Collagen triple helix formation can be nucleated at either end

Sabine Frank, Sergei Boudko, Kazunori Mizuno*, Therese Schulthess, Jürgen Engel, and Hans Peter Bächinger*

*Department of Biophysical Chemistry, Biozentrum, Universität Basel, CH-4056 Basel, Switzerland
*Shriners Hospital for Children, Research Department, and Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, OR 97201, USA

Corresponding author: Hans Peter Bächinger, Shriners Hospital for Children
3101 SW Sam Jackson Park Road
Portland, OR 97201, USA

e-mail: hpb@shcc.org
tel: 503-221-3433
fax: 503-221-3451

Running Title: Nucleation of the collagen triple helix
Summary

The directional dependence of folding rates for rod-like macromolecules such as parallel α-helical coiled-coils, DNA double-helices and collagen triple helices is largely unexplored. This is mainly due to technical difficulties in measuring rates in different directions. Folding of collagens is nucleated by trimeric non-collagenous domains. These are usually located at the C-terminus suggesting that triple helix folding proceeds from the C- to the N-terminus. Evidence is presented here, that effective nucleation is possible at both ends of the collagen-like peptide (Gly-Pro-Pro)\textsubscript{10}, using designed proteins in which this peptide is fused either N- or C-terminal to a nucleation domain, either T4-phage foldon or the disulfide knot of type III collagen. The location of the nucleation domain influences triple-helical stability, which might be explained by differences in the linker sequences and the presence or absence of repulsive charges at the carboxy-terminal end of the triple helix. Triple-helical folding rates are found to be independent of the site of nucleation and consistent with cis-trans isomerization being the rate-limiting step.

Introduction

Proteins fold along multiple pathways (1) and rod-like macromolecules may fold in two different directions depending on the site of nucleation. Experimental evidence concerning the directionality of kinetics is limited. An interesting study with a two-stranded coil-coil structure tethered by disulfide bridges at either the N- or C-terminus has
been performed (2) and for collagen NMR diffusion was proposed as a potential method to determine directionality (3). The collagen triple helix is a linear structure composed of three left-handed polyproline-II-type helices. Chains in this conformation are not stable as individual structures, but associate to form a right-handed triple helix which is stabilized by hydrogen bonds between chains (4). Formation of the triple helix is only possible if every third residue in the sequence of the chains is glycine (G). Imino acid residues favour polyproline-II-type helices and therefore the frequency of proline (P) residues in the typical collagen like sequence \((\text{Gly-X-Y})_n\) is high in X- and Y-positions. 4-hydroxyproline plays a special role in stabilizing the triple helix (5,6) and is frequently in the Y-position. Collagen triple helices are very abundant structures. They occur in all collagens as long rod-like elements and form short rods in the first component of complement C1q, lung surfactant protein SPA, mannose binding protein, scavenger receptors and many other extracellular proteins (7). Collagen triple helices are slowly folding structures with peptide cis-trans isomerization as the rate-limiting step (8). Cis-trans isomerization is often rate limiting also for globular proteins (9) and is dominating for collagens because of the many X-Pro bonds (X stands for any amino acid residue) with their high probability to form cis peptide bonds in the unfolded state. Another peculiarity for collagens is the need of nucleation domains for correct and sufficiently fast triple helical folding. It was found that trimerizing non-collagenous domains are essential for initial chain association and chain registration. This requirement for oligomerization and registration was explained by the need to establish a high local concentration at the nucleus and to prevent mismatched structures. For thermodynamic and kinetic studies, collagen-like peptides with designed oligomerization and nucleation domains have been particularly useful (10-13). Spontaneous folding of the
collagen triple helix is extremely slow, follows a third order reaction and is therefore strongly concentration dependent (11,14). For most collagens NC-1 domains are found at the C-terminus leading to the idea that triple helices may only or preferentially fold from the C-terminus. Recently it was suggested that a three-stranded coiled-coil domain at the N-terminus of collagen XIII can induce collagen folding from the N-terminus (15). Similar suggestions have been made for other collagens (16). This stimulated us to measure the rate of collagen triple helix folding in both directions applying designed proteins in which a globular oligomerization domain or a disulfide knot are attached to either the N- or C-terminus of (Gly-Pro-Pro)_{10}.

Experimental Procedures

Construction of expression plasmids and production of recombinant proteins.

The designed proteins (GPP)_{10}-Cys2 and (GPP)_{10}-foldon were expressed in E. coli as described (11,17). These models were used by these authors to study the increase in stability caused by the trimerization of the fused phage protein foldon (17) or by the linkage of the three chains by the disulfide knot Cys2 = GPPGCCC of type III collagen (11). The inverted protein with the oligomerization disulfide knot at the N-terminus Cys2-(GPP)_{10} was expressed using the same strategy but an additional Gly-Ser-spacer was inserted between Cys2 and the (GPP)_{10}-moiety. The protein foldon-(GPP)_{10} was synthesized by Fmoc chemistry with a SerGlySerGly-spacer between the domains (for exact sequences see legend of Fig. 1).
Oxidative disulfide cross-linking.

(GPP)$_{10}$-Cys$_2$ and Cys$_2$-(GPP)$_{10}$ were trimerized at 20 °C by their (GPP)$_{10}$-domains and formation of the disulfide knot was achieved by oxidation of the fully reduced material with a mixture of oxidized and reduced glutathione at a molar ratio of 9 : 1. Trimer formation was verified by mass spectroscopy and by analytical ultracentrifugation.

Mass Spectral Analysis.

For mass spectral analysis, the peptides were chromatographed on a 100 µm i.d. (inner diameter) column packed with Vydac C18 reverse-phase material (5 µm particle size). The proteins were eluted with a linear 20 minute gradient from 0.1% TFA to 80% acetonitrile/0.1% TFA at a flow rate of 1 µl/min. The outlet of the column was directed to a microspray needle, which was pulled from 100 µm i.d. 280 µm o.d. fused silica capillaries (LC Packings) on a model P-2000 quartz micropipette puller (Sutter Instrument Company). The needle was placed into an XYZ micro positioner and the voltage was applied directly to the sample stream through the capillary union (18). Spray voltages were usually between 1100 and 1400V. Mass determinations were carried out on a TSQ7000 triple quadrupole mass spectrometer (Finnigan). Formation of trimers of the predicted mass was verified for (GPP)$_{10}$-Cys$_2$ and Cys$_2$-(GPP)$_{10}$. In foldon-(GPP)$_{10}$ and (GPP)$_{10}$-foldon the three chains are not connected by covalent bonds and mass spectrometry demonstrated monomers of the predicted mass.
Analytical ultracentrifugation.

Sedimentation equilibrium experiments were performed on a Beckman Optima XL-A analytical ultracentrifuge (Beckman Instruments) equipped with 12-mm Epon double-sector cells in an An-60 Ti rotor. The peptides were analysed in 5 mM sodium phosphate buffer, pH 7.4, containing 150 mM NaCl. Sedimentation velocity runs were performed at a rotor speed of 56,000 rpm and sedimenting material was assayed by absorbance at 234 nm. Sedimentation coefficients were corrected to standard conditions (water, 20 °C; (19)). Molecular masses were evaluated from lnA versus r² plots, where A is the absorbance and r is the distance from the rotor center (19). A partial specific volume of 0.73 ml/g was used for all calculations.

CD spectroscopy.

CD experiments were acquired on a Cary 61 (Varian) or an Aviv 202 spectropolarimeter, equipped with a thermostatted 1-mm path length quartz cell. Data were normalized for concentration and path length to obtain the mean molar residue ellipticity after subtraction of the buffer contribution. Thermal stability was determined by monitoring the change in the mean molar residue ellipticity ([θ]) at a fixed wavelength of 221 nm for (GPP)₁₀-Cys₂ and Cys₂-(GPP)₁₀ and 210 nm for foldon-(GPP)₁₀ and (GPP)₁₀-foldon as a function of temperature. Data analysis was performed with the LABView (National Instruments) and Sigma Plot (Jandel Scientific) software packages.
Results and Discussion

The thermal transition profile of Cys2-(GPP)$_{10}$ was measured by the change in circular dichroism at 221 nm and a midpoint transition temperature $T_m = 67 \degree C$ was obtained which is 15 $\degree C$ lower than the value for (GPP)$_{10}$-Cys$_2$ (11) (Fig. 1a, Table 1). As previously observed for (GPP)$_{10}$-Cys$_2$, the thermal transition curves of Cys$_2$-(GPP)$_{10}$ were not fully reversible at temperatures higher than 70 $\degree C$, probably because of chemical degradation. Measurements were therefore performed in the presence of 2M guanidine hydrochloride (GuHCl) resulting in $T_m$-values of 40 and 63 $\degree C$ for and Cys$_2$-(GPP)$_{10}$ and (GPP)$_{10}$-Cys$_2$, respectively (Table 1). In the presence of guanidine HCl the transition was fully reversible according to two criteria: (i) after heating to 50 $\degree C$ for Cys$_2$-(GPP)$_{10}$ and to 70 $\degree C$ for (GPP)$_{10}$-Cys$_2$, for 5 min, the sample was cooled to 20 $\degree C$ and the value before heating returned after 2 hours, (ii) a second transition profile after this refolding was identical to the initial one. The thermal transition of foldon-(GPP)$_{10}$ was measured by circular dichroism at 210 nm, where the spectrum of foldon alone shows no change upon increasing the temperature [11]. The midpoint transition temperature $T_m = 53 \degree C$ corresponds to the (GPP)$_{10}$-triple helix and is 17 $\degree C$ lower than the value for the (GPP)$_{10}$-domain in (GPP)$_{10}$-foldon measured by the same method (Fig. 1b). Full reversibility was demonstrated by repeated scanning. Comparing the $T_m$-values with those of (GPP)$_{10}$-triple helix without oligomerization domains (Table 1), the results demonstrate that nucleation at either the N- and C-terminus stabilizes the triple helix. The mode of stabilization by foldon at the C-terminus was investigated in some detail (11,17) and the crystal structure of (GPP)$_{10}$-foldon has recently been solved (20). It was concluded that the stabilization is
entropic in nature such that the oligomerization domain creates an internal concentration of about 1 M at the junction between foldon (or the disulfide knot) and (GPP)\textsubscript{10}. A number of unfavourable interactions at the domain junction counteract the entropic stabilization. First, foldon-(GPP)\textsubscript{10} has a different linker sequence to that in (GPP)\textsubscript{10}-foldon. Second, there are repulsive carboxylate charges at the C-terminal end of the triple helix that are not present in (GPP)\textsubscript{10}-foldon and this may explain the lower transition temperature. The peptide Ac-(GPP)\textsubscript{10}-NH\textsubscript{2} has a T\textsubscript{m} of 45 °C in H\textsubscript{2}O as compared to 25 °C for (PPG)\textsubscript{10} (Mizuno and Bächinger, unpublished). Also, for Cys\textsubscript{2}-(GPP)\textsubscript{10} the linker sequence is different than that in (GPP)\textsubscript{10}-Cys\textsubscript{2} and C-terminal repulsive charges at the end of the triple helix are present, leading to a reduced stability. In all cases the internal concentrations imposed by the oligomerization domains are expected to be approximately identical.

The designed proteins with oligomerization domains either at the N- or C-terminus were then used to compare the rates of refolding. It is known that nucleation of triple helices from free (GPP)\textsubscript{10} chains is extremely slow, concentration dependent and of apparent third order (11). Connection of three chains by oligomerization domains increases the intrinsic chain concentrations to a level at which cis-trans isomerization steps in helix propagation, rather than chain finding steps, become rate limiting [11]. Given the very high concentration dependence of nucleation, it has to be expected that helix propagation will be preferentially initiated at the end at which the oligomerization domain is attached, although we cannot exclude the possibility that a small fraction of helix initiations may also occur at other sites. Figure 2 shows the time resolved phases, which follow a fast and kinetically unresolved jump of the circular dichroism signal. The fast phase mainly consists of the temperature dependent change of circular dichroism, which is visible also in temperature
regions where no transition takes place (Fig.1). It may also include a very small fast phase of triple helix formation before the first cis-peptide bond is met (8,21). The slow phase reflects the major part of the transition of the (GPP)$_{10}$-domain and long triple helices of real collagens fold entirely fold with a slow phase determined by cis-trans isomerization (8).

For Cys$_2$-(GPP)$_{10}$ and foldon-(GPP)$_{10}$ it was verified that the kinetic time courses did not change upon increasing the concentration by a factor of 4 at 10 or 20 °C for Cys$_2$-(GPP)$_{10}$ and foldon-(GPP)$_{10}$. For (GPP)$_{10}$-Cys$_2$ this was shown previously (11). Figures 2 and 3 clearly demonstrate that the first order rate constant of folding is identical for proteins carrying the oligomerization domains at opposite ends. The refolding rate constant for Cys$_2$-(GPP)$_{10}$ was 0.00037 s$^{-1}$ measured at 20 °C and a peptide concentration of 100 µM. For foldon-(GPP)$_{10}$ a value of 0.00098 s$^{-1}$ was determined for a peptide concentration of 51 µM at the same temperature. Also, the activation energies obtained from an Arrhenius plot of rate constants measured at different temperatures turned out to be identical within the limits of experimental error (Fig. 3). Values of activation energy larger than 50 kJ/mol confirm cis-trans isomerization steps as rate limiting events. Deviations from the maximum activation energies predicted for this step have been discussed (11).

In summary the results demonstrate that collagen triple helix formation can be nucleated at both ends. The rates are equal within experimental errors. In order to exclude possible effects of the nature of the nucleation domain two very different oligomerization sequences were employed. It was shown that their influence on protein stability was slightly different and also that the location of the domains at either end of the collagen-like domain was an important factor. These differences may be explained by energetic
differences in the linker sequences between the collagen triple helix and the oligomerization domains and the presence of charges at the end of the triple helix. The rate limiting step of cis-trans isomerization is however independent of the direction of folding.

References

1. Dill, K. A., and Chan, H. S. (1997) *Nat Struct Biol* **4**, 10-19
2. Krantz, B. A., and Sosnick, T. R. (2001) *Nat Struct Biol* **8**, 1042-1047
3. Buevich, A. V., and Baum, J. (2002) *J Am Chem Soc* **124**, 7156-7162
4. Traub, W., and Piez, K. A. (1971) *Adv Protein Chem* **25**, 243-352
5. Jenkins, C. L., and Raines, R. T. (2002) *Nat Prod Rep* **19**, 49-59
6. Burjanadze, T. V. (1979) *Biopolymers* **18**, 931-938
7. Hakansson, K., and Reid, K. B. (2000) *Protein Sci* **9**, 1607-1617
8. Bächinger, H. P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) *Eur J Biochem* **106**, 619-632
9. Schiene, C., and Fischer, G. (2000) *Curr Opin Struct Biol* **10**, 40-45
10. Ackerman, M. S., Bhate, M., Shenoy, N., Beck, K., Ramshaw, J. A., and Brodsky, B. (1999) *J Biol Chem* **274**, 7668-7673
11. Boudko, S., Frank, S., Kammerer, R. A., Stetefeld, J., Schulthess, T., Landwehr, R., Lustig, A., Bächinger, H. P., and Engel, J. (2002) *J Mol Biol* **317**, 459-470
12. Ottl, J., Musiol, H. J., and Moroder, L. (1999) *J Pept Sci* **5**, 103-110
13. Xu, Y., Bhate, M., and Brodsky, B. (2002) *Biochemistry* **41**, 8143-8151
14. Piez, K. A., and Sherman, M. R. (1970) *Biochemistry* **9**, 4134-4140
15. Snellman, A., Tu, H., Vaisanen, T., Kvist, A. P., Huhtala, P., and Pihlajaniemi, T. (2000) *Embo J* **19**, 5051-5059
16. Areida, S. K., Reinhardt, D. P., Muller, P. K., Fietzek, P. P., Kowitz, J., Marinkovich, M. P., and Notbohm, H. (2001) *J Biol Chem* **276**, 1594-1601
17. Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bächinger, H. P., and Engel, J. (2001) *J Mol Biol* **308**, 1081-1089
18. Davis, M. T., Stahl, D. C., Hefta, S. A., and Lee, T. D. (1995) *Anal Biochem* **67**, 4549-4556
19. van Holde, K. E. (1985) in *Physical Biochemistry*, 2nd edition Ed., pp. 93-136, Prentice Hall, Englewood Cliff, NJ
20. Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S., Landwehr, R., and Engel, J. (2002) *submitted*
21. Bächinger, H. P., Bruckner, P., Timpl, R., and Engel, J. (1978) *Eur J Biochem* **90**, 605-613

**Acknowledgements**

We thank Ariel Lustig for the analytical ultracentrifugation measurements. This work was supported by grants from the Schweizerische Nationalfonds (to J.E.) and from Shriners Hospital for Children (to H.P.B.).

**Figure Legends**

Figure 1

*Thermal stability of foldon- and disulfide knot-containing (GPP)*$_{10}$.

a. Temperature induced unfolding profiles of the collagen-like peptides Cys$_2$-(GPP)$_{10}$ (red curve) and (GPP)$_{10}$-Cys$_2$ (blue curve) monitored by CD following the change of the mean molar residue ellipticity at 221 nm ([$\Theta$]$_{221}$) at a total chain concentration of 50 µM in 5 mM sodium phosphate buffer containing 2 M GuHCl, respectively. The amino acid sequences are GSYGPPGPPCCGSGPP(GPP)$_{10}$ for Cys$_2$-(GPP)$_{10}$ and GS(GPP)$_{10}$GPPGCCGGG for (GPP)$_{10}$-Cys$_2$.  b. Temperature induced unfolding profiles of foldon-(GPP)$_{10}$ (red curve) and (GPP)$_{10}$-foldon (blue curve) monitored by CD at 210 nm in 5 mM phosphate buffer, containing 150 mM NaCl. The heating rate was 10 °C/h. The amino acid sequence of
foldon-(GPP)\textsubscript{10} is GYIPEAPRDGQAYVRKGDGEWVLLSTFLSGS(GPP)\textsubscript{10} and GS(GPP)\textsubscript{10}GSGYIPEAPRDGQAYVRKGDGEWVLLSTFL for (GPP)\textsubscript{10}-foldon.

Figure 2

*Refolding rates of (GPP)\textsubscript{10} with disulfide knots.*

Refolding rates were measured by circular dichroism at 221 nm. a. Refolding kinetics of Cys\textsubscript{2}-(GPP)\textsubscript{10} (red curve) and (GPP)\textsubscript{10}-Cys\textsubscript{2} (blue curve) at 20 °C in 5 mM phosphate buffer containing 2 M GuHCl after a 5 min unfolding at 50 °C for Cys\textsubscript{2}-(GPP)\textsubscript{10} and 70 °C for (GPP)\textsubscript{10}-Cys\textsubscript{2}. Solid lines are experimental data, dashed lines were fitted according to [11].

b. Temperature dependence of the refolding rate of Cys\textsubscript{2}-(GPP)\textsubscript{10} (red circles and line) and (GPP)\textsubscript{10}-Cys\textsubscript{2} (blue circles and line). The activation energies are 53 kJ/mol and 52.5 kJ/mole, respectively.

Figure 3

*Refolding rates of (GPP)\textsubscript{10} with foldon domains.*

Refolding rates were measured by circular dichroism at 210 nm. a. Refolding kinetics of foldon-(GPP)\textsubscript{10} (red curve) at 10 °C and (GPP)\textsubscript{10}-foldon (blue curve) at 7 °C in 5 mM phosphate buffer containing 150 mM NaCl after a 3 min unfolding at 70 °C. Solid lines are experimental data, dashed lines were fitted according to [11]. b. Temperature dependence
of the refolding rate of foldon-(GPP)₁₀ (red circles and line) and (GPP)₁₀-foldon (blue circles and line). The activation energies are 50.0 kJ/mol and 54.5 kJ/mole, respectively.

Table 1

Midpoint transition temperatures \(T_m\) of the (GPP)₁₀-domain in designed proteins and their molecular masses.

|                      | Cys₂⁻(GPP)₁₀ | (GPP)₁₀·Cys₂ | Foldon-(GPP)₁₀ | (GPP)₁₀·Foldon | (PPG)₁₀ |
|----------------------|--------------|--------------|----------------|---------------|---------|
| \(T_m\) in PBS without GuHCl* | 67 °C\(^a\)  | 82 °C\(^a\)  | 53 °C\(^b\)  | 70 °C\(^b\)  | 25 °C\(^a\) |
| \(T_m\) in PBS containing 2 M GuHCl* | 40 °C  | 63 °C  | nd.  | nd.  | <20 °C |
| calculated molecular mass of the monomer (trimer) | 3.85 kDa (11.55 kDa) | 3.46 kDa (10.38 kDa) | 5.88 kDa (17.64 kDa) | 5.88 kDa (17.64 kDa) | 2.53 kDa (7.56 kDa) |
| Molecular mass by mass spectrometry | 11.55 kDa | 10.38 kDa | 5.88 kDa\(^c\) | 5.88 kDa | 2.53 kDa |
| Molecular mass by analytical ultracentrifugation | 11.4 kDa | 10.6 kDa | 18.6 kDa | 18.8 kDa | 6.5 kDa |

\(^a\)measured by circular dichroism at 221 nm

\(^b\)\(T_m\) value of the (GPP)₁₀-moiety measured by circular dichroism at 210 nm

\(^c\)in addition, higher species of 5.88 kDa plus sodium and 5.88 kDa plus sodium and trifluoro acetic acid were obtained.

\(^*\)The \(T_m\) values were determined at total chain concentrations of 50 µM.
Collagen triple helix formation can be nucleated at either end
Sabine Frank, Sergei Boudko, Kazunori Mizuno, Therese Schulthess, Jürgen Engel and Hans Peter Bächinger

J. Biol. Chem. published online January 22, 2003

Access the most updated version of this article at doi: 10.1074/jbc.C200698200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts