CONTRIBUTION OF THE α2 CHAIN TO THE MOLECULAR STABILITY OF COLLAGEN

Wolfie TRAUB* and Peter P. FIETZEK**

*Department of Structural Chemistry, Weizmann Institute of Science, Rehovot, Israel and **Abteilung für Bindegewebsforschung, Max-Planck Institut für Biochemie, Martinsried bei München, Germany

Received 8 July 1976

1. Introduction

Collagen, the main protein of connective tissue, has long stiff rod-like molecules consisting of three polypeptide chains wound about a common axis in a triple helical structure [1]. Each chain has a little over one thousand amino acid residues, including glycine in every third position, except near the ends of the chains, and unusually large amounts of the imino acids proline and hydroxyproline which help to stabilise the molecular conformation. In many types of collagen all three chains have identical sequences, but the predominant type of collagen in most vertebrate tissues has two identical chains, designated α1, and a third, α2, of somewhat different sequence. All three chains extend over the full length of the molecule and have their N-terminal ends on the same side.

The left-handed helical symmetry of the collagen molecule relates the structurally equivalent Gly. X.Y. tripeptides by a translation of 2.9 Å and a rotation of some 108° (fig.1). Because every third residue lies near the central axis of the molecule there is room in this position for only the smallest amino acid, glycine. Residues in both the X and Y positions lie on the surface of the molecule, but differ in their backbone conformations and intramolecular stereochemical environments. The three chains are joined by systematic hydrogen bonding between NH groups of glycine residues and backbone CO groups of residues in X positions.

With the determination of amino acid sequences of α1 chains from several vertebrate collagens [2] it became clear that some amino acid residues were non-randomly distributed between X and Y positions [3–5], and explanations were proposed in terms of various steric contacts which affect molecular stability [6,7]. Some of these are stabilising interactions between specific pairs of residues in steric proximity, but on different polypeptide chains, and might involve either α1–α1 or α1–α2 chain combinations. In the last few years more than 80% of the calf-skin collagen α2 sequence has been determined and it has been found to differ from the α1 sequence in about half the X and Y positions [2]. We have now found that, in spite of these differences, the α2 sequence has much the same regularities as have been found for α1, that the combined sequences indicate the existence of a considerable number of side-chain–specific interchain interactions, and that the ends of the chains, which are separated by axial translations of one residue (fig.1), are probably in the order α1–α2–α.

2. Results

Table 1 shows the distribution of amino acid residues between X and Y position in the helical region of calfskin collagen for all 337 tripeptides of the α1 chain and some 280 tripeptides of the α2 chain. It is evident that the two chains show much the same uneven distributions for several residues. Almost all prolines and some lysines are hydroxylated after polypeptide chain synthesis by enzymes which act specifically at Y positions [8–10]. The almost complete restriction of phenylalanine, leucine, histidine and tyrosine to X positions has been explained in terms of steric hindrance for these residues at Y positions [6], and the less extreme positional of glutamine, arginine, lysine, glutamic acid and threonine

Table 1
Distribution of amino acid residues between X and Y positions for α1 and α2 chains

Amino acid	α1 chains X/Y	α2 chains X/Y
Pro	116/4	90/3
Hyp*	1/112	0/83
Phe	12/0	10/0
Leu	18/1	25/5
Arg	9/42	9/35
Lys	12/20	8/10
Gln	7/19	7/17
Asn	6/5	13/7
Glu	42/6	31/3
Thr	3/12	4/9
Ser	18/17	12/12
Asp	17/15	7/11
Ala	59/63	39/46
Val	9/8	9/20
Ile	3/4	8/8
Met	2/5	0/4
His	2/0	5/1
Hyl	0/4	0/7
Tyr	0/0	1/0
Gly	1/0	0/0

*Except for one residue of 3-hydroxyproline at an X position in the α1 chain, only 4-hydroxyproline has been found.

have been associated with specific side-chain to backbone and side-chain to side-chain interactions between polypeptide chains [7].

Side-chain to side-chain interactions imply a tendency for pairs of interacting amino acid residues to occur in steric proximity (fig.1), and table 2 shows the distribution of neighbouring residues for several pairs of amino acids which are believed to interact in this way. These include Arg at Y3 on the reference chain making salt and hydrogen-bond interactions with Glu and Asp at X2 on the clockwise chain, Lys at Y3 (REF) interacting with Glu or Asp at X5 (CW), and Arg or Asp at Y3 (REF) hydrogen bonding to the backbone NH at X2 (CW) and therefore implying a negative preference for Pro in this position.

Equivalent interactions may be made between all three pairs of chains, but, because of the staggering of the chains illustrated in fig.1, the interacting pairs of residues will depend on which chain is α2. Thus interactions between Arg at Y3 and Glu at a neighbouring X position, with aREF α2 chain, will be of the form α2(Y3)–α1(X2), α1(Y3)–α1(X5) or α1(Y3)–α2(X2) (illustrated in fig.1); with the CW chain α2, the interactions will be of the form α2(Y3)–α1(X5), α1(Y3)–α1(X2) or α1(Y3)–α2(X2); and with the α2

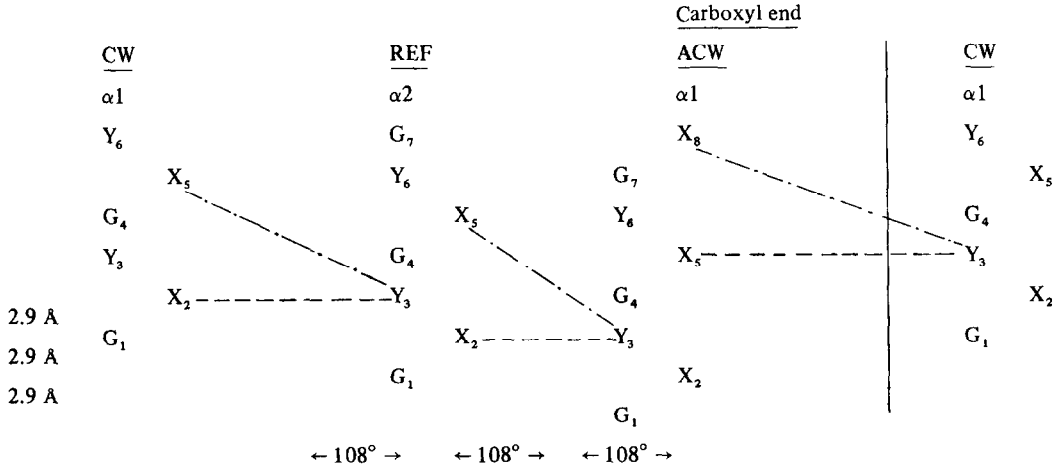

Fig.1. Schematic illustration of spatial relationships between amino acid residues in the three polypeptide chains of the collagen molecule. REF, CW and ACW indicate reference, clockwise and anti-clockwise chains, as viewed from the carboxyl end of the molecule. Equivalent residues on the three chains are related by 2.9 Å translation and 108° rotation. Note that Y3–X2 interactions relating the CW to REF and REF to ACW chains are equivalent to Y3–X5 interactions between ACW and CW chains. The figure illustrates the case where the REF chain is α2 and the CW and ACW chains are α1.

Table 2
Distribution of neighbouring amino acid residues in α1 and α2 chains

Y3	X2,5,8	Glu		Asp			Pro	
		α1 (42)	α2 (31)	α1 (17)	α2 (7)		α1 (116)	α2 (90)
Arg	2	10	7	5	1		8	8
α1(42)	5	9	4	2	2		6	8
α2(35)	8	4	5	2	1		20	13
	[R]	[5.2]	[3.9]	[2.1]	[0.9]		[14.5]	[11.3]
Lys	2	1	1	1	1	Asp(Y3)	1	2
α1(20)	5	5	5	6	1	α1(15)	1	1
α2(10)	8	6	2	2	1	α2(11)	3	2
	[R]	[2.5]	[1.1]	[1.0]	[0.2]		[5.2]	[3.5]

For each amino acid residue at Y3, the top three rows show the numbers of the various residues occurring at positions X2, X5 and X8 in the sequence for α1 and for α2 chains. Figures in ( ) show total numbers of residues in positions X or Y. [R] shows the number of pairs expected for a random distribution of residues (e.g. for α2, Arg-Glu, R = 35(31/280) = 3.9). Large deviations from [R] are underlined.

chain in the N-terminal position, ACW, the interactions will be α2(Y3)–α1(X2), α1(Y3)–α1(X2) or α1(Y3)–α2(X5).

We have tabulated appropriate sequence locations of amino acid residues for various pair preferences and calculated the number of potential interchain interactions for each of the three possible positions of the α2 chain. Table 3 shows part of such a tabulation for Arg-Glu interactions with Arg at Y3. An Arg α(813)–Glu α1(812) interaction could contribute to the molecular stability if α2 were at either the CW or ACW positions. Similarly, Arg α1(294) could contribute to stability with α2 at the REF or ACW positions, as could Arg α2(648). Taking account of all 42 arginine residues at Y positions in each α1 chain and 35 in the α2 chain, there are, respectively, 22, 23 and 24 potential Arg–Glu interchain interactions for α2 at the CW, REF and ACW positions. If the glutamic acid

Table 3
Locations of Glu X positions near Arg at Y3

α1(Y3)	α1(X2) CW/ACW	α2(X2) CW/REF	α1(X5) REF	α2(X5) ACW
Arg 813	Glu 812	–	–	–
Arg 294	–	–	Glu 296	Glu 296
α2(Y3)	REF/ACW		CW	
Arg 648	Glu 647	Glu 647	–	–

Positions for the α2 chains appropriate to the various interactions are indicated as CW, REF or ACW.

Table 4
Potential interchain interactions for various positions of α2 chains

Interaction type	α2 Position			Random
	CW	REF	ACW	distribution
Arg(Y)–Glu(X)	22 (6)	23 (9)	24 (6)	13 (4)
Arg(Y)–Asp(X)	9 (2)	6 (2)	9 (2)	5 (2)
Arg(Y)–Pro(X)	−22(−7)	−23 (−6)	−26(−8)	−38(−13)
Lys(Y)–Glu(X)	15 (2)	16 (6)	9 (2)	6 (2)
Lys(Y)–Asp(X)	7 (1)	6 (2)	12 (2)	2 (1)
Asp(Y)–Pro(X)	− 5 (0)	− 5 (−1)	− 1 (0)	−13 (−4)
Total	26 (4)	23 (12)	27 (4)	−25 (−8)

The total numbers of interactions between all three chains are shown, and the figures in ( ) show how many of these correspond to Y3–X5 interactions (Y3–X8 for interactions involving Lys). Note that Arg(Y)–Pro(X) and Asp(Y)–Pro(X) have negative pair preferences.

residues in X positions in the α1 and α2 chains (42 and 31, respectively) were randomly distributed among the 337 tripeptides in each chain in the helical region of the collagen molecule, one would expect approximately 42(42/337)+42(31/337)+35(42/337) = 13 Arg–Glu pairs for any one of the three α2 positions. Table 4 shows the numbers of potential interchain interactions, for six kinds of residue pairs, with α2 in each of the three possible positions as well as for a random distribution of residues.

3. Discussion

It is clear from table 4 that the numbers of potential interchain interactions indicated by the sequence data are appreciably greater than would be expected for a random distribution. X-ray structure analyses have shown that many intramolecular side-chain to side-chain and side-chain to backbone interactions occur in globular proteins [11,12], and the stereochemical feasibility of stabilising interactions between the pairs of residues we have considered has been demonstrated by molecular model investigations [7], though direct proof of their existence may have to await X-ray crystallographic analyses of collagen fragments [13]. Experimental support for the existence of intramolecular side-chain interactions has come from the observation that collagen in which most carboxyl groups have been blocked by methylation, or lysine residues by succinylation, has a lower denaturation temperature than the native molecule [14].

It is difficult to gauge the extent of side-chain dependent molecular stabilisation. Table 4 indicates about 50 potential interchain salt linkages, about twice the number expected from a random distribution, and 25 to 100 hydrogen bonds between Asp or Arg and backbone NH. Table 1 indicates an additional 35 to 55 Gln to backbone CO hydrogen bonds [7], so that, taking account of the unsequenced α2 region, there appear to be at least 100 interchain interactions of the types we have considered. However, as polar residues account for about one third of the X and Y positions and there may be various other types of interactions, the total may be as high as 500.

The total numbers of potential interchain interactions in table 4 do not show a clear correlation with the order of the polypeptide chains. This reflects the high degree of homology between the α1 and α2 chains [2] and the fact that they show much the same sequence regularities, though these are somewhat less strongly expressed in α2 (see tables 1 and 2). In nature, the order of the chains is determined by the assembly not of α but of procollagen chains [15]. In vitro renaturation of separated α chains has shown that artificial molecules of composition $(\alpha 1)_3$ displays much the same thermal stability as native $(\alpha 1)_2\alpha 2$ molecules, but that $(\alpha 2)_3$ molecules have a 12°C lower denaturation temperature [16]. Table 2 indicates that the latter type of molecule should indeed have fewer stabilising interactions.

The similarity of α1 and α2 sequences is consistent with the evolution of stabilising intramolecular interactions before the divergence of the two types of chain. Subsequent evolution was probably directed largely towards specific intermolecular interactions [17], but it seems unlikely that it caused any drastic reduction or unevenness in intramolecular interchain interactions. Interactions of type Y3–X5 are the most sensitive to the α2 chain position (see table 3). In table 4 the numbers of such interactions are shown in parentheses and it can be seen that the total of those of type α1(Y3)–α1(X5), corresponding to the REF position, exceeds the number expected from a random distribution by 20. This is almost twice the number corresponding to the other two positions. In view of the sequence homology, this relatively small difference in the number of interactions is what would be expected if the order of the chains were α1–α2–α1.

Acknowledgements

We are grateful to Dr Klaus Kühn for his active encouragement and for helpful discussions, to Dr Paul Bornstein for helpful comments on the manuscript, and to the Deutsche Forschungsgesellschaft for a travel grant to one of us (P.P.F.), which enabled us to get together for this work.

References

[1] Traub, W. and Piez, K. A. (1971) Advan. Protein Chem. 25, 243–352.
[2] Fietzek, P. P. and Kühn, K. (1976) Int. Rev. Connective Tissue Research 7, 1–60.
[3] Balian, G., Click, E. M. and Bornstein, P. (1971) Biochemistry 10, 4470–4478.
[4] Fietzek, P. P., Wendt, P., Kell, I. and Kühn, K. (1972) FEBS Lett. 26, 74–76.
[5] Fietzek, P. P., Rexrodt, F., Wendt, P., Stark, M. and Kühn, K. (1972) Eur. J. Biochem. 30, 163–168.
[6] Traub, W. (1974) Isr. J. Chem. 12, 435–439.
[7] Salem, G. and Traub, W. (1975) FEBS Lett. 51, 94–99.
[8] Hutton, J. J., Kaplan, A. and Udenfriend, S. (1967) Arch. Biochem. Biophys. 121, 384–391.
[9] Butler, W. T. (1968) Science 161, 796–798.
[10] Miller, R. L. (1971) Arch. Biochem. Biophys. 147, 339–342.
[11] Watson, H. C. (1969) Progr. Stereochem. 4, 299–333.
[12] Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, T. C. and Hill, R. C. (1969) J. Mol. Biol. 42, 65–86.
[13] Yonath, A., Traub, W. and Miller, E. J. (1975) FEBS Lett. 57, 57, 93–95.
[14] Rauterberg, J. and Kühn, K. (1968) Hoppe Seyler's Z. Physiol. Chem. 349, 611–622.
[15] Bornstein, P. (1974) Ann. Rev. Biochem. 43, 567–604.
[16] Tkocz, C. and Kühn, K. (1969) Eur. J. Biochem. 7, 454–462.
[17] Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. A. and Woodhead-Galloway, J. (1973) J. Mol. Biol. 79, 137–148.