Articles

A Host−Guest Set of Triple-Helical Peptides: Stability of Gly-X-Y Triplets Containing Common Nonpolar Residues†

Naina K. Shah,‡ John A. M. Ramshaw,§ Alan Kirkpatrick,§ Chirag Shah,‡ and Barbara Brodsky*,‡

Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, 675 Hoes Lane, Piscataway, New Jersey 08854-5636, and Division of Biomolecular Engineering, CSIRO, 343 Royal Parade, Parkville, 3052 Australia

Received January 9, 1996; Revised Manuscript Received May 3, 1996⊗

ABSTRACT: Host−guest peptide sets have been useful in evaluating the propensity of different amino acids to adopt an α-helical or β-sheet form, and this concept is applied here to the triple-helical conformation. A set of host−guest peptides of the form acetyl-(Gly-Pro-Hyp)$_3$-Gly-X-Y-(Gly-Pro-Hyp)$_4$-Gly-Gly$CONH_2$ was designed to evaluate the contribution of an isolated Gly-X-Y triplet to triple-helix stability in a defined environment. Peptides were synthesized to include guest triplets with the X and Y positions occupied by the most common nonpolar residues found in collagen: Pro (X position) and Hyp (Y position); Ala; Leu, the most frequent hydrophobic residue; and Phe, the only commonly occurring aromatic residue. The guest triplets of the 12 peptides synthesized represent 35% of the sequence found in the α1 chain of type I collagen. All peptides formed stable triple-helical structures, and the peptides showed a range of thermal stabilities (T_m = 21−44 °C), depending on the identity of the guest triplet. Thermodynamic calculations indicate these peptides have a range of free energy values ($\Delta\Delta G$ = 9 kcal/mol) and suggest that favorable entropy is the dominant factor in increased stability. Replacement of Ala by Leu in the X position did not affect the thermal stability, while an Ala to Leu change in the Y position was destabilizing. These data provide experimental evidence that hydrophobic residues do not stabilize the triple helical conformation. Although Leu and Phe are found almost exclusively in the X position in collagens, peptides with Leu and Phe in the Y position formed stable triple-helices. This supports the hypothesis that the X positional preference of these residues relates to their increased potential for intermolecular hydrophobic interactions rather than their destabilization of the triple-helical molecule. These studies establish the utility of host−guest peptides in defining a scale of triple-helix propensities and in clarifying the interactions stabilizing the triple-helical conformation.

Early studies indicated that collagen has a unique triple-helical structure, distinct from the folds found in globular proteins and other fibrous proteins. The collagen family now includes at least 19 distinct types of extracellular matrix molecules, each containing a large triple-helix domain (Kielty et al., 1993; Prockop & Kivirikko, 1995; Brown & Timpl, 1995). The most prevalent collagens are the fibril-forming collagens (types I, II, III, V, XI), found in D periodic fibrils in a variety of tissues (Kadler, 1994). For instance, type I collagen is found in skin, tendon, and bone, and type III collagen is found in a variety of extensible tissues. The triple helix has also been found as a domain in an increasing number of nonmatrix proteins, such as C1q, mannose binding protein, and the macrophage scavenger receptor (Hoppe & Reid, 1994; Brodsky & Shah, 1995).

The conformation of a triple helix consists of three polypeptide chains, each in a polyproline II-like helix, which are supercoiled about each other and linked together by hydrogen bonds (Ramachandran & Kartha, 1955; Rich & Crick, 1961; Fraser & MacRae, 1973). The close packing of the three chains requires that every third residue be Gly, generating a (Gly-X-Y)$_n$ repeating sequence pattern. A high content of imino acids is also required to stabilize this structure. Proline residues are commonly found in the X position of Gly-X-Y triplets in collagen. When Pro residues are incorporated in the Y position, they become posttranslationally modified to Hyp, making Gly-Pro-Hyp the most common triplet found in collagens. The X and Y residues of collagen are both accessible to solvent, with the X position more exposed than the Y position (Jones & Miller, 1991). Interchain hydrogen bonding and an extensive hydration network have been shown to be essential for stabilizing the triple-helical structure (Bella et al., 1994, 1995).

In theory there are more than 400 different possible Gly-X-Y triplets (>20 possible residues in the X position and in

† This work has been supported by NIH Grant AR 19626 (B.B.), an NSF U.S.−Australia International Cooperative Research Grant (B.B.), and the Australia/USA Bilateral Science Program (J.A.M.R.).

* To whom correspondence should be sent. Phone: (908) 235-4048. Fax: (908) 235-4783. E-mail: brodsky@rwja.umdnj.edu.

‡ University of Medicine and Dentistry of New Jersey.

§ CSIRO.

⊗ Abstract published in *Advance ACS Abstracts,* July 1, 1996.

[1] Abbreviations: CD, circular dichroism; T_m, melting temperature. Standard one-letter and three-letter abbreviations are used for the common amino acids, with hydroxyproline denoted by O (one-letter code) and Hyp (three-letter code).

S0006-2960(96)00046-3 CCC: $12.00

the Y position), but only a limited number of triplets are actually found in collagen sequences (Dölz & Heidemann, 1986). For instance, in the $(Gly\text{-}X\text{-}Y)_{338}$ sequence of the αl chain of type I collagen, only 22 triplets are found four or more times. Of these, the most frequent triplets are Gly-Pro-Hyp (12%), Gly-Pro-Ala (9%), and Gly-Ala-Hyp (6%). Some residues are asymmetrically distributed with respect to the X and Y positions. For instance, Leu and Phe are found almost exclusively in the X position (Fietzek & Kühn, 1975; Salem & Traub, 1975). Modeling studies suggest potential interactions and steric hindrance as the explanation for these positional preferences (Salem & Traub, 1975; Bansal, 1977; Bansal & Ramachandran, 1978), but direct evidence and quantitation of such interactions are lacking.

One approach to understanding the basic contributions of amino acids to triple-helix stability is to establish a scale of triple-helix propensity. For α-helices, the experimental determination of a scale of helical propensity for individual residues has been achieved using synthetic peptides, by substituting "guest" residues in a constant "host" sequence (Wojcik et al., 1990; O'Neil & DeGrado, 1990). Studies establishing a thermodynamic scale for the tendency of different amino acids to form *β*-sheets have also been reported recently, through the introduction of different guest residues into a protein host framework (Smith et al., 1994). In applying an analogous strategy to the triple helix, it is useful to consider the Gly-X-Y triplet as the basic unit, rather than individual residues.

Experimental and empirical approaches have been tried to define the relative stability of Gly-X-Y triplets. Comparisons of the triple-helical stabilities of different repeating polytripeptides, polyhexapeptides, or block polytripeptides indicated that different triplet sequences can affect stability (Segal, 1969; Doyle et al., 1971, Heidemann & Roth, 1982). Heidemann and colleagues (Heidemann & Roth, 1982; Thakur et al., 1986; Germann & Heidemann, 1988) designed a set of peptides consisting of a constant cross-linked core with varying triplets on the N-terminus. Quantitation of the stabilizing contribution of a single Gly-X-Y unit in these and other previous studies (Heidemann & Roth, 1982) was limited by the small differences in thermal stability between different peptides. It was also complicated by the inclusion of different neighboring sequences, variations in peptide length, and effects of charges at the N- or C-terminus. An empirical approach assumed Gly-X-Y sequences found in clusters with frequently occurring triplets, such as Gly-Pro-Hyp, have a high stability, while sequences found with less frequent triplets have a low stability (Thakur et al., 1986). On the basis of this cluster analysis and other data, a stability scale was proposed, classifying each Gly-X-Y triplet into one of four categories, and this scale was used to predict regions of greater and lower stability in collagen chains (Bächinger & Davis, 1991; Bächinger et al., 1993). This classification illustrates the utility of having such a scale for understanding the effects of mutations and protein design but lacks a firm experimental basis.

Here we present a design for a "host−guest" set of triple-helical peptides in which the effect of a given Gly-X-Y triplet can be evaluated in a defined and stabilizing triple-helical environment. This peptide set presents an opportunity for establishing a scale of triple-helix propensity as well as for clarifying the interactions stabilizing the triple helix. Studies are presented here for host−guest triple-helical peptides containing common nonpolar Gly-X-Y triplets. These results indicate a lack of hydrophobic stabilization for the triple helix and provide experimental information on the positional preferences of hydrophobic and aromatic residues.

MATERIALS AND METHODS

Peptide Synthesis. Peptides were synthesized on an Applied Biosystems 430A synthesizer using the standard FastMoc (Applied Biosystems) method on Fmoc-RINK resin. Side-chain protection was *tert*-butyl for the hydroxyproline residues. Acetylation was by acetic anhydride and triethylamine in dimethylformamide. Purity of peptides was greater than 95% when assessed by amino acid analysis, using a Waters HPLC system with ninhydrin detection, and by reverse-phase HPLC on a Vydac C-18 column. The identity of the peptide with the guest triplet GLA was also confirmed by liquid secondary positive-ion mass spectrometry using a VG Analytical ZAB-T instrument.

Circular Dichroism. CD spectra were recorded on an Aviv Model 62DS spectrometer, using a Hewlett-Packard Peltier thermoelectric temperature controller. Peptide solutions of concentrations 1 mg/mL were used, with peptides dried *in vacuo* over P_2O_5 for 48 h prior to weighing. Peptide solutions in either 0.1 M acetic acid, pH 2.7, or 0.15 M NaCl and 10 mM sodium phosphate, pH 7.4 (PBS), were equilibrated at 4 °C for more than 48 h prior to analysis. For wavelength scans, the signal was collected from 210 to 260 nm at 1 nm intervals, for 3 s at each point, at 2 and 60 °C. For equilibrium melting temperature (T_m) transitions, the ellipticity at 225 nm was monitored while the sample temperature was increased from 2 to 80 °C at intervals of 0.3 °C. The equilibration time at each point was 2 min, and data were collected for 10 s at each point.

More accurate peptide concentrations were determined by amino acid analysis (Waters HPLC system with ninhydrin detection) on solutions used for CD measurements. Analyses on eight host−guest peptides gave concentration values lower than obtained by weighing and were used in calculation of mean residue ellipticities.

Calculation of Thermodynamic Parameters. The equilibrium melting curves were analyzed using a two-state model, where three unfolded chains combine to form a triple helix:

$$3M \leftrightarrow T$$

The equilibrium constant is

$$K = \frac{[T]}{[M]^3} = \frac{F}{3c^2(1-F)^3}$$

where c is the monomeric peptide concentration in moles and F is the fraction folded.

The fraction folded (F) was calculated using the equation

$$F = \frac{\theta_{observed} - \theta_{monomer}}{\theta_{trimer} - \theta_{monomer}}$$

where $\theta_{observed}$ is the observed mean residue ellipticity, θ_{trimer} is the ellipticity when the peptide is fully associated, and $\theta_{monomer}$ is the ellipticity of the monomer. The monomer ellipticity exhibits a strong temperature dependence (Engel et al., 1977). The value for the ellipticity of the monomer

at low temperature was determined by extrapolating the slope of the monomer ellipticity at high temperature to lower temperatures in the transition region (Engel et al., 1977; Long et al., 1993). Linear extrapolation of the premelting baseline was performed to the get the θ_{trimer} ellipticity. The temperature where the peptide was 50% folded ($F = 0.5$) was taken as the melting temperature (T_m).

The equilibrium melting transitions were fit to a two-state trimer to monomer transition, and the van't Hoff enthalpy ΔH was determined by curve fitting with the equation (Engel et al., 1977; Marky & Breslauer, 1987)

$$K = \exp\left[\frac{\Delta H}{RT}\left(\frac{T}{T_m} - 1\right) - \ln(0.75c^2)\right]$$

The equation

$$T_m = \frac{\Delta H}{\Delta S + R \ln(0.75c^2)}$$

was used to calculate ΔS. The following equation was used to calculate ΔG:

$$\Delta G = \Delta H - T\Delta S$$

The ΔG values were calculated at the average melting temperatures of all peptides ($T = 32.2$ °C) to minimize the extrapolation of the temperature dependence of ΔH and ΔS. The precise concentrations used for calculating the thermodynamic parameters were taken from the amino acid analysis (GPO, GAO, GLO, GFO, GPL, GAL, GPF, and GFA) or from weighing (GPA, GAA, GLA, and GAF). The differences in concentration values obtained from amino acid analysis versus weighing had no effect on the enthalpy values and little effect on ΔS (<1%) and ΔG (<3%) determination.

Molecular Modeling. Molecular modeling was performed using the SYBYL 6.1 (Tripos Inc., St. Louis, MO) molecular modeling package. Coordinates for the $(GPO)_8$ host molecule were those of Fraser et al. (1979). After substitution of guest residues, the structure was refined by energy minimization using the SYBYL package.

RESULTS

Peptide Design. All peptides studied were the same length and had the following sequence, acetyl-$(GPO)_3$GXY-$(GPO)_4$GG$CONH_2$, where O is 4-hydroxyproline and GXY is the guest triplet. The residues in the X and Y positions were selected to study how various triplets affect the thermal stability of the triple helix. To assure formation of a stable triple helix, even when a particularly unfavorable guest triplet was present, the guest GXY triplet was flanked on both sides by stabilizing GPO triplets. The peptide is designed to be short enough so that the effect of a single guest triplet would not be masked by the constant part of the structure. The selection of a host peptide length of eight triplets was based on thermal stability measurements of $(Pro\text{-}Hyp\text{-}Gly)_n$ peptides of varying lengths, since $n = 10$ results in a very stable peptide ($T_m = 61$ °C) (Sakakibara et al., 1973), while $n = 5$ results in an unstable peptide ($T_m < 2$ °C) (Li, 1993). The N-terminus was acetylated and the C-terminus amidated to eliminate the charge repulsion between the three chains at both ends of the helix, which was previously seen to lower the overall thermal stability (Berg et al., 1970; Venugopal et al., 1994). A Gly-Gly sequence was included at the C-terminus to eliminate the likelihood of diketopiperazine formation during synthesis.

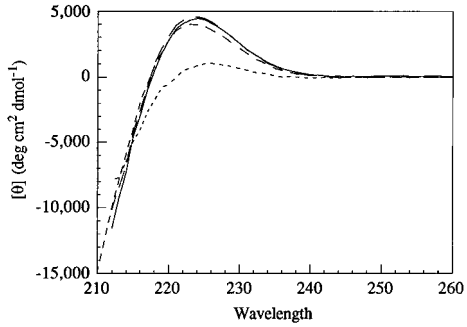

FIGURE 1: CD spectra of three host−guest peptides, GPO (—), GAO (— —), and GLO (— — —), at 10 °C, showing the characteristic collagen triple-helix maximum near 225 nm in the native state. The decrease in intensity of the peak upon denaturation is seen for GPO at 60 °C (- - -).

Selection of Guest Triplets. Initially, guest triplets were selected on the basis of their frequency in collagen. The imino acids Pro and Hyp constitute about 20% of all residues and were included in X and Y positions, respectively. Alanine, which is the residue found most frequently after Gly, Pro, and Hyp, was incorporated in both the X and Y positions. This results in the most common triplets of type I collagen, Gly-Pro-Hyp, Gly-Pro-Ala, and Gly-Ala-Hyp, together with the less frequent Gly-Ala-Ala. Leucine is the most frequently occurring hydrophobic residue and is found almost exclusively in the X position. The only aromatic residue commonly found in collagen is Phe, which shows an almost complete restriction to the X position and is often found in Gly-Phe-Hyp triplets. Placement of Leu and Phe in the X position leads to triplets typically found in collagen, Gly-Leu-Hyp, Gly-Leu-Ala, Gly-Phe-Hyp, and Gly-Phe-Ala, while their placement in the Y position leads to the rarely or never observed triplets Gly-Pro-Leu, Gly-Ala-Leu, Gly-Pro-Phe, and Gly-Ala-Phe.

Peptide Conformation and Stability. Circular dichroism studies on the 12 host−guest peptides indicate that they all form stable triple-helical structures. At low temperature, the peptides show the maximum near 223−225 nm which is characteristic of a collagen triple helix, and the peak intensity decreases at higher temperatures (Figure 1). The magnitude of the mean residue ellipticity at the maximum shows a small variation, depending on the guest triplet (4100 ± 250 deg cm^2 dmol^{-1}) for non-phenylalanine containing peptides. The phenylalanine containing peptides consistently show a higher mean residue ellipticity at 225 nm (4700 ± 200 deg cm^2 dmol^{-1}), which is likely due to a phenylalanine signal in this region (Manning & Woody, 1989).

Monitoring the change in the 223−225 nm maximum as a function of temperature results in a sharp transition (Figure 2), which previous studies indicate corresponds to the trimer to monomer transition (Long et al., 1993; Li et al., 1993). The temperature at which this transition occurs depends upon the guest triplet included, with a temperature range of about 20.7−44.3 °C in phosphate-buffered saline (PBS) (Table 1). Melting temperatures obtained at pH 2.7 were all within 0.1−1.7 °C of the values obtained at pH 7.4 (Table 1). The lack of any significant effect of pH is consistent with the presence of blocked end groups and the lack of ionizable groups in these triplets.

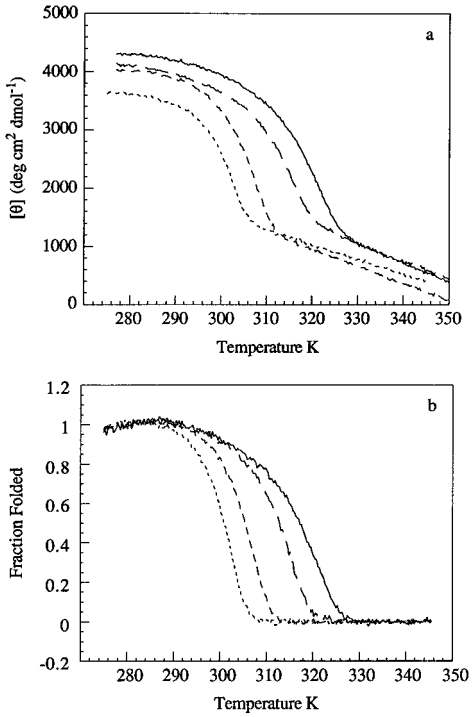

FIGURE 2: Thermal equilibrium curves for four of the host−guest peptides, GPO (—), GAO (— —), GPL (— — —), and GAL (- - -), showing their sharp transitions and the variation in thermal stability. Panels: (a) mean residue ellipiticity [θ] vs temperature; (b) fraction folded vs temperature.

Table 1: Guest Triplets Included in the 12 Host−Guest Peptides[a]

X	Y: O	Y: A	Y: L	Y: F
P	GPO 44.5 (46.0)	GPA 38.3 (39.2)	GPL 32.7 (31.7)	GPF 28.3 (27.5)
A	GAO 39.9 (41.0)	GAA 29.3 (30.0)	GAL 27.8 (27.4)	GAF 20.7 (20.6)
L	GLO 39.0 (39.5)	GLA 29.5 (30.0)		
F	GFO 33.5 (34.5)	GFA 23.4 (23.0)		

[a] The guest triplets included in the 12 host−guest peptides are indicated in a grid format, with X = P, A, L, F (vertical) and Y = O, A, L, F (horizontal). Immediately under each guest triplet is the melting temperature (°C) of that peptide in PBS, pH 7.4, as determined by fraction folded $F = 0.5$. Beneath and in parentheses is the melting temperature determined at pH 2.7.

The host peptide, where GXY = GPO, is the most stable ($T_m = 44.5$ °C). The change of a single residue, Pro to Ala, in the X position to GAO results in a decrease in stability ($T_m = 39.9$ °C), while an Hyp to Ala change in the Y position leads to a greater decrease ($T_m = 38.3$ °C). Substitution of Ala in both X and Y positions gives GAA, with a melting temperature of $T_m = 29.3$ °C. These results are consistent with the previously established importance of imino acids in stability and confirm the greater stabilizing effect of Hyp compared with Pro (Josse & Harrington, 1964; Burjanadze, 1979).

Introduction of Leu as a replacement for Ala in the X position results in little change in thermal stability, while its occurrence in the Y position results in a marked destabilization. Thus, GLA ($T_m = 29.5$ °C) has the same melting temperature as GAA ($T_m = 29.3$ °C), while GLO ($T_m = 39.0$ °C) is similar to GAO ($T_m = 39.9$ °C). The presence of a Leu in the Y position is destabilizing compared with Ala, and the degree of destabilization depends on which X position residue is adjacent to the Leu. Thus, GPL is 5.6 °C less stable than GPA, while GAL is 1.5 °C less stable than GAA. This is consistent with modeling studies predicting a greater steric hindrance if Leu in the Y position is adjacent to an imino acid in the X position (Bansal & Ramachandran, 1978).

Table 2: Thermodynamic Parameters ΔH (kcal/mol), ΔS [kcal/(mol·K)], $T\Delta S$ (kcal/mol), and ΔG (kcal/mol), Together with the Observed Melting Temperatures (T_m) in PBS, pH 7.4, for the Host−Guest Peptide Set[a]

triplet	T_m (°C)	ΔH	ΔS	$-T\Delta S$	ΔG	$\Delta\Delta G$[b]
GPO	44.5	−84.0	−0.23	71.1	−12.82	0.00
GAO	39.9	−101.2	−0.29	89.3	−12.20	−0.62
GLO	39.0	−104.5	−0.30	92.5	−12.03	−0.79
GPA	38.3	−85.5	−0.24	74.3	−11.20	−1.62
GFO	33.5	−122.8	−0.37	112.5	−10.26	−2.56
GPL	32.7	−122.9	−0.37	113.2	−10.01	−2.81
GLA	29.5	−114.4	−0.35	105.9	−8.53	−4.29
GAA	29.3	−108.4	−0.33	99.9	−8.49	−4.33
GPF	28.3	−133.1	−0.41	125.0	−8.10	−4.72
GAL	27.8	−137.1	−0.42	129.2	−7.90	−4.92
GFA	23.4	−141.7	−0.45	136.3	−5.82	−7.00
GAF	20.7	−152.2	−0.49	148.6	−3.61	−9.21

[a] ΔH was calculated from the equilibrium melting transition using the van't Hoff equation (see Materials and Methods). It is assumed that ΔH is independent of temperature in the transition region. Using this value of ΔH, ΔS was calculated at T_m. The average of the melting temperatures for all peptides ($T = 305.2$ K) was used for the calculation of $T\Delta S$ and ΔG. [b] $\Delta\Delta G$ (kcal/mol) values are given here for each peptide relative to the GPO peptide.

The presence of Phe in the X or Y position results in greater destabilization than seen for Leu. With Phe in the X position, the thermal stability of GFO ($T_m = 33.5$ °C) is less than that of GLO ($T_m = 39.0$ °C) and GAO ($T_m = 39.9$ °C), while GFA ($T_m = 23.4$ °C) is less stable than GLA ($T_m = 29.5$ °C) and GAA ($T_m = 29.3$ °C). In the Y position, there is even greater destabilization, with GPF ($T_m = 28.3$ °C) much less stable than GPA ($T_m = 38.3$ °C) and with GAF ($T_m = 20.7$ °C) much less stable than GAA ($T_m = 29.3$ °C). For the case of Phe in the Y position, the degree of destabilization does not appear to depend on the identity of the residue in the X position.

Thermodynamic Analysis. Thermodynamic parameters were calculated for the host−guest peptides (Table 2). The van't Hoff enthalpy was determined from the thermal equilibrium curves and used to calculate the entropy and free energy, assuming a two-state monomer to trimer transition. The appropriateness of a two-state model is supported by the observation of a monomer−trimer equilibrium in equilibrium ultracentrifugation studies of a peptide of similar design (Long et al., 1993; Venugopal et al., 1994), by the excellent fit of the equilibrium melting curves to the model, and by the observation of only monomer and trimer peaks in the NMR spectrum of host−guest peptides (Siegel, Brodsky, and Baum, unpublished results). In addition, Engel et al. (1977) found close agreement between the van't Hoff enthalpy and the calorimetrically determined enthalpy for $(Pro\text{-}Pro\text{-}Gly)_{10}$, a triple-helical peptide similar in length and composition to the peptides studied here.

The ranking of the relative free energy values calculated for the 12 peptides is consistent with the order of their observed thermal stabilities (Table 2). Such self-consistency supports the relevance of the thermodynamic parameters and their derivation. The establishment of a stability scale and calculation of $\Delta\Delta G$ values require that all peptides in the set adopt the same triple-helical conformation, which is supported by their characteristic triple-helical CD spectra. In addition, the sequence-independent nature of the triple-helix conformation is indicated by NMR studies (Li et al., 1993).

As previously reported for collagen, the enthalpy values for all of the triple-helical peptides are very favorable compared to that seen for other proteins, emphasizing the importance of hydrogen bonding (Privalov, 1982). Comparison of thermodynamic parameters indicates those peptides with the most favorable entropy have the lowest free energy and greatest stability. The substitution of Pro by Ala, Leu, or Phe in the X position of GPO decreases the stability as a result of a less favorable entropy term, even though there is a more favorable enthalpy term. Similar trends of entropic stabilization are seen among the four GPY peptides (GPO, GPA, GPL, and GPF) and for the five GXY peptides with no imino acid in the guest triplets. Favorable entropic effects are expected for the sterically constrained imino acids (Josse & Harrington, 1964).

DISCUSSION

The host−guest set of triple-helical peptides described here allows the stabilities and interactions of defined Gly-X-Y triplets to be compared. The peptides are all stable at reasonable temperatures, even when the guest triplet is an unfavorable one. Individual Gly-X-Y triplets clearly show different propensities for triple-helix formation, with a range of 24 °C seen for the thermal stability and a range of 9 kcal/mol for the free energy change. The large variations due to a change in one or two residues are advantageous features in the design and length of the host−guest peptides. A longer peptide might not have been so sensitive to changes in sequence, but a shorter peptide, or a peptide without the capping Gly-Pro-Hyp units, would be unlikely to form stable triple helices. Given the destabilizing effect of triplets such as GPF and GPL, it is not surprising that previously studied peptides $(Pro\text{-}Leu\text{-}Gly)_n$ or $(Pro\text{-}Phe\text{-}Gly)_n$ did not form stable triple helices in aqueous solution, although some structure can be induced in helix-promoting solvents (Scatturin et al., 1975; Tamburro et al., 1977; Brahmachari et al., 1978). The triple helices formed by the host−guest peptides are all homotrimers, formed of three identical peptide chains, and are most useful as models for homotrimeric molecules, such as type II and III collagens and the macrophage scavenger receptor.

The relative stabilities of the host−guest peptide set confirm previous results, showing the high stability of Gly-Pro-Hyp, Gly-Ala-Hyp, and Gly-Pro-Ala triplets (Segal, 1969; Heidemann & Roth, 1982). The greater stabilizing effect of hydroxyproline in the Y position compared with proline in the X position can also be seen (Josse & Harrington, 1964; Burjanadze, 1979; Privalov, 1982). However, variations in stability are seen among peptides with the same proline and/or hydroxyproline content in the guest triplet. For instance, GPA is 10 °C more stable than GPF.

Table 3: Number of Occurrences of Ala, Leu, and Phe in both the X and Y Positions of the αl Chains of Type I, III, and IV Collagens[a]

	Ala		Leu		Phe	
collagen	X	Y	X	Y	X	Y
type I	60	59	18	1	12	0
type III	47	42	19	2	8	0
type IV	17	17	60	2	31	1

[a] The preference of Leu and Phe for the X position can be seen, while Ala is evenly distributed.

In addition, the five peptides with guest triplets containing no imino acids show a 9 °C variation in T_m. The host−guest peptide results are in agreement with the most stable triplet units in the scale proposed by Bächinger and Davis (1991) but show the need for a more continuous measure of triple-helix propensity.

Hydrophobic interactions are not a stabilizing factor in the host−guest triple-helical peptide set. Collagens contain a variable amount of hydrophobic residues (Phe, Ile, Leu, Met, Val, Trp, Tyr) depending on the collagen type, ranging from 6% in types I and III to 14% in type IV. The large majority (>90%) of hydrophobic residues are found in the X position of Gly-X-Y triplets (Table 3). Previously, thermodynamic and accessibility studies have indicated that hydrophobic interactions are unlikely to be important for the collagen triple-helical conformation (Privalov, 1982; Jones & Miller, 1991). The data presented here are the first direct experimental demonstration that hydrophobic residues play no role in stability. The presence of Leu in the X position of GLA and GLO triplets confers no stability compared to the presence of Ala in that position (GAA, GAO). This suggests that hydrophobic residues in the X position of collagen are not interacting between the three chains of the triple helix and are not contributing to triple-helix stability. Leucine residues in the X position are highly exposed (about 70% accessible to solvent; Jones & Miller, 1991) and are likely to be involved in favorable hydrophobic interactions with neighboring molecules (Bansal & Ramamchandran, 1978).

Hydrophobic residues in the Y position are seen to be destabilizing to the triple helix, indicating a lack of hydrophobic interactions when they are in this position as well. This destabilization has an entropic basis, even though there is a large favorable increase in enthalpy. Modeling studies and calculations suggested unfavorable steric interactions when Phe and Leu are in the Y position rather than the X position (Salem & Traub, 1975; Bansal, 1977; Bansal & Ramachandran, 1978). A decrease in the availability of ϕ, ψ conformational space could also make the Y position less favorable as a result of side-chain entropy factors (Bansal, 1977; Cabrol et al., 1981). The Phe residue in Gly-Pro-Phe can be seen to be located in close proximity to a neighboring chain in the same molecule. In contrast, the Phe in Gly-Phe-Hyp is located on the outside of the helix and has no close contacts within the triple helix (Figure 3). The Phe in the Y position also interferes with the formation of water bridges between adjacent chains (Bansal, 1977).

This work demonstrates experimentally that triplets with Leu and Phe in the Y position can be present in stable triple helices even though they are rarely found in collagens (Table 3). For instance, the thermal stability of the GAL guest

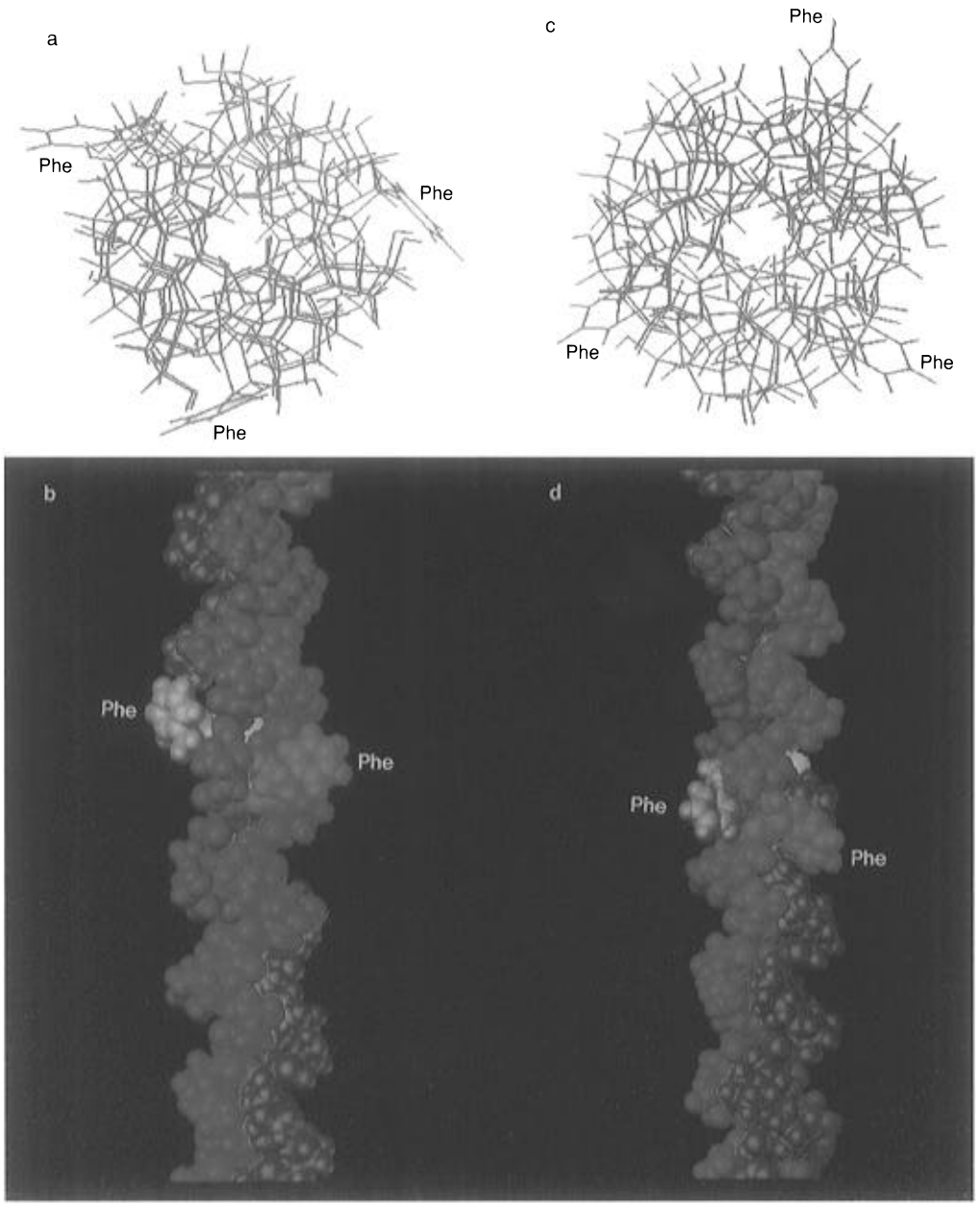

FIGURE 3: Molecular models of the Gly-Phe-Hyp and Gly-Pro-Phe host−guest peptides. The top and side views of Gly-Phe-Hyp are shown in (a) and (b), respectively, while the top and side views of Gly-Pro-Phe are shown in (c) and (d). The three chains of the triple helix are shown in red, green, and purple. In the top view the phenylalanines are marked with Phe. In the side views, the phenylalanines are shown in orange (red chain), cyan blue (green chain), and magenta (purple chain). The side views (b and d) give the appearance of being at different levels because the Phe residue in Gly-Phe-Hyp is one residue closer to the N-terminus than the Phe in Gly-Pro-Phe. This results in the observed 2.9 Å, 108° difference in position and orientation of the Phe groups.

triplet (27.8 °C), which is never found in the α1 chains of type I or III collagens, is only slightly less than the thermal stability of the GAA guest triplet (T_m = 29.3 °C), which constitutes 1−2% of all triplets (five to nine occurrences per chain). These results suggest stability is not the only factor determining the high preference of Leu and Phe residues for the X position in native triple-helical molecules.

The placement of Phe and Leu in the X position is favorable for intermolecular hydrophobic interactions (Bansal, 1977; Bansal & Ramachandran, 1978), and Phe residues in type I collagen are localized in certain regions that make them candidates for interacting and conferring rigidity to collagen fibrils (Fraser & Trus, 1986; Fraser et al., 1987). The X position has a greater solvent accessibility than the Y position, and our molecular modeling studies show that when phenylalanine occurs in the X position (Figure 3a,b), the side chain is oriented very favorably for interhelix hydrophobic interaction, with the ring face almost parallel to the axis of the triple helix. In the Y position, however, the phenylalanine is partially buried in the triple helix (Figure 3c), in close proximity to the neighboring chain (Figure 3d). The Phe in the Y position points out away from the center of the helix in an orientation unsuitable for intermolecular interaction. Modeling studies of Bansal (1977) indicate the feasibility of water-mediated interchain hydrogen bonds when Phe and Leu occur in the X position but not in the Y position. Thus, intermolecular hydrophobic interactions, along with water-mediated hydrogen bonding, may explain the positional preference of Phe and Leu for the X position in collagens.

Thermodynamic parameters derived from the equilibrium melting curves give a consistent set of values and provide a basis for calculation of $\Delta\Delta G$ values for a scale of triple-

helix potential. The host−guest triplet set described here establishes that the concept of a propensity scale, as established for α-helices (O'Neil & DeGrado, 1990) and *β*-sheets (Smith et al., 1994), is applicable to triple helices and that it is feasible to establish $\Delta\Delta G$ values from the thermodynamic parameters derived from equilibrium thermal transitions. The range of $\Delta\Delta G$ values is larger for the tripeptide set studied here than seen for variations in individual residues in their potential to form α-helices (O'Neil & DeGrado, 1990) and *β*-sheets (Smith et al., 1994). The interactions stabilizing the triple helix do not include hydrophobic interactions, which are important in stabilizing the α-helix and *β*-sheet. The entropy of side chains may be a determining factor in all of these regular structures, while the highly ordered water network seems to be unique to the triple helix.

The large number of possible Gly-X-Y sequences presents an obstacle to establishing a comprehensive scale of all triplets, but the relatively small number of tripeptides found with a high frequency makes it possible to address important aspects of triple-helix stability using a limited number of triplets. The triplets included as guests in this initial series cover almost 35% of the sequence of the α1 chain of type I collagen and 31% of the α1 chain of type III collagen, indicating it should be possible to determine the relative stabilities of a large percentage of common collagen sequences. The triple-helix potential is being measured in an atypical and stable environment consisting of all Gly-Pro-Hyp triplets, but at a future time, it is possible to incorporate two adjacent guest tripeptides Gly-X_1-Y_1-Gly-X_2-Y_2 in this design, to evaluate the effects of neighboring triplets. The extension of host−guest peptides to include polar and ionizable triplets will make it possible to predict stability along a collagen chain and to evaluate the environments of different collagen mutations. The beginning of the establishment of a propensity scale for triple helices, using these host−guest peptides, broadens our base of understanding of protein structure.

ACKNOWLEDGMENT

We thank Dr. Jean Baum and Peter Privalov for helpful discussions, Yanbo Ding for help with SYBYL, and Rachel Kramer for helping us with graphics.

REFERENCES

Bächinger, H. P., & Davis, J. M. (1991) *Int. J. Biol. Macromol. 13*, 152−156.

Bächinger, H. P., Morris, N. P., & Davis, J. M. (1993) *Am. J. Med. Genet. 45*, 152−162.

Bansal, M. (1977) *Int. J. Pept. Protein Res. 9*, 224−234.

Bansal, M., & Ramachandran, G. N. (1978) *Int. J. Pept. Protein Res. 11*, 73−81.

Bella, J., Eaton, M., Brodsky, B., & Berman, H. M. (1994) *Science 266*, 75−81.

Bella, J., Brodsky, B., & Berman, H. M. (1995) *Structure 3*, 893−906.

Berg, R. A., Olsen, B. R., & Prockop, D. J. (1970) *J. Biol. Chem. 245*, 5759−5763.

Brahmachari, S. K., Ananthanarayanan, V. S., Rapaka, R. S., & Bhatnagar, R. S. (1978) *Biopolymers 17*, 2097−2105.

Brodsky, B., & Shah, N. K. (1995) *FASEB J. 9*, 1537−1546.

Brown, J. C., & Timpl, R. (1995) *Int. Arch. Allergy Immunol. 107*, 484−490.

Burjanadze, T. V. (1979) *Biopolymers 18*, 931−938.

Cabrol, D., Broch, H., & Vasilescu, D. (1981) *Biopolymers 20*, 269−282.

Dölz, R., & Heidemann, E. (1986) *Biopolymers 25*, 1069−1080.

Doyle, B. B., Traub, W., Lorenzi, G. P., & Blout, E. R. (1971) *Biochemistry 10*, 3052−3060.

Engel, J., Chen, H. T., Prockop, D. J., & Klump, H. (1977) *Biopolymers 16*, 601−622.

Fietzek, P. P., & Kühn, K. (1975) *Mol. Cell. Biochem. 8*, 141−157.

Fraser, R. D. B., & MacRae, T. P. (1973) *Conformation in Fibrous Proteins*, Academic Press, New York.

Fraser, R. D. B., & Trus, B. L. (1986) *Biosci. Rep. 6*, 221−226.

Fraser, R. D. B., MacRae, T. P., & Suzuki, E. (1979) *J. Mol. Biol. 129*, 463−481.

Fraser, R. D. B., MacRae, T. P., & Miller, A. (1987) *J. Mol. Biol. 193*, 115−125.

Germann, H. P., & Heidemann, E. (1988) *Biopolymers 27*, 157−163.

Heidemann, E., & Roth, W. (1982) *Adv. Polym. Sci. 43*, 143−203.

Hoppe, H.-J., & Reid, K. B. M. (1994) *Protein Sci. 3*, 1143−1158.

Jones, E. Y., & Miller, A. (1991) *J. Mol. Biol. 218*, 209−219.

Josse, J., & Harrington, W. F. (1964) *J. Mol. Biol. 9*, 269−287.

Kadler, K. (1994) *Protein Profile 1*, 519−638.

Kielty, C. M., Hopkinson, I., & Grant, M. E. (1993) in *Connective Tissue and its Heritable Disorders* (Royce, P. M., & Steinmann, B., Eds.) pp 103−148, Wiley and Sons Inc., New York.

Li, M. H. (1993) Ph.D. Thesis, Rutgers University.

Li, M. H., Fan, P., Brodsky, B., & Baum, J. (1993) *Biochemistry 32*, 7377−7387.

Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J., & Brodsky, B. (1993) *Biochemistry 32*, 11688−11695.

Manning, M. C., & Woody, R. W. (1989) *Biochemistry 28*, 8609−8613.

Marky, L. A., & Breslauer, K. J. (1987) *Biopolymers 26*, 1601−1620.

O'Neil, K. T., & DeGrado, W. F. (1990) *Science 250*, 646−651.

Privalov, P. L. (1982) *Adv. Protein Chem. 35*, 1−103.

Prockop, D. J., & Kivirikko, K. I. (1995) *Annu. Rev. Biochem. 64*, 403−434.

Ramachandran, G. N., & Kartha (1955) *Nature 176*, 593−595.

Rich, A., & Crick, F. H. C. (1961) *J. Mol. Biol. 3*, 483−506.

Sakikabara, S., Inouye, K., Shudo, K., Yasua, K., Kobayashi, Y., & Prockop, D. J. (1973) *Biochim. Biophys. Acta 303*, 198−202.

Salem, G., & Traub, W. (1975) *FEBS Lett. 51*, 94−99.

Scatturin, A., Tamburro, A. M., Del Pra, A., & Bordignon, E. (1975) *Int. J. Pept. Protein Res. 7*, 425−435.

Segal, D. M. (1969) *J. Mol. Biol. 43*, 497−517.

Smith, C. K., Withka, J. M., & Regan, L. (1994) *Biochemistry 33*, 5510- 5517.

Tamburro, A. M., Scatturin, A., & Del Pra, A. (1977) *Int. J. Pept. Protein Res. 9*, 310−318.

Thakur, S., Vadolas, D., Germann, H. P., & Heidemann, E. (1986) *Biopolymers 25*, 1081−1086.

Venugopal, M. G., Ramshaw, J. A. M., Braswell, E., Zhu, D., & Brodsky, B. (1994) *Biochemistry 33*, 7948−7956.

Wojcik, J., Altmann, K. H., & Scheraga, H. A. (1990) *Biopolymers 30*, 121−134.

BI960046Y