Conformations in the Solid State and Solubility Properties of Protected Homooligopeptides of Glycine and β-Alanine[1)]

Mitsuaki NARITA,* Masamitsu DOI, Koji KUDO, and Yusuke TERAUCHI
Department of Industrial Chemistry, Faculty of Technology, Tokyo University of Agriculture and Technology, Koganei 184
(Received May 23, 1986)

IR spectroscopic conformational analyses of Boc-Gly$_n$-OBzl (n=3—7) and Boc-(β-Ala)$_n$-OBzl (n=3—8) were performed in the solid state, suggesting the occurrence of the β-sheet structure in the higher oligomers (n=5—8). Solubility data indicate that insolubilities of Boc-Gly$_n$-OBzl and Boc-(β-Ala)$_n$-OBzl in high-polar solvents begin at hexa- and heptapeptide levels, respectively. Insolubility of protected homooligopeptides of Gly and β-Ala was estimated to be caused by β-sheet aggregation. The high potential for the β-sheet formation of Boc-Gly$_n$-OBzl and Boc-(β-Ala)$_n$-OBzl ($n\geqq5$) could clearly be attributed to the great freedom of the peptide backbone dihedral angles of each of the Gly and β-Ala residues in the β-sheet structure. The implications of a replacement of a few Gly residues with β-Ala residues in surface regions of proteins are also discussed.

The solubility prediction method proposed by us is principally based on the hypothesis that the insolubility of protected peptide intermediates is essentially caused by a β-sheet aggregation consisting of peptide chains equal to or larger than an octapeptide sequence level;[2)] it provides the scope of the design of synthetic routes for peptides and proteins.[2-4)] In the solubility prediction method, the coil conformation parameters, P_c, for each amino acid residue are used for estimating the potential for the β-sheet formation. Here, the value of P_c for the Gly residue is 1.50, which is similar to the value of P_c for the Pro residue, 1.59. These P_c values represent the fact that in globular proteins Gly residues appear in coil conformation regions[5)] as frequently as Pro residues, and that Gly-rich peptides are readily soluble in organic solvents. On the other hand, considering the high flexibility of the peptide backbones of the Gly-rich peptides,[6)] peptides larger than a certain peptide size appear to be insoluble in high-polar solvents due to the β-sheet aggregation.

Therefore, in this work, an investigation of the relationship between the conformation and solubility of protected homooligopeptides of Gly and β-Ala was carried out to confirm the generality of the above hypothesis. The second purpose in this study was to demonstrate that protected homooligopeptides of Gly and β-Ala have a high potential for β-sheet formation due to the great freedom of the peptide backbone dihedral angles of each of Gly and β-Ala residues in the β-sheet structure. In relation to the creation of novel proteins with improved properties, which can not be produced by a recombinant DNA method,[7)] we have discussed the effect of the replacement of a few Ala residues with Aib residues; these have the ability to promote helical folding in helical regions of proteins.[8)] The final purpose in this study was to discuss the possibility and implications of the replacement of the Gly residue with the β-Ala residue in proteins through the investigation of the similarity of conformational behaviors between the Gly and β-Ala residues.

Experimental

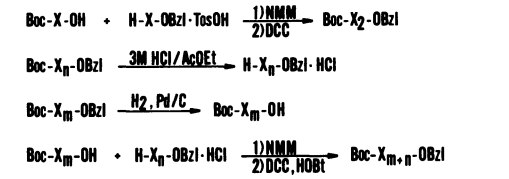
General. Uncorrected capillary melting points will be reported. Analytical instruments and conditions were described in a previous paper.[8)]

Synthesis of the Peptides 1—6. The tripeptides **1a** and **1b** were prepared by the usual stepwise elongation, starting with H-Gly-OBzl and H-β-Ala-OBzl as amino components, respectively, in the conditions described before for the preparation of oligopeptides.[8)] The higher oligopeptides **2—6** were prepared by fragment condensation in dichloromethane or DMF using DCC activation in the presence of HOBt as described before for fragment condensation.[8)] Procedures for the removal of the Boc group from the peptides (n=2—4) and of the Bzl group from the peptides (n=2—4) were essentially the same to those described previously.

IR Measurements. The IR absorption spectra of the solid samples were recorded with a Jasco Model DS-701G spectrometer in a nujol mull.

Results

Synthesis and Solubility Properties of Protected Homooligopeptides of Gly and β-Ala. The synthetic route for Boc-Gly$_n$-OBzl (n=3—7) and Boc-(β-Ala)$_n$-OBzl (n=3—8) is set out in the Scheme. The tripeptides **1a** and **1b** (n=3) were prepared by the usual stepwise elongation, starting with H-Gly-OBzl and H-β-Ala-OBzl as amino components, respectively.



X \ m+n	3(1+2)	4(2+2)	5(2+3)	6(3+3)	7(3+4)	8(4+4)
Gly	1a	2a	3a	4a	5a	
β-Ala	1b	2b	3b	4b	5b	6b

Scheme

Table 1. Synthetic Results and Physical Properties of the Peptides **1—6**

Compound	Yield/%[a]	Recrystallization Solvent	Mp θ_m/°C
1a	84	AcOEt/Hexane (1/1)	122—124
2a	76	EtOH/Water (3/1)	177—180
3a	80	b)	Over 200
4a	84	b)	Over 200
5a	78	b)	Over 200
1b	89	EtOH/Water (2/1)	120—121
2b	65	EtOH/Water (1/1)	168—171
3b	66	EtOH/Water (3/2)	Over 200
4b	88	b)	Over 200
5b	83	b)	Over 200
6b	65	b)	Over 200

a) Coupling yields in final steps. b) Purified by washing with hot MeOH.

Both amino components were obtained from the corresponding TosOH salts with an equivalent of NMM. The stepwise elongation was performed in dichloromethane using DCC activation.[9] For the preparation of the higher oligopeptides **2—6**, the removal of the Boc group[10] from the peptides (n=2—4) was performed in 3 M HCl/AcOEt (1 M=1 mol dm^{-3}), and removal of the Bzl group,[11] in AcOH/MeOH by hydrogenolysis in the presence of Pd/C as catalyst, respectively. The higher oligopeptides **2—6** were obtained by fragment condensation in dichloromethane or DMF using DCC activation in the presence of HOBt.[12] The peptides **1—6** were purified by repeated recrystallization or by washing with hot MeOH. The synthetic results and elemental analyses of the peptides **1—6** are given in Tables 1 and 2. Their solubility properties are also shown in Table 3.

Table 2. Elemental Analyses of the Peptides **1—6**

Compound	Formula	Found (Calcd)		
		C/%	H/%	N/%
1a	$C_{18}H_{25}N_3O_6$	56.94 (56.98)	6.59 (6.64)	11.06 (11.07)
2a	$C_{20}H_{28}N_4O_7$	54.93 (55.04)	6.40 (6.47)	12.73 (12.84)
3a	$C_{22}H_{31}N_5O_8 \cdot 0.5H_2O$	52.37 (52.58)	6.21 (6.42)	14.03 (13.94)
4a	$C_{24}H_{34}N_6O_9 \cdot 0.5H_2O$	51.78 (51.51)	6.05 (6.30)	15.26 (15.01)
5a	$C_{26}H_{37}N_7O_{10} \cdot H_2O$	49.90 (49.92)	6.14 (6.28)	15.70 (15.67)
1b	$C_{21}H_{31}N_3O_6$	59.98 (59.82)	7.60 (7.40)	10.00 (9.97)
2b	$C_{24}H_{36}N_4O_7$	58.73 (58.50)	7.56 (7.38)	11.34 (11.37)
3b	$C_{27}H_{41}N_5O_8$	57.55 (57.53)	7.45 (7.33)	12.39 (12.43)
4b	$C_{30}H_{46}N_6O_9 \cdot 1.5H_2O$	54.34 (54.45)	7.23 (7.46)	12.59 (12.70)
5b	$C_{33}H_{51}N_7O_{10} \cdot 0.5H_2O$	55.51 (55.45)	7.31 (7.33)	13.77 (13.72)
6b	$C_{36}H_{56}N_8O_{11} \cdot 9/4H_2O$	52.74 (52.90)	7.01 (7.46)	13.97 (13.71)

Table 3. Solubility Properties[a] of the Peptides **1—6** (c=1.0 g dl^{-1})

Compound	Solvent						
	High-polar solvent[b]	MeOH EtOH	$CHCl_3$	AcOEt	CH_2Cl_2	Acetone	CCl_4
1a	A	A	A	A	A	B	D
2a	A	A	B	B	B	B	D
3a	A	D	D	D	D	D	D
4a	D, C[c]	D	D	D	D	D	D
5a	D	D	D	D	D	D	D
1b	A	A	A	A	A	D	B
2b	A	A	A	A	B	B	D
3b	A	B	B	D	B	C	D
4b	B, A[c]	D	D	D	D	D	D
5b	D, C[c]	D	D	D	D	D	D
6b	D, C[c]	D	D	D	D	D	D

a) Solubility: A, soluble at room temperature; B, soluble at 80° or refluxing temperature; C, partially soluble at 80° or refluxing temperature; D, practically insoluble at 80° or refluxing temperature. b) DMF, NMP, DMA, DMSO, and HMPA. c) Solubility in NMP and HMPA.

IR Absorption Study in the Solid State of Protected Homooligopeptides of Gly and β-Ala. The IR absorption spectra in the amide A and amide I regions of the peptides **1a—5a** and **1b—6b** in the solid state are presented in Figs. 1 and 2, respectively. The assignments of the various bands of the peptides **1—6** were made on the basis of theoretical and experimental data for oligopeptides.[13-16] The peptides **1** and **2** (n=3 and 4) in Figs. 1 and 2 show strong bands around 3290 cm^{-1} and a few bands in the amide A region, indicating a predominant conformation with hydrogen bonds, most probably of the interchain type, but quite different from the usual antiparallel β-sheet structure, as assumed on Boc-Gly_n-OPEG (n=3 and 4).[15] However, details of the IR absorption frequencies of Boc-Gly_n-OPEG (n=3 and 4) are different from those of Boc-Gly_n-OBzl (n=3 and 4) due to the effect of the bulky and bifunctional PEG. In the case of the peptide **1b**, the absorption band of the ester carbonyl group shifts around 1715 cm^{-1}, indicating its hydrogen bonding with a N-H group. When n=5 the peptides **3a** and **3b** have a strong band around 3290 cm^{-1} in the amide A region and weak and strong bands around 1685 cm^{-1} and 1630 cm^{-1}, respectively, in the amide I region, accompanied by a shoulder around 1655—1645 cm^{-1}. These bands indicate the predominant antiparallel β-sheet and other coexistent conformations.[15] These results are quite different from the ternary helix conformation reported on CH_3CO-Gly_n-NHC_2H_5 (n=1—4) and poly$(Gly)_n$,[16] which have strong characteristic absorption bands at 3285 cm^{-1} in the amide A region, 1646 cm^{-1} in the amide I region, and 1560 cm^{-1} in the amide II region. The peptides **4—6** having n=6—8 assume predominantly the antiparallel β-sheet structure as reported on Boc-Gly_n-OPEG (n=7—9),[15] although the broad bands around 1640—1630 cm^{-1} suggests the coexistence of other conformations. Through an IR conformational analysis in the solid state of the peptides **1—6**, the conformational behaviors of each oligo(β-Ala) are very close to those of the corresponding oligo(Gly).

Discussion

By using the IR absorption spectra in the solid state of the peptides **1—6**, the occurrence of the β-sheet structure was actually established for Boc-Gly_n-OBzl (n=5—7) and Boc-$(\beta\text{-Ala})_n$-OBzl (n=5—8), and insolubilities in high-polar solvents of Boc-Gly_n-OBzl and Boc-$(\beta\text{-Ala})_n$-OBzl begin at hexa- and heptapeptide levels, respectively. As in the case of protected homooligo-L-leucines,[17] protected oligopeptides having various hydrophobic side chains,[8] and protected oligopeptides having various polar side chains,[3] insolubility of the homooligopeptides of Gly and β-Ala was estimated to be mainly caused by β-sheet aggregation. Boc-Gly_n-OBzl (n=6 and 7) and Boc-$(\beta\text{-Ala})_n$-OBzl (n=7 and 8) are practically insoluble even in the high-polar solvents such as DMF, DMA, NMP, DMSO, and HMPA (Table 3). Smaller critical sizes for the insolubility of Boc-Gly_n-OBzl and Boc-$(\beta\text{-Ala})_n$-OBzl than protected other oligopeptides[3,8,17] are probably due to the high symmetry of peptide backbones of homooligopeptides of Gly and β-Ala.

According to the solubility prediction method based on the $\langle P_c \rangle$ values of the average coil conformation

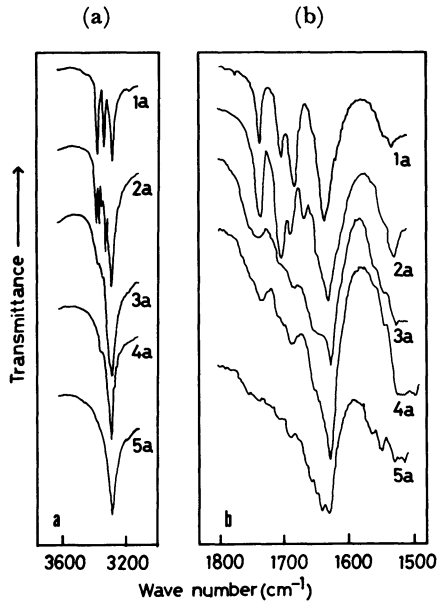


Fig. 1. IR absorption spectra of the peptides **1a—5a**. a: The amide A region; b: the amide I region.

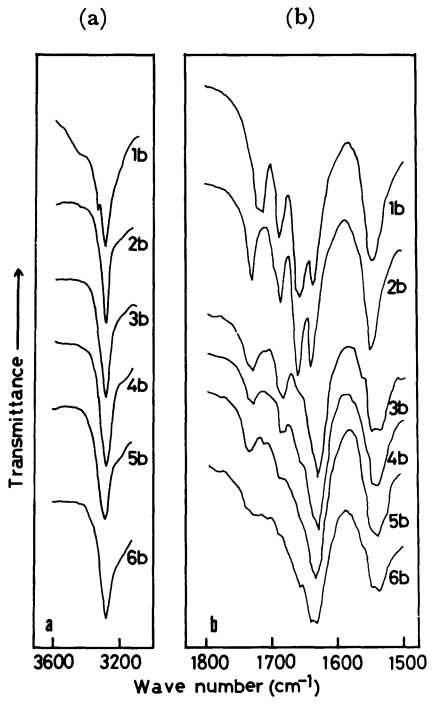


Fig. 2. IR absorption spectra of the peptides **1b—6b**. a: The amide A region; b: the amide I region.

parameters of protected homooligo(Gly)s, they are predicted to be easily soluble due to the great freedom of the backbone dihedral angles ϕ and ψ of the Gly residue.[2] Actually, the Gly residue can explore a large conformational space,[6] thus inducing a high flexibility to the peptide chain. The great freedom of the backbone dihedral angles ϕ and ψ of the Gly residue is essentially attributed to the fact that there is no substituent on the C^{α} atom of the Gly residue. The great freedom is just due to the release in the Gly residue of the steric hindrance between O_{i-1} and C^{β}_{i} (postulated C^{β} atom on C^{α}_{i}), between N_{i+1} and C^{β}_{i}, and between H_{i+1} and C^{β}_{i} as illustrated in Fig. 3a. This fact results in the large value of W_{free}, which leades the peptide chain to various conformations:

$$\Delta S_{\beta\text{-sheet}} = S_{\beta\text{-sheet}} - S_{\text{free}} = R \ln W_{\beta\text{-sheet}} - R \ln W_{\text{free}}, \quad (1)$$

where $S_{\beta\text{-sheet}}$ is the entropy value in the β-sheet structure and S_{free} that in the free structure.[18] R is the gas constant, $W_{\beta\text{-sheet}}$ the probability of a β-sheet structure and W_{free} that of the free structure. On the other hand, the high potency for a β-sheet formation of Boc-Gly_n-OBzl ($n \geqq 5$) represents the large value of $\Delta S_{\beta\text{-sheet}}$ in Eq. 1, thus inducing the large value of $W_{\beta\text{-sheet}}$ in Eq. 1. The large value of $W_{\beta\text{-sheet}}$ corresponds to the great freedom in the β-sheet structure of the backbone dihedral angles ϕ and ψ of each Gly residue of Boc-Gly_n-OBzl. If the value of $W_{\beta\text{-sheet}}$ in the β-sheet structure of Boc-Gly_n-OBzl is as small as the values of W of various other conformations, Boc-Gly_n-OBzl inevitably has various conformations, resulting in coil conformation structures. The conformational behavior of Boc-(β-Ala)$_n$-OBzl is, essentially, quite similar to that of Boc-Gly_n-OBzl. The freedom of the peptide backbone of Boc-(β-Ala)$_n$-OBzl is much greater than that of Boc-Gly_n-OBzl due to an additional release in the β-Ala residue of the steric hindrance between O_{i-1} and O_i, between O_{i-1} and C^{β}_{i}, between O_{i-1} and N_{i+1}, between H_{i+1} and O_{i-1}, and between H_i and H_{i+1} as illustrated in Fig. 3b. Due to the fact that there is no substituent on the C^{α} and C^{β} atoms of the β-Ala residue, each β-Ala residue in the β-sheet structure can explore the large conformational space, inducing the high flexibility to the peptide chain in the β-sheet structure. Using the CPK models of the antiparallel β-sheet structures of Boc-Gly_7-OBzl and Boc-(β-Ala)$_7$-OBzl, we can actually confirm the great freedom of the methylene groups in the antiparallel β-sheet structures of Boc-Gly_7-OBzl and Boc-(β-Ala)$_7$-OBzl.

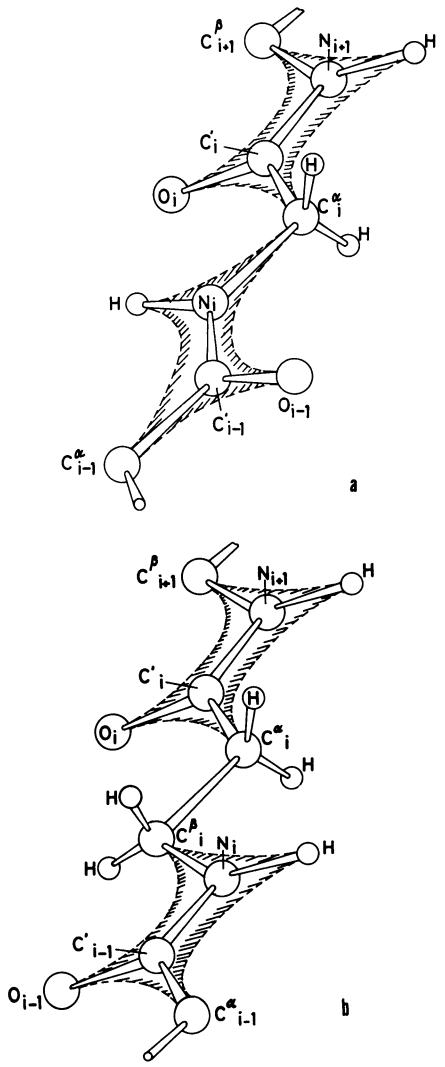


Fig. 3. Release of steric hindrance (a) in the Gly residue and (b) in the β-Ala residue.

The large value of W_{free} of the Gly residue compared with other amino acid residues corresponds to the fact that Gly residues in globular proteins have various values of the backbone dihedral angles ϕ and ψ, as shown by X-ray diffraction studies of globular proteins. Thus, the Gly residue often appears in turn structures,[19] that is, in surface regions of globular proteins. Since the β-Ala residue has a value of W_{free} that is larger than the Gly residue, it is estimated that we can suitably replace the Gly residue with the β-Ala residue in surface regions of globular proteins without changing the protein structures essentially. Moreover, active sites of proteins are usually located on the surface regions of the proteins, especially clefts or crevices, and the packing density in the active sites is usually low.[6] These facts mean that the active sites have a greater freedom than the other regions. Therefore, the replacement of a few Gly residues with β-Ala residues in surface regions of proteins offers the prospect for creating novel proteins which can not be produced by the recombinant DNA method.[7] The replacement techniques also offer the potential for altering protein functions in ways not possible by any other method, just as mentioned in protein engineering pioneered by genetic technology.[20]

References

1) Design for the Synthetic Route of Peptides and Proteins. V. For part IV of this series, see M. Narita, S. Isokawa, M. Doi, and R. Wakita, *Bull. Chem. Soc. Jpn.*, **59**, 3547 (1986). Abbreviations used are the following: Gly, glycine residue; β-Ala, β-alanine residue; Boc, *t*-butoxycarbonyl; Bzl, benzyl; IR, infrared; Aib, α-aminoisobutyric acid residue; DNA, deoxyribonucleic acid; DCC, dicyclohexylcarbodiimide; DMF, *N*,*N*-dimethylformamide; HOBt, 1-hydroxy-1*H*-benzotriazole; MeOH, methanol; TosOH, *p*-toluenesulfonic acid; NMM, *N*-methylmorpholine; AcOEt, ethyl acetate; AcOH, acetic acid; DMA, *N*,*N*-dimethylacetamide; NMP, *N*-methylpyrrolidone; DMSO, dimethyl sulfoxide; HMPA, hexamethylphosphoric triamide; PEG, polyethylene glycol.

2) M. Narita, K. Ishikawa, J.-Y. Chen, and Y. Kim, *Int. J. Peptide Protein Res.*, **24**, 580 (1984).

3) M. Narita, T. Ogura, K. Sato, and S. Honda, *Bull. Chem. Soc. Jpn.*, **59**, 2433 (1986); **59**, 2439 (1986), **59**, 2445 (1986).

4) M. Narita, S. Isokawa, M. Doi, and R. Wakita, *Bull. Chem. Soc. Jpn.*, **59**, 3547 (1986).

5) Coil conformation regions in this definition include β-turn regions.

6) G. E. Schulz and R. H. Schirmer, "Principles of Protein Structure" Springer-Verlag GmbH & Co. KG (1979).

7) K. Itakura, T. Hirose, R. Crea, and A. D. Riggs, *Science*, **198**, 1056 (1977).

8) M. Narita, M. Doi, H. Sugasawa, and K. Ishikawa, *Bull. Chem. Soc. Jpn.*, **58**, 1473 (1985); M. Narita, K. Ishikawa, H. Sugasawa, and M. Doi, *Bull. Chem. Soc. Jpn.*, **58**, 1731 (1985); M. Narita, J.-Y. Chen, H. Sato, and Y. Kim, *Bull. Chem. Soc. Jpn.*, **58**, 2494 (1985).

9) J. C. Sheehan and G. P. Hess, *J. Am. Chem. Soc.*, **77**, 1067 (1955); H. G. Khorana, *Chem. Rev.*, **53**, 145 (1953).

10) F. C. McKay and W. F. Albertson, *J. Am. Chem. Soc.*, **79**, 4686 (1957).

11) E. Abderhalden and S. Suzuki, *Hoppe-Seyler's Z. Physiol. Chem.*, **176**, 101 (1958).

12) W. König and R. Geiger, *Chem. Ber.*, **103**, 788 (1970).

13) T. Miyazawa and E. R. Blout, *J. Am. Chem. Soc.*, **83**, 712 (1961); Yu. N. Chirgadze and N. A. Nevskaya, *Biopolymers*, **15**, 627 (1976); V. N. R. Pillai and M. Mutter, *Acc. Chem. Res.*, **14**, 122 (1981), and references cited therein.

14) M. Narita, S. Isokawa, Y. Tomotake, and S. Nagasawa, *Polymer J.*, **15**, 25 (1983), and references cited therein.

15) C. Toniolo, G. M. Bonora, V. N. R. Pillai and M. Mutter, *Macromolecules*, **13**, 772 (1980).

16) M. Avignon and C. Garrigou-Lagrange, *Spectrochim. Acta*, Part A, **27**, 297 (1971).

17) M. Narita, T. Fukunaga, A. Wakabayashi, K. Ishikawa, and H. Nakano, *Int. J. Peptide Protein Res.*, **23**, 306 (1984).

18) The free structure represents all of the conformations allowed.

19) P. Y. Chou and G. D. Fasman, *Biochemistry*, **13**, 222 (1974); P. Y. Chou and G. D. Fasman, *Adv. Enzymol.*, **47**, 45 (1978).

20) K. M. Ulmer, *Science*, **219**, 666 (1983).