THE COVALENT STRUCTURE OF COLLAGEN: AMINO ACID SEQUENCE OF α1-CB3 FROM CALF SKIN COLLAGEN

Peter P. FIETZEK, Peter WENDT*, Ingrid KELL and Klaus KÜHN
Max-Planck-Institut für Eiweiss- und Lederforschung, Abteilung Kühn, D-8 München 2, Schillerstrasse 46, Germany

Received 10 July 1972

1. Introduction

The peptide α1-CB3 is derived from the α1-chain of collagen by cyanogen bromide (CNBr) cleavage and consists of 149 amino acids [1]. It occupies positions 421 through 570 of the 1055 residues of the α1-chain [2]. In a previous communication we have described the isolation and characterization of defined fragments of α1-CB3 from calf skin collagen by cleavage with trypsin, chymotrypsin, and thermolysin [3]. The present publication describes the elucidation of the complete amino acid sequence of α1-CB3, obtained exclusively by automated stepwise degradation according to Edman and Begg [4].

2. Experimental

The peptides subjected to sequencing were isolated as described by Wendt et al. [3]. Automated Edman degradation was accomplished in a "Sequencer Model 890" of Beckman Instruments, Palo Alto, Calif. The program employed resembled that of Edman and Begg [4]. Quadrol (*N,N,N',N'*-tetrabis(2-hydroxypropyl)-ethylenediamine) as well as dimethylallylamine (DMAA) were used as buffering substances. Reagents and solvents were purchased from Beckman Instruments, Palo Alta, Calif. The resulting PTH-amino acids were identified by gas- and thin-layer chromatography. A gas chromatograph "GC-45" of Beckman Instruments, Palo Alto, Calif., was employed, and the method of Pisano and Bronzert [5] was applied. Thin-layer chromatography was carried out on Silica-gel plates (F 254, Merck AG, Darmstadt, Germany), utilizing Edman's system "H" (ethylene chloride/acetic acid) [6]. PTH-amino acids were trimethylsilylated with BSA (*N, O*-bis-trimethylsilylacetamide), purchased from Analabs Inc., North Haven, Conn., USA. PTH-arginine was identified by thin-layer electrophoresis as described earlier [7]. PTH-amino acids purchased from Fluka, Bucks, Switzerland, served as standards.

Each amino acid was isolated at least twice. The peptides subjected to automated Edman degradation and their sizes are listed below together with the amounts introduced into the Sequencer, the buffers employed, and the number of useful degradation cycles achieved. For sequencing of shorter peptides it was necessary to start with greater amounts in order to counteract loss of sample during extraction [8, 9].

Peptide	Buffer	μMoles	Size*	Cycles
α1-CB3	Quadrol	0.77–0.84	149	67
α1-CB3-T 6	Quadrol	1.5	45	25
α1-CB3-C2	Quadrol	0.68–1.15	87	39
α1(125)-CBN	Quadrol	0.5 –1	56	29
α1-CB3-Th [8]	DMAA	1	27	24
α1-CB3-T 11	DMAA	2.3	20	18

* Number of amino acid residues.

* Present address: Institut für Experimentelle Chirurgie an der Chirurgischen Klinik der Technischen Universität, 8 München 80, Ismaninger Str. 22, Germany.

North-Holland Publishing Company – Amsterdam

3. Results and discussion

The peptides employed for sequencing of αl-CB3 are presented diagramatically in fig. 1. The solid bars denote that area of each peptide whose sequence was elucidated. These overlapping areas permitted complete determination of the sequence of positions 1–147 of α1-CB3. The C-terminal residue in position 149 must be homoserine since α1-CB3 is a CNBr-derived peptide. The C-terminal thermolysin-peptide Th 9, consisting of only the four residues Leu, Gln, Gly, Hse permits only the Gly for position 148. All the peptides of α1-CB3 which were obtained after cleavage with trypsin, thermolysin, and chymotrypsin [3] and whose sequence had not been determined could be allocated to their respective positions within α1-CB3 on the basis of their composition.

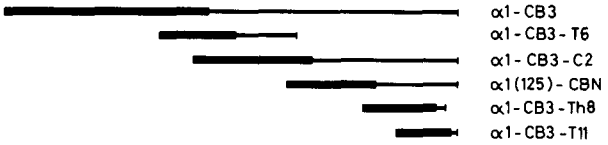

Fig. 1. Diagramatic representation of peptides employed for sequencing of α1-CB3. Solid bars indicate that area of the peptide whose sequence was elucidated.

```
                                                  10
Gly-Phe-Hyp-Gly-Pro-Lys-Gly-Ala-Ala-Gly-Glu-Hyp-Gly-Lys-Ala

                  20                                  30
Gly-Glu-Arg-Gly-Val-Hyp-Gly-Pro-Hyp-Gly-Ala-Val-Gly-Pro-Ala-

                                    40
Gly-Lys-Asp-Gly-Glu-Ala-Gly-Ala-Gln-Gly-Pro-Hyp-Gly-Pro-Ala-

                  50                                  60
Gly-Pro-Ala-Gly-Glu-Arg-Gly-Glu-Gln-Gly-Pro-Ala-Gly-Ser-Hyp-

                                    70
Gly-Phe-Gln-Gly-Leu-Hyp-Gly-Pro-Ala-Gly-Pro-Hyp-Gly-Glu-Ala-

                  80                                  90
Gly-Lys-Hyp-Gly-Glu-Gln-Gly-Val-Hyp-Gly-Asp-Leu-Gly-Ala-Hyp-

                                   100
Gly-Pro-Ser-Gly-Ala-Arg-Gly-Glu-Arg-Gly-Phe-Hyp-Gly-Glu-Arg-

                 110                                 120
Gly-Val-Glu-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Pro-Arg-Gly-Ala-Asn-

                                   130
Gly-Ala-Hyp-Gly-Asn-Asp-Gly-Ala-Lys-Gly-Asp-Ala-Gly-Ala-Hyp-

                  140
Gly-Ala-Hyp-Gly-Ser-Gln-Gly-Ala-Hyp-Gly-Leu-Gln-Gly-Hse
```

Fig. 2. Amino acid sequence of α1-CB3 from calf skin collagen.

The sequence presented in fig. 2 exhibits a structure which is based on the tripeptide unit Gly–X–Y, typical for the helical areas of the collagen molecule. Likewise typical is the observed alternation along the polypeptide chains of apolar and polar sequence regions. Frequently, the apolar regions are found to contain the tripeptide units Gly–Pro–Hyp and Gly–Pro–Ala which contribute particularly to the stability of the triple helical conformation. The polar regions are generally distinguished by the already earlier observed [10] close vicinity of basic and acidic amino acids such as Lys–Asp–Gly–Glu (residues 32–35) or Glu–Arg–Gly–Glu (residues 50–53). The distribution of polar and apolar amino acids along the peptide chain is directly reflected by the pattern of light and dark bands of the cross-striation pattern of segments long spacing, as demonstrated earlier [10]. A similarly good correlation between amino acid sequence and electron microscopical appearance was also observed for α1-CB3 (not shown).

It had been observed earlier [7, 11] that certain amino acids are not distributed evenly among positions X or Y of the tripeptide units Gly–X–Y. Thus, all three phenylalanine residues present in α1-CB3 as well as all others found in the α1-chain so far are located in position X. Similarly, leucine was so far also found only in position X. Two out of the three leucine residues in α1-CB3 are also located in position X. The content of glutamine of α1-CB3 is relatively high. All six glutamine residues in α1-CB3 occurred in position Y. So does generally the majority of glutamine residues found in the α1-chain. Whether or not this preference of certain amino acids for positions X or Y will also be found in other, as yet unresolved sequences of the α1-chain, awaits further study.

As in other regions of the α1-chain, hydroxylation of proline in position Y is not always complete. The degree of hydroxylation varies in these cases of incomplete hydroxylation between 30 and 70%. More precise determinations by automated Edman degradation are usually prevented by problems of background and overlap [9].

Butler isolated from α1-CB3 of rat skin collagen the C-terminal as well as the lysine containing tryptic peptides [12]. He also determined the sequences of positions 1–5 and 118–123. Apart from an exchange of alanine against asparagine in position 119, the sequences given by Butler are identical with the cor-

responding sequences of calf skin collagen. However, the amino acid composition of the tryptic peptides suggests additional substitutions between calf and rat skin collagen. T2 (positions 7–14) and T11 (positions 130–149) of rat skin collagen thus contain a threonine residue while an alanine is present in the corresponding calf peptide.

Acknowledgement

The authors wish to thank the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 51) for financial support.

References

[1] J. Rauterberg and K. Kühn, European J. Biochem. 19 (1971) 398–407.
[2] J. Rauterberg and K. Kühn, FEBS Letters 1 (1968) 230–232.
[3] P. Wendt, P.P. Fietzek and K. Kühn, FEBS Letters 26 (1972) 69.
[4] P. Edman and G. Begg, European J. Biochem. 1 (1967) 80–91.
[5] J.J. Pisano and T.J. Bronzert, J. Biol. Chem. 244 (1969) 5597–5607.
[6] P. Edman, in: Protein Sequence Determination, ed. B. Saul Needleman (Springer Verlag, Berlin, Heidelberg, New York, 1970) p. 211.
[7] P.P. Fietzek, F. Rexrodt, P. Wendt, M. Stark and K. Kühn, European J. Biochem., in press.
[8] P.P. Fietzek, I. Kell and K. Kühn, in preparation.
[9] P.P. Fietzek and K. Kühn, Topics in Current Chemistry 29 (1972) 1–57.
[10] K. v.d. Mark, P. Wendt, F. Rexrodt and K. Kühn, FEBS Letters 11 (1970) 105–108.
[11] G. Balian, E.M. Click and P. Bornstein, Biochemistry 10 (1971) 4470–4478.
[12] W.T. Butler, J. Biol. Chem. 244 (1969) 3415–3417.