAAGVM (B) and GAAVM (E). Views show that the GAAVM region adopts a “pseudo-triple helix,” which preserves the 1-residue stagger between the 3 chains and introduces close packing and hydrophobic interactions of the three Val residues at the site of the break. Residues are color-coded similarly to the enlarged ribbon diagrams. The AGV segment in GAAGVM is colored as Ala15 (dark blue), Gly16 (yellow), and Val17 (red), and the AVM segment in GAAVM is colored as Ala13 (dark blue), Val16 (red), and Met17 (light blue). C and F, Ramachandran plots for model structures of GAAGVM (C) and GAAVM (F). To highlight the central region of the peptides, only residues 6–21 in GAAGVM and residues 6–17 in GAAVM are plotted and shown in black triangles. Ramachandran contour map for Pro residues is shown in the background, and typical secondary structures are indicated ($\alpha$, $\alpha$-helix; $\beta$, $\beta$-sheet; and PPII, polyproline II and collagen).
GAAVM model is not unique, it provides a representative example of the structure that can be adopted by the peptide containing the break that fits the experimental data, and, as described below, it explains the observed consensus sequence of 4-residue breaks. The NMR data on the GAAVM peptide in solution is consistent with the rod-like nature of the triple helices and the highly localized nature of the structural perturbation of previous x-ray structures on two peptides with small breaks (5, 33). The increased conformational flexibility at the break site reported for the crystal structure is consistent with the faster hydrogen exchange rates at Gly13 in GAAGVM versus GAAVM (Ref. 13 and data not shown).

**Biological Implications**—The GAAVM break studied here represents one of the most common types of break in type IV collagen and other non-fibrillar collagens, with 4 non-Gly residues in a row, Gly-X-Y-Gly-AAA/AAA-AAA-Gly-X-Y. The consensus pattern for such breaks (denoted as G4G) includes a small residue at the AA2 position and a hydrophobic residue at the AA3 position (13). The GAAVM break contains the consensus sequence, with an Ala residue at the AA2 position and a Val residue at the AA3 position. The model structure derived from NMR data suggests that the placement of a hydrophobic residue at the AA3 position of the break promotes novel hydrophobic packing near the central axis of the superhelix, whereas the presence of a smaller AA2 residue is required just before the hydrophobic residue to permit the distortion required for this hydrophobic packing. The presence of a small hydrophobic core formed by the Val residues at position AA3 suggests that hydrophobic stabilization can partly compensate for the Gly packing normally found at the same position and helps explain why this and similar breaks can be present in non-fibrillar collagens with melting temperatures similar to those of fibrillar collagens.

The structural consequences of the GAAVM break serve as a prototype for a typical break of this kind in non-fibrillar collagens, which are homotrimers, such as type VII collagen in anchoring fibrils and type X collagen in hypertrophic cartilage (1). The G4G breaks in these homotrimer collagens are predicted to resemble the locally perturbed triple helix structure reported here for the model peptides. Other non-fibrillar collagens, such as type IV collagen in basement membranes, are heterotrimeric, consisting of two or three distinct chain types. For type IV collagen molecules, a G4G break in one chain is sometimes opposite a G4G break in the other chains, giving a structure similar to that seen in the model peptide. However, in most cases the G4G break in one type IV collagen chain is opposite another kind of break or an uninterrupted sequence in the other chains, and it is not clear whether such “mixed breaks” will have larger or different effects on the triple helix structure.

**Acknowledgments**—We thank Jianxi Xiao for preliminary NMR experiments on the GAAVM peptide and Dr. Seho Kim for help with implementation of triple resonance experiments. We thank the New York Structural Biology Center for access to their 900-MHz NMR and for help in implementing the G-matrix Fourier Transform NMR experiments.

**REFERENCES**

1. Kielty, C. M., and Grant, M. E. (2002) Connective Tissue and Its Heritable Disorders, Molecular, Genetic and Medical Aspects, 2nd Ed., pp. 159–222, Wiley Liss, New York
2. Miles, A. J., Knutson, J. R., Skubitz, A. P., Furcht, L. T., McCarthy, J. B., and Fields, G. B. (1995) *J. Biol. Chem.* 270, 29047–29050
3. Rich, A., and Crick, F. H. (1961) *J. Mol. Biol.* 3, 483–506
4. Ramachandran, G. N. (1967) in Treatise on Collagen (Ramachandran, G. N., ed) pp. 103–184, Academic Press, New York
5. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) *Science* 266, 75–81
6. Schmitt, J. C., and Yurchenco, P. D. (1989) *Curr. Opin. Cell Biol.* 1, 983–988
7. Hudson, B. G., Reeder, S. T., and Tryggyvason, K. (1993) *J. Biol. Chem.* 268, 26033–26036
8. Hudson, B. G., Tryggyvason, K., Sundaramoorthy, M., and Neilson, E. G. (2003) *N. Engl. J. Med.* 348, 2543–2556
9. Yurchenco, P. D., and Ruben, G. C. (1988) *Ann. J. Pathol.* 132, 278–291
10. Zhou, J., Hertz, J. M., Leinonen, A., and Tryggyvason, K. (1992) *J. Biol. Chem.* 267, 12475–12481
11. Brazel, D., Oberbaumer, L., Dieringer, H., Babel, W., Glanville, R. W., Deutzmann, R., and Kuhn, K. (1987) *Eur. J. Biochem.* 168, 529–536
12. Baum, J., and Brodsky, B. (1999) *Curr. Opin. Struct. Biol.* 9, 122–128
13. Mohs, A., Popiel, M., Li, Y., Baum, J., and Brodsky, B. (2006) *J. Biol. Chem.* 281, 17197–17202
14. Kim, S., and Szymerski, T. (2003) *J. Am. Chem. Soc.* 125, 1385–1393
15. Kim, S., and Szymerski, T. (2004) *J. Biol. Chem.* 279, 117–130
16. Ikura, M., Kay, L. E., and Bax, A. (1990) *Biochemistry* 29, 4659–4667
17. Chubb, R. T., Thanabal, V., and Wagner, G. (1992) *J. Magn. Reson.* 97, 213–217
18. Namazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., and Kay, L. E. (1994) *J. Am. Chem. Soc.* 116, 11655–11666
19. Kay, L. E., Ikura, M., and Bax, A. (1991) *J. Magn. Reson.* 91, 84–92
20. Witterkind, M., and Mueller, L. (1993) *J. Magn. Reson. B* 101, 201–205
21. Grzesieks, S., and Bax, A. (1992) *J. Am. Chem. Soc.* 114, 6291–6293
22. Fesik, S. W., and Zuiderweg, E. R. (1988) *J. Magn. Reson.* 78, 588–593
23. Messers, B. A., Wider, G., Otting, G., Weber, C., and Wittrich, K. (1989) *J. Magn. Reson.* 85, 608–613
24. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc.* 111, 1515–1517
25. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc.* 115, 7772–7777
26. Kay, L. E., Nicholson, L. K., Delagio, F., Bax, A., and Torchia, D. A. (1992) *J. Magn. Reson.* 97, 359–375
27. Johnson, B. A., and Blevins, R. A. (1994) *J. Biol. Chem.* 269, 603–614
28. Zhu, G., and Bax, A. (1992) *J. Magn. Reson.* 98, 192–199
29. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principle and Practice*, pp. 518–528, Academic Press, San Diego
30. Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (1999) *Nat. Struct. Biol.* 6, 454–457
31. Ponder, J. W., and Case, D. A. (2003) *Adv. Protein Chem.* 66, 27–85
32. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) *J. Mol. Biol.* 285, 1735–1747
33. Bella, J., Liu, J., Kramer, R., Brodsky, B., and Berman, H. M. (2000) *J. Mol. Biol.* 302, 298–311
34. Li, M. H., Fan, P., Brodsky, B., and Baum, J. (1993) *Biochemistry* 32, 7377–7387
35. Liu, X. Y., Kim, S., Dai, Q. H., Brodsky, B., and Baum, J. (1998) *Biochemistry* 37, 15528–15533
36. Kramer, R. Z., Vitagliano, L., Bella, J., Berisio, R., Mazzarella, L., Brodsky, B., Zagari, A., and Berman, H. M. (1998) *J. Mol. Biol.* 280, 623–638
37. Nagarajan, V., Kamitori, S., and Okuyama, K. (1999) *J. Biochem. (Tokyo)* 125, 310–318
38. Kramer, R. Z., Venugopal, M. G., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (2000) *J. Mol. Biol.* 301, 1191–1205
39. Delagio, F., Grzesieks, S., Vuister, G. W., Zou, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* 6, 277–293