# Relationship between the Individual Collagenases of *Clostridium histolyticum*: Evidence for Evolution by Gene Duplication†

Michael D. Bond and Harold E. Van Wart*

ABSTRACT: The relationship between the six collagenases ($\alpha$, $\beta$, $\gamma$, $\delta$, $\epsilon$, and $\zeta$) isolated and characterized in the preceding papers [Bond, M. D., & Van Wart, H. E. (1984) *Biochemistry* (preceding two papers in this issue)] has been investigated. Chemical modification reactions establish that all six enzymes contain essential carboxyl, tyrosine, and lysine residues. Circular dichroism spectra of the peptide bond region show that the secondary structures of the collagenases are very similar. Ouchterlony double-immunodiffusion experiments carried out with antiserum prepared against $\beta$-collagenase indicate that all six collagenases are cross-reactive. Reverse-phase high-pressure liquid chromatography elution profiles of tryptic digests of these collagenases and sodium dodecyl sulfate electrophoresis gels of the peptides formed on reaction with cyanogen bromide have been obtained. The results indicate that the class I collagenases have extensive sequence homology with each other and that the class II collagenases have extensive sequence homology with each other but that the enzymes in the two classes have substantially different sequences. In addition, the data show that $\beta$-collagenase probably consists of domains that have homologous amino acid sequences, which may have arisen by full or partial intragenic gene duplication. This may account for the unusually high molecular weight of this and the other collagenases. Finally, on the basis of the similarities between the collagenases in the two classes, it is suggested that one class evolved from the other by gene duplication followed by independent evolution by point mutations to yield enzymes with different substrate specificities.

In the preceding papers (Bond & Van Wart, 1984a,b), the purification, separation, and characterization of six collagenases (EC 3.4.23.3)[1] from *Clostridium histolyticum* have been described. On the basis of the biochemical and physicochemical properties of these collagenases, they have been assigned to two classes. The class I ($\alpha$, $\beta$, and $\gamma$) and class II ($\delta$, $\epsilon$, and $\zeta$) enzymes differ with respect to their activities, stabilities, and amino acid compositions, but all six enzymes share many similarities. The purpose of the present study was to investigate the relationship between the enzymes within each class and also the relationship between the enzymes in the two different classes.

The observation of multiple forms of an enzyme can be due to numerous phenomena. In the case of *Clostridial* collagenases, the most often cited explanation is that the bacterium produces one collagenase and that the other forms are derived from it by proteolysis. Without experimental evidence to the contrary, this explanation seems reasonable since it is clear that the bacterium produces numerous extracellular proteinases capable of carrying out the postulated proteolysis. However, the data in the preceding papers (Bond & Van Wart, 1984a,b) strongly support the existence of at least two types of collagenases. With this in mind, a series of additional experiments has been carried out to reveal both the similarities and differences between the mechanisms, three-dimensional structures, and sequences of the six enzymes.

## Materials and Methods

*Materials.* All collagenases and substrates are those described in the preceding papers (Bond & Van Wart, 1984a,b). Trypsin (TPCK) and chymotrypsin were obtained from Worthington Chemical Co., thermolysin and papain were from Sigma Chemical Co., and *Staphylococcus aureus* protease was from Pierce Chemical Co. Tetranitromethane, acetylimidazole, diethyl pyrocarbonate, *N*-ethyl-5-phenylisoxazolium-3′-sulfonate (Woodward's Reagent K), *N*-ethylmaleimide, and 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide were purchased from Aldrich Chemical Co., phenylmethanesulfonyl fluoride, (*p*-chloromercuri)benzoate, EDTA, trinitrobenzenesulfonate, and pepstatin A from Sigma Chemical Co., and 8-hydroxyquinoline-5-sulfonic acid and cyanogen bromide from Eastman Organic Chemicals.

*Chemical Modification Reactions.* Collagenases were incubated with the reagents shown in Table I for 1 h at 23 °C as described earlier (Bond et al., 1981). For the reversible inhibitors 8-hydroxyquinoline-5-sulfonic acid, pepstatin, and EDTA, the same concentration of reagent was added to the assay tube as was present in the incubation tube. In some reactions, hydroxylamine was added after 1 h and assays carried out after an additional 30 min.

*Circular Dichroism Measurements.* Spectra of the collagenases were obtained with a Jasco J-500C spectropolarimeter at 23 °C with a path length of 0.1 cm. The enzymes were dissolved in 1 mM Tris and 1 mM $CaCl_2$, pH 7.5, at a concentration of 0.8 mg/mL. The results are expressed as molar ellipticities, $[\theta]$, in units of deg cm$^2$/dmol of residues by using a mean residue molecular weight of 112 for all samples.

*Immunology.* Antiserum was raised against $\beta$-collagenase by subcutaneous, intramuscular and intradermal injections of a total of 600 $\mu$g of enzyme in Freund's complete adjuvant into a white, male rabbit. Secondary injections were give 3 weeks later of an additional 600 $\mu$g of enzyme in Freund's

† From the Department of Chemistry and the Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306. *Received October 4, 1983.* This work was supported by Research Grant GM27939 and Research Career Development Award AM01066 (to H.E.V.W.) from the National Institutes of Health, U.S. Public Health Service.

[1] Abbreviations: collagenase, *Clostridium histolyticum* collagenase; FALGPP, 2-furanacryloyl-L-leucylglycyl-L-prolyl-L-proline; FALGPA, 2-furanacryloyl-L-leucylglycyl-L-prolyl-L-alanine; [$^{14}CH_3$]collagen, [*methyl*-$^{14}$C]collagen; Tricine, *N*-[tris(hydroxymethyl)methyl]glycine; Tris, tris(hydroxymethyl)aminomethane; Mes, 2-(*N*-morpholino)-ethanesulfonic acid; nkat, nanokatal.

0006-2960/84/0423-3092$01.50/0 © 1984 American Chemical Society

Table I: Effect of Chemical Modification Reactions on the Activity of *Clostridium histolyticum* Collagenases

| reagent[a] | concn (mM) | buffer[b] | % act.[c] α | β | γ | δ | ε | ζ |
|---|---|---|---|---|---|---|---|---|
| none | | | 100 | 100 | 100 | 100 | 100 | 100 |
| 8-hydroxyquinoline-5-sulfonic acid | 1 | Tricine, pH 7.5 | 0 | 0 | 0 | 0 | 0 | 0 |
| ethylenediaminetetracetic acid | 1 | Tricine, pH 7.5 | 0 | 0 | 0 | 0 | 0 | 0 |
| *p*-(chloromercuri)benzoate | 1 | Tricine, pH 7.5 | 79 | 80 | 74 | 61 | 67 | 78 |
| *N*-ethylmaleimide | 1 | Tricine, pH 7.5 | 107 | 101 | 108 | 94 | 100 | 95 |
| pepstatin A | 1 μg/mL | Tricine, pH 7.5 | 104 | 91 | 96 | 90 | 103 | 99 |
| phenylmethanesulfonyl fluoride | 1 | Tricine, pH 7.5 | 103 | 88 | 103 | 94 | 92 | 86 |
| butanedione | 10 | borate, pH 8.0 | 98 | 96 | 101 | 99 | 100 | 98 |
| 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide | 10 | Mes, pH 6.0 | 7 | 21 | 25 | 15 | 8 | 13 |
| | 50 | | 3 | 3 | 0 | 2 | 0 | 3 |
| *N*-ethyl-5-phenylisoxazolium-3′-sulfonate | 30 | Tris, pH 7.5 | 3 | 6 | 6 | 7 | 0 | 2 |
| trinitrobenzenesulfonate | 10 | Tricine, pH 7.5 | 0 | 0 | 0 | 0 | 0 | 0 |
| diethyl pyrocarbonate | 20 | Tricine, pH 7.5 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 20 + 50 mM $NH_2OH$ | | 0 | 0 | 0 | 0 | 0 | 0 |
| tetranitromethane | 5 | Tricine, pH 7.5 | 3 | 0 | 0 | 0 | 0 | 0 |
| acetylimidazole | 10 | Tricine, pH 7.5 | 5 | 0 | 7 | 0 | 3 | 2 |
| | 10 + 50 mM $NH_2OH$ | | 69 | 50 | 70 | 20 | 41 | 38 |

[a] All reactions were carried out at 23 °C for 1 h. [b] All buffers contain 10 mM $CaCl_2$ (omitted for the reaction with ethylenediaminetetraacetic acid) and 400 mM NaCl. Buffer concentrations were 50 mM, except for the reaction with *N*-ethyl-5-phenylisoxazolium-3′-sulfonate for which the Tris concentration was 500 mM. [c] Activities were measured with [$^{14}CH_3$]collagen as substrate.

incomplete adjuvant. Thirty milliliters of blood was collected weekly and allowed to clot overnight, and the serum was decanted and centrifuged before use. The Ouchterlony double-immunodiffusion assays were carried out in 1.5% agar gels soaked overnight in 4% poly(ethylene glycol) with 20-μL sample wells.

*Tryptic Digests and Cyanogen Bromide Reactions.* For the tryptic digests, samples (100–400 μg) were heated to 100 °C for 15 min in 5 mM EDTA and 100 mM Tris, pH 8.1, and then incubated at a concentration of 1 mg/mL with a 4% w/w solution of Worthington trypsin-TPCK in 100 mM Tris and 15 mM $CaCl_2$, pH 8.1, at 37 °C for 10 h. The cyanogen bromide reactions were carried out according to the method of Epstein (1974) by addition of 1.5 mg of freshly prepared cyanogen bromide in 0.5 mL of 70% formic acid to 0.5 mg of each sample. The mixture was flushed with nitrogen and incubated at 37 °C for 10 h, after which time the formic acid was removed by lyophilization.

*Gel Electrophoresis.* All gel electrophoresis experiments were carried out as described in the preceding paper (Bond & Van Wart, 1984b) except that 10–15% gradient gels were used to analyze the peptides formed in the cyanogen bromide reactions. After the gels were fixed (25% trichloroacetic acid) and stained, each lane was scanned with an E-C Apparatus Corp. densitometer, and peak areas were integrated with a Hewlett-Packard Model 3390A integrator. Several tracings were obtained from different portions of the lanes to eliminate artifacts.

*High-Pressure Liquid Chromatography.* The tryptic digests of the collagenases were applied to a Bio-Sil ODS-5S reverse-phase column (250 × 4 mm) in 0.1% trifluoroacetic acid (Pierce, sequanal grade) and eluted with a linear gradient of acetonitrile (Baker, HPLC grade) containing 0.1% trifluoroacetic acid at a flow rate of 1.0 mL/min at 23 °C. The instrument consisted of a Beckman Model 421 controller, two Beckman 110A pumps, an Altex Model 210 injector, and a Varian Model UV-50 detector using a wavelength of 220 nm.

## Results

*Identification of Active Site Constituents.* We have previously established that the zinc atom in β-collagenase is essential for activity by demonstrating reversible inhibition with chelating agents (Bond et al., 1981). In addition, by carrying out a series of chemical modification reactions, it has been established that β-collagenase contains catalytically essential carboxyl, tyrosine, and lysine residues (Bond et al., 1981). To compare the mechanisms of these collagenases, these same experiments have been carried out on all six enzymes (Table I).

All six collagenases are totally inhibited by 8-hydroxyquinoline-5-sulfonic acid. Chelation of the active site zinc is responsible for this inhibition, since it is totally reversed by addition of excess zinc but not by addition of excess calcium (data not shown). EDTA also abolishes the activity of all six collagenases. Since activity is restored upon addition of excess calcium but not excess zinc, the source of the inhibition is the chelation of enzyme-bound calcium ions. These ions probably function to maintain the tertiary structure of the enzyme, much as in thermolysin.

None of the collagenases are significantly inactivated by the cysteinyl reagents *p*-(chloromercuri)benzoate or *N*-ethylmaleimide, the seryl reagent phenylmethanesulfonyl fluoride, the arginyl reagent butanedione, or the aspartyl proteinase inhibitor pepstatin A. However, the carboxyl reagents 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide and *N*-ethyl-5-phenylisoxazolium-3′-sulfonate (Woodward's Reagent K) inactivate all six collagenases, providing support for a catalytic aspartic or glutamic acid residue. The inactivation of these collagenases by tetranitromethane indicates that they all have an essential tyrosine residue. This is supported by the loss of activity on treatment with acetylimidazole and its partial restoration on deacylation with hydroxylamine. The loss in activity of all collagenases on treatment with trinitrobenzenesulfonate and diethyl pyrocarbonate and the failure of hydroxylamine to restore activity demonstrate that these enzymes also have an essential lysine residue. Collectively, these experiments show that all six collagenases have very similar functional constituents at the active site that include a zinc atom as well as essential carboxyl, tyrosine, and lysine residues.

*Circular Dichroism Spectra.* To investigate the relationship between the secondary structures of these collagenases, circular dichroism spectra in the peptide bond (Figure 1) region have been recorded. The shapes of the curves, which are known to reflect the distribution of secondary structure (Greenfield

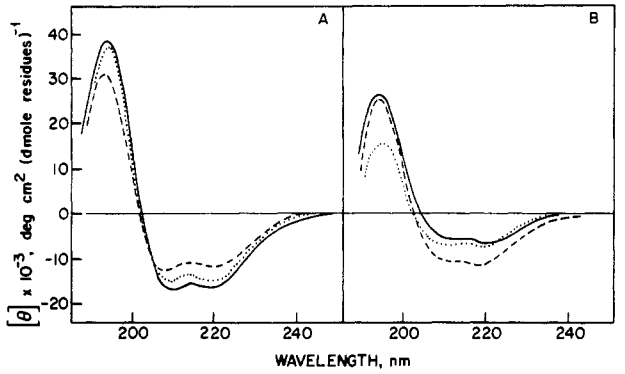


FIGURE 1: Far-ultraviolet circular dichroism spectra of the (A) class I collagenases, (--) $\alpha$, (—) $\beta$, and (···) $\gamma$, and (B) the class II collagenases, (—) $\delta$, (--) $\epsilon$, and (···) $\zeta$. $[\theta]$ has been calculated per decimole of residues from a mean residue weight of 112. The samples were dissolved in 1 mM Tris and 1 mM $CaCl_2$, pH 7.5, at 23 °C.

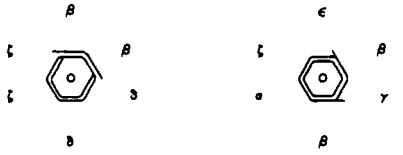


FIGURE 2: Diagram showing the precipitin lines formed in Ouchterlony double-immunodiffusion gels on reaction of antiserum prepared against $\beta$-collagenase (center well) with purified collagenases (outer wells) for 24 h. The antiserum was undiluted, and the collagenase concentration was 100 $\mu$g/mL.

& Fasman, 1969), are very similar for all of the class I collagenases and have extrema at 191, 209, and 221 nm (Figure 1A). By the procedure of Greenfield & Fasman (1969), $\beta$-collagenase is calculated to contain approximately 45% $\alpha$-helix, 25% $\beta$-structure, and 30% of the unordered conformation.

The spectra of the three class II collagenases are also very similar to each other and generally similar to those of the class I collagenases. They have extrema at almost the same wavelengths as the class I collagenases, but the molar ellipticities at each wavelength are smaller, and there are small differences in shape (Figure 1B). The curves for the class II collagenases indicate that they contain a higher percentage of $\beta$-structure at the expense of $\alpha$-helix. $\epsilon$-Collagenase is calculated to contain 28% $\alpha$-helix, 42% $\beta$-structure, and 30% of the unordered conformation. The circular dichroism spectrum reported by Heindl et al. (1980) for the *Clostridial* collagenase isolated by Emod & Keil (1977) is very similar to that of the class II collagenases reported here.

The great similarity of the circular dichroism spectra of $\alpha$-, $\beta$-, and $\gamma$-collagenases is particularly striking in view of their large differences in molecular weight. For the spectra of these three proteins to be the same implies that the 36 000 molecular weight fragment by which the $\gamma$ and $\beta$ enzymes differ and the 47 000 fragment by which the $\alpha$ and $\beta$ enzymes differ must have the same distribution of secondary structures as $\alpha$- and $\gamma$-collagenases themselves. This is the first of several pieces of experimental data to suggest that $\beta$-collagenase consists of more than one identical or nearly identical domain and that $\alpha$- and $\gamma$-collagenases differ primarily by lacking one or more of these domains.

*Immunological Experiments.* To explore the relationship between the three-dimensional structures of these enzymes, their degree of immunological identity has been examined. Thus, antiserum was prepared against $\beta$-collagenase, and a series of Ouchterlony double-immunodiffusion experiments on agar gels was carried out to examine the cross-reactivity

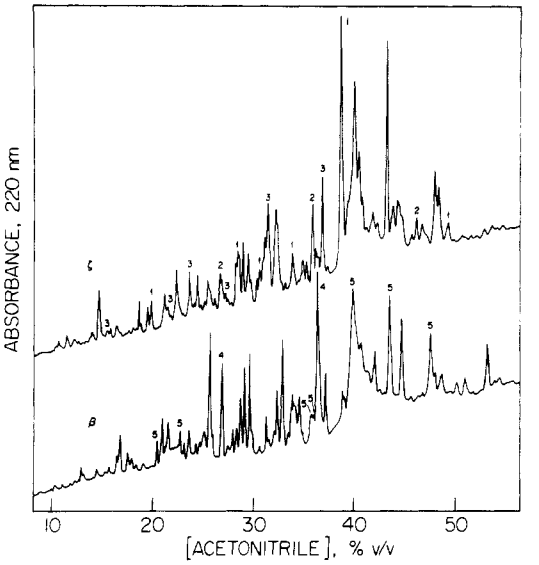


FIGURE 3: Reverse-phase high-pressure liquid chromatography elution profiles for tryptic digests of $\zeta$- and $\beta$-collagenases. The numbers 1, 2, and 3 designate peaks that are less intense or missing entirely in $\delta$, $\epsilon$, or both of these collagenases, respectively. The numbers 4 and 5 designate peaks that are less intense or missing entirely in $\alpha$-collagenase or $\alpha$- and $\gamma$-collagenases, respectively. Profiles were obtained by elution from a Bio-Sil ODS-5S column (250 × 4 mm) with a linear acetonitrile–water gradient at a flow rate of 1.0 mL/min at 25 °C.

of this serum with the other collagenases. The pattern of precipitin lines, which do not photograph well, are shown in schematic form in Figure 2. The solid lines indicate precipitin lines between the serum in the center well and the enzymes in the outer wells.

There are at least three different precipitin lines in these gels, indicating that the antigen, $\beta$-collagenase, is immunologically heterogeneous. This is consistent with the electrofocusing data, which reveal that there are subspecies of $\beta$-collagenase. When there are identical enzymes in the adjacent wells, the same number of precipitin lines form and they are confluent, indicating immunological identity. When there are nonidentical collagenases in adjacent wells, there is at least one line of confluence. In some cases, spurs are formed on a given precipitin line, demonstrating that the antibody subpopulation itself is heterogenous and that one of the enzymes shares fewer immunological determinants with $\beta$-collagenase than the other. This behavior indicates partial immunological identity. There are reactions of immunological identity or partial identity between all six collagenases with at least one subpopulation of antibodies, implying that there is an evolutionary relationship between all six of these enzymes (Salton, 1977).

*Tryptic Digests.* The best means of determining the relationship between a group of proteins is to compare their amino acid sequences. This is outside the scope of the present study. However, some information about the similarities in the sequences of these collagenases has been obtained by digestion of the denatured enzymes with trypsin followed by analysis of the peptides formed on a reverse-phase high-pressure liquid chromatography column. The resulting fingerprints for $\beta$- and $\zeta$-collagenases are shown in Figure 3.

The chromatograms obtained for all three class I collagenases are very similar to one another, as are those for the class II enzymes very similar to one another. Thus, only the chromatograms for $\beta$- and $\zeta$-collagenases are shown. However, the chromatograms for collagenases in the two different classes are markedly different. The numbers 1, 2, and 3 that appear above the profile for $\zeta$-collagenase identify peaks that are either

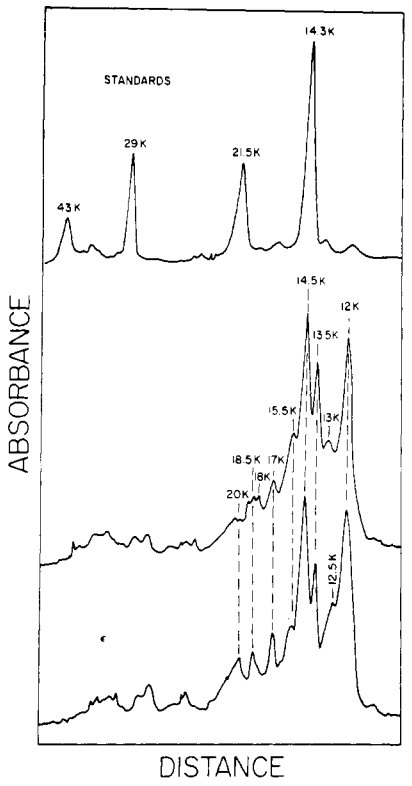


FIGURE 4: Densitometer tracings of sodium dodecyl sulfate–polyacrylamide gels (10–15% gradient) containing the peptides formed on reaction of $\epsilon$- and $\zeta$-collagenases with cyanogen bromide. A tracing for the molecular weight standards ovalbumin (43 000), carbonic anhydrase (29 000), soybean trypsin inhibitor (21 500), and lysozyme (14 300) is shown on top for reference.

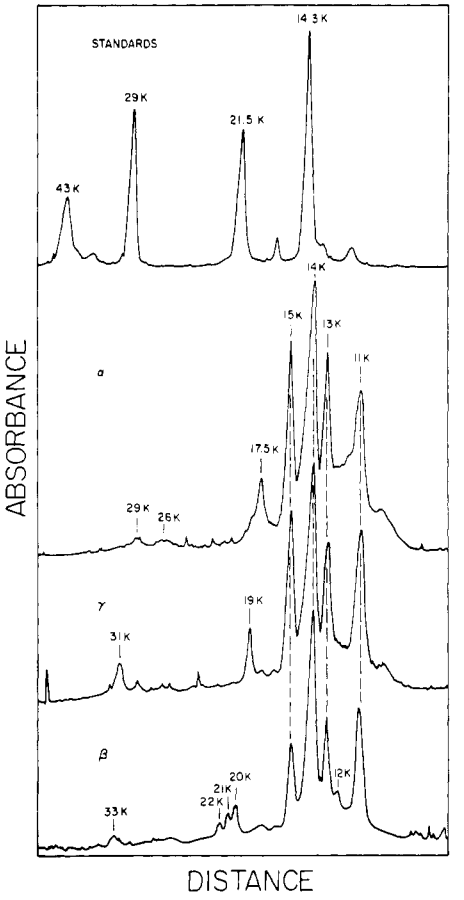


FIGURE 5: Densitometer tracings of sodium dodecyl sulfate–polyacrylamide gels (10–15% gradient) containing the peptides formed on reaction of $\alpha$-, $\beta$-, and $\gamma$-collagenases with cyanogen bromide. A tracing for the molecular weight standards ovalbumin (43 000), carbonic anhydrase (29 000), soybean trypsin inhibitor (21 500), and lysozyme (14 300) is shown on top for reference.

less intense or absent in $\delta$, $\epsilon$, or both of these collagenases, respectively. The numbers 4 and 5 that appear above the chromatogram of $\beta$-collagenase identify peaks that are less intense or absent in $\alpha$- or $\alpha$- and $\gamma$-collagenases, respectively. These results minimally establish that there is substantial sequence homology between the three class I collagenases, substantial homology between the three class II collagenases, but marked sequence differences between the enzymes in the two classes.

While it is extremely difficult to use these chromatograms to quantitatively assess the degree of homology between these collagenases, it is still possible to draw several conclusions. Starting with the class II collagenases, the fact that $\epsilon$- and $\delta$-collagenases differ from $\zeta$-collagenase by only 9 and 12 peaks, respectively, out of about 40 demonstrates that the homology between the three enzymes is extensive. Clearly, some of these differences are attributable to the fact that the three enzymes have different molecular weights. Thus, some of the peptides missing in $\delta$- and $\epsilon$-collagenases are undoubtedly those for sections of these proteins only present in $\zeta$-collagenase. Accordingly, there are no peptides present in $\delta$-collagenase that are not present in $\epsilon$- and $\zeta$-collagenases and none present in $\epsilon$-collagenase that are not present in $\zeta$-collagenase.

The similarities in the profiles for the class I enzymes are much more surprising because of the large differences in the molecular weights of the three enzymes. Thus, in spite of the fact that $\alpha$- and $\gamma$-collagenases differ from $\beta$-collagenase in molecular weight by 47 000 and 36 000, respectively, there are only nine and seven peaks, respectively, in the chromatograms that differ for the three enzymes. Gel electrophoretic analysis of the digests confirms that the trypsinization was complete and rules out the possibility that $\beta$-collagenase has a high molecular weight core deficient in lysine and arginine or otherwise resistant to digestion. The similarities of the tryptic-digest chromatograms of these three enzymes are reminiscent of their similar circular dichroism spectra. The most likely explanation for these data is that $\beta$-collagenase consists of domains with homologous amino acid sequences and that $\alpha$- and $\gamma$-collagenases differ from $\beta$-collagenase primarily by lacking one or more of these domains.

*Cyanogen Bromide Reactions.* Another way of comparing the primary structures of these collagenases is to cleave the chains with cyanogen bromide and compare the sizes of the peptides formed. Thus, each of the collagenases was reacted with cyanogen bromide for various lengths of time, and the products were applied to sodium dodecyl sulfate–polyacrylamide gradient gels. These experiments revealed that the most complete reaction was obtained only when EDTA was included in the reaction mixture and that a reaction time of 10 h was optimal. Densitometer tracings of the gels containing $\epsilon$- and $\zeta$-collagenases are shown in Figure 4, and those for $\alpha$-, $\beta$-, and $\gamma$-collagenases are shown in Figure 5. Also shown in these figures are the positions of a series of protein standards from which the molecular weights of the cyanogen bromide fragments have been estimated to the nearest 500-dalton increment.

The densitometer tracings for the peptide cyanogen bromide fragments of $\epsilon$- and $\zeta$-collagenases shown in Figure 4 are qualitatively very similar. $\delta$-Collagenase was not examined because of the very small quantity of this enzyme that was available. There are at least eight peptides with molecular weights in the 10 000–20 000 range. In addition, there are

numerous weak bands with molecular weights above 20 000 that most likely result from incomplete cleavages, as is so often observed when the cyanogen bromide technique is applied to proteins with so many methionines (Spande et al., 1970). It is very difficult to draw any quantitative conclusions from these data, but it appears that the number of peptides resolved in the low molecular weight region is fewer than expected for ζ-collagenase, which contains 15 methionines (Bond & Van Wart, 1984b). Thus, it may be that there are many different peptides with molecular weights in the 12 000–20 000 range that are hidden under the profile or that several of the methionines in the enzyme are so close together that the cleavages produce peptides too small to detect electrophoretically. Alternatively, these collagenases may contain long stretches of internal sequences that are homologous to one another and produce cyanogen bromide fragments of the same size. At the very least, the traces in Figure 4 provide evidence that the distribution of methionine residues throughout ε- and ζ-collagenases is very similar.

Figure 5 shows the tracings for the three class I collagenases, which are easier to interpret since these enzymes contain fewer methionine residues. The tracing for α-collagenase contains five major peptides, consistent with the finding that it has four methionine residues (Bond & Van Wart, 1984b). The sum of the molecular weights of these five bands is 71 500, which is in good agreement with the $M_r$ of 68 000 measured for the intact enzyme (Bond & Van Wart, 1984b). Two minor bands with $M_r$ of 29 000 and 26 000 are also apparent and are probably due to incomplete cleavages (i.e., 29 000 = 15 000 + 14 000 or 17 500 + 11 000; 26 000 = 15 000 + 11 000 or 14 000 + 13 000; although other combinations are possible).

The tracing observed for γ-collagenase is extremely similar to that observed for α-collagenase with two exceptions. First, the 17 500 peptide for α-collagenase is replaced by a peptide with an $M_r$ of 19 000 in γ-collagenase. Second, the intensity of the $M_r$ 11 000 band in γ- compared to α-collagenase is approximately 50% larger than the other four bands. One possibility is that there is a second peptide with an $M_r$ of approximately 11 000 that falls under this band in γ-collagenase. This would be consistent with the finding that γ-collagenase has five methionine residues. The sum of the molecular weights of these peptides would then be 84 000, which is in reasonable agreement with the value of 79 000 for the intact enzyme (Bond & Van Wart, 1984b). Alternatively, the increase in intensity of the 11 000 band may simply be the result of a greater yield in the reaction. The fifth methionine, then, could be close to one of the other methionines and produce a peptide too small to be resolved on the gel. In this case, the predicted $M_r$ of 73 000 is also in reasonable agreement with that of the intact enzyme.

The tracing for β-collagenase is surprisingly similar to that for γ-collagenase, but a careful comparison reveals three differences. First, the band at 19 000 for γ-collagenase is replaced by a small triplet corresponding to peptides with $M_r$ of 20 000, 21 000, and 22 000. This triplet has approximately the same normalized area as the 19 000 band for γ-collagenase and the 17 500 band for α-collagenase. One possibility is that these peptides in the three proteins are those at either the N- or C-terminus and that the three enzymes differ slightly in the length of this peptide. Interestingly, isoelectric focusing experiments have established that β-collagenase consists of three major subspecies. These subspecies may differ by the lengths of either their N- or C-termini, and this may be responsible for the triplet observed for β-collagenase.

A second difference is that β-collagenase has a peptide with an $M_r$ of 12 000 that is not present in γ-collagenase. This may be one of the two peptides which had an $M_r$ of 11 000 for γ-collagenase but which has an $M_r$ of 12 000 in β-collagenase. In support of this interpretation, the integrated intensity of this band compared to that at 15 000 is very close to that expected if it were hidden under the other 11 000 band in γ-collagenase.

The third difference between the tracings for γ- and β-collagenases is that the relative areas of the bands at 14 000, 13 000, and 11 000 have approximately doubled compared to that at 15 000. This observation, coupled with the results of the tryptic digests, indicates that β-collagenase probably contains two domains that have homologous or duplicate amino acid sequences, each of which contains peptides corresponding to the 14 000, 13 000, and 11 000 bands in this gel. The explanation of the data given above is based upon the presence of nine cyanogen bromide peptides for β-collagenase. This is in agreement with the amino acid analysis for this enzyme, which shows that it has eight methionine residues (Bond & Van Wart, 1984b). Summation of the molecular weights of these peptides gives a value of about 124 000 for intact β-collagenase, which is in reasonable agreement with the measured value of 115 000, considering the errors involved in estimating molecular weights by these procedures.

*Limited Proteolysis Experiments.* The similarities in the tryptic digests and cyanogen bromide fragment patterns suggest the possibility that α- and γ-collagenases are derived from β-collagenase by proteolysis. To investigate this possibility, a series of limited proteolysis experiments have been carried out. The first series of experiments was designed to test whether an enzyme present in the crude *Clostridial* collagenase preparation derived from the culture fluid could carry out the proteolytic conversions cited above. Thus, β-collagenase has been incubated with crude *Clostridial* enzyme (25 μg, 1 h) at 35 °C under several sets of conditions. Separate incubations were carried out in the presence and absence of calcium ions and with crude enzyme samples that either were or were not previously activated with dithiothreitol. Samples were removed from this incubation mixture at various times and applied to sodium dodecyl sulfate electrophoresis gels (not shown) to search for the production of protein bands with molecular weights corresponding to α- or γ-collagenases. These experiments show that limited proteolysis of β-collagenase produces small quantities of proteins with $M_r$ of 87 000, 67 000, and 61 000. Other conditions employed, including increasing the reaction time, did not enhance the yield of these proteins. Thus, while it is possible that the protein with a molecular weight of 67 000 is α-collagenase, we have been able to detect only small amounts of it in these experiments.

It is possible that the proteolysis of β-collagenase to form α- or γ-collagenases is carried out by an enzyme present in the culture fluid during fermentation, but that this enzyme is not present in the ammonium sulfate precipitate that is used as the crude starting material. This could account for the failure of the experiments outlined above to produce sizable quantities of α- or γ-collagenases. Thus, to test whether the β to γ or β to α conversions could be effected enzymatically, β-collagenase was subjected to limited proteolysis by several common proteases with different specificities. The results of this series of experiments are shown in Figure 6. Limited proteolysis with trypsin, chymotrypsin, papain, *Staphylococcus aureus* protease, and thermolysin produces a whole variety of protein bands with different molecular weights. Interestingly, four of the five enzymes generate at least one protein band

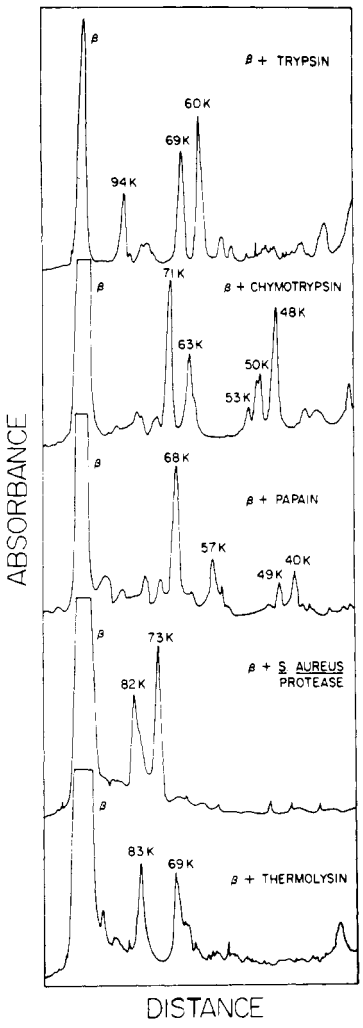


FIGURE 6: Densitometer tracings of sodium dodecyl sulfate–polyacrylamide gels (7.5%) showing, from top to bottom, the products of the reaction of pure β-collagenase in 10 mM Tris, pH 7.5, at 23 °C with trypsin (0.4 μg, 1 h), chymotrypsin (0.01 μg, 10 min in the presence of 10 mM $CaCl_2$), papain (2 ng, 1 h in the presence of 10 mM $CaCl_2$), *S. aureus* protease (0.2 μg, 20 min) and thermolysin (5 μg, 1 h).

with a molecular weight very close to that of α-collagenase but none with molecular weights close to that of γ-collagenase. Only the digestion with papain gives a 68 000 band as the major product. Thus, these experiments lend some support to the hypothesis that α-collagenase is derived from β-collagenase by proteolysis. However, an alternative explanation is that α- and β-collagenases are separate but homologous proteins coded for by different genes and that β-collagenase contains a 47 000 domain absent in α-collagenase. If this 47 000 domain in β-collagenase were inherently susceptible to proteolysis, it would yield a protein with the same molecular weight as α-collagenase. The sequence of this protein would be homologous to, but not identical with, that of α-collagenase.

## Discussion

The biochemical and physicochemical properties of the six *Clostridial* collagenases described in this and the preceding paper provide a sizable body of information with which to decipher the relationship between these enzymes. All six enzymes are very similar in many ways, yet the differences between them allow their grouping into two distinct classes. With regard to their similarities, these include the fact that they all hydrolyze native collagen, gelatin, and synthetic tetrapeptides that contain amino acid sequences found in collagen. On the basis of their active site constituents, all six enzymes appear to carry out these reactions by a similar mechanism. They all contain 1 mol of zinc/mol of protein, which acts as a catalytic metal, and also contain variable amounts of calcium ions that are required for activity, presumably by stabilizing their tertiary structures. Chemical modification reactions establish that all six enzymes are inactivated by the same reagents and implicate carboxyl, tyrosine, and lysine residues in catalysis.

There are also many similarities between the primary, secondary, and tertiary structures of these enzymes. All six consist of single polypeptide chains with molecular weights that are unusually high for bacterial neutral metallohydrolases. All six enzymes have reasonably similar amino acid compositions and isoelectric points, and none contain a significant amount of carbohydrate. Circular dichroism measurements show that the amounts of the various types of secondary structure (α-helix, β-structure, and unordered conformation) are similar in all of these collagenases. Importantly, antibodies grown against β-collagenase cross-react with the other five collagenases. This indicates that all six have portions of their three-dimensional structures that are similar enough to constitute common antigenic determinants and, hence, probably have a common evolutionary origin.

In spite of all the similarities cited above, there are a number of very real differences between some of these collagenases. Whenever these differences occur, they all serve to divide the six collagenases into two classes. The class I collagenases have higher collagenase and gelatinase activities but lower activities toward the synthetic substrates FALGPA and FALGPP. In fact, the substrate specificities of these two classes of enzymes show many marked differences (D. R. Steinbrink and H. E. Van Wart, unpublished data). The class I enzymes are also more stable to freeze–thawing. While the amino acid compositions of all six enzymes are similar, the differences within each class are significantly smaller than those between the two classes. The strongest evidence in favor of this classification comes from an analysis of the cyanogen bromide cleavage and tryptic digestion patterns. In particular, the tryptic-digest reverse-phase high-pressure liquid chromatography fingerprints show that there is extensive sequence homology between the enzymes in each class but that the sequences of the enzymes in the two classes are markedly different.

In considering the relationship between these enzymes, it is perhaps first appropriate to compare them to other bacterial proteinases. The properties enumerated above leave little doubt that these collagenases are related to a group of bacterial enzymes referred to as neutral metallohydrolases. This group includes the neutral proteinases from *Bacillus thermoproteolyticus*, *Bacillus subtilis*, and a series of other bacteria. Although the properties of each of these enzymes have not been fully established, they all have pH optima in the 7–9 region and are all believed to contain a single atom of zinc per molecule located at the active site. Many of these enzymes require calcium ions for activity and stability. Most of these enzymes lack cysteine, and all have molecular weights close to the 35 000–40 000 region (Matsubara & Feder, 1971). All of these neutral proteinases have very similar substrate specificities (Morihara, 1974) and require that the residue that contributes the amino side of the scissile peptide bond be bulky or hydrophobic. *Clostridial* collagenases clearly belong to this group of proteinases. However, they are atypical by virtue of their unique substrate specificities, their unusually high molecular weights, and their cysteine content.

The similarities of the gel electrophoretic tracings and high-pressure liquid chromatography elution profiles for the three class I enzymes establish that there are sequence homologies for these enzymes. In addition, these similarities taken together with the wide variation in the molecular weights of these enzymes indicate that there must be an unusual relationship between their primary sequences. The possibility that $\beta$-collagenase contains a large core that is not susceptible to tryptic digestion has been ruled out. Importantly, the fact that all of these collagenases have a single zinc atom strongly implies that they all have a single active site. This rules out the possibility that $\beta$-collagenase consists of subunits that have somehow become covalently linked together. Hence, a more likely explanation for these data is that $\beta$-collagenase contains more than one domain that has homologous or duplicate amino acid sequences and that one or more of these domains is (are) also present in $\alpha$- and $\gamma$-collagenases.

The deduction that $\beta$-collagenase consists of domains that have homologous amino acid sequences taken together with its unusually high molecular weight suggests that the enzyme evolved by complete or partial gene duplication. Higher than expected molecular weights may also be due to gene fusion, but this would not explain the evidence for homologous domains. Our data are insufficient to establish how large the region or regions of duplication might have been. However, $\alpha$- and $\gamma$-collagenases are approximately twice, and $\beta$-collagenase approximately 3 times, the typical $M_r$ of 35 000–40 000 for a bacterial neutral metallohydrolase, suggesting the possibility that these collagenase evolved from a smaller neutral proteinase by sequential tandem duplications. If this were the case, subsequent point mutations must have eradicated all but one of the active sites. Alternatively, these collagenases could have evolved from a series of partial gene duplications in which the elongations did not transfer a complete or viable active site.

Given the fact that there is a substantial amount of sequence homology between the $\alpha$, $\beta$-, and $\gamma$-collagenases, three possibilities for the origins of these three enzymes can be put forth. The first possibility is that all three enzymes are derived by proteolysis from a protein coded for by a single gene. A second possible explanation is that all three enzymes are coded for by separate genes and that the homologies are present in the genes. A third possibility is that there is a single gene that codes for all three collagenases and that the $\alpha$, $\beta$, and $\gamma$ enzymes are produced by transcription with different start or stop signals. An ultimate resolution of this question must await an examination of the gene structure of this bacterium. However, the choice between the possible explanations for the existence of multiple class I collagenases in no way detracts from the evidence that $\beta$-collagenase contains domains with homologous amino acid sequences.

The hypotheses cited above to explain the presence of multiple class I collagenases with sequence homologies may also pertain to the class II collagenases. The tryptic digests of the three class II collagenases are very similar; however, since the differences in molecular weight between the $\delta$-, $\epsilon$-, and $\zeta$-collagenases are much smaller, there is no compelling reason to conclude that these enzymes contain domains with homologous amino acid sequences. Indeed, it was only the similarities of the tryptic digests of the class I collagenases *and* their large differences in molecular weight that enabled the internal sequence homology to be detected.

The next question that arises is that of the relationship between the two different classes of collagenases. There are clearly significant differences between their sequences. However, their immunological cross-reactivity and the similarities in their amino acid compositons, structures, and mechanisms imply that there is a close evolutionary relationship between them. The most likely explanation for the observed similarities is that one class evolved from the other by gene duplication followed by subsequent divergent evolution to form a new class of enzymes with different substrate specificities. Gene duplication is an important mechanism for evolving new proteins since it allows the organism to accumulate point mutations in one copy of a gene without risking the loss of function associated with the old one.

This situation is strikingly similar to the well-known case of the pancreatic proteinases (Neurath et al., 1967; Smith, 1970). For example, trypsin and chymotrypsin are thought to have originated from a common gene that duplicated and evolved by point mutations to give enzymes with different specificities. This diversity in function confers a metabolic advantage to an animal that is dependent upon protein degradation for survival. Since *Clostridium histolyticum* is a pathogenic anaerobe that uses collagenases as a means to invade the host, the organism would likewise benefit from having two types of collagenases with different substrate specificities. This hypothesis is consistent with our observation that the two classes of collagenases digest connective tissue synergisticaly (M. D. Bond and H. E. Van Wart, unpublished data).

It is interesting that recognized examples of gene duplication in the evolution of proteins are much more prevalent in eucaryotes than in procaryotes (Dayhoff, 1976). To our knowledlge, the data presented here are the first to implicate gene duplication as a mechanism for the evolution of a bacterial proteinase. Interestingly, gene duplications are implicated in both the evolution of one class of collagenases from smaller proteins *and* in the evolution of one class of collagenases from the other. It remains unclear why *Clostridium histolyticum* amongst other bacteria that produce neutral proteinases should have utilized gene duplication in the evolution of its collagenases.

## Acknowledgments

We are indebted to Dr. Randolph Rill for the use of his reverse-phase column, to Dr. Ken Roux for help preparing the antisera to $\beta$-collagenase, and to Dr. Herbert Taylor for valuable discussions about genetic mechanisms for gene duplication.

**Registry No.** Collagenase, 9001-12-1; L-tyrosine, 60-18-4; L-lysine, 56-87-1.

## References

Bond, M. D., & Van Wart, H. E. (1984a) *Biochemistry* (first paper of three in this issue).

Bond, M. D., & Van Wart, H. E. (1984b) *Biochemistry* (second paper of three in this issue).

Bond, M. D., Steinbrink, D. R., & Van Wart, H. E. (1981) *Biochem. Biophys. Res. Commun. 102*, 243–249.

Dayhoff, M. O. (1976) *Atlas of Protein Sequence and Structure*, Vol. 5, Supplement 2, pp 2–5, National Biomedical Research Foundation, Silver Spring, MD.

Emod, I., & Keil, B. (1977) *FEBS Lett. 77*, 51–56.

Epstein, E. H. (1974) *J. Biol. Chem. 249*, 3225–3231.

Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108–4116.

Heindl, M. C., Fermandjian, S., & Keil, B. (1980) *Biochim. Biophys. Acta 624*, 51–59.

Matsubara, H., & Feder, J. (1971) *Enzymes, 3rd Ed. 3*, 721–795.

Morihara, K. (1974) *Adv. Enzymol. Relat. Areas Mol. Biol. 41*, 179–243.

Neurath, H., Walsh, K. A., & Winter, W. P. (1967) *Science (Washington, D.C.) 158*, 1638–1644.

Salton, M. (1977) *Immunochemistry of Enzymes and Their Antibodies*, pp 60–62, Wiley, New York.

Smith, E. L. (1970) *Enzymes, 3rd Ed. 1*, 267–339.

Spande, T. F., Witkop, B., Degani, Y., & Patchornik, A. (1970) *Adv. Protein Chem. 24*, 205–210.

# Interaction of Specific Platelet Membrane Proteins with Collagen: Evidence from Chemical Cross-Linking†

Nicolas J. Kotite, James V. Staros, and Leon W. Cunningham*

ABSTRACT: Two recently developed membrane-impermeant cross-linkers, 3,3′-dithiobis(sulfosuccinimidyl propionate) (DTSSP) and bis(sulfosuccinimidyl) suberate ($BS^3$), have been used to examine the interaction of human platelets with collagen. Reaction of human platelets with either of the two cross-linking reagents at micromolar concentrations completely inhibited platelet aggregation in response to collagen but not in response to thrombin. Platelet adhesion to collagen was, however, not affected by these reagents. Inhibition of collagen-induced platelet aggregation by DTSSP or $BS^3$ appears to be due to cross-linking and not simply to the chemical modification of membrane proteins, since the homologous monofunctional reagent sulfosuccinimidyl propionate had no effect on platelet aggregation. Inhibition of platelet aggregation by $BS^3$ was accompanied by a decrease in the intensity of glycoprotein bands IIb, IIIa, and IV when analyzed on sodium dodecyl sulfate ($NaDodSO_4$)–polyacrylamide gels. In order to determine if collagen is directly interacting with a specific platelet membrane glycoprotein, $^3H$-labeled platelets were allowed to adhere to collagen and then cross-linked with various concentrations of DTSSP. Proteins which remain associated with collagen after lysis and washing were analyzed on $NaDodSO_4$ gels. At concentrations of 16–50 μM DTSSP, glycoproteins IIb and IIIa appeared to be specifically cross-linked to collagen. These results suggest that the glycoprotein IIb–IIIa complex, which has previously been implicated as the fibrinogen receptor in activated platelets, may also be directly involved in collagen-induced platelet aggregation.

The sequence of events initiated when injury to a blood vessel exposes the subendothelial layer to circulating platelets presumably first involves specific interactions of platelet membrane components, primarily glycoproteins, with the matrix. In addition, through the intervention of soluble proteins including von Willebrand factor (Weiss et al., 1978; Kao et al., 1979; Fujimoto et al., 1982), and fibrinogen (Marguerie et al., 1979; Bennett & Vilaire, 1979), platelets adhere to one another to form the platelet "plug". Despite intensive study, the identities of the platelet membrane glycoproteins involved in these events have been determined only partially. Thus, for example, the role of a glycoprotein IIb–IIIa complex as a fibrinogen receptor, after platelet activation, seems relatively well established (Nachman & Leung, 1982; Di Minno et al., 1983; Bennett et al., 1983). Both glycoprotein I and glycoprotein V have been implicated as sites of interaction of thrombin (Tollefsen et al., 1974; Workman et al., 1977; Larsen & Simons, 1981; Berndt & Phillips, 1981), but the identification of the membrane component responsible for specific interaction with collagen, or with von Willebrand factor, is less clear (Jamieson & Okumura, 1978; Jenkins et al., 1983; Chiang & Kang, 1982).

Chemical cross-linking reagents have been applied in a limited way to the analysis of topography and function of the platelet membrane (Larsen & Simons, 1981; Davies & Palek, 1982; Lahav et al., 1982). The availability of a new family of such reagents, 3,3′-dithiobis(sulfosuccinimidyl propionate) (DTSSP)[1] and bis(sulfosuccinimidyl) suberate ($BS^3$), which exhibit considerable stability toward hydrolysis and which are membrane impermeant (Staros, 1982) has led us to examine their effect upon cross-linking of membrane glycoproteins to one another and to collagen fibrils. In the course of these investigations, it was found that the exposure of platelets to micromolar levels of these cross-linking reagents led to the specific inhibition of collagen-induced aggregation but did not affect either thrombin-induced aggregation or adhesion to collagen.

We report here also the effects of these cross-linking reagents on platelet membrane glycoproteins as evidenced by changes in the electrophoretic pattern and the selective cross-linking of certain membrane glycoproteins of adhering platelets to fibrillar collagen.

## Experimental Procedures

*Materials.* Tritiated sodium borohydride (22 Ci/mM) was obtained from Research Products International, Mount Pro-

† From the Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232. *Received October 18, 1983.* This research was supported by National Institutes of Health Grants 5 R01 HL24553, 5 R01 AM25489, and 1 R01 AM31880.

[1] Abbreviations: DTSSP, 3,3′-dithiobis(sulfosuccinimidyl propionate); $BS^3$, bis(sulfosuccinimidyl) suberate; SSP, sulfosuccinimidyl propionate; $NaDodSO_4$, sodium dodecyl sulfate; MbA, methylenebis(acrylamide); DATD, *N,N′*-diallyltartardiamide; DTT, dithiothreitol; TEMED, *N,N,N′,N′*-tetramethylethylenediamine; ACD, acid–citrate–dextrose; EDTA, disodium ethylenediaminetetraacetate; HPLC, high-performance liquid chromatography; $HOSu(SO_3)$, *N*-hydroxysulfosuccinimide; Tris, tris(hydroxymethyl)aminomethane; NEM, *N*-ethylmaleimide; BSA, bovine serum albumin; $M_r$, relative molecular weight.

0006-2960/84/0423-3099$01.50/0 © 1984 American Chemical Society