THE COVALENT STRUCTURE OF COLLAGEN. AMINO ACID SEQUENCE OF THE N-TERMINAL REGION OF α2-CB4 FROM CALF AND RAT SKIN COLLAGEN

Peter P. FIETZEK, Ingrid KELL and Klaus KÜHN

Max-Planck-Institut für Eiweiss- und Lederforschung, Abteilung Kühn, 8 München 2, Schillerstr. 46, Germany

Received 10 July 1972

1. Introduction

Elucidation of the primary structure of collagen has rapidly progressed in recent years (reviewed by Traub and Piez [1]. Thus, approx. 65% of the amino acid sequence of the α1-chain have been established [1–9]. The distribution of methionyl residues within the α2-chain is much less favourable, cyanogen bromide cleavage giving rise to three very large and three very small peptides. The large peptides are of equal length of approx. 330 amino acid residues, and the three short peptides comprise three, fourteen and thirty residues, respectively [10]. So far, only the sequences of the small peptides have been determined [1, 11, 12]. The automated degradation procedure introduced by Edman and Begg [13] now rendered the large peptides likewise amenable to sequence work. In the present communication we report the sequence of the first 42 N-terminal amino acid residues of α2-CB4 from calf and rat skin collagen [10, 14]. The entire peptide α2-CB4 comprises 330 residues and extends over the N-terminal third of the α2-chain [14, 15]. Comparison of these sequences of rat and calf collagen revealed four interspecies substitutions, all of them involving apolar residues. We also compared the first 60 N-terminal residues of the α1- and α2-chains of rat skin collagen.

2. Materials and methods

Collagen from calf skin and from lathyritic rat skin was prepared as described earlier [16, 17]. The α2-chains were isolated by CM-cellulose chromatography essentially as described by Piez et al. [18].

The α2-chains were cleaved with cyanogen bromide and the peptide α2-CB4 separated from the bulk of the other peptides by chromatography on phosphocellulose as previously described [10, 14]. Further purification of α2-CB4 was achieved by rechromatography on CM-cellulose and Agarose 1.5, 200–400 mesh, as described [19].

Automated Edman degradation was performed in a Beckman Sequencer Model 890 (Beckman Instruments, Palo Alto, Calif.), using a modification of the procedure of Edman and Begg [13]. As described before [8], the resulting PTH-amino acids were identified by gas–liquid [20] and thin-layer chromatography [21] with the exception of PTH-arginine which was identified by thin-layer electrophoresis [8].

3. Results and discussion

The sequences of the first 42 amino acid residues of α2-CB4 from rat and from calf skin collagen are presented in fig. 1. The typical collagen sequence with glycine in every third position is readily apparent. Only four interspecies substitutions are observed, barely interrupting the extensive homology in this region of the α2-chain. All substitutions involve apolar residues only and do not alter the net charge in this region of the molecule.

The peptide α2-CB4 is preceded at its N-terminal end by the two peptides α2-CB1 and α2-CB0, comprising a total of 17 residues whose sequence in rat skin collagen is known [1]. Thus, the sequence at the N-terminal end of the α2-chain from rat skin collagen is now established for positions one through fifty-nine.

North-Holland Publishing Company – Amsterdam

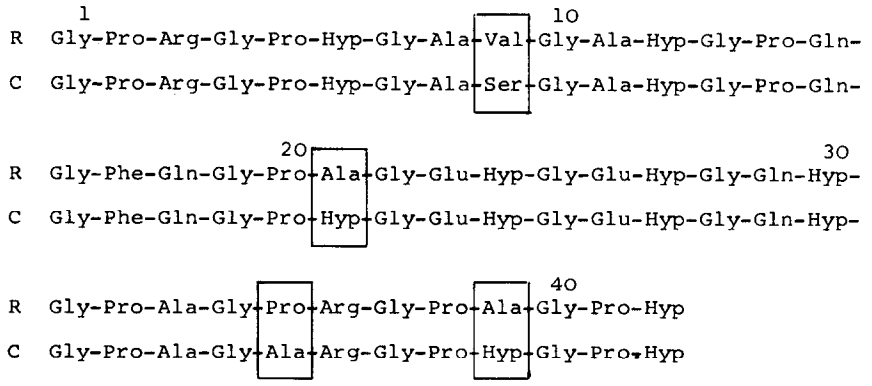

Fig. 1. The amino acid sequence of the N-terminal region of α2-CB4 from rat and calf skin collagen. Nonidentical residues found at the same position in the two peptides are enclosed in boxes. R = rat, C = calf.

The amino acid sequence of the corresponding region of the α1-chain from rat skin collagen is also known [1, 2, 4]. Therefore, for the first time, corresponding sequences of the α1- and the α2-chains of rat skin collagen may now be compared (fig. 2). Greatest homology is obtained when the two chains are aligned in such a way as to bring the beginning of the helical portions of the chains, indicated by the repetitive tripeptide structure Gly–X–Y, into register. This then reveals the α2-chain to be one residue shorter than the α1-chain. Residues 1–12 and 2–12 represent the nonhelical cross-link regions of the α1- and the α2-chain, respectively. They are known to exhibit great variability in amino acid composition as well as sequence [1]. The helical portion of the sequence presented here consists of 48 residues containing 10 interchain substitutions. The latter occur in the X– as well as in the Y– position of the tripeptide unit Gly–X–Y. All interchain substitutions involve apolar residues. The polar residues arginine and glutamic acid occupy like positions in the α1- and α2-chain, consistent with the identical charge pattern of α1- and α2-chains in this area of the molecule. A clustering of basic and acidic amino acid residues is not found in this area of the molecule, in contrast to the situation in the C-terminal peptide of the chain, α1-CB6 [5].

This area of the collagen molecule displays more interchain than interspecies substitutions, whereas ten interchain substitutions occur in rat skin collagen,

Fig. 2. The amino acid sequence of the N-terminal region of rat skin collagen. Nonidentical residues found at the same position of the helical region are enclosed in boxes.

only four interspecies substitutions are found between rat and calf skin collagen in the $\alpha 2$-chains and none in the $\alpha 1$-chains [9]. These data are consistent with the assumption, based on gross amino acid composition of the entire $\alpha 1$- and $\alpha 2$-chains, that a higher degree of homology exists between corresponding chains from different species than between $\alpha 1$- and $\alpha 2$-chains of the same species.

Acknowledgements

We wish to thank Miss L. Reynolds and Miss Amira Djonlagic for experimental assistance and the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 51) for financial support.

References

[1] W. Traub and K. Piez, in: Advances in Protein Chemistry Vol. 25 (Academic Press Inc., New York, London, 1971) p. 243.
[2] P. Bornstein, Biochemistry 6 (1967) 3082.
[3] W.T. Butler, Biochemistry 9 (1970) 44.
[4] W.T. Butler and L. Ponds, Biochemistry 10 (1971) 2076.
[5] K.v.d. Mark, P. Wendt, F. Rexrodt and K. Kühn, FEBS Letters 11 (1970) 105.
[6] G. Balian, E.M. Click and P. Bornstein, Biochemistry 10 (1971) 4470.
[7] J. Rauterberg, P.P. Fietzek, F.W. Rexrodt, U. Becker, M. Stark and K. Kühn, FEBS Letters 21 (1972) 75.
[8] P.P. Fietzek, F.W. Rexrodt, P. Wendt, M. Stark and K. Kühn, European J. Biochem., in press.
[9] P.P. Fietzek, P. Wendt, I. Kell and K. Kühn, FEBS Letters, 26 (1972) 74.
[10] P.P. Fietzek and K.A. Piez, Biochemistry 8 (1969) 2129.
[11] J.H. Highberger, A.H. Kang and J. Gross, Biochemistry 10 (1971) 610.
[12] P.P. Fietzek, H. Furthmayr, I. Kell and K. Kühn, Biochem. Biophys. Res. Commun., submitted for publication.
[13] P. Edman and G. Begg, European J. Biochem. 1 (1967) 80.
[14] P.P. Fietzek, M. Münch, D. Breitkreutz and K. Kühn, FEBS Letters 9 (1970) 229.
[15] J. Vuust, J.M. Lane, P.P. Fietzek, E.J. Miller and K.A. Piez, Biochem. Biophys. Res. Commun. 38 (1970) 703.
[16] K. Kühn, G. Schuppler and J. Kühn, Z. Physiol. Chem. 338 (1964) 10.
[17] P. Bornstein and K.A. Piez, Biochemistry 5 (1966) 3803.
[18] K.A. Piez, E.A. Eigner and M.S. Lewis, Biochemistry 2 (1963) 58.
[19] P.P. Fietzek and K. Kühn, in preparation.
[20] J.J. Pisano and T.J. Bronzert, J. Biol. Chem. 244 (1969) 5597.
[21] P. Edman, in: Protein Sequence Determination, ed. Saul B. Needleman (Springer Verlag, Berlin, Heidelberg, New York, 1970) p. 211.