HOMOLOGOUS REGIONS OF COLLAGEN α1(I) AND α1(II) CHAINS: APPARENT CLUSTERING OF VARIABLE AND INVARIANT AMINO ACID RESIDUES.

William T. Butler, Edward J. Miller, John Edward Finch, Jr.
Institute of Dental Research
University of Alabama Medical Center
University Station
Birmingham, Alabama 35294

Tadashi Inagami
Department of Biochemistry
Vanderbilt University
Nashville, Tennessee 37203

Received January 7,1974

Summary: The amino acid sequence of 75 residues from the central portion of the α1(II) chain of bovine cartilage collagen is reported and compared to that of a homologous region of the α1(I) chain of bovine skin collagen. The comparison suggests that collagen α chains contain regions with a high level of sequence variability alternating with other regions displaying few, or no, interchain sequence differences. The presence of two residues of galactosylhydroxylysine in the α1(II) sequence, absent in α1(I), suggest reasons for the relatively high carbohydrate content of α1(II).

INTRODUCTION

The polypeptides contained in triple-stranded collagen molecules consist of several chain types. The most ubiquitious molecular species of collagen contains two chains of one type, α1(I), and a single chain of another type, α2. In contrast the collagen unique to cartilage is comprised of three identical polypeptide chains, termed α1(II), which are similar in chromatographic properties to, but differ from, α1(I) in amino acid composition and carbohydrate content (1). A fourth type of α chain, recently detected in soft tissues and named α1(III), is found in collagen molecules with the chain composition, $\{\alpha 1(III)\}_3$ (2,3). The collagen of basement membranes appears to consist of three identical α1-like polypeptide chains which represent yet another type of α chain

(4). The similarity in size and composition suggests that all α chains are homologous proteins. Indeed the homology between α1(I) and α2 was convincingly demonstrated by comparison of the amino acid sequences of appropriate regions of these chains (5,6).

Several of the CNBr peptides derived from α1(II) have compositions and sizes similar to ones from α1(I) (7,8). Since they arise from the same relative regions (1,9), it seems safe to conclude that they represent homologous amino acid sequences of the respective α chains. Two such peptides, α1(I)-CB3 from bovine skin collagen and α1(II)-peptide 8 from bovine nasal cartilage collagen, each contain 149 amino acids and are derived from the central portion of the respective chains. We report here the amino acid sequence of the first seventy-five residues of α1(II)-peptide 8 and compare the results with those reported for α1(I)-CB3 (10).

MATERIALS AND METHODS

The determination of amino acid sequences was performed with a Beckman Automatic Sequencer (Model 890-B), operated at 57°, utilizing 0.5 M Quadrol buffer for the coupling reaction and 15 ml ethyl acetate washes. The PTH-amino acids were identified by gas-liquid chromatography (11) with 4 foot glass columns of "10%" SP-400 by methods described in the Beckman Sequencer Manual. Each residue was examined with and without silylation. The identities of the PTH-amino acids were confirmed by thin-layer chromatography on flexible plates of silica gel (12). PTH-arginine, in the aqueous layer of the Edman degradation procedure, was identified by the method of Inagami (13).

The preparation of α1(II)-peptide 8 from bovine nasal cartilage has been described (8). For further purification the samples were routinely subjected to ion-exchange-chromatography on CM-cellulose

Abbreviations: cyanogen bromide, CNBr; phenylthiohydantoin, PTH; carboxymethylcellulose, CM-cellulose; galactosylhydroxylysyl, gal-hyl.

columns at pH 4.8 (14). To obtain smaller fragments of α1(II)-peptide 8, the peptide was degraded with trypsin and the resultant "tryptides" were purified by molecular sieve and ion-exchange chromatographic procedures (15). The largest of these tryptic peptides, T6, contained forty-five residues, and was almost identical in composition to that of a tryptic peptide from α1(I)-CB3 (15); peptide T6 represents residues 52 through 96 of α1(II)-peptide 8.

RESULTS AND DISCUSSION

The amino acid sequence of the NH_2-terminal half of α1(II)-peptide 8 was determined by performing 54 cycles of automated Edman degradation on uncleaved α1(II)-peptide 8, followed by 24 cycles on peptide T6. The results of these experiments are illustrated in Figure 1, along with

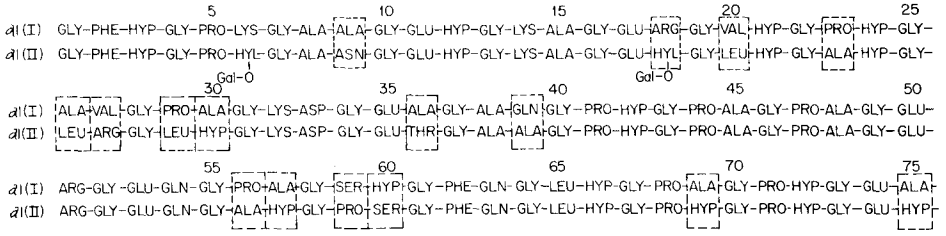

Fig. 1. Comparison of the amino acid sequence of the NH_2-terminal half of α1(I)-CB8 of bovine skin collagen reported by Fietzek et al. (10) with that of α1(II)-peptide 8 from the collagen of bovine nasal cartilage.

---

a comparison of similar results reported for α1(I)-CB3. The gal-hyl residues at positions 6 and 18 were not identified as PTH- derivatives but were determined on amino acid analysis of alkaline hydrolysates of appropriate tryptic peptides. The glycosidic bond to the δ-hydroxyl group of hydroxylysine is stable to heating at 108° in 2 N NaOH (16); the resultant gal-hyl compound elutes in a position between leucine and tyrosine, when subjected to single-column amino acid analysis (17).

The α1(II) chain contains about five times as much hydroxylysine and ten times as much hydroxylysine-linked carbohydrate as does the α1(I) chain (18). The present results (Figure 1) offer a partial explanation for this difference. The lysyl residue at position 6 of α1(I)-CB3, though a substrate for hydroxylation (15), is for some reason not hydroxylated*. The equivalent residue in cartilage collagen is fully hydroxylated, thus generating a site for enzymic attachment of carbohydrate not present in α1(I). Formation of additional hydroxylysine sidechains is not the only reason for the relatively high level of carbohydrate in cartilage collagen, however, since some of the same lysyl residues in dentin collagen are extensively hydroxylated without a significant increase in the number of hexose moieties (15).

A second manner whereby additional hydroxylysine residues suitable for attachment of carbohydrate may be generated is exemplified by the gal-hyl at position 18. An arginyl residue in the third position of a collagen triplet in α1(I) is substituted by lysine in α1(II) and in the latter chain the lysyl residue is fully hydroxylated and glycosylated. The replacement of arginines by lysines is probably not extensive, however, since the amount of hydroxylysine plus lysine and of arginine in α1(II) is not significantly different from that of α1(I) (7).

Comparison of the covalent structures from this segment of the α1(I) and α1(II) chains (Figure 1) shows conclusively that the two sequences are homologous, since they are identical in fifty-eight of seventy-five positions. The sequence differences often occur in clusters; in other words, in the positions where the two chains differ, there are often interchain amino acid sequence differences in nearby, or adjacent

---

* Sequence studies of α1(I) chains have identified some sites where only lysine occurs, other sites where both lysine and hydroxylysine occur and still others where only hydroxylysine is found. Since hydroxylation of lysine, to form hydroxylysine, is a post-ribosomal event, we refer to the different relative amounts as partially or completely hydroxylated lysines.

residues, as well. For example, in the span from residues 18 through 30, sequence differences occur at seven positions, a frequency which is unusually high, especially when one considers that glycine residues must remain invariant to maintain the helical configuration of the collagen molecule. Conversely there are groups of amino acids in the sequences exhibiting no interchain differences, the most noteworthy span being residues 40 through 55.

Some of the structures which are probably selected for during evolution of the α chains are those promoting the formation of the triple-chain helix, formation of fibrils and the formation of covalent cross-links. The above observations suggest the evolutionary preservation of common structural features of various collagens, which reside in clusters of invariant amino acid residues. The data also suggest that cartilage collagen has been subjected to additional selective pressures which are expressed in the clusters of variable amino acid residues. If the above hypothesis proves to be valid, the identification of other variable and invariant regions should aid in locating sites in collagen α chains involved in the above interactions.

ACKNOWLEDGEMENTS: This work was supported by USPHS Grant DE-02670 from the National Institute of Dental Research.

REFERENCES

1. Miller, E. J., Clin. Ortho. 92, 260-280 (1973).
2. Miller, E. J., Epstein, E. H., Jr. and Piez, K. A., Biochem. Biophys. Res. Commun. 42, 1024-1029 (1971).
3. Chung, E. and Miller, E. J., Science, in press.
4. Kefalides, N. A., Biochem. Biophys. Res. Commun. 45, 226-234 (1971).
5. Piez, K. A., Balian, G., Click, E. M. and Bornstein, P., Biochem. Biophys. Res. Commun. 48, 990-995 (1972).
6. Fietzek, P. P., Kell, I. and Kühn, K., FEBS Letters 26, 66-68 (1972).
7. Miller, E. J., Biochemistry 10, 3030-3035 (1971).
8. Miller, E. J. and Lunde, L. G., Biochemistry 12, 3153-3159 (1973).

9. Miller, E. J., Woodall, D. L. and Vail, M. S., J. Biol. Chem. 248, 1666-1671 (1973).
10. Fietzek, P. P., Wendt, P., Kell, I. and Kühn, K., FEBS Letters 26, 74-76 (1972).
11. Pisano, J. J., Bronzert, T. J. and Brewer, H. B., Jr., Anal. Biochem. 45, 43-59 (1972).
12. Inagami, T. and Murakami, K., Anal. Biochem. 47, 501-504 (1972).
13. Inagami, T., Anal. Biochem. 52, 318-321 (1973).
14. Bornstein, P., Biochemistry 8, 63-71 (1969).
15. Butler, W. T., Biochem. Biophys. Res. Commun. 48, 1540-1548 (1972).
16. Butler, W. T. and Cunningham, L. W., J. Biol. Chem. 241, 3882-3888 (1966).
17. Miller, E. J. and Piez, K. A., Anal. Biochem. 16, 320-326 (1966).
18. Miller, E. J., Biochemistry 10, 1652-1659 (1971).