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. 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. 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. The 42, 33 and 22 kDa proteins were shown to have the sequence Val_{101}–Asn_{469}, Phe_{320}–Pro_{369}, and Ile_{270}–Asn_{469} of preprocollagenase, respectively. The preprocollagenase 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_{369}, and were shown to possess activity similar to that of their full-length recombinant counterparts. 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_{260} (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) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometric techniques were investigated because of their ability to determine the molecular weights of proteins with high accuracy, high sensitivity and rapidity. 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.
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, was placed under control of the T7 polymerase promoter in plasmid pET11a and modified to introduce a termination codon after Pro201 of the pro-COL sequence. The resulting plasmid, pKV50, codes for a truncated procollagenase with an N-terminal methionine residue immediately preceding the native procollagenase residues FPAFL... 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, 1vvm 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 (proCOL-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 were stored frozen at -20°C.

Production of [15N, 13C] pro-COL-t was carried out as described above except that cells were grown in Minimal Medium (6 g/L NaH2PO4·H2O, 3 g/L KH2PO4, pH 7.5, 1 g/L [15N] NH4Cl, 0.25 g/L MgSO4·7H2O, 20 mg/L CaCl2, 4 g/L [13C] 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 ethylenediaminetetraacetic acid (EDTA), 2.0 mM di-thiothreitol (DTT), 0.2 mM phenylmethylsulfonyl fluoride (PMSF), and 1 mg/mL lysozyme in a Dounce homogenizer (VWR Scientific, Rochester, NY). Additionally, Brij-30 to 0.5%, 20 mM MgCl2, and 35 units/mL of DNase 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 mL tricine (pH 7.5), 0.2 M NaCl, 10 mM CaCl2, 0.02% NaN3, 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 CaCl2, 0.02% NaN3) with 2.0 mM DTT and 0.2 mM PMSF, and incubated at room temperature for 1 h. The solution was centrifuged, and pro-COL-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 CaCl2, 0.1 mM ZnOAc, and 0.02% NaN3, 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 pro-COL-t was activated with 1 mM (p-aminophenyl)mercuric acetate (APMA) overnight at 37°C. 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 CaCl2, and 0.02% NaN3. 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 [15N, 13C] 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$_2$ and (2): 7-methoxyxycoumarin-4-y1 acetyl-Pro-Leu-Gly-Leu [3-(2',4'-dinitrophenyl)-l,-2,3-diamino propionyl] Ala-Ala-NH$_2$ 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 $\rightarrow$ E-S $\rightarrow$ $k_{cat}$ 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 TPCN-treated trypsin just prior to the enzymatic assay. Samples of mCL-t were assayed directly. Digital output of fluorescence detection (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:

$$F = F_{max}[1 - e^{-kt}] + F_{min}$$ (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 $F_{max}$, since $\frac{dF}{dt} = F_{max}k e^{-kt}$, and thus $\frac{dF}{dt}$ ($t = 0$) = $F_{max}k$. $v_0$ was linear with respect to enzyme and substrate concentration, which confirmed that $[S] < K_m$ and that intermolecular quenching of product was not occurring. The fluorescence response of the product, $\mu$M product/$F$, was determined by the ratio of the known initial concentration of substrate to the fluorescence at full hydrolysis, $[S_0]/F_{max}$. Thus,

$$k_{cat}/K_m = v_0([S_0]/F_{max})/(IS_0([E])) = v_0/(F_{max}[E])$$ (2)

where [E] is the enzyme concentration.

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](http://example.com/fig1.png)

Figure 1. Electrospray ionization mass spectrum of mCL-t. Flow rate 1 $\mu$L/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(109) 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 Sephacyrl 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(109) 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]+ ions in the MALDI spectrum (laboratory 1) 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 sequences10 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 sequence10 of the protein.

Figure 2. MALDI spectra of mCL-t obtained at (a) laboratory I and (b) laboratory II. [M + H]+ and [M + 2H]+ 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 sequences10,17 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

<table>
<thead>
<tr>
<th>Protein/Technique</th>
<th>Measured MW (Da)</th>
<th>Calculated MW* (Da)</th>
<th>Difference (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCL-t ESI</td>
<td>18840 ± 6</td>
<td>18812</td>
<td>28</td>
</tr>
<tr>
<td>MALDI (lab I)*</td>
<td>18842 ± 2</td>
<td>18812</td>
<td>30</td>
</tr>
<tr>
<td>MALDI (lab II)*</td>
<td>18843 ± 3</td>
<td>18812</td>
<td>31</td>
</tr>
<tr>
<td>mCL-t(−)Val ESI</td>
<td>18740 ± 6</td>
<td>18713</td>
<td>27</td>
</tr>
<tr>
<td>MALDI (lab I)*</td>
<td>18743 ± 2</td>
<td>18713</td>
<td>30</td>
</tr>
<tr>
<td>MALDI (lab II)*</td>
<td>18747 ± 2</td>
<td>18713</td>
<td>34</td>
</tr>
</tbody>
</table>

* Molecular weight was calculated based on the published sequence.17

** Data from [M + H]+ and [M + 2H]+ ion contributions. The uncertainties are estimated values from our experience with several hundred protein samples.

*** 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

<table>
<thead>
<tr>
<th>Base</th>
<th>381 (105)</th>
<th>410, 411 (115)</th>
<th>664 (200)</th>
<th>689 (208)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid</td>
<td>GGA (Gly)</td>
<td>AGG (Arg)</td>
<td>GAT (Asp)</td>
<td>AGA (Arg)</td>
</tr>
<tr>
<td>Goldberg et al.</td>
<td>GGG (Gly)</td>
<td>ACC (Thr)</td>
<td>CAT (His)</td>
<td>ACA (Thr)</td>
</tr>
<tr>
<td>Whitham et al.</td>
<td>GGG (Gly)</td>
<td>ACC (Thr)</td>
<td>CAT (His)</td>
<td>ACA (Thr)</td>
</tr>
<tr>
<td>mCL-t clone</td>
<td>GGG (Gly)</td>
<td>ACC (Thr)</td>
<td>CAT (His)</td>
<td>AGA (Arg)</td>
</tr>
</tbody>
</table>

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

Figure 4. (a) A deconvoluted ESI mass spectrum of mCL-t enriched with (^15N + ^13C) isotopes. (b) MALDI spectrum of mCL-t enriched with (^15N + ^13C) 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,16 are 18900 and

Table 3. Comparison of specific activity (kcat/Km[M−1 s−1]) of native HFC and mCL-t

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Native HFC</th>
<th>mCL-t</th>
<th>N-terminal collagen</th>
<th>oHFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>4000 ± 1300</td>
<td>6100 ± 900</td>
<td>653 ± 300</td>
<td>830 ± 100</td>
</tr>
<tr>
<td>(2)</td>
<td>15900 ± 1100</td>
<td>16000 ± 1000</td>
<td>—</td>
<td>14800 ± 500</td>
</tr>
</tbody>
</table>

*a Data from the present investigation.
*b Data from Murphy et al.18
*c Data from Knight et al.18
18801 Da for mClt-1 and mClt-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. In one case, the truncation of collagenase was at the same site (P396) as reported here, but very few details of expression, purification, or characterization were reported. In the other, Murphy et al. expressed a truncated procollagenase protein in mouse myeloma cells, which was eight residues shorter (truncated at the C-terminus) than proClt-t. This enzyme was designated 'N-terminal collagenase' by Murphy et al. The same group also produced the preparation of a recombinant full length collagenase (rHFC). None of the published investigations 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 (mClt-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). For both substrates, mClt-t showed equal or higher specific activity than native full-length collagenase. When compared with the 'N-terminal collagenase' of Murphy et al. using substrate (1), the specific activity of mClt-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. 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. 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 mClt-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 mClt-t

For use in determining the three-dimensional structure of mClt-t using nuclear magnetic resonance spectroscopy (NMR), 15N and (15N+13C) substituted mClt-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 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 mClt-t with fully incorporated 15N, 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 mClt-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 15N in mClt-t. mClt-t has 237 N-atoms which were expected to be substituted with 15N. Therefore, the expected molecular weight of 15N-mClt-t is 19083 Da and that of 15N-mClt-t(-)Val is 18892 Da. The average measured molecular weights of 19075 and 18877 Da support an incorporation of 97-98% 15N in mClt-t.

The mClt-t sample labeled with both 15N and 13C isotopes (designated 15N13C-mClt-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 mClt-t. Therefore, a
molecular weight increase of 1071 Da is expected for the complete (15N + 13C) 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