organic compounds

Boc–Pro–Hyp–Gly–OBzl and Boc–Ala–Hyp–Gly–OBzl, two repeating triplets found in collagen

Mitsunobu Doi,* Kumiko Imori, Noriko Sakaguchi and Akiko Asano

Osaka University of Pharmaceutical Sciences, 4-20-1 Nasahara, Takatsuki, Osaka 569-1094, Japan
Correspondence e-mail: doit@gly.oups.ac.jp

The protected tripeptides benzyl N-{2-[N-(tert-butoxycarbonyl)prolyl]-4-hydroxyprolyl}glycinate or Boc–Pro–Hyp–Gly–OBzl, $C_{24}H_{33}N_3O_7$, and benzyl N-{2-[N-(tert-butoxycarbonyl)alanyl]-4-hydroxyprolyl}glycinate or Boc–Ala–Hyp–Gly–OBzl, $C_{22}H_{31}N_3O_7$, are the minimum repeating triplets found in collagen. Within the crystal structure of each are two independent peptide molecules with similar structures. The peptides are arranged antiparallel to one another and interact through hydrogen bonds involving the main chains and the 4-hydroxyprolyl groups. The structures exhibit characteristics of a triple helix, but the peptides tend to assume a sheet-like structure.

Comment

Collagen contains a characteristic amino acid sequence in which a Gly residue appears at every third position (X–Y–Gly). Moreover, the Pro–Hyp–Gly triplet (Hyp is 4-hydroxyproline) is frequently found as a repeating sequence, and earlier structural studies carried out with synthetic polypeptides [(Pro–Hyp–Gly)$_n$] indicate that this repeating sequence
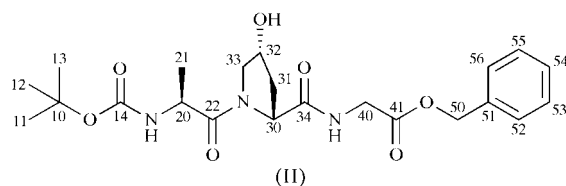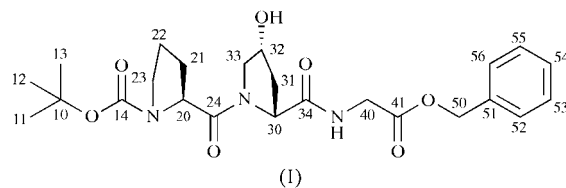
assumes a triple-helical structure (Okuyama et al., 1998, 2004; Berisio et al., 2001). We have used the protected triplet tert-butyloxycarbonyl–Pro–Hyp–Gly benzyl ester or Boc–Pro–Hyp–Gly–OBzl, (I), to investigate the structural characteristics of the collagen minimum repeating sequence. In addition, because (Ala–Hyp–Gly)$_n$ also shows collagen-like optical properties (Rao & Adams, 1979), the structure of Boc–Ala–Hyp–Gly–OBzl, (II), was used to investigate the contribution made by imino acids to the structure of the sequence.

For both (I) and (II), two independent peptide molecules (molecules A and B) are present in the crystal structure; they are arranged antiparallel to one another and are slightly bent
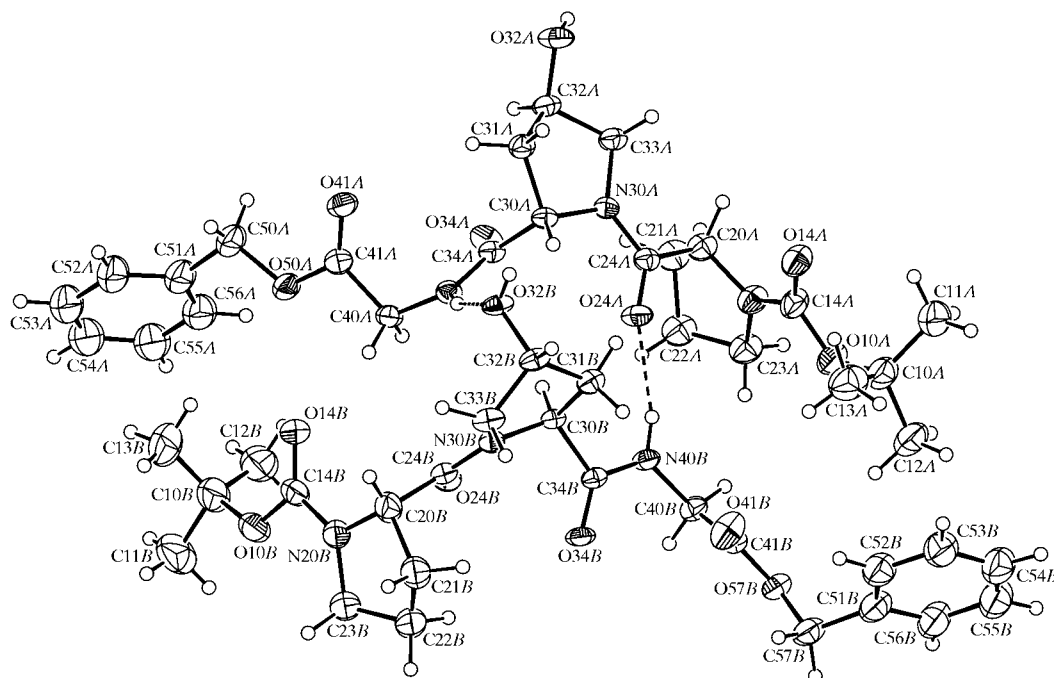

Figure 1
The structure of (I), showing molecules A and B present in the asymmetric unit. Displacement ellipsoids are drawn at the 40% probability level and dashed lines represent hydrogen bonds.

Figure 2
The structure of (II), showing molecules A and B present in the asymmetric unit. Displacement ellipsoids are drawn at the 40% probability level and dashed lines represent hydrogen bonds.
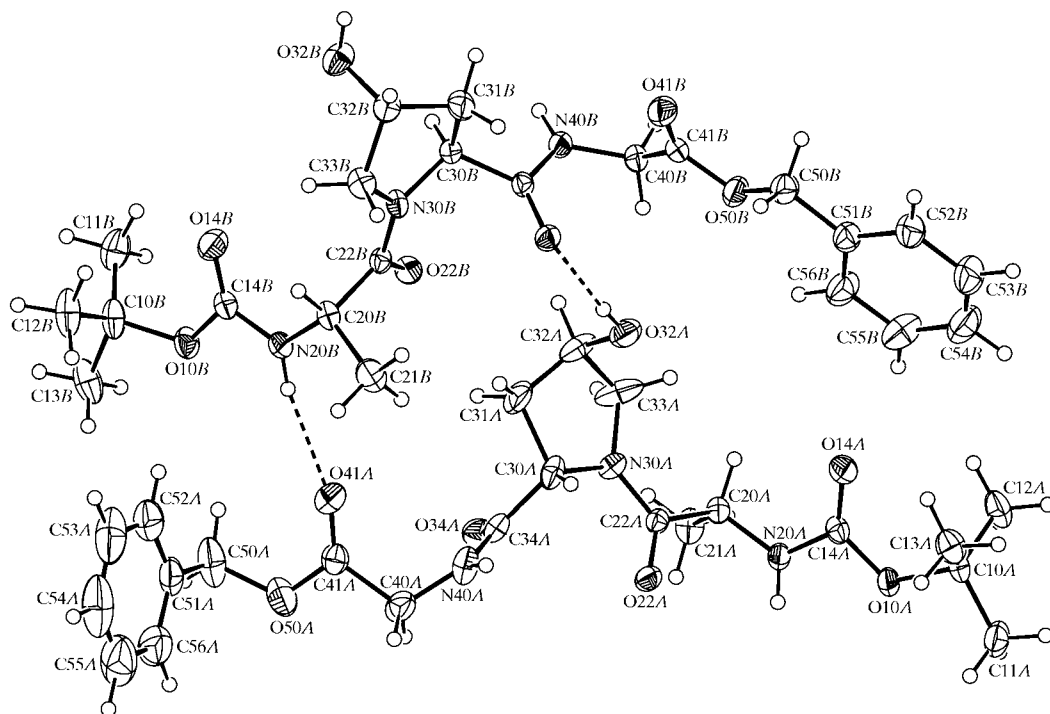

Figure 3
A stereoview of the superposition of (IA), (IB), (IIA) and (IIB). Molecular fitting was performed for the main chains using *ProFit* (Martin, 1996). The figure was drawn using *RasMol* (Bernstein, 2000) and *POV-RAY3.5* (POV-Team, 1991). The r.m.s. deviations are 0.233–0.586 Å. From left to right at the base, the molecules are in the sequence (IIB), (IA), (IIA) and (IIB).

Figure 4
The sheet-like structure of (I). H atoms have been omitted for clarity. Broken lines represent hydrogen bonds.

(Figs. 1 and 2). Within the triple helix of (Pro–Hyp–Gly)$_{10}$, the averaged φ and ψ values are approximately −70/162, −57/150 and −71/173° for the Pro, Hyp and Gly residues, respectively (Okuyama *et al.*, 2004). Neither (I) nor (II) assumes a triple-helical structure, though the torsion angles suggest conformational similarity to a triple helix, despite slight differences at the Ala residues of (II) (Table 1). The torsion angles also show structural similarity among (IA), (IB), (IIA)
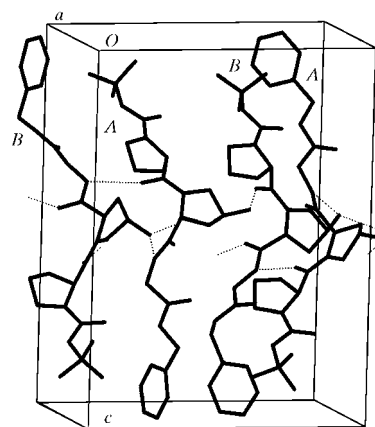
and (IIB). Molecular fitting revealed large deviations at the terminal *tert*-butyloxycarbonyl and benzyl ester groups (Fig. 3), but the conformation of the main chain was conserved among the four molecules. In addition, Hyp is an imino acid showing ring puckering, with amplitudes q_2 (Cremer & Pople, 1975) of 0.340 (5)–0.409 (5) Å. The ring-puckering parameter φ_2 (N30—C30—C31—C32—C33) of 80.6 (7)° in (IA) differs from that of the other three molecules (Table 1). Nevertheless, the main chain conformation around the Hyp unit of (IA) is similar to that of the other molecules.
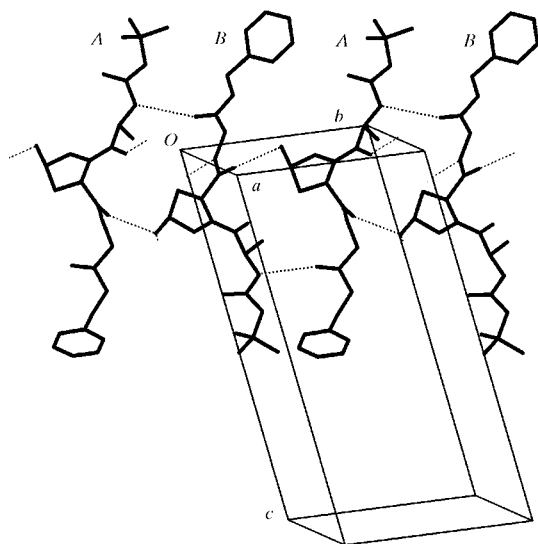
Figure 5
The sheet-like structure of (II). H atoms have been omitted for clarity. Broken lines represent hydrogen bonds.

These peptide molecules, with their similar conformations, make up a sheet-like structure in the solid state (Figs. 4 and 5). Within the packing of (I), four hydrogen bonds are formed (Table 2); two are formed between the main chains (N40$A \cdots$O32B and N40$B \cdots$O24A) and the remaining two involve the 4-hydroxy groups (O32). Atom O32B interacts with atoms N40A and O34B^{i}, acting as a hydrogen-bond acceptor and donor, respectively. By contrast, atom O32A acts only as a donor. Such hydroxy-group-related interactions contribute to the stabilization of the sheet-like structure of (I) (Fig. 4). Within (II), on the other hand, amide groups are situated within hydrogen-bonding range of the Ala and Gly residues (N20 and N40), so that four of the six hydrogen bonds formed are between the main chains (Table 2), thus stabilizing the sheet-like structure (Fig. 5). Increases in such main chain–main chain interactions might be expected to affect the peptide structure, but superimposition showed only a slight deviation between Pro of (I) and Ala of (II) (Fig. 3). A different hydrogen-bonding mode is also observed for the hydroxy groups of (II) (O32A and O32B). Within the triple helix of (Pro–Hyp–Gly)$_{10}$, the 4-hydroxy groups of Hyp participate in hydrogen bonds with solvent molecules and intertriple-helices but not in interstrand interactions (Berisio *et al.*, 2001; Kishimoto *et al.*, 2005). It was also suggested that the 4-position of Hyp shows an inductive effect from 4-fluoroproline-incorporated collagen-like peptide (Holmgren *et al.*, 1998). It is noteworthy that the 4-hydroxy groups in both (I) and (II) play different roles in the hydrogen bonding of molecules A and B.

Experimental

Compounds (I) and (II) were synthesized by the conventional liquid phase method and purified by silica-gel column chromatography. Crystals were grown from methanol and ethyl acetate solutions.

Compound (I)

Crystal data

C$_{24}$H$_{33}$N$_3$O$_7$
$M_r = 475.53$
Monoclinic, $P2_1$
$a = 8.922\ (2)$ Å
$b = 15.208\ (3)$ Å
$c = 18.692\ (4)$ Å
$\beta = 91.523\ (3)°$
$V = 2535.1\ (9)$ Å^3

$Z = 4$
$D_x = 1.246$ Mg m^{-3}
Mo $K\alpha$ radiation
$\mu = 0.09$ mm^{-1}
$T = 200\ (2)$ K
Plate, colorless
$0.45 \times 0.30 \times 0.08$ mm

Data collection

Bruker SMART APEX area-detector diffractometer
ω scan
Absorption correction: empirical (using intensity measurements) (*SADABS*; Sheldrick, 1996)
$T_{min} = 0.883$, $T_{max} = 0.993$

21865 measured reflections
5790 independent reflections
4804 reflections with $I > 2\sigma(I)$
$R_{int} = 0.028$
$\theta_{max} = 27.1°$

Refinement

Refinement on F^2
$R[F^2 > 2\sigma(F^2)] = 0.056$
$wR(F^2) = 0.162$
$S = 1.11$
5790 reflections
627 parameters
H atoms treated by a mixture of independent and constrained refinement

$w = 1/[\sigma^2(F_o^2) + (0.0823P)^2 + 1.2083P]$
where $P = (F_o^2 + 2F_c^2)/3$
$(\Delta/\sigma)_{max} = 0.004$
$\Delta\rho_{max} = 0.51$ e Å^{-3}
$\Delta\rho_{min} = -0.46$ e Å^{-3}

Compound (II)

Crystal data

C$_{22}$H$_{31}$N$_3$O$_7$
$M_r = 449.50$
Triclinic, $P1$
$a = 5.8816\ (3)$ Å
$b = 10.6294\ (6)$ Å
$c = 19.707\ (1)$ Å
$\alpha = 75.240\ (1)°$
$\beta = 82.560\ (1)°$
$\gamma = 81.780\ (1)°$

$V = 1173.6\ (1)$ Å^3
$Z = 2$
$D_x = 1.272$ Mg m^{-3}
Mo $K\alpha$ radiation
$\mu = 0.10$ mm^{-1}
$T = 200\ (2)$ K
Plate, colorless
$0.45 \times 0.35 \times 0.20$ mm

Data collection

Bruker SMART APEX CCD area-detector diffractometer
ω scan
Absorption correction: empirical (using intensity measurements) (*SADABS*; Sheldrick, 1996)
$T_{min} = 0.927$, $T_{max} = 0.981$

13595 measured reflections
5142 independent reflections
5049 reflections with $I > 2\sigma(I)$
$R_{int} = 0.015$
$\theta_{max} = 27.1°$

Refinement

Refinement on F^2
$R[F^2 > 2\sigma(F^2)] = 0.038$
$wR(F^2) = 0.100$
$S = 1.02$
5142 reflections
593 parameters
H atoms treated by a mixture of independent and constrained refinement

$w = 1/[\sigma^2(F_o^2) + (0.0684P)^2 + 0.1501P]$
where $P = (F_o^2 + 2F_c^2)/3$
$(\Delta/\sigma)_{max} = 0.002$
$\Delta\rho_{max} = 0.43$ e Å^{-3}
$\Delta\rho_{min} = -0.16$ e Å^{-3}

H atoms were placed at calculated positions (C—H = 0.95–1.00 Å and N—H = 0.88 Å), with isotropic displacement parameters [U_{iso}(H) = 1.5U_{eq}(C) for methyl H atoms and 1.2U_{eq}(C,N,O) for all other H atoms], and included in the structure-factor calculations. H atoms of hydroxy groups were found in a difference Fourier map by

organic compounds

Table 1
Selected torsion angles and ring-puckering parameters ($^\circ$, Å) of Hyp.

Bond		(IA)	(IB)	(IIA)	(IIB)	1V7H*
C14—N20—C20—C22	(φ1)	−71.1 (6)	−69.9 (7)	−100.6 (5)	−76.1 (6)	−70.0
N20—C20—C22—N30	(ψ1)	164.0 (4)	174.2 (4)	154.2 (4)	144.2 (5)	162.3
C22—N30—C30—C34	(φ2)	−73.8 (4)	−63.1 (4)	−66.1 (7)	−53.8 (6)	−57.0
N30—C30—C34—N40	(ψ2)	148.8 (3)	171.7 (3)	144.7 (5)	153.6 (4)	149.6
C34—N40—C40—C41	(φ3)	−70.9 (5)	−66.2 (5)	−73.1 (8)	−58.6 (6)	−71.1
N40—C40—C41—O50	(ψ3)	176.5 (4)	179.2 (3)	161.7 (6)	152.9 (4)	173.4
q_2**		0.340 (5)	0.409 (5)	0.365 (9)	0.380 (7)	0.353
φ_2**		80.6 (7)	285.6 (6)	282 (1)	267.1 (9)	280.1

*PDB entry number: averaged torsion angles are from Okuyama *et al.* (2004). **Five-membered ring-puckering parameters are defined by Cremer & Pople (1975) and calculated by *PLATON* (Spek, 2003).

Table 2
Hydrogen bond geometry (Å, $^\circ$).

D—H$\cdots A$	D—H	$D\cdots A$	H$\cdots A$	D—H$\cdots A$
(I)				
N40A—H40$A\cdots$O32B	0.88	3.033 (5)	2.18	162
N40B—H40$B\cdots$O24A	0.88	2.914 (4)	2.06	163
O32B—H32$B\cdots$O34B^i	0.75 (7)	2.729 (4)	1.99 (8)	172 (8)
O32A—H32$A\cdots$O24B^{ii}	0.80 (8)	2.849 (5)	2.12 (8)	152 (8)
(II)				
O32A—H32$A\cdots$O34B	0.76 (8)	2.807 (7)	2.05 (8)	175 (8)
N20B—H20$B\cdots$O41A	0.88	3.073 (8)	2.22	164
N40A—H40$A\cdots$O32B^{iii}	0.88	2.844 (7)	2.05	149
N20A—H20$A\cdots$O41B^{iv}	0.88	3.196 (7)	2.32	172
N40B—H40$B\cdots$O22A^v	0.88	2.816 (6)	1.94	173
O32B—H32$B\cdots$O34A^{vi}	0.8 (1)	2.666 (8)	1.8 (1)	172 (12)

Symmetry codes: (i) $-x, y + \frac{1}{2}, -z + 1$; (ii) $-x + 1, y + \frac{1}{2}, -z + 1$; (iii) $x, y + 1, z$; (iv) $x + 1, y + 1, z$; (v) $x, y - 1, z$; (vi) $x - 1, y - 1, z$.

taking the hydrogen-bonding network into consideration, and refined. Restraints for displacement parameters were applied for (I) and (II) using SIMU and DELU instructions in *SHELXL97* (Sheldrick, 1997). In the absence of significant anomalous scattering effects, Friedel pairs were merged, and the absolute configuration was assumed from that of the starting material.

For both compounds, data collection: *SMART* (Bruker, 1998); cell refinement: *SAINT-Plus* (Bruker, 1998); data reduction: *SAINT-Plus*; program(s) used to solve structure: *SHELXS97* (Sheldrick, 1990) for (I) and *SHELXD* (Sheldrick & Gould, 1995) for (II); program(s) used to refine structure: *SHELXL97* (Sheldrick, 1997); molecular graphics: *PLATON* (Spek, 2003); software used to prepare material for publication: *SHELXL97*.

References

Berisio, R., Vitagliano, L., Mazzarella, L. & Zagari, A. (2001). *Biopolymers*, **56**, 8–13.
Bernstein, H. J. (2000). *Trends Biochem. Sci.* **25**, 453–455.
Bruker (1998). *SAINT-Plus* (Version 5) and *SMART* (Version 5). Bruker AXS Inc., Madison, Wisconsin, USA.
Cremer, D. & Pople, J. A. (1975). *J. Am. Chem. Soc.* **97**, 1354–1358.
Holmgren, S. K., Taylor, K. M., Bretscher, L. E. & Raines, R. T. (1998). *Nature (London)*, **392**, 666–667.
Kishimoto, T., Morihara, Y., Osanai, M., Ogata, S., Kamitakahara, M., Ohtsuki, C. & Tanihara, M. (2005). *Biopolymers*, **79**, 163–172.
Martin, A. C. R. (1996). *ProFit.* Version 2.5. University College London, England.
Okuyama, K., Hongo, C., Fukushima, R., Wu, G., Narita, H., Noguchi, K., Tanaka, Y. & Nishino, N. (2004). *Pept. Sci.* **76**, 367–377; erratum in *Biopolymers* (2005), **80**, 714–715.
Okuyama, K., Nagarajan, V., Kamitori, S. & Noguchi, K. (1998). *Chem. Lett.* pp. 385–386.
POV-Team (1991). *POV-RAY3.5.* Persistence of Vision Ray Tracer, Williamstown, Victoria, Australia. (URL: http://www.povray.org.)
Rao, N. V. & Adams, E. (1979). *Biochem. Biophys. Res. Commun.* **86**, 654–660.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (1996). *SADABS.* University of Göttingen, Germany.
Sheldrick, G. M. (1997). *SHELXL97.* University of Göttingen, Germany.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.