

Sequence Dependence of the Folding of Collagen-like Peptides

SINGLE AMINO ACIDS AFFECT THE RATE OF TRIPLE-HELIX NUCLEATION*

(Received for publication, October 19, 1998)

Michael S. Ackerman^{‡§}, Manjiri Bhate^{‡§}, Nigel Shenoy[‡], Konrad Beck[‡], John A. M. Ramshaw[¶],
and Barbara Brodsky^{‡||}

From the [‡]Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854 and the [¶]Commonwealth Scientific and Industrial Research Organization Division of Molecular Science, Parkville, Victoria 3052, Australia

The refolding of thermally denatured model collagen-like peptides was studied for a set of 21 guest triplets embedded in a common host framework: acetyl-(Gly-Pro-Hyp)₃-Gly-Xaa-Yaa-(Gly-Pro-Hyp)₄-Gly-Gly-amide. The results show a strong dependence of the folding rate on the identity of the guest Gly-Xaa-Yaa triplet, with the half-times for refolding varying from 6 to 110 min (concentration = 1 mg/ml). All triplets of the form Gly-Xaa-Hyp promoted rapid folding, with the rate only marginally dependent on the residue in the Xaa position. In contrast, triplets of the form Gly-Pro-Yaa and Gly-Xaa-Yaa were slower and showed a wide range of half-times, varying with the identity of the residues in the triplet. At low concentrations, the folding can be described by third-order kinetics, suggesting nucleation is rate-limiting. Data on the relative nucleation ability of different Gly-Xaa-Yaa triplets support the favorable nature of imino acids, the importance of hydroxyproline, the varying effects of the same residue in the Xaa position *versus* the Yaa position, and the difficulties encountered when leucine or aspartic acid are in the Yaa position. Information on the relative propensities of different tripeptide sequences to promote nucleation of the triple-helix in peptides will aid in identification of nucleation sites in collagen sequences.

The collagen triple helix is the basic structural motif found in all fibril-forming collagens as well as some host-defense proteins such as C1q, mannose-binding protein, and macrophage scavenger receptor (1, 2). The triple-helix conformation consists of three extended polyproline II-like chains supercoiled around each other as determined by x-ray fiber diffraction, crystallography, and NMR (3–7). The three chains are staggered by one residue with respect to each other and stabilized by interchain hydrogen bonding (5, 8, 9). This conformation requires that every third residue must be a glycine, generating a repeating (Gly-Xaa-Yaa)_n pattern, and that a high proportion of residues are the imino acids proline and hydroxyproline. Gly-Pro-Hyp is

indeed the most common and stabilizing tripeptide found in collagens.¹

The folding of the collagen triple helix *in vivo* is a multistep process involving chain association, registration, nucleation, and propagation (10–12). There is also evidence for the involvement of chaperones (13, 14). Fibril-forming collagens are synthesized at the rough endoplasmic reticulum membrane in a precursor form, procollagen, containing both N- and C-terminal propeptides terminating the long central triple helix. Proper chain selection and registration is initiated by the association of the C-propeptide domains into trimers followed by nucleation of the correctly aligned triple helix (15, 16) and propagation in a C- to N-terminal direction (17). In the unfolded state, most proline residues in the Yaa position of Gly-Xaa-Yaa triplets are enzymatically hydroxylated, and the resulting hydroxyproline (Hyp) residues are required for the formation and stabilization of the triple helix (18).

Folding studies on mature collagens are complicated by their length and varied sequences; these complicating features can be reduced by the use of natural or synthetic peptides. Collagen fragments have been used to better define the folding process, e.g. the observation of third-order kinetics for a 36-residue cyanogen bromide fragment of collagen type I (19). Synthetic peptide models of the triple helix allow the sequence dependence of folding to be investigated systematically by varying both the design and the composition of the Gly-Xaa-Yaa sequences. Investigations on synthetic model peptides, such as (Pro-Pro-Gly)_n, which adopt a stable triple-helical structure, have allowed quantitation of third-order rate constants, the effect of length and the testing of sophisticated theoretical models including folding intermediates resulting from incorrectly staggered chains (20, 21). Here we present data on the refolding of a set of homologous peptides that contain one variable Gly-Xaa-Yaa guest triplet embedded in a Gly-Pro-Hyp-rich host sequence. This design allows the analysis of the effect of a single triplet sequence on different triple-helix properties. Work in our laboratory has shown that all host-guest peptides analyzed so far form stable triple helices, with melting temperatures dependent on the identity of the guest triplet (22–24). This study reports folding rates for a set of host-guest peptides that provide information on the relative propensities of different tripeptide sequences to promote nucleation of the triple helix.

EXPERIMENTAL PROCEDURES

Peptide Synthesis—Peptides were synthesized on an Applied Biosystems 430A synthesizer using the standard FastMoc (Applied BioSys-

* This work has been supported by the National Institutes of Health Grant AR19626 (to B. B.), a National Science Foundation U. S.-Australia International Cooperative Research grant (to B. B.), and the Australia/U. S. A. Bilateral Science Program (to J. A. M. R.). 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.

§ Equal contributors to this work.

|| To whom correspondence should be addressed: Dept. of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, 675 Hoes Lane, Piscataway, NJ 08854-5636. Tel.: 732-235-4048; Fax: 732-235-4783; E-mail: brodsky@rwj.umdj.edu.

¹ Standard three-letter nomenclature is used to describe peptide sequences with Hyp representing 4-hydroxyproline. In the standard one-letter code for amino acids, O is used to represent 4-hydroxyproline. Host-guest peptides are named by the sequence of their guest triplet.

tems) method on a *N*-(9-fluorenyl)methoxycarbonyl (Fmoc)-RINK resin as described previously (22, 23). Peptides were purified to >90% purity by reversed-phase high performance liquid chromatography on a C-18 column eluted with a binary 10–30% (v/v) acetonitrile/water gradient containing 0.1% trifluoroacetic acid. Peptide identity was confirmed by laser desorption mass spectrometry and amino acid analysis.

Sample Preparation—Peptides were placed *in vacuo* over P₂O₅ for more than 48 h before weighing, dissolved in phosphate-buffered saline (10 mM sodium phosphate, 150 mM NaCl, pH 7.0), and stored at 4 °C.

Refolding Experiments—Measurements were performed on an AVIV Model 62DS CD² spectrophotometer equipped with a thermoelectric temperature control in 1-mm-path length quartz cells. Peptides were denatured in glass test tubes at 70 °C for 20 min and then rapidly cooled to 15 °C, unless otherwise specified, by quenching in an ice-water bath before transfer into the cuvette, kept at 15 °C. Quenching time was ~5 s, and the time needed for sample transfer was ~15 s. The ellipticity at 225 nm was monitored as a function of time over at least 30 min and until the fraction of folded peptides exceeded 0.5 (see below), with data intervals and averaging times of 1 to 60 s, depending on concentration and folding speed. Experiments repeated for some peptides indicated a deviation between their half-times of less than 10%.

Data Analysis—The fraction of folded peptide (F) is defined as

$$F = (\theta_{\text{obs}} - \theta_m) / (\theta_t - \theta_m) \quad (\text{Eq. 1})$$

where θ_{obs} , θ_t , and θ_m represent the observed, the triple helix, and monomer ellipticity, respectively. θ_t was measured directly before denaturation at the temperature used for refolding. θ_m was determined by extrapolating the initial data points to time zero. This θ_m value was slightly lower than that resulting from linear extrapolation of the monomer ellipticity observed in the high temperature region of equilibrium melting curves, but the use of either value gave similar results. To compare the data of different peptides independent of the folding mechanism, the time ($t_{1/2}$) at which $F = 0.5$ was determined.

The concentration of monomer [A] at any given time was calculated as

$$[A] = (1 - F)[A_0] \quad (\text{Eq. 2})$$

with [A]₀ denoting the initial monomeric peptide concentration, which was assumed to be equal to the total peptide concentration. Data were fitted to a single-step first (Eq. 3)-, second (Eq. 4)-, or third (Eq. 5)-order kinetics:

$$\ln[A] = \ln[A_0] - k_1 t \quad (\text{Eq. 3})$$

$$1/[A] = 1/[A_0] + 2k_2 t \quad (\text{Eq. 4})$$

$$1/[A]^2 = 1/[A_0]^2 + 6k_3 t \quad (\text{Eq. 5})$$

Rate constants k_i were calculated after linearization from the slope resulting from linear least squares fit. Curves were categorized as of *i*th order based on maximum linear correlation coefficients.

RESULTS

Peptide Design and Stability—Host-guest peptides of the form acetyl-(Gly-Pro-Hyp)₃-Gly-Xaa-Yaa-(Gly-Pro-Hyp)₄-Gly-Gly-amide provide a useful template to evaluate the contribution of individual Gly-Xaa-Yaa triplets to triple-helix properties (22–24). To assure formation of a stable triple-helix, the guest triplet is flanked by stabilizing Gly-Pro-Hyp triplets. The N and C termini are blocked by acetylation and amidation, respectively, to ensure that the only ionizable groups, if any, would be those introduced in the guest triplet and to eliminate charge repulsion at the ends of the triple helix. The peptide length is designed to be short enough so that the effects of a single guest triplet would not be masked by the constant part of the structure but long enough to ensure triple-helix stability (22). Because imino acids are found at high frequency in triple helices, guest triplets of the form Gly-Xaa-Hyp, Gly-Pro-Yaa, and Gly-Xaa-Yaa were considered, with the Xaa and Yaa residues occupied by the most common nonpolar residues, and charged residues found in collagens. In the following, we refer to the different peptides by their guest triplet sequence. This

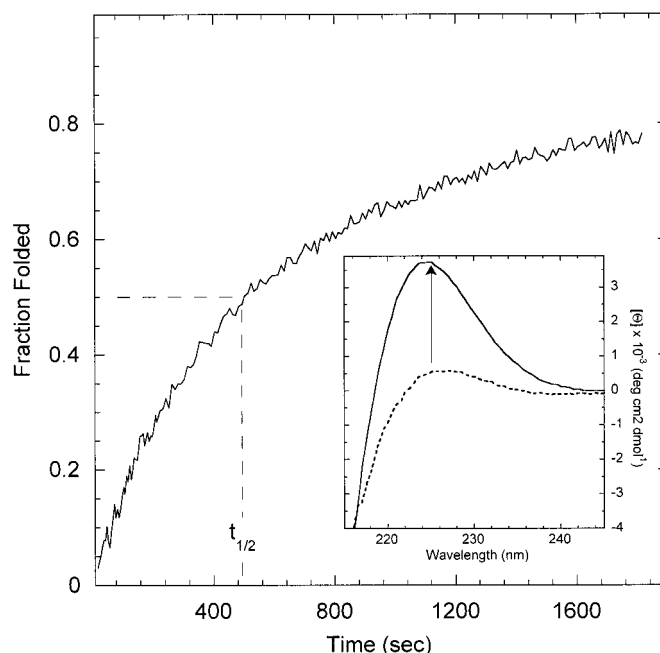


FIG. 1. Refolding curve of host-guest peptide Gly-Ala-Hyp. The refolding of peptide Gly-Ala-Hyp (concentration = 1 mg/ml) was monitored by following the recovery of ellipticity at 225 nm at 15 °C after denaturation at 70 °C for 20 min. The signal was normalized to the fraction of folded peptide according to Equation 1. The time $t_{1/2}$ (~8.3 min) at which half the molecules are folded ($F = 0.5$) is indicated by a dashed line. The inset shows the CD spectra of the same sample recorded at 15 °C (—) and 70 °C (- -). The arrow marks the transition followed in the refolding experiment.

design allows analysis of the effects on folding of a single residue within a defined triple-helical environment.

CD measurements of all host-guest peptides indicate triple-helical structures at low temperature. The spectra show a characteristic maximum near 225 nm with a mean residue ellipticity in the order of 4,000 deg cm² dmol⁻¹, which decreases upon unfolding (Fig. 1, inset). Equilibrium unfolding curves exhibit a highly cooperative behavior with melting temperatures ranging from 20 to 45 °C, depending on the identity of the guest triplet (22–25). The curves can be fitted to a two-state trimer to monomer transition, an assumption supported by analytical ultracentrifugation experiments performed on closely related peptides (26).

Folding of Host-Guest Peptides—Refolding rates were measured for a total of 21 host-guest peptides (concentration = 1 mg/ml). Despite the variations in stability of the host-guest peptides, a common folding temperature of 15 °C was selected. At this temperature all peptides showed a fraction of folded peptide close to one in their equilibrium melting curves, and little dependence of the folding rate on temperature of folding was observed at 5, 10, 15, and 20 °C (data not shown). As an example, the signal recovery for peptide Gly-Ala-Hyp at 15 °C is shown in Fig. 1, which also illustrates the determination of $t_{1/2}$ values as a common measure to compare the refolding behavior of the different peptides. The $t_{1/2}$ values for the host-guest peptides varied between 6 and 110 min, revealing that the folding rate critically depends on the sequence of the guest triplet (Table I). For example, the folding half-times of peptides Gly-Pro-Hyp, Gly-Ala-Hyp, Gly-Pro-Ala, and Gly-Ala-Ala are 6.0, 8.3, 12, and 21 min, respectively, at 15 °C (Fig. 2). The fast folding of Gly-Pro-Hyp is decreased slightly by the substitution of Pro by Ala in the Xaa position and somewhat more when the Hyp is replaced by an Ala.

Guest triplets with Hyp in the Yaa position (Gly-Xaa-Hyp) fold fastest and show only a small dependence on the identity of

² The abbreviation used is: CD, circular dichroism.

TABLE I
Sequence dependence of refolding

Refolding of the indicated peptides (concentration = 1 mg/ml) after denaturation at 70 °C for 20 min was monitored at 15 °C, and half-times $t_{1/2}$ (in min) were determined. Peptides are sorted in complementary sets.

Gly-Xaa-Hyp	$t_{1/2}$	Gly-Pro-Yaa	$t_{1/2}$	Gly-Xaa-Yaa	$t_{1/2}$
Gly-Pro-Hyp	6.0	Gly-Pro-Pro	11		
Gly-Ala-Hyp	8.3	Gly-Pro-Ala	12	Gly-Ala-Ala	21
Gly-Leu-Hyp	12	Gly-Pro-Leu	40	Gly-Leu-Ala	35
				Gly-Ala-Leu	86
Gly-Asn-Hyp	8.3	Gly-Pro-Asn	38		
Gly-Asp-Hyp	8.6	Gly-Pro-Asp	98	Gly-Asp-Ala	83
Gly-Gln-Hyp	9.4	Gly-Pro-Gln	22		
Gly-Glu-Hyp	6.1	Gly-Pro-Glu	27		
		Gly-Pro-Lys	29		
		Gly-Pro-Arg	15	Gly-Ala-Arg	15

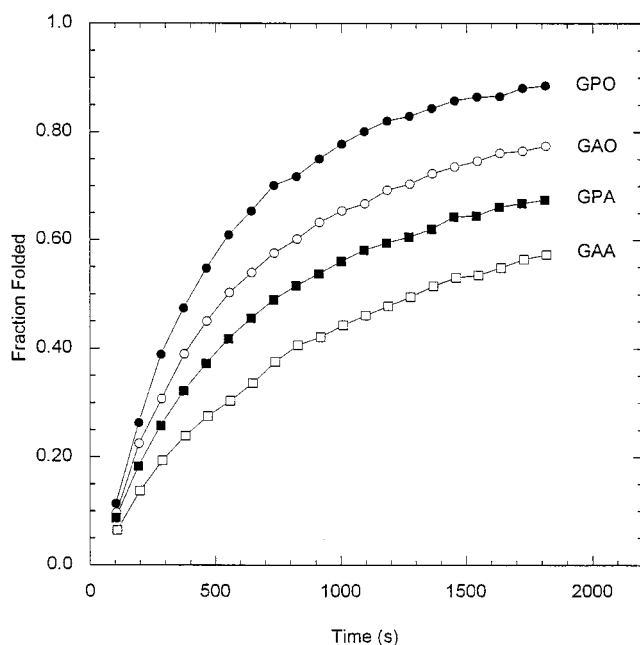


FIG. 2. Effect of alanine in positions Xaa and Yaa on refolding kinetics. Recovery of the CD signal at 225 nm was monitored at 15 °C after 20 min denaturation at 70 °C, and the signal was converted to the fraction of folded peptide for Gly-Pro-Hyp (●), Gly-Ala-Hyp (○), Gly-Pro-Ala (■) and Gly-Ala-Ala (□) (concentration = 1 mg/ml).

the residue in the Xaa position. Gly-Pro-Yaa guest triplets with Pro in the Xaa position show a larger range of folding rates and a stronger influence of the identity of the Yaa residue. In the Gly-Pro-Yaa peptides, the residues glutamine and glutamic acid have a very similar influence on the kinetic behavior ($t_{1/2}$ = 22 min for Gln and 27 min for Glu), whereas an exchange of aspartic acid for asparagine drastically slows down the refolding ($t_{1/2}$ = 38 min for Asn and 98 min for Asp). The combination of nonimino acids in both the Xaa and Yaa positions (Gly-Xaa-Yaa) leads to a broad distribution of $t_{1/2}$ values, with peptides Gly-Ala-Leu and Gly-Asp-Ala folding particularly slowly. In general, leucine was observed to decrease the folding rate compared with Ala, with the effect being most striking in the Yaa position.

Kinetic Analysis of Triple-helix Folding—To understand the influence of the single guest triplet on the folding mechanism, the folding data at 1 mg/ml was analyzed by fitting to first-, second-, and third-order kinetics (Equations 3–5). Considering the full folding process, some peptides showed best fits when assuming a first-order kinetics, whereas others showed best fits when either second- or third-order kinetics were assumed. Analyses of the folding curves of several peptides in which only the late data points were fitted to third-order kinetics (Equa-

tion 5) revealed that these regions show a reasonably straight line in $1/[A]_0^2(t)$ plots (data not shown). This suggests that at a sufficiently low monomer concentration, the third-order step, which relates to chain association and/or nucleation, becomes the rate-limiting step for trimer formation.

Folding via a third-order kinetic process should show a profound dependence on concentration, whereas the first-order kinetics of propagation should be independent of concentration. Measurements of the kinetics of folding for a set of 8 peptides at a concentration of 0.4 mg/ml (Fig. 3A) indicated that at this lower concentration all but the very early time points ($t < 10$ to 300 s) followed a straight line when plotted as $1/[A]_0^2(t)$, giving a good fit to third-order kinetics (Fig. 3B). This indicates that at a sufficiently low monomer concentration, chain association/nucleation becomes the rate-limiting kinetic step in triple-helix folding. The apparent third-order rate constants derived from the slopes varied from 470 to 16,000 $M^{-2} s^{-1}$, with S.D. of approximately $\pm 10\%$. The lowest third-order rate constants were found for peptides Gly-Ala-Leu and Gly-Pro-Asp, and the largest was found for Gly-Pro-Hyp (Table II). The Gly-Pro-Hyp peptide had a slower folding rate than Gly-Pro-Hyp, showing the effect of a hydroxyl group at one position. The large variation in the rate constants suggests that for these relatively short peptides, the chain association/nucleation step critically depends on the identity of the single guest triplet, and at this low concentration, any other processes like helix propagation have only a marginal contribution.

DISCUSSION

The folding of collagen molecules *in vivo* is an intricate and coordinated process (10, 12, 27). The formation of a trimer requires association and registration of three chains, which is mediated for fibril-forming collagens by the globular C-terminal propeptide domain (10–12) (Fig. 4). C-propeptide trimerization constrains the three chains at the C terminus such that nucleation can occur at the C termini of the triple-helical region. Nucleation is considered to be the process in which a series of tripeptide units from the three chains adopt appropriate collagen-like φ, ψ angles and form interchain hydrogen bonds. At the C termini of fibril-forming collagens are 5–6 sequential triplets of the form Gly-Xaa-Hyp (often Gly-Pro-Hyp units), which are thought to constitute all or part of the nucleation site. For example, in collagen type III, C-terminal Gly-Xaa-Hyp triplets are required for nucleation, and deletion experiments showed that as long as two Gly-X-Hyp triplets are retained, nucleation is effective, and triple-helix folding is complete (16). Following nucleation, propagation proceeds in a zipper-like manner from the C to N terminus. The rate-limiting step of propagation is *cis-trans* isomerization at imino acid peptide bonds, which have a significant proportion of *cis* bonds in unfolded chains (17, 28, 29).

The studies reported here on a set of host-guest peptides demonstrate that the folding rate of the triple helix critically depends on the sequence of a single guest triplet. Analysis of our data indicates that the folding proceeds via a mechanism involving more than a single reaction step and that the folding involves a third-order process that becomes rate-limiting at low concentrations. The steps requiring the involvement of three polypeptide chains are chain association and nucleation. Although the association process, which is limited by diffusion, is unlikely to be significantly affected by sequence, triple-helix nucleation is known to be facilitated by the presence of conformationally restricted imino acids and is thus expected to be sequence-dependent (30) (Fig. 4). In the present study, the magnitude of the third-order rate constant is strongly affected by the identity of the guest triplet and is greater for imino acid-containing triplets, suggesting it is the nucleation step

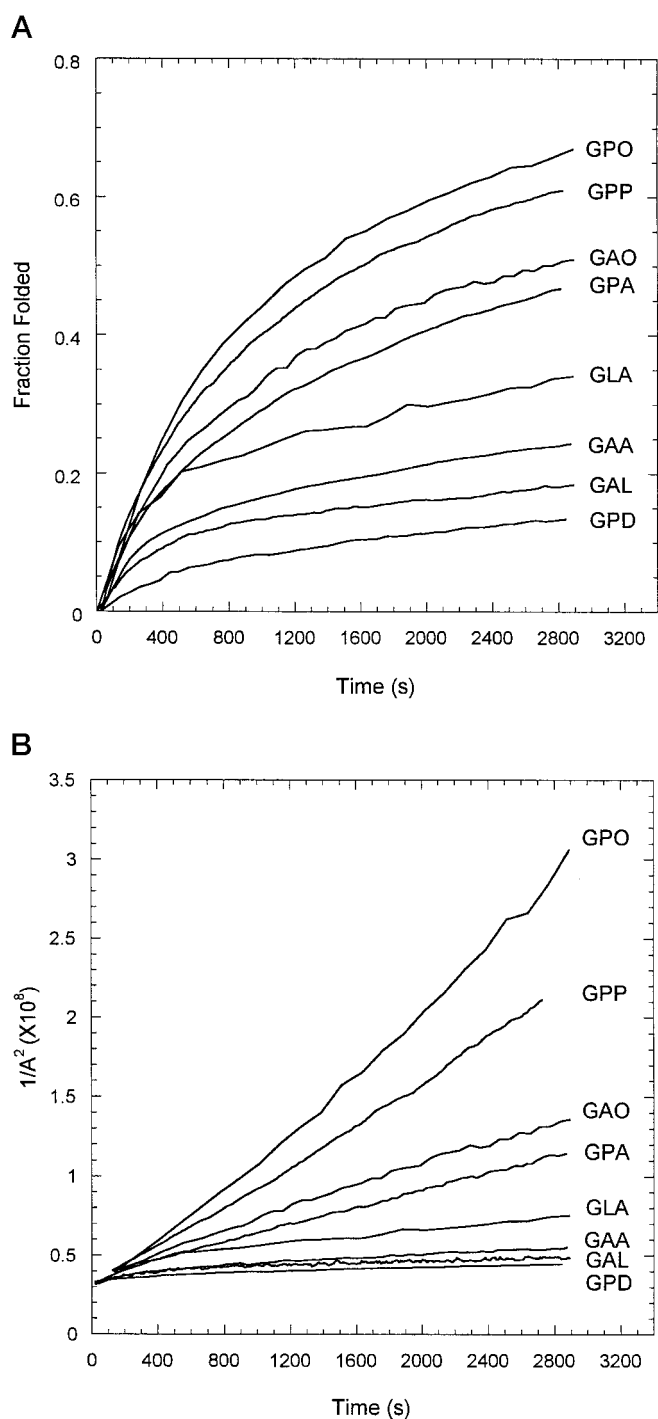


FIG. 3. Refolding kinetics of host-guest peptides at low concentration. A, recovery of the CD signal at 225 nm was monitored at 15 °C after denaturation, and the signal was converted to the fraction of folded peptide at a concentration of 0.4 mg/ml for (from top to bottom) Gly-Pro-Hyp, Gly-Pro-Pro, Gly-Ala-Hyp, Gly-Pro-Ala, Gly-Leu-Ala, Gly-Ala-Ala, Gly-Ala-Leu, and Gly-Pro-Asp. B, curves linearized according to Equation 5 are shown for the same peptide set. Third-order rate constants (Table II) were derived from the slope of the fitted lines.

itself reflected by these values. NMR studies on peptides with specific ^{15}N labels indicated that nucleation can occur at (Gly-Pro-Hyp) $_n$ sites at either end of a peptide (31). For the Gly-Pro-Hyp-enriched host-guest peptides, it is realistic to assume that nucleation could begin at any tripeptide unit in the chain (32) (Fig. 4). Previous findings suggest the nucleation domain is as long as six triplets in noncovalently linked peptides (20, 21), making this a dominant event in short peptides.

TABLE II
Effect of sequence on third-order rate constants
Refolding of the indicated peptides (concentration = 0.4 mg/ml) after denaturation at 70 °C for 20 min was monitored at 15 °C. Third-order rate constants k_3 were determined according to Equation 5 as outlined in Fig. 4B.

Peptide	k_3 $\text{M}^{-2} \text{s}^{-1}$
Gly-Pro-Hyp	16,000
Gly-Pro-Pro	12,000
Gly-Ala-Hyp	6,000
Gly-Pro-Ala	4,700
Gly-Leu-Ala	1,500
Gly-Ala-Ala	950
Gly-Ala-Leu	500
Gly-Pro-Asp	470

The third-order rate constants and half-times of folding yield information concerning the relative propensity of different residues in the Xaa and Yaa position to initiate triple-helix nucleation. Entropic factors are likely to play an important role in nucleation because imino acids are sterically constrained to dihedral angles similar to those found in collagen. Gly-Pro-Hyp is the fastest folding triplet, and all Gly-Xaa-Hyp triplets are very favorable. The Hyp residue appears more favorable than Pro, as seen in the faster folding rate of Gly-Pro-Hyp *versus* Gly-Pro-Pro. It has been suggested that the OH of Hyp has an inductive effect, leading to a decrease in the *cis:trans* isomer ratio compared with Pro in the unfolded state (33). A decrease in the *cis* isomer concentration could accelerate the propagation step. In addition, the decreased *cis:trans* ratio could promote nucleation by making it more likely to find a stretch of contiguous all-*trans* tripeptide units and by creating a more rigid monomer chain (34). Although Pro in the Xaa position also can lead to favorable folding, the identity of the nonimino acid residue in the Yaa position of Gly-Pro-Yaa triplets has a very strong influence. For example, Gly-Pro-Ala is a fast folding peptide, whereas Gly-Pro-Asp has the slowest folding rate observed.

In addition to entropic factors, the influence of specific side chains in promoting nucleation may relate to steric factors, electrostatic interactions, and hydrogen bonding. The difficulty in packing bulky residues such as Leu in the Yaa position (35), which is less exposed than the Xaa position, may contribute to the slow folding of Gly-Pro-Leu and Gly-Ala-Leu peptides. Despite its large side chain, arginine in the Yaa position is favorable in promoting chain nucleation as well as for stabilization (24), and both features may be related to its ability to form multiple hydrogen bonds combined with its restricted mobility (36, 37). Gly-Pro-Asp is the slowest folding peptide, suggesting an unfavorable effect of aspartic acid in the Y position. It was previously observed that aspartic acid in the Yaa position had a destabilizing effect on the triple helix. Both the decreased folding rate and low stability may be related to the restricted rotational freedom of Asp in the triple helix, hindering its participation in interchain hydrogen bond formation (23).

The systematic exchange of single guest triplets embedded in an otherwise constant environment allows their influence on folding to be related to their contribution to triple-helix stability. The relationship between folding half-times and the melting temperatures of the host-guest peptide concentrations set at 1 mg/ml was considered (Fig. 5). All Gly-Xaa-Hyp peptides have fast folding rates and high stabilities, with a small range for both $t_{1/2}$ values and melting temperature values. The Gly-Pro-Yaa peptides show a broad range for both folding half-times (12–98 min) and melting temperatures (30–45 °C), with the more stable peptides tending to fold faster. For Gly-Xaa-Yaa triplets with no imino acids, four peptides with similar

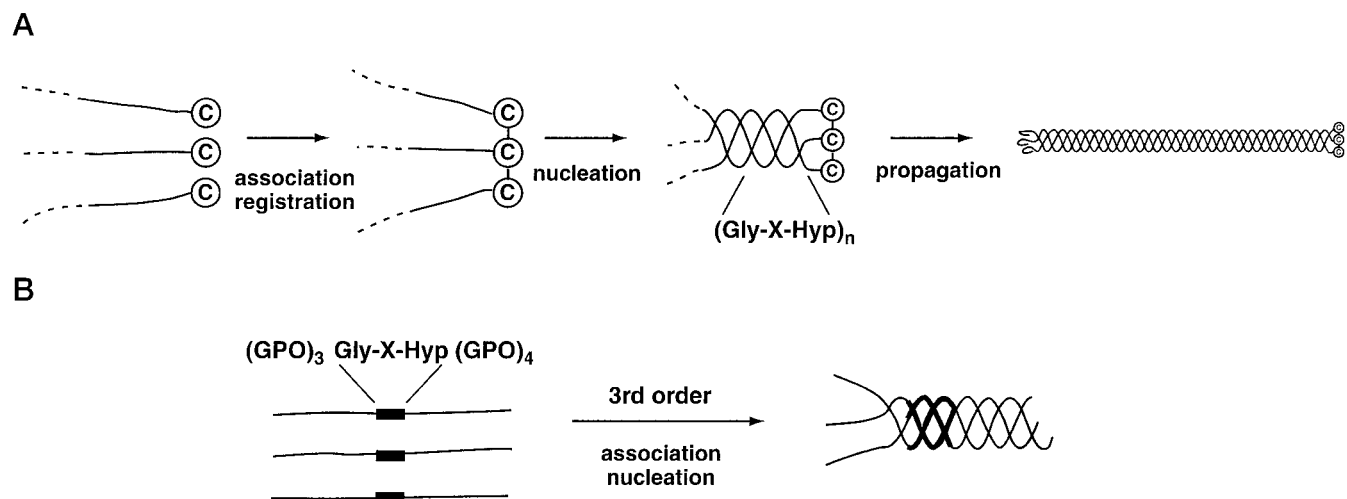


FIG. 4. **Schematic of triple-helix folding.** A, folding of collagen, showing association of C-propeptides followed by nucleation of three chains at the C terminus and then propagation. B, folding of host-guest peptides, showing association and nucleation of three nonlinked peptide chains. Nucleation is shown at the C terminus, but it is likely that it can occur at any site within the peptide. The guest triplet is indicated by a *bold line*.

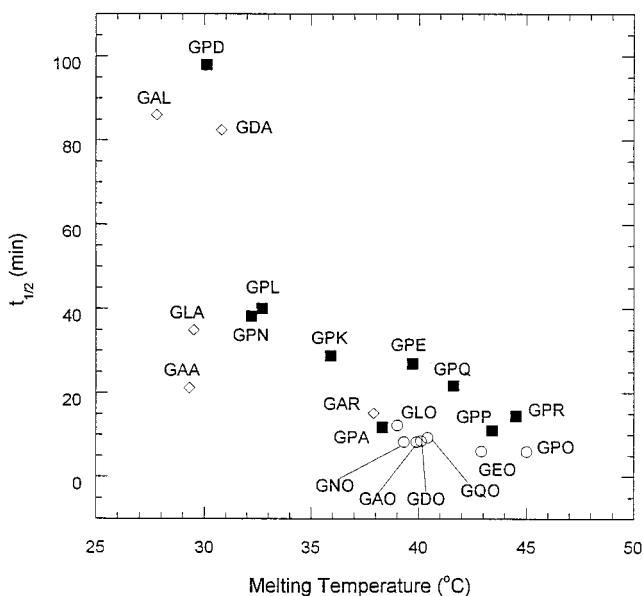


FIG. 5. **Relationship between refolding rate and thermal stability.** Refolding half-times $t_{1/2}$ are plotted versus melting temperatures, both determined at 1 mg/ml. Letters adjacent to the symbols denote the residues in the Gly-Xaa-Yaa guest triplet (standard one letter code) within the peptide sets Gly-Xaa-Hyp (\circ), Gly-Pro-Yaa (\blacksquare), and Gly-Xaa-Yaa (\diamond). Melting temperatures are from Shah *et al.* (22), Chan *et al.* (23), and Yang *et al.* (24) except for Gly-Asn-Hyp, Gly-Gln-Hyp, Gly-Pro-Asn, Gly-Pro-Gln, and Gly-Asp-Ala (Shenoy, N., Beck, K., and Brodsky, B., unpublished data).

thermal stabilities were found to have very different folding times. This suggests the interactions determining folding differ from those important for stability for tripeptides with no imino acids.

The host-guest peptide set shows the wide range in effectiveness of Gly-Xaa-Yaa tripeptides in a fixed Gly-Pro-Hyp environment to facilitate or depress nucleation. The nucleation step of peptides differs from that of collagen in that this step occurs in three independent peptides, whereas triple-helix nucleation in collagen occurs in a molecule that is linked together by the association of disulfide-linked C-propeptides (Fig. 4). Despite this difference, it is likely that the propensity of individual tripeptides to nucleate a peptide triple helix can be applied to the ability of different sequences at the C terminus of collagen to serve as a nucleation site. In addition, it is possible that

interruptions in the $(\text{Gly-Xaa-Yaa})_n$ -repeating sequence, as found normally in basement membrane collagen or for osteogenesis imperfecta Gly \rightarrow Xaa mutations in type I collagen, may terminate propagation (38, 39), making a renucleation event necessary to complete triple-helix formation. Information on the propensity of different Gly-Xaa-Yaa triplets to promote nucleation will aid in the identification of such renucleation sequences.

Acknowledgments—We are grateful to Alan Kirkpatrick for peptide synthesis and Nick Bartone for peptide characterization. We also thank Drs. Naina K. Shah, Wei Yang, and Ellen Doss for helpful comments, Chirag Shah for preliminary results, and Ms. Virginia C. Chan for experimental guidance in the early stages of this project. Furthermore, we appreciate the continuous discussions with Dr. Jean Baum.

REFERENCES

1. Brodsky, B., and Ramshaw, J. A. M. (1997) *Matrix Biol.* **15**, 545–554
2. Beck, K., and Brodsky, B. (1998) *J. Struct. Biol.* **122**, 17–29
3. Rich, A., and Crick, F. H. C. (1961) *J. Mol. Biol.* **3**, 483–506
4. Fraser, R. D. B., MacRae, T. P., and Suzuki, E. (1979) *J. Mol. Biol.* **129**, 463–481
5. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) *Science* **266**, 75–81
6. Mayo, K. H. (1996) *Biopolymers* **40**, 359–370
7. Baum, J., and Brodsky, B. (1997) *Folding & Design* **2**, R53–R60
8. Privalov, P. (1982) *Adv. Protein Chem.* **35**, 1–104
9. Bella, J., Brodsky, B., and Berman, H. M. (1995) *Structure* **3**, 893–906
10. Engel, J., and Prockop, D. J. (1991) *Annu. Rev. Biophys. Biophys. Chem.* **20**, 137–152
11. McLaughlin, S. H., and Bulleid, N. J. (1997) *Matrix Biol.* **16**, 369–377
12. Bulleid, N. J. (1996) *Cell. Dev. Biol.* **7**, 667–672
13. Wilson, R., Lee, J. F., and Bulleid, N. J. (1998) *J. Biol. Chem.* **273**, 9637–9643
14. Nagata, K. (1996) *Trends. Biochem. Sci.* **21**, 23–26
15. Lees, J. F., Tasab, M., and Bulleid, N. J. (1997) *EMBO J.* **16**, 908–916
16. Bulleid, N. J., Dalley, J. A., and Lees, J. F. (1997) *EMBO J.* **16**, 6694–6701
17. Bächinger, H. P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) *Eur. J. Biochem.* **106**, 619–632
18. Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., and Prockop, D. J. (1973) *Biochim. Biophys. Acta* **303**, 198–202
19. Piez, K. A., and Sherman, M. S. (1970) *Biochemistry* **9**, 4134–4140
20. Sutoh, K., and Noda, H. (1974) *Biopolymers* **13**, 2477–2488
21. Weidner, H. (1975) *Biopolymers* **14**, 763–780
22. Shah, N. K., Ramshaw, J. A. M., Kirkpatrick, A., Shah, C., and Brodsky, B. (1996) *Biochemistry* **35**, 10262–10268
23. Chan, V. C., Ramshaw, J. A. M., Kirkpatrick, A., Beck, K., and Brodsky, B. (1997) *J. Biol. Chem.* **272**, 31441–31446
24. Yang, W., Chan, V. C., Kirkpatrick, A., Ramshaw, J. A. M., and Brodsky, B. (1997) *J. Biol. Chem.* **272**, 28837–28840
25. Ramshaw, J. A. M., Shah, N. S., and Brodsky, B. (1998) *J. Struct. Biol.* **122**, 86–91
26. Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J., and Brodsky, B. (1993) *Biochemistry* **32**, 11688–11695
27. Brodsky, B. (1990) in *The Protein Folding Problem* (Gierasch, L., and King, J., eds) pp. 55–62, American Association for the Advancement of Science, Washington, D. C.
28. Bächinger, H. P., Bruckner, P., Timpl, R., and Engel, J. (1978) *Eur. J. Biochem.* **90**, 605–613
29. Bächinger, H. P. (1987) *J. Biol. Chem.* **262**, 17144–17148
30. Von Hippel, P. H., and Harrington, W. F. (1959) *Biochim. Biophys. Acta* **36**, 427–447

31. Liu, X., Siegel, D. L., Fan, P., Brodsky, B., and Baum, J. (1996) *Biochemistry* **35**, 4306–4313
32. Weidner, H., Engel, J., and Fietzek, P. P. (1974) in *Peptides, Polypeptides, and Proteins* (Blout, E. R., Bovey, F. A., Goodman, M., and Lotan, N., eds) pp. 419–435, Wiley, New York
33. Eberhardt, E. S., Panasik, N., Jr., and Raines, R. T. (1996) *J. Am. Chem. Soc.* **118**, 12261–12266
34. Bansal, M., and Ananthanarayanan, V. S. (1988) *Biopolymers* **27**, 299–312
35. Bansal, M. (1977) *Int. J. Pept. Protein Res.* **9**, 224–234
36. Richardson, J. S., and Richardson, D. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., ed) pp. 1–98, Plenum Publishing Corp., New York
37. Vitagliano, L., Nemethy, G., Zagari, A., and Scheraga, H. A. (1993) *Biochemistry* **32**, 7354–7359
38. Bächinger, H. P., Morris, N. P., and Davis, J. M. (1993) *Am. J. Med. Genet.* **45**, 152–162
39. Liu, X., Kim, S., Dai, Q.-H., Brodsky, B., and Baum, J. (1998) *Biochemistry* **37**, 15528–15533