DO MAMMALIAN COLLAGENASES AND DNA RESTRICTION ENDONUCLEASES SHARE A SIMILAR MECHANISM FOR CLEAVAGE SITE RECOGNITION?

Robert A. Brown, David W.L. Hukins and Jacqueline B. Weiss*

Departments of Rheumatology and Medical Biophysics,
University of Manchester, Manchester, England

Trevor M. Twose

Department of Biochemistry, I.C.I. Pharmaceuticals Ltd.,
Alderley Park, Cheshire, England

Received December 8,1976

Summary: Despite the apparent uniformity of the collagen molecule, vertebrate and invertebrate collagenases cleave it in one region only. We suggest that the enzyme recognises the cleavage site by the arrangement of the imino acids proline and hydroxyproline on either side of a region where the helical conformation of the collagen molecule is less stable. This less stable region could fold out of the rigid collagen molecule allowing the two recognition sites to be simultaneously attached to identical subunits in the same collagenase molecule. Class II DNA restriction endonucleases are confronted by a similar recognition problem in cleaving the DNA molecule at a specific site and it is generally accepted that here recognition is achieved by a sequence of bases with two-fold symmetry. We postulate that collagenase may, like the DNA restriction enzyme, be active in the dimeric form and that it recognises its substrate site by a similar two-fold symmetric arrangement of imino acid residues.

Collagen is a long rigid molecule (1) with a fairly uniform amino acid sequence in which glycine occurs as every third residue sequentially, proline accounts for 12% of all the amino acid residues and hydroxyproline for approximately 8% in Type I collagen (2). Thus, the triplet sequence Gly-X-Y repeats regularly throughout the length of the molecule and X and Y may be proline or hydroxyproline respectively, although most amino acids are sterically capable of being incorporated into the X or Y position (3). The whole molecule is made up of three so-called α chains, each coiled in a left handed minor helix which together coil in a right handed major helix and in which freedom of rotation about the backbone bonds within the constituent α chains is restrained by the presence of the imino acids, proline and hydroxyproline, thus stabilising both the minor helices and major helix of the interwoven chains (3). In three of the four known collagen types the three α chains are the same, but in Type I collagen, which is the predominant collagen of most vertebrate tissues, there are

* To whom reprint requests should be sent.

two identical α_1(I) chains and a dissimilar α_2 chain (4). Small terminal peptides at both ends of each chain are not in a helical conformation and do not have glycine as every third residue.

Mammalian collagenases from a variety of sources as well as tadpole tail fin collagenase are all able to cleave the collagen molecule in only one site along the whole helical body (5). This site on the α_1(I) chain is between two amino acids which are contiguous in at least two other places within the helix and on the α_2 chain is between two different amino acids (6). Other proteolytic enzymes are unable to cleave the helical portion of the molecule but can attack the terminal (non-helical) peptides. We are concerned here with recognition of the cleavage site by the collagenase molecule and postulate that it may resemble the mechanism by which Class II DNA restriction endonucleases distinguish the site at which they cleave the DNA molecule (7).

Collagenase cleavage site

Weiss (5), has indicated that the collagenase cleavage site is at the end of a region where the helical conformation of the collagen molecule is less stable and that at either end of this region the imino acid residues proline and hydroxyproline have a distinctive distribution (Fig. 1). This less stable helical region consists of 16 amino acid residues containing only one imino acid. After cleavage by collagenase it has been shown that pepsin at 18°C, which cannot attack the stable triple helix is able to remove these 16 residues (8). The six triplets on either side of this region show what Weiss termed "inverse symmetry" with regard to their proline and hydroxyproline content (5). Non-hydroxylated and hydroxylated proline have been shown to alter the conformational stability of the triple helix (9)(10). Evidence for a considerable change in melting temperature, T_m of collagen fibrils containing differing proportions of hydroxyproline and proline is itself conclusive, but recent evidence using synthetic polypeptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and $(Pro\text{-}Pro\text{-}Gly)_{10}$ showed the former to have a T_m 30°C higher than the latter (11). It has also been suggested that hydroxyproline, unlike proline, serves a double function in collagen, the OH group bonding via a water molecule to link peptide bonds in a single triple helix or alternatively to form hydrogen bonds linking two triple helices together (12).

In all but one case of vertebrate or invertebrate collagenases the actual substrate cleavage occurs at the amino terminal end of this "weak" region (5) but one collagenase from human gastric mucosa has been shown to

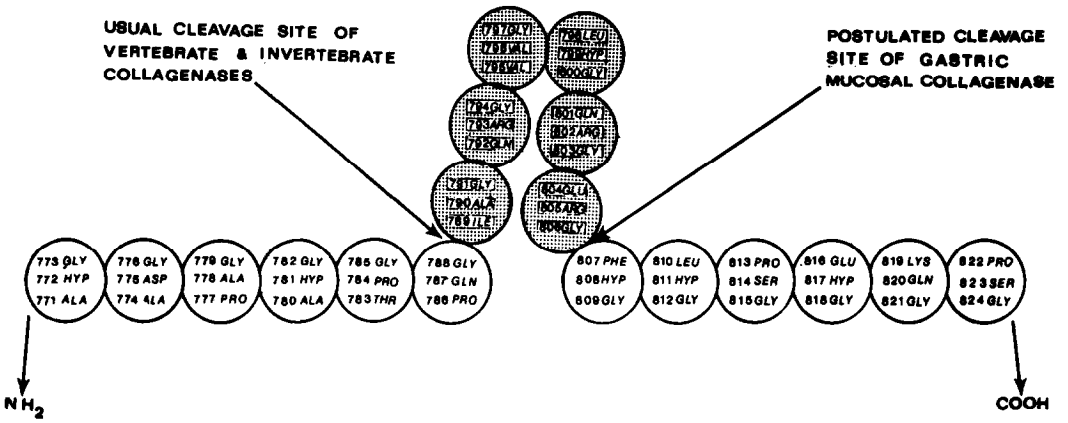

Fig. 1. Collagen sequence in vicinity of collagenase cleavage site. Each circle represents a Gly-X-Y triplet and identifies the sequence and numbering scheme referred to in the text. The less stable helical region is shown folded out of the rest of the sequence to emphasise how the two postulated recognition sites can approach each other. Sequence information for the α_1(I) chain is reviewed by Fietzek and Kühn (1976).(2)

cleave only at the carboxyl terminal end of the same region (13). The cleavage sequence itself is not stringent since the α_1(I) polypeptide chain is cleaved between Gly-Ileu and the α_2 polypeptide chain between Gly-Leu by the same enzyme (6). Certain other differences of sequence occur in this region but they are not unique. For example, one of the three threonine residues in the α_1(I) chain which occur in the X position is found in the triplet immediately preceding the cleavage site as is one of the only four proline residues present in the Y position. This latter point is of importance and will be discussed subsequently. No such abnormality is evidenced in any of the triplets at the other end of the postulated non-triple helical sequence where gastric collagenase cleaves.

The above information relates to the α_1(I) chain since the sequences for this region of the α_2 chain or of Types II and III collagens are not yet available. However, the evidence for the "weak" region was obtained by electron microscopic examination of segment long spacing aggregates (in which the molecules are stacked one above the other, presenting a laterally extended view of the whole molecule) and of course these include the α_2 chain. It is interesting that electron microscopic evidence was used to postulate this "weak" region of some 16 residues before it was confirmed by

sequence analysis (8): there being one only imino acid residue present in the five triplets involved. Although there are other series of adjacent triplets lacking in imino acid present along the chain, no others are surrounded by distinctive arrangements of imino acid residues. Reversibly denatured collagen is also cleaved only at the same single locus as the native molecule but the enzyme has a remarkably low specificity for this substrate and it has been pointed out that this is strong presumptive evidence for a requirement by the enzyme of a substrate with helical structure (14).

DNA restriction endonuclease cleavage site recognition

In a sense the problem of recognition confronting mammalian collagenase is not dissimilar to that of Class II DNA restriction endonuclease enzymes in their cleavage of the DNA molecule. It is now generally accepted that cleavage sites along DNA possess two-fold rotational symmetry, that is short sections of sequence are related to sequence on the complementary chain by one of the diad axes (two-fold rotational axes) generated by the symmetry of the Watson-Crick base pairs. Such restriction enzymes of the Ru class appear to make only one cleavage per 1000 base pairs and the cleavage sites are generally 4 to 6 base pairs long.(15)

The enzyme molecule itself has been reported to be a dimer.(16) It is postulated that the two identical subunits recognise the same sequence of bases on opposite strands of the DNA double helix. Presumably the enzyme subunits are related by a diad axis to enable them to recognise the same symmetry related base sequences on the two polynucleotide chains (7).

We note that both the restriction endonuclease and collagenase require a divalent cation for their activity. Removal of extrinsic Ca^{2+} from all collagenases results in a reversible loss of enzymic activity (17). In the case of collagenase it has been shown that Ca^{2+} does not stabilise the substrate itself and that binding of enzyme to substrate does not require Ca^{2+} but its role in the stabilisation or activation of the enzyme complex, like that of Mg^{2+} for the restriction endonuclease, is not yet known.

Collagenase cleavage site recognition

We suggest that the arrangement of imino acid residues around the region of the collagen molecule where the helical conformation is less stable because of the low content of imino acids is responsible for collagenase recognising its cleavage site. Furthermore we propose that an

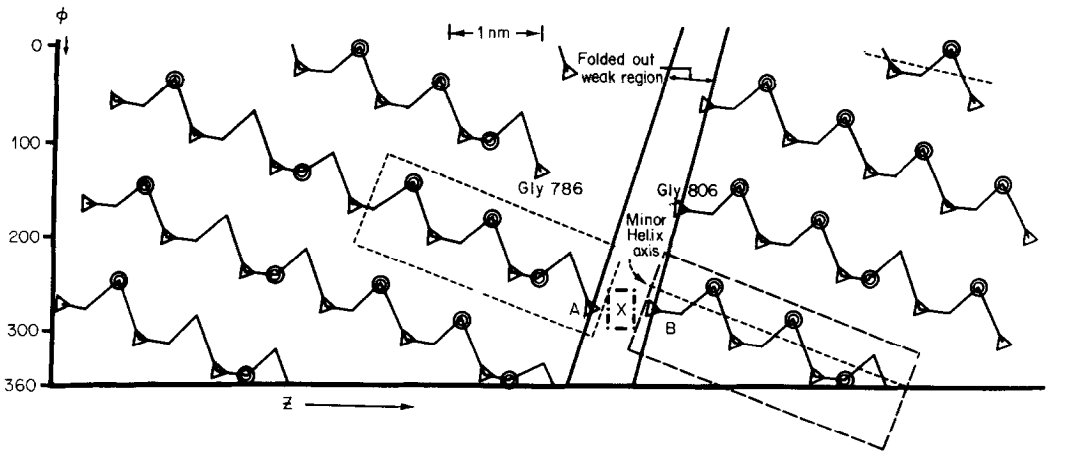

Fig. 2. Three dimensional structure of collagen molecule. The sequence information given in Fig. 1 is plotted on a helix net. The necessary information on the helix structure is given by Rich and Crick (1961).(22) Gly = Δ ; an imino acid = ⊙ . In the figure the less stable helical region is considered to fold out from the rest of the molecule and the recognition site nearest the carboxyl terminus has been assigned an arbitrary origin relative to the end of the other site given by a rotation of 0. and a translation of the order of 0.3 nm. An axis which approximates to a diad relating the two recognition sites in their arrangement of glycine and imino acids is marked with a cross.

essential feature of the arrangement is two-fold rotational symmetry analagous to that in the base sequence of DNA at Class II restriction endonuclease cleavage sites. It is clear from Fig. 1 that the arrangement of glycine and imino acid residues starting from glycine 788 moving towards the amino terminus is about the same as that starting from glycine 806 and moving towards the carboxyl terminus, if we do not distinguish between proline and hydroxyproline. The movement of one imino acid (either proline 784 or hydroxyproline 811) by one position in the sequence would make the two arrangements identical.

A complication arises in that wherever hydroxyproline occurs as the imino acid in one recognition site, proline occurs in the symmetry related position in the other and vice versa. We may explain this observation as follows. Both regions must contain hydroxyproline as well as proline to enhance the stability of their helical conformations as compared to that of the "weak" region. However it would appear that hydroxylation of proline is generally only feasible at the Y position in the Gly-X-Y triplet (2).

These conditions, together with the two-fold symmetry relating the arrangement of imino acids in the two regions, are simultaneously satisfied by alternation of proline and hydroxyproline in the two regions. This alternation has been termed "inverse symmetry" (5). We note that the restriction on the site of proline hydroxylation means that the anomalous position of proline 784 or hydroxyproline 811 could only be overcome by moving the proline residue to position 783. This proline residue is therefore in the Y position and is one of only four such residues present in the Y position in the whole molecule. It may be that this eccentricity leads to usual cleavage site at one end of the "weak" region rather than the other.

In the linear sequence of Fig. 1 the two arrangements of imino acids are related by a diad axis and Fig. 2 shows that this relationship can be almost retained in three dimensions. The relationship could be improved by rotating about the axis of one of the minor helices which then slightly disturb the triple helical conformation. Some disturbance of the triple helix is not unexpected especially during the early stage of cleavage by collagenase. We expect that the weak helical region could fold out of this conformation allowing both the imino acid-containing recognition sites to bind to the collagenase molecule. Because of the uncertain conformation of the weak region, the relationship between the two recognition sites is arbitrary. The free energy required to distort the weak helical region could be offset by the presumed total decrease in free energy brought about by cleavage of the collagen molecule by collagenase. This conformational distortion is analagous to the distortion of a stable C1 pyranose ring into a less stable conformation when lysosyme cleaves its substrate (18).

Our suggestion is consistent with the observation that collagenase exists as oligomers. Two highly purified preparations of collagenase studied in this context have been shown to exist as monomers and dimers or as monomers, dimers and tetramers (19)(20). We believe that the collagenase molecule must consist of at least two identical subunits, probably related by a diad axis, in order, like the DNA restriction endonucleases, to recognise each nearly identical site on the collagen molecule simultaneously.

Acknowledgments: This work was supported in part by a grant from the Arthritis Rheumatism Council. Mr. A. Brooks kindly prepared the drawings.

References:

1. Harrington, W.F. and von Hippel, P.H. (1961) Advan. Protein Chem. 16, 1-138.
2. Fietzek, P.P. and Kühn, K. (1976) Int. Rev. Conn. Tiss. Res. 7, 1-56.
3. Traub, W. and Piez, K.A. (1971) Advan. Protein Chem. 25, 243-352.

4. Martin, G.R., Byers, P.H. and Piez, K.A. (1975) Advan. Enzymol. 42, 167-191.
5. Weiss, J.B. (1976) Int. Rev. Conn. Tiss. Res. 7, 102-149.
6. Gross, J., Harper, E., Harris, E.D.Jr., McCroskery, P., Highberger, J.H., Corbett, C. and Kang, A.H. (1974) Biochem. Biophys. Res. Commun. 61, 605-612.
7. Kelly, T.J. and Smith, H.O. (1970) J. Mol. Biol. 51, 393-409.
8. Leibovich, S.J. and Weiss, J.B. (1973) Connect. Tiss. Res. 2, 11-19.
9. Jimenez, S., Harsch, M. and Rosenbloom, J. (1973) Biochem. Biophys. Res. Commun. 52, 106-111.
10. Berg, R.A. and Prockop, D.J. (1973) Biochem. Biophys. Res. Commun. 52, 115-120.
11. Traub, W. (1974) Israel J. Chem. 12, 435-439.
12. Ramachandran, G.N., Bansal, M. and Ramakrishnam, C. (1975) Curr. Sci. 44, 1-3.
13. Woolley, D.E., Tucker, J.S., Green, G. and Evanson, J.M. (1976) Biochem J. 153, 119-126.
14. McCroskery, P.A., Wood, S.Jr. and Harris, E.D.Jr. (1973) Science 182, 70-71.
15. Nathans, D. and Smith, H.O. (1975) Ann. Rev. Biochem. 44, 273-289.
16. Greene, P.J., Bellaver, M.C. and Boyer, H.W. (1974) Methods Mol. Biol. 7, 87-
17. Seltzer, J.L., Welgus, H.G., Jeffrey, J.J. and Eisen, H.Z. (1976) Arch. Biochem. Biophys. 173, 355-361.
18. Phillips, D.C. (1967) Proc. Nat. Acad. Sci. U.S. 57, 484-495.
19. Harris, E.D. Jr., DiBona, D.R. and Krane, S.M. (1969) J. Clin. Invest. 48, 2104-2113.
20. Birkedal-Hansen, H., Taylor, R.E. and Fullmer, H.M. (1976) Biochim. Biophys. Acta 420, 428-432.
21. Rich, A. and Crick, F.H.C. (1961) J. Mol. Biol. 3, 483-506.