Vol. 83, No. 1, 1978 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 14, 1978 Pages 43-49

THE AMINO ACID SEQUENCE OF CHICK SKIN COLLAGEN α1-CB7: THE PRESENCE OF A PREVIOUSLY UNRECOGNIZED TRIPLET*

John H. Highberger, Clare Corbett, Andrew H. Kang and Jerome Gross

Developmental Biology Laboratory
Department of Medicine
Massachusetts General Hospital and Harvard Medical School,
Boston, Massachusetts 02114

and

The Veterans Administration Hospital
and Departments of Biochemistry and Medicine,
University of Tennessee Medical School,
Memphis, Tennessee 38104

Received May 16,1978

SUMMARY: Because alignment of the amino acid sequences of chick skin collagen α2-CB3 (1) with the relevant portion of chick skin collagen α1-CB7 (2) suggested that a Gly-X-Y triplet may have been missed in the latter, the peptide TM-2, produced by tryptic digestion of maleylated α1-CB7, was reinvestigated. Cleavage by trypsin at the unblocked lysine at position 18, and isolation of the resulting COOH-terminal peptide, showed this to be a 15-residue peptide containing a previously unrecognized Gly-Pro-Hyp triplet. Sequencing of the peptide showed this to occupy positions 4 through 6, or 56 through 58 of α1-CB7. The latter thus has 271 instead of 268 residues, and the α1[I] chain 1055 instead of 1052.

INTRODUCTION: Of the peptides released by CNBr cleavage of the collagen α1[I] chain, α1-CB7 has emerged as the site of a number of important biologic functions. These include the single specific peptide bond (Gly-Ile) cleaved by animal collagenase (3), the binding site for the cell attachment protein CAP (4), and probably fibronectin, and the locus which appears to promote myoblast differentiation (5). Thus, it is critical that the sequence of this region be correctly known.

Recently a comparison of newly determined amino acid sequences of segments of the α2 chain with the relevant portions of α1 suggested that the published sequences of α1-CB7 (6,2) might be in error by having overlooked one Gly-X-Y triplet. Were this to be the case, it would lie in the region between lysine

* This is publication No. 755 of the Robert W. Lovett Memorial Group for the Study of Diseases Causing Deformities. This work was supported by National Institutes of Health Grants AM03564 (J.G.) and AM 16506 (A.H.K.) from the NIAMDD and by research funds from the Veterans Administration.

0006-291X/78/0831-0043$01.00/0

at position 52 and the next basic residue, arginine, 12 residues farther along. This portion of chick skin collagen α1-CB7, included in the fragment designated as TM-2 (2) has therefore been re-examined. We report the presence, nature, and location of a previously unrecognized triplet.

MATERIALS AND METHODS: Chick skin collagen was prepared from 3-week-old white Leghorn lathyritic chicks, and α1-CB7 was prepared from the α1 chain of this collagen, as previously described (7). The TM-2 peptide, beginning at position 35 of α1-CB7, was prepared as described (2) by maleylating α1-CB7, digesting with trypsin, unblocking, and fractionating on a 2.5 x 110 cm column of Sephadex G-50s. The resulting chromatogram was identical with that shown in Figure 3 of reference (2).

The peptide TM-2 was isolated from the pooled fractions of Peak 3 of the Sephadex G-50 chromatogram. Some preparations were purified by rechromatography on a 0.9 x 10 cm phosphocellulose column, eluted at 42° with a linear gradient formed between 250 ml of 0.001 M Na acetate, pH 3.8, and the same volume of the same buffer containing 0.6 M NaCl. The column was monitored by continuous recording of absorbance at 230 nm. Other preparations were purified by rechromatography of the Sephadex G-50s fractions on a 2 x 100 cm column of Biogel P-10 eluted with 0.1N acetic acid.

Recycling chromatography of the tryptic digest of TM-2 was performed using a ReCyChrom Selector valve, Type 4911B (LKB), manually operated, with a peristaltic pump eluting the 1.6 x 60 cm column of Biogel P-4 in reverse flow in the arrangement recommended by the manufacturer. The eluant was 0.1N acetic acid, and the absorbance at 226 nm was continuously recorded.

Samples for amino acid analysis were hydrolyzed for 24 hours at 110°, in glass distilled constant boiling HCl under nitrogen. The analyses were run on a Beckman 121 analyzer with Autolab System AA Integrator, using the buffer system of Trelstad and Lawley (8). No corrections were made for destruction of amino acids during hydrolysis.

Sequencing was done in a Beckman up-dated 891B Sequencer. Peptide samples were modified by treating, in the cup, with 1,5-naphthalene disulfonic acid according to Foster et al. (9). The samples were dried using the peptide application subroutine (Beckman No. 031371). Degradation was carried out using the 0.1M Quadrol Peptide Program (Beckman No. 110576). All residues were subjected to identification both by gas-liquid chromatography using SP-400 (Beckman) column packing (10), and by back hydrolysis in HI (11) with identification on the amino acid analyzer.

RESULTS AND DISCUSSION: The amino acid analysis of TM-2 purified by rechromatography on phosphocellulose, given in the first column of Table I, shows it to contain 33 residues instead of the 30 originally supposed. The three additional residues over those of the originally published analysis (Table II of reference 2) are one each of glycine. proline, and hydroxyproline.

In addition to its arginine, which must be the COOH-terminal residue from the method of preparation, TM-2 contains one residue of lysine. Since the alignment experiments of Dixit et al. (1) indicate that the three extra resi-

Table I

Analyses of TM-2 and its Component Tryptic Peptides

	TM-2 Purified on Phosphocellulose	TM-2 used for Tryptic Digestion	Tryptic Digestion Products of TM-2 Peak 1	Peak 2	Peak 3
Hyp	3.45	3.21	3.00	1.88	1.14
Asp	1.16	1.25	0.51	0.99	0.23
Thr	1.67	1.55	1.60	0.91	0.89
Ser		0.38	0.18	0.13	0.12
Glu	1.42	1.48	0.78	0.26	0.96
Pro	5.98	5.04	4.85	2.77	2.69
Gly	11.00	11.00	11.00	6.00	5.00
Ala	5.27	4.90	5.25	2.07	2.63
Val		0.15	0.16		
Ile	0.86	0.76	0.61	0.88	0.14
Leu	0.80	0.81	0.59	0.84	0.18
Lys	1.06	1.15	0.95	0.97	0.23
Arg	0.96	0.98	0.74		0.74
Total Residues	33	32	32	18	15

All values shown are averages of closely agreeing determinations on two different preparations. A blank indicates that less than 0.1 residue was found.

dues are between this lysine and the arginine, that is, in the COOH-terminal component peptide which would be produced by cleaving TM-2 with trypsin at its lysyl residue, we proposed to make this cleavage, isolate the component peptides and identify the nature and position of the residues by sequencing the COOH-terminal component, which should have 15 residues if the alignment results are correct. Because the two component peptides have the same net formal charge and differ in size by only 3 residues we resorted to molecular sieve recycling chromatography on an appropriate resin.

Because the yield of TM-2 purified on phosphocellulose was so low (25 mol% on the starting α1-CB7), the slightly less pure preparations obtained by re-

chromatography on Biogel P-10 were used for the tryptic digestions. These were obtained in an average yield of 56.5 mol%, and the analysis of such material is shown in the second column of Table I.

Tryptic cleavage of the TM-2 lysyl residue proved to be difficult to achieve, probably because of the protective effects of the immediately preceding aspartyl and the next but one glutamyl residues. The digestion was carried out for 18-20 hours at 37° in 0.1M Tris HCl, pH 7.9, containing 10^{-3}M $CaCl_2$ and 0.05% (W/V) Na azide. Trypsin (Worthington TPCK-treated, 3 x crystallized) was used at 3% of the substrate weight. At the end of the first digestion period a similar aliquot of fresh enzyme was added, and the digestion repeated. At the end of the second digestion period the solution was acidified to pH 4 with 1M acetic acid, and immediately lyophilized. In this way cleavages of approximately 80 mol% were obtained.

The lyophilized material was dissolved in the minimal volume of 0.1N acetic acid, and run on a Biogel P.4 column by recycling chromatography as described in the Methods section. A typical chromatogram is shown in Figure 1. The last three columns of Table I give the analysis of the pooled fractions collected on the fourth and last cycle. The material collected in Peak 1, which is obviously heterogeneous, is primarily uncleaved TM-2. Peak 2 consists of the NH_2-terminal component with lysine as its COOH-terminal residue, and containing 18 residues. Peak 3 contains the COOH-terminal, arginine-containing component, and, as shown in the last column of Table I, it contains 15 residues of which one each of glycine, proline, and hydroxyproline were not shown in the previous sequence. Both component peptides were obtained in average yields of 57 mol% on the basis of the TM-2 digested.

The 15-residue peptide was sequenced as described in the Methods section. The results are presented in Figure 2, as a semi-logarithmic plot of the yields of the major amino acid found at each degradation cycle by back hydrolysis and amino acid analysis, against the degradation number. Only 14 cycles are shown since the identification of the COOH-terminal arginine was not quantitative.

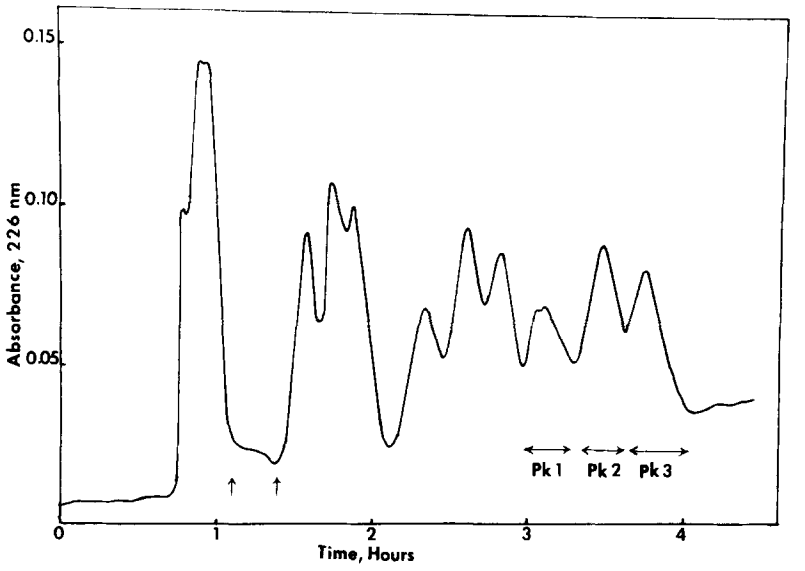

Figure 1. Recycling chromatogram of tryptic digest of 965 nmoles α1-CB7 peptide TM-2 on 1.6 x 60 cm column of Biogel P-4 in 0.1N acetic acid. Flow rate, 45 ml per hour. Fractions between the vertical arrows were removed from the system. Horizontal arrows indicate how fractions were pooled for the analyses shown in Table I.

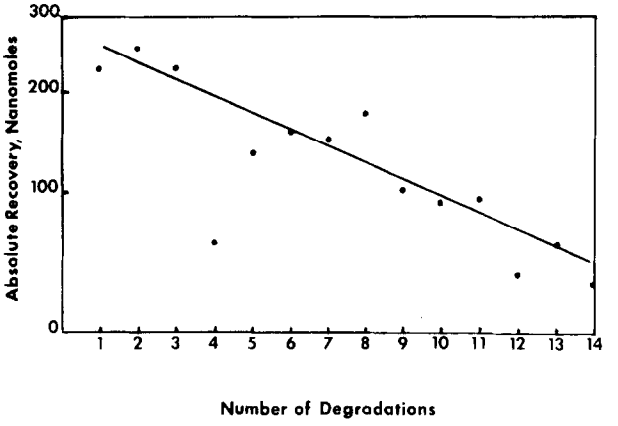

Figure 2. Semi-logarithmic plot of absolute yields of amino acids at each degradation cycle of 600 nmoles of COOH-terminal component of α1-CB7 peptide TM-2.

The complete degradation was carried out twice, first with 600 nmoles of peptide, and second with 336 nmoles of a separate preparation. In each case degradation was carried through 15 cycles, and the same sequence was found. This is shown as residues 53 through 67 of Figure 3. The repetitive yield

```
                                10                                  20                                  30
Hyp-Gly-Glu-Arg-Gly-Ala-Ala-Gly-Leu-Hyp-Gly-Ala-Lys-Gly-Asp-Arg-Gly-Asp-Hyp-Gly-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro-Gly-Lys

                                40                                  50                                  60
Asp-Gly-Leu-Arg-Gly-Leu-Thr-Gly-Pro-Ile-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Ala-Hyp-Gly-Asp-Lys-Gly-Glu-Ala-Gly-Pro-Hyp-Gly-Pro

                                70                                  80                                  90
Ala-Gly-Pro-Thr-Gly-Ala-Arg-Gly-Ala-Hyp-Gly-Asp-Arg-Gly-Glu-Hyp-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Phe-Ala-Gly-Pro-Hyp-Gly-Ala

                               100                                 110                                 120
Asp-Gly-Gln-Hyp-Gly-Ala-Lys-Gly-Glu-Thr-Gly-Asp-Ala-Gly-Ala-Lys-Gly-Asp-Ala-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Pro-Thr-Gly-Ala

                               130                                 140                                 150
Hyp-Gly-Pro-Ala-Gly-Glx-Val-Gly-Ala-Hyp-Gly-Pro-Hyl-Gly-Ala-Arg-Gly-Ser-Ala-Gly-Pro-Hyp-Gly-Ala-Thr-Gly-Phe-Hyp-Gly-Ala

                               160                                 170                                 180
Ala-Gly-Arg-Val-Gly-Pro-Hyp-Gly-Pro-Ser-Gly-Asn-Ile-Gly-Leu-Hyp-Gly-Pro-Hyp-Gly-Pro-Ala-Gly,Lys,Glx,Gly,Ser,Lys-Gly-Pro

                               190                                 200                                 210
Arg-Gly-Glu-Thr-Gly-Pro-Ala-Gly-Arg-Hyp-Gly-Glu-Hyp-Gly-Pro-Ala-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Glu-Lys-Gly-Ser-Hyp-Gly-Ala

                               220                                 230                                 240
Asp-Gly-Pro-Ile-Gly-Ala-Hyp-Gly-Thr-Pro-Gly-Pro-Gln-Gly-Ile-Ala-Gly-Gln-Arg-Gly-Val-Val-Gly-Leu-Hyp-Gly-Gln-Arg-Gly-Glu

                               250                                 260                                 270
Arg-Gly-Phe-Hyp-Gly-Leu-Hyp-Gly-Pro-Ser-Gly-Glu-Hyp-Gly-Lys-Gln-Gly-Pro-Ser-Gly-Ala-Ser-Gly-Glu-Arg-Gly-Pro-Hyp-Gly-Pro

271
Met
```

Figure 3. The amino acid sequence of chick skin collagen α1-CB7, corrected by insertion of Gly-Pro-Hyp at positions 56-58, and by interchange of the Thr and Ala at positions 64 and 66. The peptide TM-2 has the sequence shown from position 35 through 67 and the component peptide sequenced in this work is 53 through 67. The sequence has also been corrected by putting Ser at 207 in place of the Ala previously shown (13).

based on Gly at steps 1 and 13 was 95.01%, on Pro at 5 and 11 was 97.29%, and on Ala at 3 and 14 was 92.14%.

The results of this work therefore indicate that the alignment experiments of Dixit et al. (1) were correct in suggesting that a triplet had been missed in the original determination of the amino acid sequence of chick skin collagen α1-CB7. This probably came about because the TM-2 sequenced in that case was much less pure than was realized, and because it was sequenced as the whole peptide, rather than as the component peptides.

The corrected sequence of chick skin collagen α1-CB7 is given in Figure 3. It is to be noted that, in addition to the presence of the new triplet, and the consequent differences in numbering beyond position 55, this also differs from that given previously in the interchange of the Thr and Ala at positions 64 and 66 respectively.

It is therefore clear that α1-CB7 contains 271 rather than the 268 residues previously supposed, and the α1(I) chain 1055 residues rather than 1052 (12).

ACKNOWLEDGEMENT: The authors thank Mrs. Elizabeth O'Connor and Mr. Wing Yu for expert technical assistance.

REFERENCES:

1. Dixit, S.N., Seyer, J.M. and Kang, A.H. (1977) Eur. J. Biochem. 81,599-607.
2. Highberger, J.H., Corbett, C., Kang, A.H. and Gross, J. (1975) Biochemistry 14,2872-2881.
3. Gross, J., Harper, E., Harris, E.D., McCroskery, P.A., Highberger, J.H., Corbett, C. and Kang, A.H. (1974) Biochem. Biophys. Res. Commun. 6,605-612.
4. Kleinman, H.K., McGoodwin, E.B. and Klebe, R.J. (1976) Biochem. Biophys. Res. Commun. 72,426-432.
5. Hauschka, S.D. and White, N.K. (1972) In: Research Concepts in Muscle Development and the Muscle Spindle, Banker, B., Preybylski, R., Van der Meulen, J. and Victor, M., eds. pp. 53-71, Excerpta Medica, Amsterdam.
6. Fietzek, P.P., Roxrodt, F.W., Hopper, K.E. and Kuhn, K. (1973) Eur. J. Biochem. 38,396-400.
7. Kang, A.H., Piez, K.A. and Gross, J. (1969) Biochemistry 8,1506-1514.
8. Trelstad, R.L. and Lawley, K.R. (1976) Anal. Biochem. 70,287-289.
9. Foster, J.A., Bruenger, E., Hu, C.L., Albertson, K. and Franzblau, C. (1973) Biochem. Biophys. Res. Commun. 53,70-74.
10. Pisano, J.J., Bronzert, T.J. and Brewer, H.B.J. (1972) Anal. Biochem. 45, 43-59.
11. Smithies, O., Gibson, D., Fanning, E.M., Goodfliesch, R.M. and Gilman, J.G. (1971) Biochemistry 10,4912-4921.
12. Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. and Woodhead-Galloway, J. (1973) J. Mol. Biol. 79,137-148.
13. Highberger, J.H., Corbett, C. and Gross, J. (1978) In preparation.