RAPID COMMUNICATIONS IN MASS SPECTROMETRY, VOL. 9, 563–569 (1995)

Examination of Recombinant Truncated Mature Human Fibroblast Collagenase by Mass Spectrometry: Identification of Differences with the Published Sequence and Determination of Stable Isotope Incorporation

Swapan K. Chowdhury,[1,*] Karen J. Vavra,[2] Patricia G. Brake,[1] Tracey Banks,[2] Joseph Falvo,[2] Robert Wahl,[2] Jamshid Eshraghi,[1] George Gonyea,[1] Brian T. Chait[3] and Christina H. Vestal[4]

[1] Sanofi Research, Sanofi Winthrop, Inc., 9 Great Valley Parkway, Malvern, PA 19355, USA
[2] Life Sciences Research Laboratories, Eastman Kodak Co., Rochester, NY 14650-2117, USA
[3] Rockefeller University, New York, NY 10021, USA
[4] PerSeptive Biosystems, Vestec Mass Spectrometry Products, 9299 Kirby Drive, Houston, TX 77054, USA

Human fibroblast collagenase belongs to a family of matrix metalloproteinases which have been implicated in a number of connective tissue disorders ranging from rheumatoid arthritis to tumor invasion. To examine the active site of this enzyme by biophysical studies, a 19 kDa recombinant truncated mature collagenase (mCL-t) was prepared. Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry have been utilized for the characterization of mCL-t. The molecular weights measured by these techniques identified the presence of two closely related protein components separated by approximately 100 Da. Edman sequence analysis demonstrated that the two protein components differ from each other by an amino terminal valine, consistent with the mass spectrometric data. In addition, the molecular weight of mCL-t determined by mass spectrometry did not agree with that calculated from the reported sequence. To identify the origin of this discrepancy, the DNA sequence of the mCL-t clone was examined. Several differences were noted between the DNA sequence of mCL-t and the published collagenase gene sequence. When these differences were taken into account, the measured molecular weights were found to be in good agreement with that calculated for the modified sequence. In separate experiments, both ESI and MALDI mass spectrometry have been used to determine molecular weights of mCL-t samples enriched with stable isotopes ^{15}N and ($^{15}N + ^{13}C$). The measured molecular weights demonstrated a 97% (^{15}N) and 99% ($^{15}N + ^{13}C$) incorporation of labeled isotopes in the two samples.

Human fibroblast collagenase (HFC) is a member of a family of matrix metalloproteinases implicated in the proteolytic breakdown of cartilage associated with both rheumatoid arthritis and osteoarthritis.[1] As a result, there is considerable interest in determining the three-dimensional protein structure by X-ray crystallography and multi-dimensional NMR spectroscopy, to elucidate the structure of the enzyme active site, and to develop inhibitors for the treatment of osteoarthritis and rheumatoid arthritis.

HFC is secreted as a 52 kDa latent proenzyme which is reduced to a 42 kDa active enzyme through an autolytic intermolecular cleavage event known as activation.[2] During storage of purified native procollagenase, new 33 kDa and 22 kDa fragments are generated in addition to the 42 kDa active enzyme through spontaneous autocatalytic events.[2,3] The 42, 33 and 22 kDa proteins were shown[2] to have the sequence Val_{101}–Asn_{469}, Phe_{20}–Pro_{269}, and Ile_{270}–Asn_{469} of pre-procollagenase, respectively. The pre-procollagenase sequence starts at the initiator Met_1, while the procollagenase lacks the 19 residue signal sequence. The 33 kDa fragment becomes catalytically active following activation. Both the 42 kDa and the activated 33 kDa truncated collagenase fragments have a common sequence of Val_{101}–Pro_{269}, and were shown to possess activity similar to that of their full-length recombinant counterparts.[4–7] The active site of the enzyme must lie within this common sequence. Therefore, it is possible to generate a smaller truncated enzyme without creating major alterations in the active site.

Because of its large size and instability, investigation of the active site of the native collagenase using biophysical studies is difficult. We therefore generated, using genetic engineering techniques, an *Escherichia coli* clone that expressed human fibroblast procollagenase truncated at Pro_{269} (Scheme 1). This enzyme is designated as proCL-t. From this protein a 19 kDa mature collagenase with a sequence Val_{101}–Pro_{269}, of high specific activity, was purified. This enzyme, mCL-t, is the subject of the present investigation. Furthermore, mCL-t fully labeled with ^{15}N or ($^{15}N + ^{13}C$) was prepared for multi-dimensional NMR studies.

To examine the purity and homogeneity, and to determine the molecular weight of mCL-t for the verification of its sequence, electrospray ionization (ESI)[8] and matrix-assisted laser desorption/ionization (MALDI)[9] mass spectrometric techniques were investigated because of their ability to determine the molecular weights of proteins with high accuracy, high sensitivity and rapidity.[8,9] The high accuracy molecular weight determinations were also utilized to quantitatively determine the stable isotope incorporation in labeled mCL-t. The present investigation also provides a direct comparison of ESI and MALDI techniques for analyses performed on the same protein samples.

* Author for correspondence.

```
        101                 111                 121
Whitham V L T E G* N P R W E Q T H L T* Y R I E N Y T P D L P R A D V
Goldberg         G                   R
mCL-t            G                   T

        131                 141                 151
Whitham D H A I E K A F Q L W S N V T P L T F T K V S E G Q A D I M
Goldberg
mCL-t

        161                 171                 181
Whitham I S F V R G D H R D N S P F D G P G G N L A H A F Q P G P G
Goldberg
mCL-t

        191                 201                 211
Whitham I G G D A H F D E H* E R W T N N F T* E Y N L H R V A A H E L
Goldberg                D                   R
mCL-t                   D                   R

        221                 231                 241
Whitham G H S L G L S H S T D I G A L M Y P S Y T F S G D V Q L A Q
Goldberg
mCL-t

        251                 261
Whitham D D I D G I Q A I Y G R S Q N P V Q P
Goldberg
mCl-t
```

Scheme 1. The numbering of the sequence of mCL-t given here corresponds to the sequence of human pre-procollagenase (Met_1 to Asn_{469}). The mCL-t sequence obtained in the present investigation is compared with the sequences reported by Whitham[17] and Goldberg.[10] The present sequence differs from the Whitham sequence at positions 200 (D vs. H) and 208 (R vs. T), and from the Goldberg sequence at 115 (T vs. R).

MATERIALS AND METHODS

Protein production, purification and characterization

The coding region of a cDNA clone containing the human fibroblast collagenase gene, obtained from G. Goldberg,[10] was placed under control of the T7 polymerase promoter in plasmid pET11a,[11] and modified to introduce a termination codon after Pro_{269} of the pre-pro-CL sequence. The resulting plasmid, pKV150, codes for a truncated procollagenase with an N-terminal methionine residue immediately preceding the native procollagenase terminus FPATL. . . , and a C-terminal proline at residue 269.

Fermentations were performed in 15L Bioengineering AG (Wald, Switzerland) fermentors. For preparation of unlabeled truncated collagenase, the fermentor, containing 9.4 L LB/M9 media (10 g/L Bactotryptone,
5 g/L Bacto-yeast extract, 5 g/L NaCl, and 100 μg/mL ampicillin) was inoculated with a 500 mL overnight seed culture of pKV150 in *Escherichia coli* host BL21/DE3 (Novagen, Madison, WI, USA), and the culture was grown at 37 °C, 1 vvm and 500 rpm until the optical density, measured at 600 nm, reached a value between 1.1 and 1.2. Production of pro-truncated collagenase (proCL-t) was induced by addition of isopropyl β-D thiogalactopyranoside (IPTG) to 0.5 mM, and the culture was harvested at the beginning of the stationary phase (up to 3 h after induction). Cells were collected by centrifugation, and the cell pellets stored frozen at −20 °C.

Production of [^{15}N, ^{13}C] proCL-t was carried out as described above except that cells were grown in Minimal Medium (6 g/L $Na_2HPO_4 \cdot H_2O$, 3 g/L KH_2PO_4, pH 7.5, 1 g/L [^{15}N] NH_4Cl, 0.25 g/L $MgSO_4 \cdot 7H_2O$, 20 mg/L $CaCl_2$, 4 g/L [^{13}C] D-glucose, and 100 mg/L ampicillin) in both seed flasks and fermentors.

The cell pellet (20–35 g) was homogenized in 35 mL of 0.1 M tricine (pH 8.0), 10 mM ethylenediaminetetracetic acid (EDTA), 2.0 mM dithiothreitol (DTT), 0.2 mM phenylmethylsufonyl fluoride (PMSF), and 1 mg/mL lysozyme in a Dounce homogenizer (VWR Scientific, Rochester, NY). Additionally, Brij-35 to 0.5%, 20 mM $MgCl_2$, and 35 units/mL of DNaseI were added and, after further incubation at 37 °C, the solution was centrifuged to collect inclusion bodies. The above procedure was repeated twice and the pellet then washed twice with 50 mM tricine (pH 7.5), 0.2 M NaCl, 10 mM $CaCl_2$, 0.02% NaN_3, 2.0 mM DTT, and 0.2 mM PMSF, resuspended in 50 mL freshly made urea buffer (20 mM tricine (pH 7.5), 6 M urea, 5 mM $CaCl_2$, 0.02% NaN_3) with 2.0 mM DTT and 0.2 mM PMSF, and incubated at room temperature for 1 h. The solution was centrifuged, and proCL-t was purified from the supernatant liquid by chromatography on a Mono-Q column eluted in a 0–0.25 M NaCl gradient in urea buffer. Pooled fractions were diluted dropwise while stirring into a 10-fold excess of 50 mM tricine (pH 7.5) containing 0.4 M NaCl, 10 mM $CaCl_2$, 0.1 mM ZnOAc, and 0.02% NaN_3, at 4 °C, and the resulting refolded protein dialyzed against this buffer. The protein was concentrated to 2–4 mg/mL and stored at −70 °C.

Purified proCL-t was activated with 1 mM (*p*-aminophenyl)mercuric acetate (APMA) overnight at 37 °C.[12] The resultant mature truncated collagenase (mCL-t) was then purified on a Sephacryl S100 HR column, equilibrated in 20 mM tricine (pH 7.5) containing 0.4 M NaCl, 5 mM $CaCl_2$, and 0.02% NaN_3. Fractions were concentrated and stored at −70 °C.

Protein concentration was determined by absorbance at 280 nm, and a purity of greater than 95% was verified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and/or reverse-phase high-performance liquid chromatography (HPLC). Typically, 150 mg of unlabeled and 23 mg of [^{15}N, ^{13}C] labeled mCL-t were recovered from 9.9 L of cells.

Enzyme assays

The k_{cat}/K_m values for collagenase and commercially available intramolecularly quenched substrates (1): 2,4 dinitrophenyl-Pro-Leu-Gly-Leu-Trp-Ala-(D-Arg)-NH_2[13] and (2): 7-methoxycoumarin-4-yl acetyl-Pro-Leu-Gly-Leu [3-(2′,4′-dinitrophenyl)-L-2,3-diamino propionyl] Ala-Arg-NH_2[18] were determined from full hydrolysis curves. (k_{cat} is the rate at which the enzyme substrate complex (ES) proceeds towards product (E + P) in the reaction: $E + S = ES \rightarrow^{k_{cat}} E + P$. K_m is the Michaelis–Menten constant, generally expressed in units of M.) Fluorescence was measured using excitation at 328 nm and emission at 384 nm for assays with (2), and excitation at 280 nm and emission at 360 nm for assays with (1) The assay buffer was 50 mM tricine (pH 7.5) containing 0.2 M NaCl, 10 mM $CaCl_2$, 0.05% Brij-35, and 0.02% NaN_3. Samples of native fibroblast procollagenase were activated with TPCK-treated trypsin just prior to the enzymatic assay. Samples of mCL-t were assayed directly. Digital output of fluorescence (Perkin–Elmer (Norwalk, CT, USA) LS-5 fluorescence spectrophotometer) was collected every 30 s for 7–10 min. Fluorescence vs time data were fitted by nonlinear regression (Enzfitter) to the equation:

$$F = F_{max}\{1 - e^{-kt}\} + F_{min} \quad (1)$$

where F is fluorescence, F_{max} is fluorescence at full hydrolysis, and F_{min} is fluorescence at time 0. The initial velocity of fluorescence increase, v_0, was calculated from $F_{max}k$, since $dF/dt = F_{max}k\,e^{-kt}$, and thus dF/dt $(t = 0) = F_{max}k$. v_0 was linear with respect to enzyme and substrate concentration, which confirmed that $[S] \ll K_m$ and that intermolecular quenching of product was not occurring. The fluorescence response of the product, μM product/F, was determined by the ratio of the known initial concentration of substrate to the fluorescence at full hydrolysis, $[S_o]/F_{max}$. Thus,

$$k_{cat}/K_m = v_o([S_o]/F_{max})/([S_o][E]) = v_o/(F_{max}[E]) \quad (2)$$

where $[E]$ is the enzyme concentration.

Mass spectrometry

Electrospray ionization mass spectrometry

Electrospray ionization (ESI) mass spectrometry[8] was performed using a Finnigan TSQ-700 triple quadrupole mass spectrometer (Finnigan MAT, San Jose, CA, USA) equipped with an Analytica (Analytica of Branford, CT, USA) electrospray ion source, which has been described previously.[14] Briefly, the protein sample was dissolved in a solution comprised of 47.5% distilled water, 47.5% methanol, and 5% acetic acid, to obtain a final protein concentration of approximately 5 μM. The sample solution was infused at a flow rate of 1 μL/min using a Harvard Apparatus (Southnatick, MA, USA) (Model 22) syringe pump. Dry nitrogen (10 mL/min) at a temperature of 110 °C was used in the counter-current mode, to assist the evaporation of droplets formed by the electrospray process. The spray needle assembly was operated at ground potential and the gold-plated cylindrical electrode at −3.6 to −3.8 kV.[14] The mass spectrometer was scanned from m/z 200 to 2000 in 2 s. The spectra were averaged for 1 min, which corresponds to the accumulation of 20 scans.

Matrix-assisted laser desorption/ionization mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI)[9] was performed using two instruments: a commercial mass spectrometer (Vestec Model Bench Top II, Vestec, Houston, TX (laboratory I)) and an instrument built in-house at the Rockefeller University (laboratory II).[15] The procedures adopted to acquire spectra at the two laboratories were similar, as follows: 2 μL of the protein sample in 0.1% trifluoroacetic acid (TFA) (~15 μM) was mixed with 10 μL of a saturated solution of sinapinic acid (matrix) in 70% of 0.1% aqueous TFA and 30% acetonitrile. An aliquot of 0.5–1.0 μL of the matrix/protein solution was then applied to a metal probe-tip, dried, inserted into the ionization chamber of the mass spectrometer, and irradiated with a laser beam at 337 nm generated from a nitrogen laser (laboratory I) or at 355 nm generated from a Nd-YAG laser (laboratory II). Spectra were obtained by summing the transient records of 100–200 individual laser shots.

RESULTS AND DISCUSSION

Mass spectrometric characterization of mCL-t

To determine the purity, homogeneity, and molecular weight, mCL-t was examined by both electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometric methods. The ESI spectrum of mCL-t is shown in Fig. 1. The molecular weight profile derived from the raw data is given as an inset in Figure 1. Two major protein components were observed to be present in mCL-t (Fig. 1). The molecular weights of the two components obtained from the derived spectrum (inset) are 18840 and

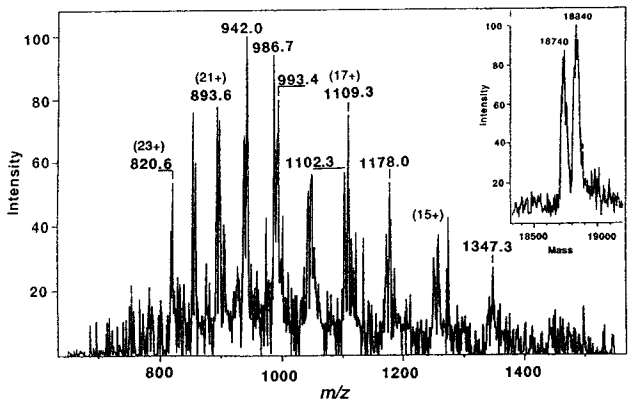

Figure 1. Electrospray ionization mass spectrum of mCL-t. Flow rate 1 μL/min, scan range m/z 200 to 2000, at 2s/scan, averaging time = 1 min. Data acquired in profile mode. *Inset:* The molecular weight profile of mCL-t obtained from the spectrum of Fig. 1 using a Finnigan MAT deconvolution algorithm. The centroids of the two peaks are at 18840 and 18740 Da, respectively.

18740 Da. The difference between these two values is 100 Da, consistent with addition of an amino terminal valine (residue mass 99 u) as determined by amino-terminal (Edman) sequencing (see below).

The amino-terminal sequence of mCL-t was determined from preparations separated by SDS-PAGE and blotted prior to sequencing. Truncated mature collagenase (mCL-t) samples contained two amino-termini in roughly equal amounts. These were found to be VLTEGN. . . and LTEGN. . . These sequences start at positions $Val_{(101)}$ and $Leu_{(102)}$ of the pre-procollagenase (Scheme 1), respectively, and are the N-termini normally generated from native procollagenase by activation with APMA.[16] Chromatography of mCL-t on Sephacryl S100 shows a bimodal elution profile (data not shown). Sequence analysis of the two peaks revealed that the column partly resolves the species containing Val_{101} and Leu_{102} at their amino-termini. Mass spectrometric molecular weight determinations thus confirm the results obtained from Edman sequencing. The spectrum shown in Fig. 1 indicates nearly equal amounts of the two proteins in mCL-t, in good agreement with the Edman sequencing data.

Although the difference between the measured molecular weights of the two protein components supports the lack of N-terminal valine in one of the proteins, the individual molecular weights did not match those calculated from the expected sequence of mCL-t (18812.6) and mCL-t(−)Val (18713.5) obtained from the sequence of proCL reported by Whitham *et al.*[17] The molecular weight of mCL-t derived from the ESI measurement is approximately 28 higher than the value calculated from this published sequence.[17] It should also be noted that the measured molecular weights of the two protein components also differ from the values deduced from the DNA sequence of Goldberg *et al.*[10] (see later). To determine the origin of the discrepancy between the measured and the calculated molecular weights, the mCL-t sample was further examined by MALDI-MS. The MALDI spectra of mCL-t obtained in two laboratories are given in Fig. 2, and the molecular weights deduced from the MALDI spectra together with the values obtained from ESI mass spectrometry are summarized in Table 1. The molecular weights, derived from the mass-to-charge ratios of $[M+H]^+$ and $[M+2H]^{2+}$ ions in the MALDI spectrum (laboratory I) are 18842 ± 2 and 18743 ± 2 Da for mCL-t and mCL-t(−)Val, respectively (Fig. 2(a)). These values are again approximately 30 Da higher than those calculated for the published sequences[17] of mCL-t and mCL-t(−)Val. Similarly, the molecular weights of the two protein components determined in laboratory II (Fig. 2(b)) are 18843 ± 3 and 18747 ± 2 Da. These values are 31 and 34 Da higher than the corresponding calculated values (Table 1). The MALDI measurement at the second laboratory was performed to determine the consistency of measurements and to assure that the observed differences were not results of instrumental artifacts. Good agreement between the data from the two laboratories indicated that the observed differences in molecular weights are independent of the instrumentation used. The lack of agreement between the measured molecular weights of the protein components with the corresponding calculated values prompted us to re-examine the DNA sequence[10] of the protein.

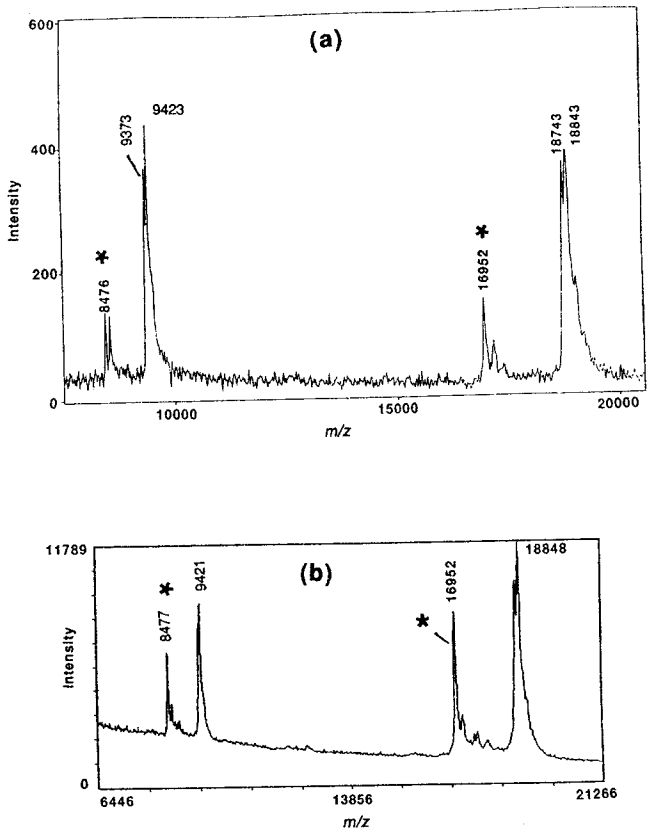

Figure 2. MALDI spectra of mCL-t obtained at (a) laboratory I and (b) laboratory II. $[M+H]^+$ and $[M+2H]^{2+}$ ions of horse apomyoglobin were used to calibrate the mass spectrometer. Both spectra demonstrate the presence of two protein components in mCL-t at nearly equal abundances.* denotes the ion peaks from horse myoglobin used for calibration.

The coding region of the plasmid was sequenced and compared with the two published sequences of Goldberg *et al.*[10] and Whitham *et al.*[17] It was discovered that the present gene sequence differs from the published sequences[10, 17] at five locations, four of which alter the predicted amino acid sequence. The differences in the reported gene sequences are compared in Table 2. The sequence of the mCL-t clone agreed with the Goldberg sequence at three sites (bases 381, 664, and 689) and with the Whitham sequence at two sites

Table 1. Molecular weight (MW) of mCL-t determined by ESI and MALDI methods

Protein/Technique	Measured MW (Da)	Calculated MW[a] (Da)	Difference (Da)
mCL-t			
ESI	18840 ± 6	18812	28
MALDI (lab I)[b]	18842 ± 2	18812	30
MALDI (lab II)[b,c]	18843 ± 3	18812	31
mCL-t(−)Val			
ESI	18740 ± 6	18713	27
MALDI (lab I)[b]	18743 ± 2	18713	30
MALDI (lab II)[b,c]	18747 ± 2	18713	34

[a] Molecular weight was calculated based on the published sequence.[17]
[b] Data from $[M+H]^+$ and $[M+2H]^{2+}$ ion contributions. The uncertainties are estimated values from our experience with several hundred protein samples.
[c] Data from two different measurements.

Table 2. Location of the nucleotide and corresponding amino acid residue differences between the present clone and published collagenase sequences

Base[a]	381	410, 411	664	689
Amino acid[b]	(105)	(115)	(200)	(208)
Goldberg *et al.*	GGA (Gly)	AGG (Arg)	GAT (Asp)	AGA (Arg)
Whitham *et al.*	GGG (Gly)	ACC (Thr)	CAT (His)	ACA (Thr)
mCL-t clone	GGG (Gly)	ACC (Thr)	GAT (Asp)	AGA (Arg)

[a] Bases that were observed to be different in the published DNA sequences. The numbering begins with the initiator codon of pre-procollagenase.
[b] The numbering of amino acid residues is based on the sequence of Whitham *et al.*[17]

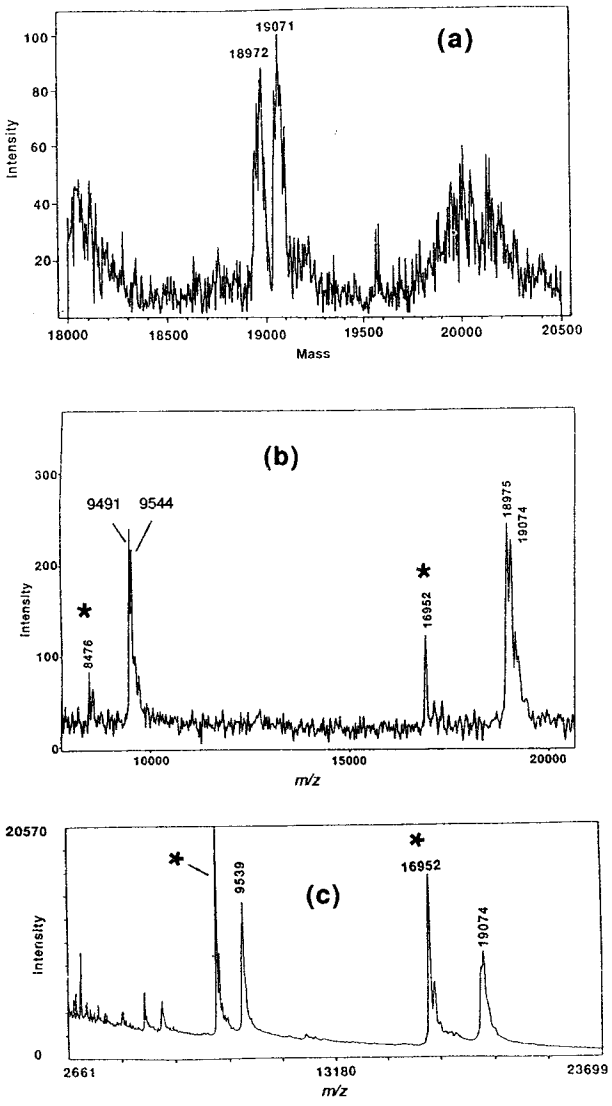

Figure 3. (a) A deconvoluted ESI mass spectrum of mCL-t enriched with ^{15}N isotopes. Experimental conditions were the same as those used to obtain the spectra in Figs 1 and 2. (b) MALDI spectrum of mCL-t enriched with ^{15}N isotope acquired in laboratory I, and (c) in laboratory II. Experimental conditions are the same as those used to obtain Fig. 2.* denotes the ion peaks from horse myoglobin used for calibration.

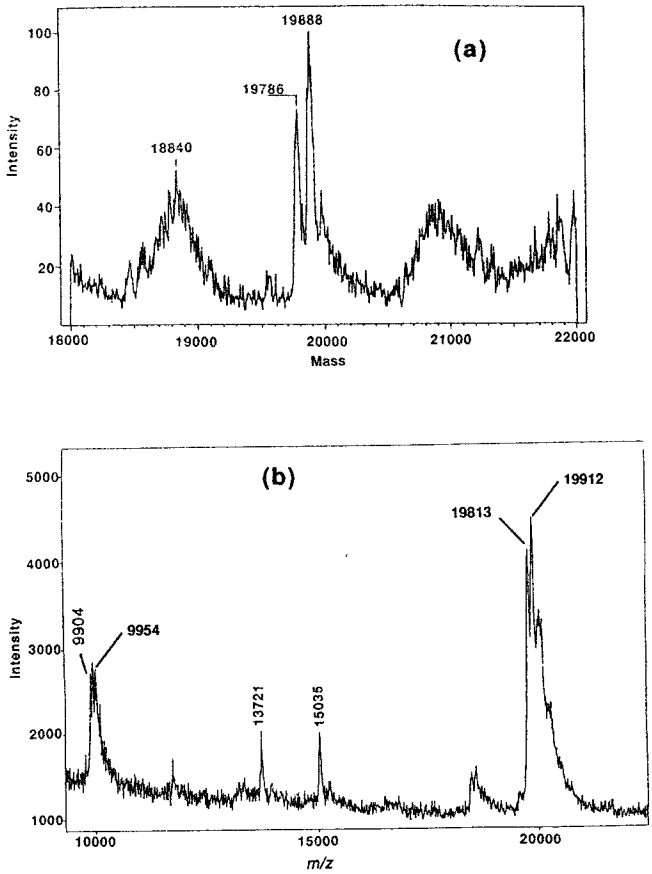

Figure 4. (a) A deconvoluted ESI mass spectrum of mCL-t enriched with ($^{15}N + ^{13}C$) isotopes. (b) MALDI spectrum of mCL-t enriched with ($^{15}N + ^{13}C$) isotopes, obtained in laboratory I.

(bases 410 and 411), giving rise to a protein with a glycine at position 105, threonine at 115, aspartic acid at 200 and arginine at 208 (Scheme 1). Thus the mCL-t protein sequence differs from the Whitham sequence,[17] for example, at position 200 (Asp vs. His) and at 208 (Arg vs. Thr) (see Scheme 1 and Table 2). These differences result in an increase of 33 Da in the molecular weight of the present protein. The measured molecular weight obtained from the three separate mass spectrometric determinations with two different ionization techniques (ESI and MALDI) provided the first indication of the existence of such differences from the published sequences for mCL-t, and supports the expected molecular weight increase, within the measurement accuracy, predicted by the present DNA sequence.

It should also be noted that the calculated molecular weight values of the two protein components, according to the Goldberg DNA sequence,[10] are 18900 and

Table 3. Comparison of specific acitvity ($k_{cat}/K_m[M^{-1}s^{-1}]$) of native HFC and mCL-t

Substrate	Native HFC[a]	mCL-t[a]	N-terminal collagenase[b]	rHFC[c]
(1)	4000 ± 1300	6100 ± 900	653	830
(2)	15900 ± 1100	16000 ± 1000	—	14800

[a] Data from the present investigation.
[b] Data from Murphy *et al.*[6]
[c] Data from Knight *et al.*[18]

Table 4. Molecular weights and the amount of stable isotope incorporation in ^{15}N and ($^{15}N + ^{13}C$) enriched mCL-t determined by MALDI and ESI methods

	MALDI[a] (lab I)	MALDI[a] (lab II)	ESI	Average MW[b]	Calculated MW[c]	% enriched
^{15}N-						
mCL-t	19079 ± 6	19076 ± 6	19071 ± 6	19075 ± 10	19083	97 ± 2
mCL-t(−)Val	18976 ± 4	18982 ± 4	18972 ± 6	18977 ± 8	18982	98 ± 2
$^{15}N + ^{13}C$-						
mCL-t	19909 ± 2	—	19888 ± 5	19899 ± 5	19916	99 ± 1
mCL-t(−)Val	19809 ± 3	—	19786 ± 5	19798 ± 6	19810	99 ± 2

[a] Data from $[M+H]^+$ and $[M+2H]^{2+}$ ion contributions. The uncertainties are estimated values from our experience with several hundred protein samples.
[b] Average of MALDI and ESI data.
[c] The calculated (isotopically averaged) molecular weight corresponds to mCL-t with complete incorporation of ^{15}N and ($^{15}N + ^{13}C$).

18801 Da for mCL-t and mCL-t(−)Val, respectively. The measured average molecular weights of 18842 and 18743 Da from the three determinations are lower by 58 Da. This difference agrees with the observed replacement of an Arg with a Thr (difference of 55 Da) from the DNA sequence (Table 2).

Two reports of recombinant preparations of truncated collagenase have appeared previously in the Literature.[4,6] In one case,[4] the trucation of collagenase was at the same site (Pro_{269}) as reported here, but very few details of expression, purification, or characterization were reported. In the other, Murphy *et al.*[6] expressed a truncated procollagenase protein in mouse myeloma cells, which was eight residues shorter (truncated at the C-terminus) than proCL-t. This enzyme was designated 'N-terminal collagenase' by Murphy *et al.*[6] The same group also reported the production of a recombinant full length collagenase (rHFC).[18] None of the published investigations[4,6,18] used mass spectrometry to characterize their recombinant collagenase preparations.

Because of our interest in using this truncated collagenase as a model for the full-length enzyme, it was important to demonstrate that its activity was not compromised either by truncation or by the amino acid differences observed here. Thus, the specific activities of the present preparation (mCL-t) using two substrates have been determined, and were compared with those of the native procollagenase (HFC) and with published data for other recombinant collagenases (Table 3).[4,6,18] For both substrates, mCL-t showed equal or higher specific activity than native full-length collagenase. When compared with the 'N-terminal collagenase' of Murphy *et al.*[6] using substrate (1), the specific activity of mCL-t was found to be about 9-fold higher (Table 3). Unfortunately, the 'N-terminal collagenase' was not assayed with the substrate (2). The specific activity of native procollagenase (HFC) can also be compared with the published data for the corresponding recombinant protein (rHFC). For substrate (1), a specific activity value more than 4-fold higher was measured for our native, mature, full-length HFC, compared to that reported for the recombinant full-length enzyme (rHFC) by Knight *et al.*[18] This difference may be due to differences in assay conditions, since the specific activity of our native preparation, using substrate (2), is similar to that reported for the full-length recombinant enzyme.[18] In addition, inhibition constants for a number of inhibitors were found to be identical for native and truncated collagenase (data not shown). Our study is the first to compare the specific activity of a recombinant collagenase species with a native collagenase. We conclude from these studies that the activity of mCL-t is indistinguishable from that of native full-length collagenase, making it a suitable model for biophysical studies.

Determination of the amount of labeled isotope incorporation in mCL-t

For use in determining the three-dimensional structure of mCL-t using nuclear magnetic resonance spectroscopy (NMR), ^{15}N and ($^{15}N + ^{13}C$) substituted mCL-t samples were also prepared. Before the multidimensional NMR examination is conducted, it is desirable to determine the extent of isotope incorporation in these proteins. We used both electrospray ionization and matrix-assisted laser desorption mass spectrometry for the determination of the amount of isotope incorporation[19] in the recombinant proteins because of the high accuracy in molecular weight measurements obtainable using these techniques.

The molecular weight profile obtained from the ESI spectrum, and the MALDI spectra from the two laboratories, of mCL-t with fully incorporated ^{15}N, are shown in Fig. 3(a), (b) and (c) respectively. The three spectra again show the presence of two components with an approximate molecular weight difference of 100 Da, as was observed with the unlabeled mCL-t (Table 1). The molecular weights deduced from Fig. 3(a)–(c) are given in Table 4, together with values calculated for the complete incorporation of ^{15}N in mCL-t. mCL-t has 237 N-atoms which were expected to be substituted with ^{15}N. Therefore, the expected molecular weight of ^{15}N-mCL-t is 19083 Da and that of ^{15}N-mCL-t(−)Val is 18982 Da. The average measured molecular weights of 19075 and 18877 Da support an incorporation of 97–98% ^{15}N in mCL-t.

The mCL-t sample labeled with both ^{15}N and ^{13}C isotopes (designated $^{15}N^{13}C$-mCL-t) was also investigated by ESI and MALDI methods. The derived ESI and MALDI spectra obtained from this preparation of collagenase are shown in Fig. 4(a) and (b), respectively. The measured molecular weights are compared with the calculated values in Table 4. There are 834 C-atoms and 237 N-atoms in mCL-t. Therefore, a

molecular weight increase of 1071 Da is expected for the complete ($^{15}N + ^{13}C$) enrichment in mCL-t, and 1065 Da for mCL-t(−)Val. The measured average molecular weights of 19899 and 19798 Da for the two proteins, with and without N-terminal valine, indicate 99% isotopic incorporation into the proteins. Compared with conventional methods, mass spectrometric measurements provide a close reflection of the amount of stable isotopes incorporated in recombinant proteins because of the high accuracy with which the molecular weights can be determined.

CONCLUSION

Determination of accurate and precise molecular weights by mass spectrometry provides a rapid verification of the protein primary structure. Observation of a difference between the measured molecular weight and that calculated from the expected sequence provides clues as to the presence of amino acid difference(s) or modification(s) in the protein. Together with the data from cDNA sequence analysis, precise location of such differences can be obtained. The identification of amino acid sequence differences in mCL-t, compared to previously published reports, is an excellent example of the utility of accurate molecular weight measurements. Comparison of the data obtained from ESI and MALDI measurements on the same protein samples leads us to the following conclusions: (i) ESI using a quadrupole mass spectrometer can provide higher resolution compared to MALDI with a linear time-of-flight spectrometer. This higher resolution can be useful in resolving proteins in a mixture of close molecular weights. (ii) In the present experimental configuration, larger amounts of samples are consumed in electrospray ionization than those in MALDI. Approximately 20 pmol of sample was used for a complete electrospray analysis, compared to 10 pmol in MALDI with only 1 pmol loaded on the MALDI probe. (iii) Both techniques provide similar precision for the molecular weight determination (±0.03%). Independent MALDI determinations of the molecular weights using two different instruments provided very similar values. (iv) This level of accuracy and precision can also be used to quantitatively determine stable isotope enrichment.[19]

Acknowledgements

We wish to thank Dennis Carlton, Tricia Pulvino, Ellen Schneider, Eileen Roth, Ronald Lirette, Doris Dixon, Carol Ryerson and Richard Ciccarelli for technical help and critical review of the project. Native collagenase was obtained from Henning Birkedal-Hansen, and the collagenase cDNA clone from Gregory Goldberg. B.T.C. acknowledges support from NIH (Grant number RR00862).

REFERENCES

1. C. H. Evans, in *Drugs in Inflammation*, Birkhauser Verlag, Basel 1991.
2. B. Birkedal-Hansen, W. G. I. Moore, R. E. Taylor, A. S. Bhown and H. Birkedal-Hansen, *Biochemistry* **27**, 6751–6758 (1988).
3. I. M. Clark and T. E. Cawston, *Biochem. J.* **263**, 201–206 (1989).
4. C. L. Lowry, G. McGeehan and H. LeVine, III, *Proteins: Struct. Funct. Genet.* **12**, 42–48 (1992).
5. H. I. Marcy, L. L. Eiberger, R. Harrison, H. K. Chan, N. I. Harrison, W. K. Hagmann, P. M. Cameron, D. A. Boulton and J. D. Hermes, *Biochemistry* **30**, 6476–6483 (1991).
6. G. Murphy, J. A. Allan, F. Willenbrock, M. I. Cockett, J. P. O'Connell and A. J. P. Docherty, *J. Biol. Chem.* **267**, 9612–9618 (1992).
7. Q.-Z. Ye, L. L. Johnson, D. J. Hupe and V. Baragi, *Biochemistry* **31**, 11231–11235 (1992).
8. (a) J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Mass Spectrom. Rev.* **9**, 37–70 (1990); (b) R. D. Smith, J. A. Loo, R. R. Ogorzalek Loo, M. Busman and H. R. Udseth, *Mass Spectrom. Rev.* **10**, 359–455 ((1991); (c) T. R. Covey, R. F. Bonner, B. I. Shushan and J. D. Henion, *Rapid Commun. Mass Spectrom.* **2**, 249–256 (1988); (d) S. K. Chowdhury, V. Katta and B. T. Chait, *Rapid Commun. Mass Spectrom.* **4**, 81–87 (1990).
9. (a) M. Karas, U. Bahr and F. Hillenkamp, *Int. J. Mass Spectrom. Ion Processes* **92**, 231–242 (1989), (b) F. Hillenkamp, M. Karas, R. C. Beavis and B. T. Chait, *Anal. Chem.* **63**, 1193A–1203A (1991), (c) B. T. Chait and S. B. H. Kent, *Science* **257**, 1885 (1992).
10. G. I. Goldberg, S. M. Wilhelm, A. Kronberger, E. A. Bauer, G. A. Grant and A. Z. Eisen, *J. Biol. Chem.* **261**, 6600–6605 (1986).
11. F. W. Studier, A. H. Rosenberg, J. J. Dunn and J. W. Dubendorff, *Methods Enzymol.* **185**, 60–89 (1990).
12. G. P. Stricklin, J. J. Jeffery, W. T. Roswit and A. Z. Eisen, *Biochemistry* **22**, 61–68 (1983).
13. M. S. Stack and R. D. Gray, *J. Biol Chem.* **264**, 4277–4281 (1989).
14. J. Eshraghi and S. K. Chowdhury, *Anal. Chem.* **65**, 3528–3533 (1993).
15. R. C. Beavis and B. T. Chait, *Rapid Commun. Mass Spectrom* **3**, 233–237 (1989).
16. K. Suzuki, J. J. Enghild, T. Morodomi, G. Salvesen and H. Nagase, *Biochemistry* **29**, 10261–10270 (1990).
17. S. E. Whitham, G. Murphy, P. Angel, H.-J. Rahmsdorf, B. J. Smith, A. Lyons, T. J. R. Harris, J. J. Reynolds, P. Herrlich and A. J. P. Docherty, *Biochem. J.* **240**, 913–916 (1986).
18. C. G. Knight, F. Willenbrock and G. Murphy, *FEBS Letts* **296**, 263–266 (1992).
19. T. G. Heath and V. Thanabal, *Biotechnol. Tech.* **7**, 367–372 (1993).