Sequence Specificity of Human Skin Fibroblast Collagenase

EVIDENCE FOR THE ROLE OF COLLAGEN STRUCTURE IN DETERMINING THE COLLAGENASE CLEAVAGE SITE

(Received for publication, October 15, 1986)

Gregg B. Fields‡, Harold E. Van Walt†, and Henning Birkedal-Hansen§

From the ‡Department of Chemistry and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306 and the §Department of Oral Biology and Institute of Dental Research, University of Alabama School of Dentistry, Birmingham, Alabama 35294

The sequence specificity of human skin fibroblast collagenase has been investigated by measuring the rate of hydrolysis of 16 synthetic octapeptides covering the P₁ through P₄' subsites of the substrate. The choice of peptides was patterned after potential collagenase cleavage sites (those containing either the Gly-Leu-Ala or Gly-Ile-Ala sequences) found in types I, II, and III collagen. The initial rate of hydrolysis of the P₁-P₄' bond of each peptide has been measured by quantitating the concentration of amino groups produced upon cleavage after reaction with fluorescamine. The reactions have been carried out under first-order conditions ([S] ≈ Kₘ) and kₘ₄/Kₘ₄ values have been calculated from the initial rates. The amino acids in subsites P₁ (Pro, Ala, Leu, or Asn), P₂ (Gln, Leu, Hyp, Arg, Asp, or Val), P₄' (Ile or Leu), and P₄'' (Gln, Thr, His, Ala, or Pro) all influence the hydrolysis rates. However, the differences in the relative rates observed for these octapeptides cannot in themselves explain why fibroblast collagenase hydrolyzes only the Gly-Leu and Gly-Ile bonds found at the cleavage site of native collagens. This supports the notion that the local structure of collagen is important in determining the location of the mammalian collagenase cleavage site.

The triple helical region of interstitial collagens is highly resistant to all proteinases except specific collagenases (1, 2). In all higher organisms, these collagens are catabolized by the so-called tissue collagenases which have a characteristic and highly specific mode of action. They cleave all three α chains of native types I, II, and III collagens at a single locus by hydrolyzing the peptide bond following the Gly residue of the partial sequences Gly-[Ile or Leu]-[Ala or Leu] located approximately three-fourths from the NH₂ terminus (3–5). It has been pointed out, however, that there are Gly-[Ile or Leu] sequences at other sites within the triple helical domain of these collagens that are not cleaved (3, 6). In fact, it is possible to identify a total of 10 different sites in the α(I) chains of chick, rat, and calf skin (7–11), the α(2)I chain of calf skin (4, 12), and the α(III) chain of human liver collagens (13) that contain Gly-[Ile or Leu]-Ala sequences that are not cleaved by tissue collagenases in the native collagens.

The basis for the selective hydrolysis of the Gly-[Ile or Leu] bond at the cleavage site must lie in the influence of the surrounding residues in each α chain. On the one hand, it is possible that tissue collagenases have a large active site and a very restrictive sequence specificity so that only the extended sequences found at the cleavage site are recognized by the enzyme. One postulate to explain the cleavage site specificity in terms of collagen sequence has also considered the symmetry in the distribution of imino acids around the scissile bond (14). Alternatively, the sequence surrounding the cleavage site might indirectly endow the scissile Gly-[Ile or Leu] bond with hyper-reactivity by altering the local conformation in this region. For example, a local deficiency of imino acids could destabilize a short segment of the triple helix, allowing the enzyme access to this bond (Refs. 15–18 and references cited therein). Another possibility is that the cleavage site region of collagen adopts a presently unknown, but specific, secondary structure distinct from the triple helix that is recognized by the enzyme. It is of fundamental importance to determine whether it is the sequence specificity of tissue collagenases or a local conformational feature of the collagenase cleavage site that is responsible for this unique substrate specificity.

In the present study, the action of human skin fibroblast collagenase on a series of synthetic octapeptides has been investigated. The sequences of these peptides have been specifically chosen with reference to those in the potential, but non-cleavable, collagenase cleavage sites in native rat, calf, and chick type I and human type III collagens. By quantitating the effects of single amino acid substitutions on the rates of hydrolysis of these synthetic octapeptides, definitive information on the sequence specificity of the enzyme has been obtained. The results of these single substitutions on the hydrolysis rates allow us to assess the degree to which the sequence specificity of this enzyme alone determines the location of cleavable sites in native collagens.

EXPERIMENTAL PROCEDURES

Materials—Procollagenase was purified to homogeneity from serum-free cultures of human skin fibroblasts in a three-step procedure involving consecutive chromatography over zinc-chelate-Sepharose, heparin-Sepharose, and Ultrogel ACA 44 columns (1). The enzyme consisted of a 57/54-kDa doublet and was free from gelatinase activity. Fmoc-Arg(Mtr)₂ and Fmoc-His(Boc)OPfp were purchased from

1 H. Birkedal-Hansen, B. Birkedal-Hansen, R. E. Taylor, and H. Y. Lin, manuscript in preparation.
2 The abbreviations used are: Fmoc, 9-fluorenylmethoxycarbonyl (Fmoc-amino acid denotes N-Fmoc throughout this paper and side chain blocking groups all follow the amino acid in parentheses, all amino acids are of the L-configuration); Mtr, N′4-methoxy-2,3,6-trimethylbenzenesulfonyl; Boc, t-butyloxycarbonyl; OPfp, pentafluorophenyl ester; Tricine, N-tris(hydroxymethyl)methylglycine; dansyl, 5-dimethylaminonaphthalene-1-sulfonfyl.

6221
For several peptides, \( u_0 \) was measured as a function of \([S]\) and the \( k_{cat} \) and \( K_M \) values for the reactions were determined from double-reciprocal plots. \( [E_0] \) was measured spectrophotometrically using \( \epsilon_{290} = 6.5 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1} \) which assumes a molecular weight of 51,929 for procollagenase (24).

The site of hydrolysis of all peptides was determined by dansylation of the reaction products followed by high pressure liquid chromatographic analysis. Aliquots (100 \( \mu l \)) of the reaction mixtures were removed at 0 and 48 h and reacted with 100 \( \mu l \) of an 8 mM solution of dansyl chloride in acetonitrile. The samples were applied to a Rainin Microsorb C5 3-\( \mu l \) column (4.6 x 250 mm) and eluted with a linear gradient prepared by mixing 50 mM sodium phosphate, pH 6.5, and 50% acetonitrile in water. In all cases, only two peptides were observed. The site of hydrolysis was determined by comparing the retention time of the two peptides with that of a dansylated tetrapeptide of known composition that corresponded to either the first or last 4 residues of the substrate. All substrates were hydrolyzed exclusively at the P1\( ' \) bond.

The kinetic parameters for the hydrolysis of rat tendon type I collagen by fibroblast collagenase were measured at 30 \( ^\circ \text{C} \) in the same buffer by using a newly developed radioassay (25, 26).

**RESULTS**

The choice and placement of the amino acids in the peptides that have been synthesized have been guided by sequences found in types I, II, and III collagens, all of which are substrates for human skin fibroblast collagenase. The whole or partial sequences surrounding the tissue collagenase cleavage sites in the \( \alpha_1(1) \), \( \alpha_1(11) \), and \( \alpha_1(111) \) chains of several collagens have been determined (3-7, 10, 13, 27, 28) and are listed in Table I. For those cases in which the entire chain has been sequenced, the residues that form the scissile bond have been numbered, where the number indicates the position in the chain starting from the Gly residue at the NH\( 2 \) terminus of the first Gly-X-Y triplet that is believed to be the start of the triple helical region. For the \( \alpha_1(1) \) and \( \alpha_1(11) \) chains, this first triplet is taken to be Gly-Pro-Met (4, 7, 10) while for the \( \alpha_1(111) \) chains it is assumed to be Gly-Tyr-Hyp (13).

The labeling of the substrates of the \( \alpha_1(1) \), \( \alpha_1(11) \), and \( \alpha_1(111) \) chains, either a Gly\( ^7 \)Ile\( ^7 \)Leu\( ^7 \) or Gly\( ^7 \)Ile\( ^7 \)Leu\( ^7 \) bond is hydrolyzed. It is known that tissue collagenases cleave a Gly-Ile-Leu bond in \( \alpha_1(11) \) chains, but not enough of the sequence has been completed to establish the location in the chain. In all of these cases, the scissile bond is followed by either an Ala or a Leu residue. Thus, the sequences of the chains of types I, II, and III collagens immediately surrounding the tissue collagease cleavage site are Gly-[Ile or Leu]-[Ala or Leu].

The complete sequences of several collagen chains have been examined in order to locate loci that are not in the collagease cleavage site, but which contain the partial sequences Gly-[Ile or Leu]-[Ala or Leu]. Examination of the partial sequence of the \( \alpha_1(111) \) chain from rat (10, 11) and of the complete sequences of the \( \alpha_1(1) \) chains from calf and chick skin (7-10), the \( \alpha_2(1) \) chain from calf skin (4, 12), and the \( \alpha_1(11) \) chain from human liver (13) reveals 10 loci that contain either the Gly-Ile-Ala or Gly-Leu-Ala sequences that are not cleaved by tissue collagenases. The sequence of 8 residues centered on each of these sites is listed in Table II, where amino acid substitutions relative to the cleavage site in the \( \alpha_1(1) \) chain of chick or calf skin are underlined. These sequences span the P\( 7 \) through P\( ' \) sites of these potential substrates. No loci containing the Gly-Ile-Leu or Gly-Leu-Leu sequences were found. By comparing the 10 sequences...
by fibroblast collagenase are uniformly high, accurate values approximately been synthesized. Two peptides the peptides. An example of a double-reciprocal plot is shown not exceed from experiments in which the substrate concentration did

| Collagen chain | Gly · Pro · Gln · Gly²⁷⁶ · Ile²⁷⁸ · Ala · Gly · Gln |
|---------------|-------------------------------------------------|
| Calf α1(I)    | 27                                              |
| Chick α1(I)   | 7                                               |
| Calf α2(I)    | 4, 10                                           |
| Chick α2(I)   | 5                                               |
| Bovine cartilage α1(II) | Gly · Ile · Ala · Gln | 3 |
| Human liver α1(III) | Gly · Pro · Leu · Gly²⁷⁶ · Ile²⁷⁸ · Ala · Gln | 13 |
| Human skin α1(III) | Gly · Leu · Ala · Leu · Gln | 3 |
| Calf α1(III)  | Gly · Pro · Leu · Gly²⁷⁶ · Ile²⁷⁸ · Ala · Gly · Leu | 28 |

**Table I**

Collagen sequences cleaved by tissue collagenases

**Table II**

Types I and III collagen sequences containing the Gly-Ile-Ala or Gly-Leu-Ala triplets but not cleaved by tissue collagenases

Amino acid substitutions relative to the α1(I) chain of chick or calf skin collagen are underlined.

| Collagen chain | Gly · Ala · Hyp · Gly²⁷⁸ · Ile²⁷⁸ · Ala · Gly · Ala |
|---------------|-------------------------------------------------|
| Rat, calf, and chick α1(I) | 7, 8, 11 |
| Calf and chick α1(I) | 7, 9 |
| Chick α1(I) | 7 |
| Chick and calf α2(I) | 4, 10, 12 |
| Calf α2(I) | 4, 10 |
| Calf α2(II) | 4, 10 |
| Human liver α1(III) | Gly · Leu · Hyp · Gly²⁷⁸ · Ile²⁷⁸ · Ala · Gln · His | 4 |
| Human liver α1(III) | Gly · Leu · Hyp · Gly²⁷⁸ · Ile²⁷⁸ · Ala · Gly · Pro | 13 |
| Human liver α1(III) | Gly · Ala · Arg · Gly²⁷⁸ · Leu²⁷⁶ · Ala · Gln · Pro | 13 |
| Human liver α1(III) | Gly · Leu · Hyp · Gly²⁷⁸ · Leu²⁷⁶ · Ala · Gln · Thr | 13 |

**Fig. 1.** Double-reciprocal plots for the hydrolysis of (A) Gly-Pro-Gln-Gly-Ile-Ala-Gly-Gln (I) and (B) rat tendon type I collagen by human skin fibroblast collagenase. Assays were carried out in 50 mM Tris-Cl, 0.2 M NaCl, 10 mM CaCl₂, pH 7.5, at 30°C.

listed in Table II to that of the cleavage site in the α1(I) chain, 13 substitutions have been identified: Ala, Asn, or Leu for Pro in subsite P,; Hyp, Asp, Val, Leu, or Arg for Gln in subsite P, Leu for Ile in subsite P,; and Ala, Pro, His, or Thr for Gln in subsite P,’. In order to investigate the specificity of fibroblast collagenase toward these sequences, a series of octapeptides (2–14), each constituting a single substitution relative to Gly-Pro-Gln-Gly-Ile-Ala-Gly-Gln (1), have been synthesized. Two peptides (15 and 16) with multiple substitutions were also prepared. Earlier work by Nagai and associates (30, 31) indicated that an octapeptide is a suitably sized synthetic substrate for studies with a vertebrate collagenase.

The proper kinetic parameter with which to assess sequence specificity is kcat/KM. In an initial series of kinetic measurements, the initial rates of hydrolysis of several peptides were measured as a function of substrate concentration and kcat and KM values were determined individually from double-reciprocal plots. The maximum substrate concentrations (approximately 2 mM) studied were limited by the solubilities of the peptides. An example of a double-reciprocal plot is shown in Fig. 1A for peptide 1 to demonstrate that the plot is linear with no kinetic anomalies and that Michaelis-Menten kinetics are obeyed. For this substrate, kcat and KM are found to be 730 h⁻¹ and 3.5 mM, respectively. The parameters for this and six other peptides are summarized in Table III. In all cases, the KM values are high (>2 mM) and the kcat values fall in the 260–1200 h⁻¹ range. Thus, these parameters were determined from experiments in which the substrate concentration did not exceed KM and the accuracy of the numbers listed in Table II must be viewed in the light of this limitation.

Since the KM values for the hydrolysis of these octapeptides by fibroblast collagenase are uniformly high, accurate values of kcat/KM for all 16 peptides were determined by measuring the initial rates at a substrate concentration of 0.2 mM at which the reaction was first-order in substrate ([S]₀ ≪ KM). Equation 1 was used to calculate the kcat/KM values which are summarized in Table IV in units of micromolar⁻¹ h⁻¹. All of the single substitutions examined influence the hydrolysis rates measurably. Substitution of Ala (2), Leu (3), or Asn (4) for Pro in subsite P, lowers the rates to 50, 14, and 17% of that of peptide 1, respectively. When the Gln residue in
subsite P2 is replaced by Leu (5), Hyp (6), Arg (7), Asp (8), or Val (9), the rates are 150, 11, 17, 30, and 32% of those of peptide 1. Substitution of Leu (10) for Ile at subsite P1' increases the relative rate to 130%. Replacement of the Gln residue of 1 at subsite P1 with Thr (11), His (12), Ala (13), or Pro (14) changes the rates to 160, 91, 36, and 77% of that of peptide 1, respectively. Thus, it is clear that the identity of the residues in subites P3, P2, P1', and P1'' affect the rates at which these octapeptides are hydrolyzed by fibroblast collagenase.

Peptides 2–14 all differ from 1 by single substitutions at one of four subsites. All of the noncleavable collagen sequences listed in Table II, however, differ from the cleavable sequences by two or three substitutions. Only 15 and 16 are directly comparable to these noncleavable sequences. In order to estimate the effects of the multiple substitutions shown in Table II on the rates of hydrolysis of these sequences relative to that of peptide 1, it is of interest to determine whether the effects of the single substitutions on the rates are independent of one another (i.e., the effects on the rates are multiplicative). From transition state theory, this amounts to assuming that the alterations of \( \Delta G^2 \) for the reaction produced by each amino acid substitution are additive.

To test this assumption, the rates of hydrolysis of the multiply substituted peptides 15 (Leu for Gin in subsite P2, and Pro for Gin in subsite P1') and 16 (Hyp for Gin in subsite P2, Leu for Ile in subsite P1', and Ala for Gin in subsite P1'') have been determined. The relative hydrolysis rates predicted for peptides 15 and 16 based on the assumption that the effects of single substitutions are multiplicative are 120 and 12% of that of 1 and are in reasonably good agreement with the measured values of 120 and 16%, respectively. Thus, this assumption appears to be reasonable and has been used to calculate the relative rates of hydrolysis of the noncleavable sequences in collagen listed in Table II. The results are expressed relative to the cleavable sequence from the al(I) chain (peptide 1) and are summarized in Table V.

The calculated effects of these multiple substitutions range from changes of 1.2 to 120% of the rate observed for peptide 1. Based on these results, it is apparent that sequence specificity alone cannot explain the substrate specificity of human skin fibroblast collagenase toward native types I and III collagens. If the sequence specificity of the enzyme was the sole factor in determining the location of the cleavage site in collagens, the data in Table V indicate that \( \alpha \) chains in at least four of the 10 loci would be hydrolyzed at a rate that is at least 10% of the rate for the cleavable sequence and the other six at more than 1.2% of this rate. In fact, there is no evidence to indicate that the loci listed in Table II are cleaved at all in intact, triple helical collagens. Thus, the sequence specificity of the enzyme is not restrictive enough to account for its hydrolysis of native interstitial collagens at a single site. Instead, it appears that some local conformational feature of collagen at the cleavage site controls the specificity of this enzyme, since potentially susceptible sequences that are cleaved in small synthetic peptides are protected in the native collagens.

In order to provide a point of reference for the kinetic parameters measured here for the octapeptides, the values for the hydrolysis of soluble rat tendon type I collagen have also been determined. Using a newly developed soluble radioassay (25, 26), the double-reciprocal plot shown in Fig. 1 has been obtained. The values of \( k_{cat} \) and \( K_M \) are 10 h\(^{-1}\) and \( ^{\mu} \text{M} \), respectively, where the \( k_{cat} \) value refers to the hydrolysis across all three chains of triple helical collagen at the collagen.
Collagenase cleavage site. For comparison, the parameters for the hydrolysis of types I and III collagens (32) and the α1(I), α2(I), and α1(III) gelatins (33) taken from the literature are listed in Table VI. Also shown are the data for the octapeptides with sequences that most closely resemble those of the cleavage sequence in the α1(I), α2(I), and α1(III) chains of types I and III collagens. It is clear from these data that the $k_{cat}/K_m$ values for the hydrolysis of collagens and gelatins by fibroblast collagenase are much higher than those for the octapeptides primarily because $K_m$ is much lower. Thus, there is clearly a major difference in the way the enzyme binds to a cleavable sequence when it is embedded in an intact α chain in either gelatin or collagen compared to an octapeptide.

**DISCUSSION**

This study presents data on the sequence specificity of human skin fibroblast collagenase toward synthetic octapeptides. The 16 peptides that have been synthesized have been carefully chosen with reference to loci found in types I and III collagens that contain the Gly-Ile-Ala or Gly-Leu-Ala sequences. More importantly, however, our data show that the identity of the amino acids directly adjacent to this core does not play a dominant role in determining the location of the cleavage site in native collagen through sequence specificity.

The foregoing conclusions suggest that the sequence surrounding the noncleavable loci determines their susceptibility to collagenase in native collagens indirectly by altering the local conformation of collagen itself. Since all of the noncleavable sequences identified have the repeating Gly-X-Y primary structure, all have the potential to adopt the traditional triple helix. The hypothesis has been put forth, however, that there are “locally unstable” regions of the triple helix brought about by a local deficiency of imino acids (15). The presence of a cleavable sequence in an unstable region could expose the scissile bond to the enzyme and account for the observed specificity. However, the local imino acid content of the collagen chains surrounding the noncleavable sequences listed in Table II are not very different from that in the cleavage region. Thus, this local deficiency per se cannot be the sole basis for the specificity of the enzyme toward native collagens. Apparently, there is a presently unrecognized local conformational feature of collagen that endows the cleavage region with hyper-reactivity toward collagenase.

It is of interest to compare the kinetic parameters for the hydrolysis of the α1(I), α2(I), and α1(III) chains in both collagen and gelatin with those for the octapeptides which are “locally unstable” regions of the triple helix brought about by a local deficiency of imino acids (15). The data presented here confirm that fibroblast collagenase is capable of cleaving many loci containing the Gly-[Ile or Leu]-Ala sequences. More importantly, however, our data show that the identity of the amino acids directly adjacent to this core does not play a dominant role in determining the location of the cleavage site in native collagen.

### Table VI

|                  | $k_{cat}$ | $k_{cat}/K_m$ | $T$  | Ref. |
|------------------|-----------|---------------|------|------|
| **Type I collagen** (calf skin) | 0.80 | 34.2 | 42.7 | 25 | 32 |
| **Type I collagen** (rat skin) | 0.90 | 19.5 | 21.7 | 25 | 32 |
| **Type I collagen** (rat tendon) | 0.80 | 10 | 13 | 30 | This study |
| α1(I) gelatin (guinea pig skin) | 7.0 | 230 | 33 | 37 | 33 |
| α2(I) gelatin (guinea pig skin) | 3.7 | 750 | 200 | 37 | 33 |
| Gly-Pro-Gln-Gly-Ile-Ala-Gly-Gln (1) | 3500 | 730 | 0.22 | 30 | This study |
| Gly-Pro-Gln-Gly-Leu-Ala-Gly-Gln (10) | 2800 | 970 | 0.35 | 30 | This study |
| Type III collagen (human) | 1.4 | 565 | 463.6 | 25 | 32 |
| α1(III) gelatin (human placenta) | 1.6 | 71 | 37 | 33 |
| Gly-Pro-Leu-Gly-Ile-Ala-Gly-Gln (5) | 3600 | 1200 | 0.53 | 30 | This study |

[2] G. B. Fields, K. A. Mockhtiar, and H. E. Van Wart, unpublished data.
not available, the closest set that can be assembled to date is summarized in Table VI. Several problems arise with respect to interpretation of these data. First, not all of the parameters were obtained under the same conditions or with collagen types from the same source. Second, the parameters for the hydrolysis of the $\alpha_1(I)$ and $\alpha_1(III)$ chains in gelatin were obtained by measuring the rate of disappearance of these chains as the result of hydrolysis at multiple sites. Thus, the parameters do not reflect a single proteolytic event. Third, the different sequences reported for the $\alpha_2(I)$ chains (Table I) makes direct comparison with a single octapeptide problematic. In spite of these limitations, one trend is clear. It is not immediately obvious why the $K_M$ values of collagenase toward the disordered $\alpha_1(I)$ and $\alpha_2(I)$ gelatin chains are at least 40-fold lower than those for peptides 1 and 10 which contain similar cleavage sequences, respectively. The $k_{cat}$ values, however, are similar in magnitude. This makes the $k_{cat}/K_M$ values much greater for the gelatin chains. It can also be seen that the $K_M$ values toward the native collagens are even lower than for the gelatins.

It is not immediately obvious why the $K_M$ values of collagenase are so much lower for the long ($\sim1000$ residue) gelatins and collagens. It is possible that fibroblast collagenase binds nonproductively to many sites along these substrates, thus lowering $K_M$ compared to the octapeptides. Another possibility is that binding to these substrates is enhanced because of their secondary structures. The long gelatin chains that are generally rich in imino acids might induce the sequences around the cleavage sites to adopt the poly-Pro(II) secondary structure and this might enhance binding. The triple-helical structure of the native collagens might likewise be a specific recognition feature. It will be necessary to examine the action of collagenase on longer substrates with alternate secondary structures to investigate these possibilities.

Acknowledgment—We thank Satiah Mallya for determining the kinetic parameters for the reaction of fibroblast collagenase with rat tendon type I collagen.

REFERENCES

1. Seifert, S., and Harper, E. (1971) in The Enzymes (Beyer, P. D., ed) 3rd Ed, Vol. III, pp. 649-697, Academic Press, Orlando, FL.
2. Woolley, D. B. (1984) in Extracellular Matrix Biochemistry (Piez, K., and Reddi, A. H., eds) pp. 119-158, Elsevier/North-Holland, New York.
3. Miller, E. J., Harris, E. D., Jr., Chung, E., Finch, J. E., Jr., McCrookery, P. A., and Butler, W. T. (1976) Biochemistry 15, 787-792.
4. Hofmann, H., Fietzek, P. P., and Kuhn, K. (1978) J. Mol. Biol. 125, 137-165.
5. Dixit, S. N., Mainardi, C. L., Seyer, J. M., and Kang, A. H. (1979) Biochemistry 18, 5416-5422.