Incorporation of Norleucine at Methionine Positions in Recombinant Human Macrophage Colony Stimulating Factor (M-CSF, 4–153) Expressed in *Escherichia coli*: Structural Analysis†

Zafar I. Randhawa,*,‡ H. Ewa Witkowska,§ James Cone,‡ James A. Wilkins,‡,∥ Patricia Hughes,‡ Kazuya Yamanishi,⊥ Setsuko Yasuda,⊥ Yoshihiro Masui,⊥ Patrick Arthur,‡ Cathy Kletke,‡ Francis Bitsch,§,# and Cedric H. L. Shackleton§

Otsuka America Pharmaceutical Inc., 9900 Medical Center Drive, Rockville, Maryland 20850, Otsuka Pharmaceutical Co. Ltd., Cellular Technology Institute, 463-10 Kagasuno Kawauchi-cho, Tokushima, 777-01, Japan, and Children's Hospital Oakland Research Institute, 747 52nd Street, Oakland, California 94609

Received December 1, 1993; Revised Manuscript Received February 8, 1994®

ABSTRACT: Expression of the 17.5-kDa truncated form of human recombinant macrophage colony stimulating factor (rM-CSF, 4–153) in *Escherichia coli* is complicated by the replacement of methionine residues by norleucine. In order to detect and quantitate this mistranslational event, the intact and the *S*-carboxyamidomethylated proteins were analyzed by amino acid analysis, automated Edman amino acid sequencing, and electrospray mass spectrometry. In addition, the endoproteinase Glu-C generated peptides were subjected to amino acid sequencing, high-performance liquid chromatography, and electrospray ionization mass spectrometry. The extent of norleucine substitution in different batches of rM-CSF varied between 0% and 20%. The relative instability of methionine residues needs to be considered when calculating the extent of norleucine substitution at methionine positions. The mass spectrometry of the intact rM-CSF allowed for examination of the distribution of multiply substituted methionine to norleucine species, and it enabled detection and quantitation of the norleucine incorporation down to the ~3% level. Selective ion chromatograms of molecular ions of interest obtained in reversed-phase high-performance liquid chromatography/electrospray ionization mass spectrometry of proteolytic fragments offered a reliable and fast method of detection and quantitation of norleucine-containing peptides. Norleucine residues were uniformly distributed among all four methionine positions (10, 27, 61, and 65). A substitution of methionine by its structural norleucine analog does not have any effect on the activity of the refolded rM-CSF dimers.

Human macrophage colony stimulating factor (M-CSF[1]) has been shown to promote the growth of macrophage colonies from bone marrow cell culture (Stanley et al., 1983). A truncated form of the molecule has been made in recombinant *Escherichia coli* as an insoluble product in inclusion bodies and has been refolded into a bioactive homodimer (Halenbeck et al., 1989). The amino acid sequence derived by automated Edman degradation of M-CSF (4–153), as well as disulfide placement and intermediates involved in folding pathways, has been reported (Yamanishi et al., 1993; Wilkins et al., 1993). In contrast to other hematopoietic growth factors, M-CSF is biologically active only in a disulfide-linked dimer form. The three-dimensional structure of dimeric human recombinant M-CSF has been reported recently (Pandit et al., 1992). Each M-CSF (4–153) monomer contains four methionine residues at positions 10, 27, 61, and 65 (position numbers refer to the amino acid sequence of a full-length human protein). It has been shown that *E. coli* can biosynthesize norleucine (Nle) and incorporate it into recombinant proteins at positions normally occupied by methionine (Cowie et al., 1959; Trupin et al., 1966; Kerwar & Weissback, 1970; Bogosian et al., 1989).

In order to determine the extent of Nle incorporation, and to establish which of the four methionine residues are involved in the incorporation during large-scale production of M-CSF in *E. coli*, we undertook the task of its complete structural characterization. We used a variety of analytical tools (amino acid analysis, amino-terminal sequencing, and a variety of mass spectrometry methods), realizing that the estimation of Nle can pose unique challenges when the purified product contains a very small amount of Nle. The total Nle content can be quantitated by amino acid analysis since the PTC derivative of Nle is well-separated from other amino acid residues on C-18 RP-HPLC columns (Randhawa et al., 1992). By combining amino acid analysis and N-terminal sequencing of the intact protein or RP-HPLC-purified fragments obtained from enzymatic digestion, we have examined the extent of Nle incorporation in M-CSF produced under a variety of fermentation conditions. Norleucine incorporation initially was determined by amino acid analysis. Selected samples (with 3%, 6%, 12%, and 20% Nle) were subjected to

† Some of the cost of this work was met by the NIH (Grant DK 34400, CHLS PI). The electrospray MS instrument was provided by the NIH through the Shared Instrumentation program (Grant RR06505).

* Author to whom correspondence should be addressed.

‡ Otsuka America Pharmaceutical Inc.

§ Children's Hospital Oakland Research Institute.

∥ Present address: Alexion Pharmaceutical Inc., 25 Science Park, New Haven, CT 06511.

⊥ Otsuka Pharmaceutical Co. Ltd.

Present address: Lawrence Berkeley Laboratory, Bldg. 70A-3663, 1 Cyclotron Rd., Berkeley, CA 94720.

® Abstract published in *Advance ACS Abstracts,* March 15, 1994.

[1] Abbreviations: M-CSF, macrophage colony stimulating factor; Nle, norleucine; Met, methionine; ESMS, electrospray ionization mass spectrometry; LC/MS, liquid chromatography followed by mass spectrometry; PEG, poly(ethylene glycol); PTC, phenylthiocarbamyl; PTH, phenylthiohydantoin; RP-HPLC, reversed-phase high-performance liquid chromatography; Glu-C, endoproteinase Glu-C; CAM, carboxyamidomethylation; DTT, dithiothreitol; *E. coli*, *Escherichia coli*; TFA, trifluoroacetic acid; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; AAA, amino acid analysis; AAS, amino acid sequencing; PBS, phosphate-buffered saline.

0006-2960/94/0433-4352$04.50/0

electrospray ionization mass spectrometry (ESMS) and LC/MS in order to establish the percent Nle incorporation at each of the four methionine residues in reduced carboxyamidomethylated M-CSF monomers. The results presented here evaluate the usefulness and limitation of chemical methods (amino acid analysis and N-terminal sequencing), as well as the mass spectrometry methods (ESMS and LC/MS), in assessing the level of norleucine in purified samples of a recombinant protein expressed in *E. coli*.

MATERIALS AND METHODS

Material. The methods for the production of M-CSF were modified from those published previously (Yamanishi et al., 1991). In order to produce a more homogeneous form of M-CSF, we cloned a gene encoding residues 4–153, as opposed to 3–153 as reported earlier. We also placed this gene under the control of transcriptional regulation elements taken from the bacteriophage λ. The expression vector was placed in the protease-deficient host SG21171 [htpR–, clp A–] (P. Arthur, unpublished work; *E. coli* SG 21171 was kindly provided by Dr. Sussan Gottesman of NIH). The M-CSF expression strain was grown in 14-, 41-, or 450-L Chemap fermentors. Unless specified otherwise in the text, the medium was defined and fermentations were carried out in a fed-batch mode. The glucose feeding was linked to the dissolved oxygen control. The temperature in the culture was shifted from 32 to 38–42 °C for 3 h to induce M-CSF production (P. Hughes, unpublished work). At the end of fermentation, cells were harvested by either centrifugation or tangential flow filtration. The resulting *E. coli* paste was disintegrated using a Rannie homogenizer, and inclusion bodies were harvested from the cell lysate by centrifugation. The refolding, purification, and characterization of M-CSF recovered from *E. coli* were performed as previously described (Yamanishi et al., 1991). A more homogeneous form of M-CSF (4–153) was produced, refolded, and purified by a modified scheme for large-scale production (Wilkins et al., 1993). Each purified monomer had greater than 85% purity, and each refolded M-CSF dimer had greater than 95% purity as determined by RP-HPLC assay, N-terminal sequencing, and SDS–PAGE analysis (data not shown here).

Reduction and Alkylation. M-CSF dimer (1 mg in 0.5 mL of 50 mM PBS containing 5 M guanidine hydrochloride and 0.1 M Tris-HCl) was adjusted to pH 8.5 and reduced with 20 μL of 0.1 M DTT at 37 °C for 1 h under argon. Following reduction, 20 μL of 0.5 M iodoacetamide (Sigma) was added under argon and allowed to react for 0.5 h in the dark at 37 °C. The residual iodoacetamide was reacted with 40 μL of 0.1 M DTT. CAM-M-CSF was dialyzed against 5% acetic acid using 3500 MWCO Spectra/Por dialysis tubing (Spectrum) in a cold room. The dialyzed sample was lyophilized and stored at 0–4 °C until use.

Digestion with Glu-C Proteinase. One-half milligram of lyophilized CAM-M-CSF protein was dissolved in 250 μL of 50 mM ammonium bicarbonate (pH 7.8). Glu-C proteinase (Promega) was added to achieve an enzyme:substrate ratio of 1:50, and the mixture was incubated under argon at 37 °C for 16 h. The reaction was terminated with the addition of 20 μL of 1% TFA. The digestion mixture was lyophilized, and peptides were dissolved in 0.5 mL of 0.1% TFA.

Peptide Separation. The Glu-C digests were fractionated by narrow-bore reversed-phase HPLC. For N-terminal sequencing analysis, the digest (0.1 mg in 100 μL of 0.1% TFA) was applied to a Brownlee C18 column (2.1 × 220 mm) using a Hewlett-Packard HPLC (Model 1090) equipped with a built-in diode array detector. A linear gradient from 0 to 60% was run over a period of 30 min at a flow rate of 225 μL/min (buffer A was 0.1% TFA and buffer B was 99.9% CH_3CN in 0.1% TFA). The remainder of the digest was analyzed by LC/MS, and details are provided in the mass spectrometry sections (Materials and Methods).

Determination of Norleucine. The total Nle content in each sample was determined by four independent methods, viz., amino acid analysis (ABI's PTC method), amino acid sequence by automated Edman degradation, and two types of mass spectrometry (ESMS and LC/ESMS).

Amino Acid Analysis. The desalted sample (intact dimer, CAM-M-CSF- or Glu-C-cleaved, and HPLC-fractionated peptide of M-CSF) was spotted onto a glass fiber disk and hydrolyzed with 6 N HCl (vapor phase) for 1.25 h at 155 °C using ABI's autohydrolyzer. The hydrolyzate was derivatized using a precolumn derivatizer (Model 420-A derivatizer, ABI, Foster City, CA) and analyzed using an on-line PTC analyzer (Model 130-A). The data were processed by an ABI Model 920-A data reduction module. The separation was optimized to obtain base-line separation of PTC-Nle from the neighboring PTC-Leu and PTC-Phe by employing a shallower gradient (Randhawa et al., 1992).

Amino Acid Sequence Determination. A known amount (0.1–5 nmol, based on AAA) of either intact M-CSF or the HPLC-fractionated methionine- and norleucine-containing peptide was directly spotted onto a biobrene-coated glass fiber disk in a 30-μL aliquot. The amino-terminal sequence was determined by using an ABI Model 477-A protein microsequencer equipped with an on-line Model 120-A PTH analyzer. The data were analyzed by a Model 900-A data reduction module. A gradient at the recommended flow rate of 210 μL/min (ABI's normal-1 gradient) starting with 11% buffer B instead of 12% buffer B and changing to 39% B instead of 38% B at 18 min provided base-line resolution of PTH-Nle from the neighboring PTH-Leu zone (Randhawa et al., 1992).

Determination of Bioactivity by the Dependent Cell Line NSF-60 Assay. The biological activity of M-CSF was determined by measuring the effect of samples on the growth of NSF-60 cells, as previously described (Wilkins et al., 1993).

Reversed-Phase HPLC/Electrospray Ionization Mass Spectrometry (LC/ESMS). Separation of all peptide mixtures was performed on a Michrom microbore HPLC system (Michrom Biosources, Pleasanton, CA) at a flow rate of 40 μL/min and 35 °C, using a 1- × 150-mm, 300-Å pore size, 5-μm particle size Vydac C-18 column. Endoproteinase Glu-C peptides derived from 200 pmol of protein were separated using solvents and a linear gradient described in the section on peptide separation above. The splitting of the flow (10:1) was achieved directly in the UV cell, allowing 4 μL/min to be analyzed by the mass spectrometer. Typically, 2 μL/min 0.2% TFA in MeOH was added as a sheath liquid prior to nebulization. The eluate was introduced into the mass spectrometer through a modified VG BioQ probe (Bitsch et al., 1993), and 4-s scans between m/z 300 and 1300 were acquired. Instrument conditions for mass spectrometric analysis are given below. A mixture of horse heart myoglobin and PEG 550 was used for mass-scale calibration. Quantitation of the relative abundance of the peptides containing methionine and norleucine residues (E-3, E-(4+5), E-7, and E-8) was done by calculating peak areas in selective ion chromatograms constructed for molecular ions specific for these peptides.

Electrospray Ionization Mass Spectrometry (ESMS). ESMS was performed on a VG BioQ quadrupole mass

spectrometer. The instrument was controlled and data was analyzed using LabBase software, and maximum entropy analysis of electrospray mass spectra of intact proteins was accomplished using MaxEnt software (both from VG Biotech/Fisons, Altrincham, UK). The electrospray ion source was operated at 3.2 kV. The nozzle-to-skimmer bias value was typically 50–55 V. For the analysis of intact rM-CSF, spectra were scanned from m/z 1035 to 1235 Da in 10 s, and several scans were summed to obtain the final spectrum. Mass-scale external calibration employed the multiply charged ions from a separate introduction of horse heart myoglobin (average mass 16 951 Da), and the resolution was adjusted so that the m/z 998 peak from myoglobin was 1.5–2 Da wide at 10% height. Samples of intact CAM-rM-CSF (4–153) were dissolved in 40% formic acid/40% MeOH and diluted with 50% MeOH/1% acetic acid at least 20-fold to give a final protein concentration of 10–25 pmol/μL. MaxEnt processing of raw data was employed to quantitate ESMS signals produced by various CAM-M-CSF variants, and peak areas were used as a measure of the species concentration in solution (Figure 4). A ratio between the peak area representing a species with n Nle (n = 0–4) and the sum of the peak areas of all M-CSF species was approximated as the species relative abundance in the intact protein preparation.

Calculation of Binomial Probabilities. Binomial probabilities of various M-CSF species were calculated according to the premise that total % Nle (a fraction of all methionine positions occupied by norleucine residues) represents an incidence rate (probability factor p) of independent events of the Nle for Met substitution. The results of total % Nle derived from different analytical techniques, within one standard deviation from the average value, were used for calculations to give the representative ranges of binomial probabilities. Specific equations used to calculate binomial probabilities were as follows:

$$0\ \text{Nle} = (1-p)^4;\quad 1\ \text{Nle} = 4p(1-p)^3;$$
$$2\ \text{Nle} = 4p^2(1-p)^2;\quad 3\ \text{Nle} = 4p^3(1-p);\quad 4\ \text{Nle} = p^4$$

RESULTS

The Strategy. We first optimized the separation of Nle (as a PTC derivative) on a C-18 RP-HPLC column using an ABI amino acid analyzer for the quantitation of total Nle in the M-CSF samples. Due to the relative instability of methionine during *in vacuo* acid hydrolysis, the total % Nle was calculated using an average residue value derived from the mean of several acid-stable amino acid residues. We also optimized the separation of the PTH derivative of Nle and used N-terminal sequencing to calculate the percentage of methionine replaced by Nle at position 10 in M-CSF (4–153), i.e., in Edman cycle 7. Again, the percentage of Nle released during Edman cycle 7 was calculated using the amount of Met, which was standardized according to the isoleucine response value obtained from Edman cycle 8. The ESMS data provided masses representing M-CSF species containing varying numbers of norleucine residues. The successful application of ESMS in examining modified M-CSF monomers (18 kDa) and native dimers (35 kDa) was reported earlier (Randhawa et al., 1991). The LC/MS of Glu-C proteinase digests not only provided a separation profile for each CAM-M-CSF sample but also gave on-line analysis of the masses of methionine- and Nle-containing fragments. From the reconstructed selected ion chromatograms, the ability to detect small amounts of Nle (3%) was demonstrated. In Figure 1, we present a comparison between the results of quantitation

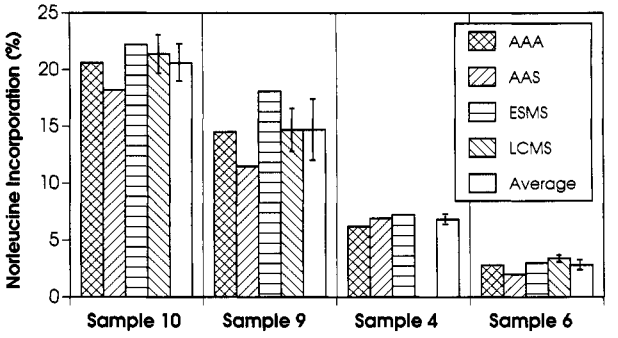

FIGURE 1: Nle incorporation in rM-CSF (4–153) samples 10, 9, 4, and 6, as determined by AAA, N-terminal sequencing (AAS) calculation method 2, ESMS of intact CAM monomers, and LC/MS of Glu-C digests. No LC/MS analysis was performed for sample 4. Standard deviations shown for LC/MS reflect the averaging of results obtained for the Nle incorporation at the different Met positions 10, 27, 61, and 65 (contributions at positions 10 and 27 were assumed to be identical since they are present in the same peptide). Open bars represent the average values of total Nle incorporation calculated from all of the above methods, and their heights equal 100p (p is a frequency parameter used for binomial predictions, as described in Materials and Methods).

Table 1: Amino Acid Composition of M-CSF (4–153) Samples 6 and 10 (with the Lowest and the Highest Nle Incorporation, Respectively)

amino acid	theor residue	sample 6		sample 10	
		pmol	comp	pmol	comp
Asp	40	1247	39.3	1292	41.3
Glu	46	1458	45.9	1461	46.7
Ser	24	850	26.8	802	25.6
Gly	4	230	7.2	199	6.4
His	6	194	6.1	180	5.7
Arg	10	300	9.5	298	9.5
Thr	18	753	23.4	694	22.2
Ala	10	352	11.1	327	10.4
Pro	6	219	6.9	213	6.8
Tyr	8	168	5.3	179	5.7
Val	18	471	14.8	436	13.9
Met	8	178	5.6	171	5.5
Nle		7	0.2	50.6	1.6
Met+Nle[a]	(8)	185	5.8	221.6	7.1
Ile	14	388	12.2	376	12.0
Leu	32	1001	31.5	978	31.2
Phe	18	556	17.5	535	17.1
Lys	22	641	20.2	699	22.3
total[b]	284	9013	283.8	8890	283.9
pmol			31.7[c]		31.3[c]

[a] Theoretically, Met+Nle should equal eight residues. [b] Cys and Trp values are not included in the composition. [c] Amount of calculated average residue in picomoles. This value of average residue was obtained by dividing the total pmol value by the number of expected residues (284) in the M-CSF (4–153) dimer.

of Nle incorporation obtained by different methodologies. The average value of total Nle incorporation derived from all techniques was interpreted as corresponding to the likelihood of the Met→Nle substitution event. It was further applied as a frequency parameter p for the calculation of binomial probabilities of various Nle-carrying species (Materials and Methods).

Assessment of the Norleucine Content in the Intact rM-CSF Species

I. Amino Acid Analysis. Amino acid analysis was used as a primary technique to quantitate the level of norleucine incorporated into a truncated form of M-CSF expressed in *E. coli*. In order to accurately determine the amount of Nle, a separation of PTC-Nle from the neighboring PTC-Leu and

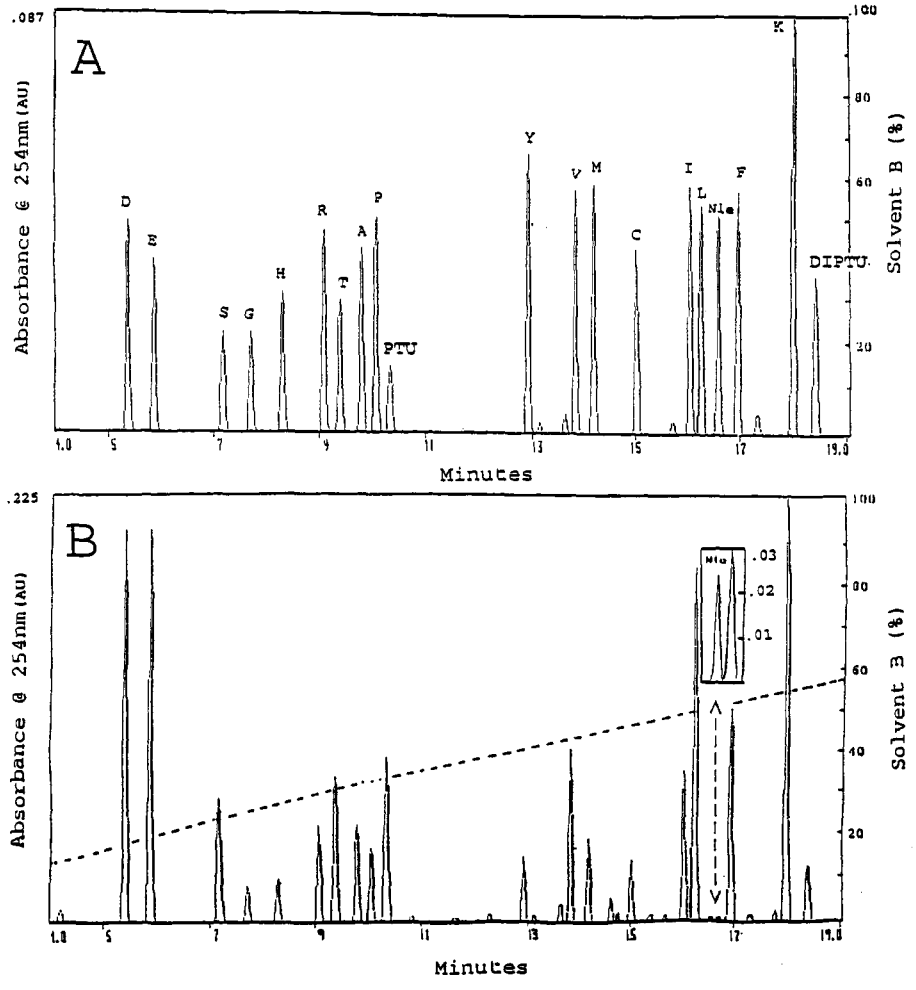

FIGURE 2: HPLC analysis of PTC amino acid derivatives: (A) standard amino acids and norleucine containing 250 pmol of each component; (B) acid hydrolyzate from M-CSF sample 6 (with the lowest amount of Nle). The insert shows an amplified A(254 nm) response recorded during the emergence of Nle from the column. The C-18 PTC column (ABI) was equilibrated with solvent A (0.05 M sodium acetate, pH 5.4) at 34 °C, and the PTC amino acid zones were separated using a modified gradient with solvent B (70% ACN/30% 0.032 M sodium acetate, pH 6.1) at a constant flow rate of 0.3 mL/min. The HPLC gradient is shown with interrupted lines.

PTC-Phe was achieved (Figure 2A) by modifying the Model 130-A PTC amino acid analysis gradient (Randhawa et al., 1992). This allowed for a reliable quantitation of total % Nle in preparations ranging from high (~20% in sample 10) to very low (~3% in sample 6) levels of the Met→Nle substitution, while using about 30 pmol of protein (Table 1). The % Nle is defined as the ratio between the measured amount of norleucine and the methionine content expected from the amino acid sequence. When calculating total % Nle using AAA data, the relative instability of the methionine residue (Savige & Fontana, 1977; Schecter et al., 1975), in contrast with the stability of the norleucine residue (Piez, 1968), in the conditions of protein hydrolysis has to be taken into account. Therefore, the average residue value derived from acid-stable amino acids rather than the actual amount of released methionine should be used as a representation of methionine content in protein, to which the amount of released norleucine is to be compared. Average residue values derived from a set of either three (Leu, Phe, and Glu) or seven (Leu, Phe, Glu, Asp, His, Lys, and Arg) stable amino acid residues give very close results for total % Nle (Table 2). In contrast, when the sum of released methionine and norleucine is used as a measure of all methionine positions in the protein molecule, elevated values of % Nle (between 10–30%) are obtained because chemically modified Met residues remain unaccounted for (Table 2).

Table 3 summarizes the Nle (%) data for 10 M-CSF samples analyzed by different methods. In each of the 10 samples, the M-CSF was produced under different fermentation conditions with varying levels of norleucine substitution. The extent of norleucine incorporation varied between 0 and 20% by ABI's PTC method and between 0 and 19% by Water's Pico-Tag method for the same samples (data not shown here). The amino acid analysis data shown in Table 2 clearly point out that while the levels of Nle in samples 6 and 10 are different (2.8% and 20.6% Nle, respectively), the amount of Ile released under our *in vacuo* acid hydrolysis conditions is very similar (Table 1). These data can be interpreted to suggest that the rate of hydrolysis of slow-releasing peptide bonds involving Met/Nle–Ile bonds between positions 11 and 12 and peptide bonds Ile–Met/Nle between positions 60 and 61 in the M-CSF sequence (Yamanishi et al., 1992) is also similar; therefore, the percentage of calculated Nle in these samples is not influenced by the hydrolysis conditions. In each of 10 samples analyzed by either or both amino acid analysis methods, the purity of each M-CSF sample was >85% for the monomer and >95% for the purified, refolded dimer samples. The purity of each was determined by using a combination of N-terminal sequencing and SDS–PAGE patterns gel-stained with Coomassie blue or silver-staining methods (data not shown here).

II. N-Terminal Amino Acid Sequencing. Although the automated Edman degradation method has been commonly

Table 2: Amount of Norleucine Determined[a] in Selected M-CSF Samples by Different Methods of Calculation

	sample		
	6	9	10
Met (pmol)	178	145	171
Nle (pmol)	7	32	51
method of calculation			
Nle (%) of Met+Nle[b]	3.9	18.00	23.0
(pmol average residue)	(23.1)	(22.1)	(27.7)
Nle (%) of AR from 3 AA[c]	2.8	14.5	20.6
(pMol average residue)	(31.4)	(27.4)	(31.0)
Nle (%) of AR from 7 AA[d]	2.8	14.7	20.4
(pmol average residue)	(30.9)	(27.0)	(31.3)
Nle (%) of AR from 16 AA[e]	2.8	14.6	20.4
(pmol average residue)	(31.7)	(27.2)	(31.3)

[a] Determined by the amino acid analysis method. [b] The Nle (%) value was calculated from the pmol of Nle, which was multiplied by 100 and divided by total (pmol) Met + Nle. [c] Average residue (AR) value of three very stable amino acid residues (viz., L, F, and E taken from Table 1) was multiplied by 8 (number of methionine residues in M-CSF) to obtain the theoretical methionine in each sample. Nle (%) values expressed as % of the total theoretical methionine. [d] Data were calculated in the same way as shown in the above row. The AR value of seven stable residues (viz, L, F, E, D, H, K, and R) was taken from Table 1). [e] The AR value for 16 amino acid residues (excluding C and W) was taken from Table 1, and Nle (%) of theoretical amount of Met was calculated as described above.

Table 3: Comparison between Norleucine Values of M-CSF Samples Produced in Different Fermentation Size Tanks and Purified as Monomers and/or Dimers after Refolding

M-CSF sample no.	sample purity (%)[a]	norleucine (%) determined by amino acid analysis[b]	amino acid sequence calculation method 1[c]	amino acid sequence calculation method 2[d]
Fermentation Size: 14 L				
1[e]	97	0	0	0
2[e]	99	0	0	0
3[f]	86	5.0	7.7	6.7
Fermentation Size: 41 L				
4	98	6.2	8.5	6.9
5[f]	84	19.1	20.2	19.2
Fermentation Size: 450 L				
6	98	2.8	2.1	2.0
7	95	12.3	12.5	10.9
8	98	13.4	13.9	11.8
9	97	14.5	15.3	11.5
10	98	20.6	19.3	18.2
range	84–98	0–20.6	0–20.2	0–19.2
average	95	9.4	9.9	8.7

[a] The purity of each sample was determined by the combination of SDS–PAGE and N-terminal sequencing. [b] Each value in this column was determined by the amino acid analysis method using the precolumn derivatization procedure and PITC chemistry (ABI). The data are expressed as the percent of Nle using an average residue (Table 2). [c] Calculation method 1 shows the % Nle calculated from Met+Nle released in Edman cycle 7. [d] Calculation method 2 shows Nle (%) calculated from the corrected value of Ile taken from Edman cycle 8 (instead of the amount of Met released in cycle 7) (see text). [e] This represents the use of complex medium fermentation conditions. All of the remaining samples (3–10) were grown in minimal fermentation medium. [f] Samples 3 and 5 were analyzed from purified M-CSF monomer. The rest of the samples (1, 2, 4, and 6–10) were analyzed as refolded bioactive M-CSF dimer.

used to obtain the N-terminal sequence of proteins and peptides at low picomole levels, the use of the method to quantitate any residue with greater accuracy has been difficult due to the (uncontrolled) recovery of the protein and an approximate method for determining the repetitive yield during sequencing. We addressed this issue in an attempt to obtain more

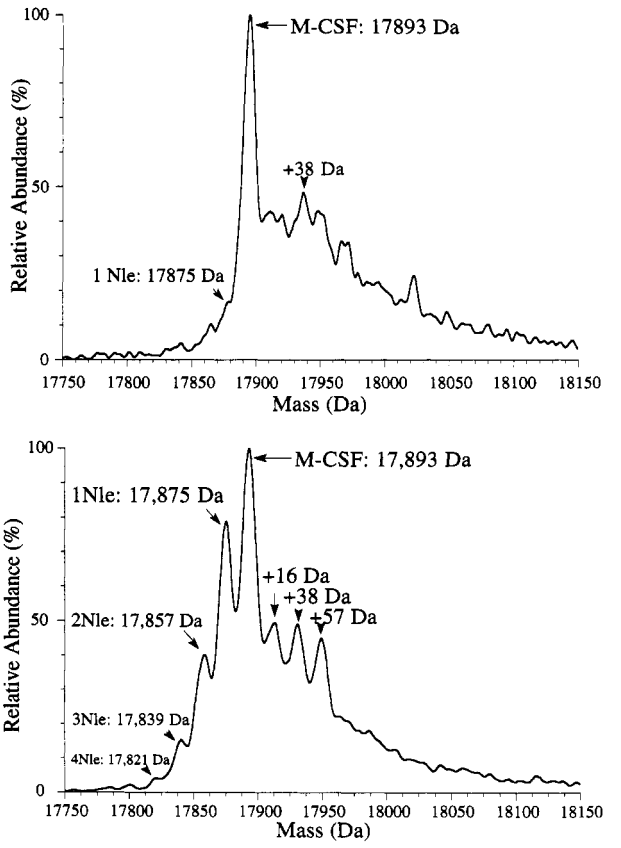

FIGURE 3: Deconvoluted electrospray mass spectra of CAM-rM-CSF (4–153) sample 6 with 2.8% Nle (top) and sample 10 with 20% Nle (bottom). Series of multiply protonated molecular ions are generated from each protein species in the process of electrospray ionization. They are sorted in a quadrupole mass spectrometer according to their mass (m) over charge (z) values (m/z). There is one charge difference in the number of charges carried by any two adjacent molecular ions derived from the same species. From this assumption and from the measured m/z values, the absolute number of charges of every molecular ion and, consequently, the molecular mass of a species are calculated. In the process of further data processing, all multiply charged molecular ions are mathematically transformed to deliver a deconvoluted spectrum shown above, where the molecular weight of the species is presented in the real mass scale. Average molecular masses of an authentic rM-CSF (4–153) and various norleucine-containing species are given.

quantitative information on the % Nle released at Edman cycle 7 (sequence position 10) in *E. coli*-derived M-CSF (4–153) samples. Table 3 summarizes the data obtained from N-terminal sequence analysis of 10 M-CSF samples, which varied in their total % Nle content from 0 to 20.8%. The total % Nle data were calculated using two methods: (1) from total Met + Nle (pmol) released during Edman degradation, and (2) using the corrected repetitive yield value of the neighboring Ile residue from Edman cycle 8. In each case there is good agreement between the % Nle values in their respective category. However, due to the relative instability of methionine during Edman degradation, as a result of exposure to harsh chemical conditions (such as TFA, etc.), the total % Nle value calculated from total Met + Nle was found to be ~10% higher using calculation method 1 (Table 3). Therefore, the value of Ile (a stable residue) from Edman cycle 8 was used to calculate the total % Nle released in Edman cycle 7, after making necessary correction of the theoretical yield in cycle 7 (calculation method 2, Table 3). It can be concluded from the data in Table 3 that the Nle value calculated from the N-terminal sequencing method is comparable to the Nle value obtained by amino acid analysis methods. These data provide strong evidence that N-terminal

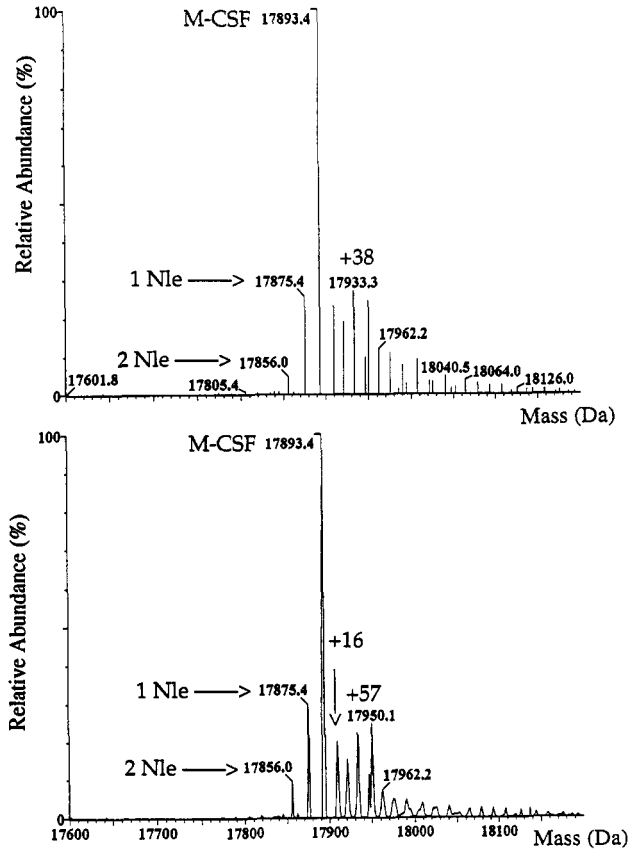

FIGURE 4: MaxEnt output of ESMS analysis of intact CAM-rM-CSF (sample 4). Data were acquired as described in the legend for Figure 3 and processed using MaxEnt software to produce the zero-charge mass spectrum on a molecular mass scale (Ferrige, 1992). The profile MaxEnt data and the corresponding peak area representation are shown on the bottom and the top, respectively. Stick heights in the top spectrum are quantitative representations of the intensity of the signals produced by all detected compounds.

sequencing methods alone can provide very reliable data for % Nle incorporation, and the data presented here also support the earlier premise that the Nle incorporation at any of the four methionine positions in M-CSF monomers during synthesis in *E. coli* is uniform and random (Cowie et al., 1959; Bogosian et al., 1989).

III. Electrospray Ionization Mass Spectrometry. Electrospray ionization mass spectrometry (ESMS) was used both to confirm the authenticity of recombinant proteins (by demonstrating agreement between the average molecular mass measured and that expected from the amino acid sequence) and to evaluate the relative abundance of different protein species. The molecular masses of all species were found to be within 1 Da of the expected values (CAM-M-CSF, 17 893 Da). No evidence of the presence of species with (additional) N-terminal methionine or norleucine was found. The methionine to norleucine substitution results in a −18-Da difference in the molecular mass of the variant protein. This mass differential (0.1%) is large enough to be detected and measured by electrospray mass spectrometry that allows for molecular mass measurement with 0.01% accuracy at the level of an ~18-kDa protein (Van Dorsselaer et al., 1990). Therefore, the presence or absence of norleucine residues in the CAM-rM-CSF molecule demonstrated itself directly by the presence or absence of consecutive peaks preceding the authentic rM-CSF peak by $n \times 18$ Da on a mass scale (n is the number of misincorporated norleucine residues; Figures 3 and 4).

Table 4: Amino Acid Sequence of Norleucine and/or Methionine Peptides Purified by RP-HPLC from Glu-C Endoproteinase-Digested Reduced CAM-M-CSF (4–153)

			corresponding retention time (min) of HPLC peptide/amount (pmol) released during Edman			
no.	sequence position	amino acid residue	19.61[d] min	18.54[d] min	Edman cycle no.	19.22[d] min
1	4	S	51	97	–	–
2	5	E[a]	147	261	–	–
3	6	Y	101	161	1	94
4	7	C[b]	+	+	2	+
5	8	S	37	266	3	37
6	9	H[c]	ND	ND	4	ND
7	10	M	21	110	5	26
		Nle	32	0		25
8	11	I	51	130	6	57
9	12	G	41	167	7	44
10	13	S	15	37	8	17
11	14	G	41	77	9	66
12	15	Hc	ND	20	10	ND
13	16	L	32	69	11	23
14	17	Q	26	58	12	21
15	18	S	8	22	13	7
16	19	L	27	53	14	37
17	20	Q	25	44	15	30
18	21	R	ND	22	16	ND
19	22	L	25	45	17	31
20	23	I	20	38	18	15
21	24	D	7	18	19	5
22	25	S	5	11	20	5
23	26	Q	14	22	21	14
24	27	M	5.9	15	22	5.1
		Nle	8.8	0		7.6
25	28	E	4	13	23	4
residues			4-28	4-28		6-28
designation			E-3+4	E-3+4		E-4
Met-10:Nle-10			0.4:0.6	1:0		0.5:0.5
Met-27:Nle-27			0.4 0.6	1:0		0.5:0.6

[a] Peptide bond partially cleaved by Glu-C enzyme. [b] Cysteine residue was reduced and carbamidomethylated. [c] PTH–histidine coeluted with PTH–alanine; therefore, accurate quantitation of His cannot be made. [d] Mass spectrometry analysis data shown in Table 5.

Apart from the expected rM-CSF and its Nle-substituted species, the higher molecular mass species were noted in all samples (Figures 3 and 4). Figure 4 shows the refined MaxEnt spectra (Ferrige et al., 1992) from which a tentative assignment of some of the higher mass components could have been made. In all samples, ions indicative of potassium adducts (+38 Da), oxidation (+16 Da), and additional carboxyamidomethylation (+57 Da) of the authentic rM-CSF molecule were detected. All rM-CSF variants are likely to carry the same modifications/adducts at the same level (except for an oxidation that would strongly depend upon the methionine content). Electrospray mass spectrometry is generally regarded as a nonquantitative technique. However, very close structural resemblance of different rM-CSF variants allows for the assumption that the concentration of a species directly correlates with the intensity of the molecular ion signals that it generates. We applied the MaxEnt algorithm (Ferrige et al., 1992) to quantitate peak areas corresponding to various rM-CSF species, and we thus derived species abundance to calculate % Nle (Randhawa et al., 1991), as well as the distribution of rM-CSF variants. Incidentally, the mass differences between the most prominent modifications/adducts are very close to the mass differences between rM-CSF variants. Therefore, the shares of higher mass components derived from differently mistranslated components could not be unambiguously defined and thus quantitated. This com-

Table 5: Calculated and Determined Molecular Masses of Methionine and Norleucine Peptides Purified from Reduced Carbamidomethylated M-CSF (4–153) Sample 10

Glu-C-cleaved fragment	amino acid sequence of M-CSF fragment	sequence position	average mass (Da)		HPLC (t_R, min)
			calcd	obsd	
E-3+4 (2 Met)	S-E[a]-Y-C[b]-S-H-M-I-G-S-G-H-L-Q-S-L-Q-R-L-I-D-S-Q-M-E	4–28	2907.2	2907.0	18.53[c]
E-3+4 (1 Met + 1 Nle)	S-E[a]-Y-C[b]-S-H-M/Nle-I-G-S-H-L-Q-S-L-Q-R-L-I-D-S-Q-M/Nle-E	4–28	2889.2	2889.0	19.61[c]
E-3+4 (2 Nle)	S-E[a]-Y-C[b]-S-H-Nle-I-G-S-H-L-Q-S-L-Q-R-L-I-D-S-Q-Nle-E	4–28	2871.2	2870.8	20.21
E-4 (1 Met + 1 Nle)	Y-C[b]-S-H-M/Nle-I-G-S-G-H-L-Q-S-L-Q-R-L-I-D-S-Q-M/Nle-E	6–28	2673.0	2672.0	19.22[c]
E-4 (2 Met)	Y-C[b]-S-H-M-I-G-S-G-H-L-Q-S-L-Q-R-L-I-D-S-Q-M-E	6–28	2691.0	2690.7	18.53
E-4 (2 Nle)	Y-C[b]-S-H-Nle-I-G-S-G-H-L-Q-S-L-Q-R-L-I-D-S-Q-Nle-E	6–28	2655.0	2654.3	20.21
E-7 (1 Met)	Q-L-K-D-P-V-C-Y-L-K-K-A-F-L-L-V-Q-D-I-M-E	42–62	2552.1	2551.8	28.58
E-7 (1 Nle)	Q-L-K-D-P-V-C-Y-L-K-K-A-F-L-L-V-Q-D-I-Nle-E	42–62	2534.1	2533.7	29.57
E-8 (1 Met)	D-T-M-R-F-R-D-N-T-P-N-A-I-A-I-V-Q-L-Q-E	63–82	2332.6	2332.4	15.15
E-8 (1 Nle)	D-T-Nle-R-F-R-D-N-T-P-N-A-I-A-I-V-Q-L-Q-E	63–82	2314.6	2313.7	15.98

[a] Peptide bond partially cleaved by Glu-C enzyme. [b] Cysteine residue was reduced and carbamidomethylated. [c] The amino acid sequence is shown in Table 4.

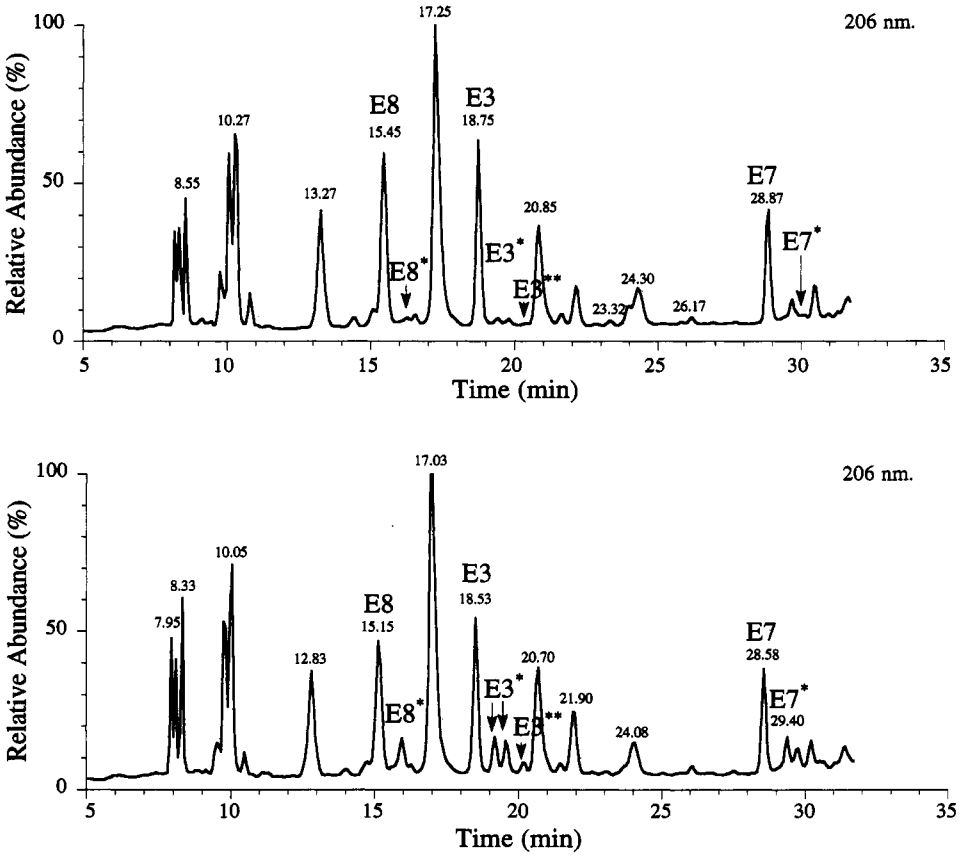

FIGURE 5: Reversed-phase HPLC profiles of reduced carboxyamidomethylated aliquots of Glu-C protease-digested M-CSF samples with varying norleucine content: (top) sample 6 (2.8% Nle) and (bottom) sample 10 (20.6% Nle). Each lyophilized digestion mixture was reconstituted in 20% CH_3CN in 0.1% TFA, and 200 pmol was applied to a Vydac C-18 (300 Å, 5 μm, 1 × 150 mm) column at a 40 μL/min flow rate. The peptides were eluted with a linear gradient of CH_3CN and detected by their absorbances at 206 nm. Eluate was split at the flow cell 1:10 in order to deliver 10% sample to a VG Bio-Q mass spectrometer for on-line mass spectrometric analysis. Each of the methionine-containing peptides is identified with an E followed by a number identifying the location of the peptide in the M-CSF sequence, and each corresponding norleucine-containing peptide is marked with an asterisk. Peptide E-(3+4) coelutes with its E-3 counterpart.

plication forced us to simplify calculations by only accounting for the protonated ions and ignoring the interference of alkali adducts and products of chemical modifications. This approach resulted in some underestimation of the number of rM-CSF species with none or one Nle residue, whereas the quantitation of multiply substituted variants was least affected. Thus obtained estimates of total % Nle were generally in good agreement, although they slightly exceeded the results obtained by either AAA or AAS. The actual distribution of differently mistranslated species was compared with the theoretically expected binomial probabilities, which were calculated using total norleucine incorporation (total % Nle) as a frequency parameter p (see Materials and Methods): the relative abundance of multiply substituted species was generally higher than that expected, at the expense of singly and/or nonsubstituted species. We believe it to be a real phenomenon rather than an artifact resulting from the oversimplified calculations, since any errors due to the underestimation of the amount of the authentic and singly substituted rM-CSF (see above) would result in the lower rather than higher binomial probability predictions for multiply substituted species.

Assessment of the Norleucine Content at All Methionine Positions of M-CSF

In order to directly compare norleucine incorporation at all methionine positions, endoproteinase Glu-C peptides derived

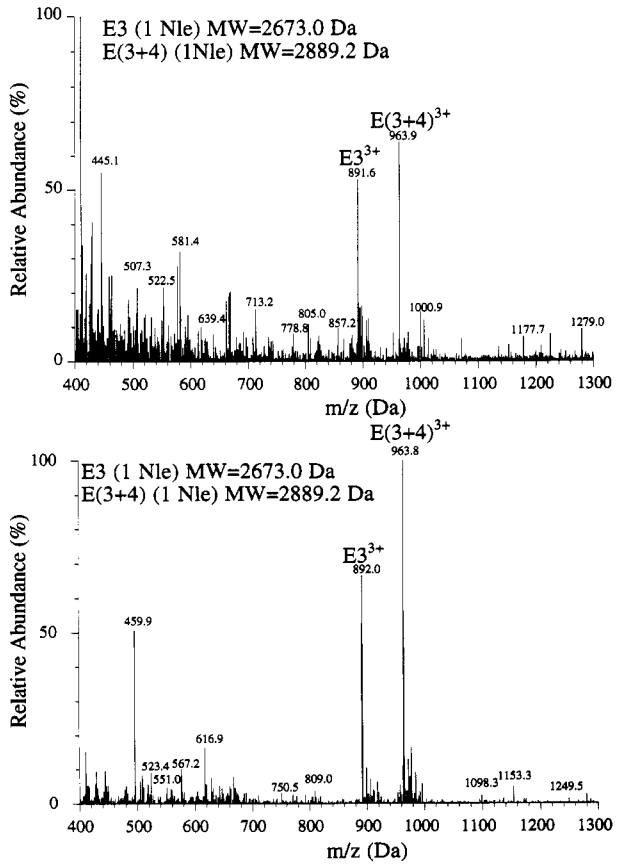

FIGURE 6: Electrospray mass spectra of peptides E-3* and E-(3+4)* (both containing 1 Nle), on-line separated by microbore HPLC derived from M-CSF sample 6 (top) and sample 10 (bottom), with total Nle levels of 2.8% and 20.6%, respectively, by AAA (Table 3).

from reduced and alkylated rM-CSF were analyzed by Edman sequencing and on-line reversed-phase HPLC/electrospray mass spectrometry.

I. Edman Sequencing of HPLC-Separated Proteolytic Fragments. All Nle- and Met-containing peptides were separated by reversed-phase HPLC on a Vydac C-18 reversed-phase HPLC column and sequenced by Edman degradation. The level of Nle incorporation was found to be constant for all positions of substitution in the protein chain (Table 4). Table 4 also shows an example of amino acid sequencing of the fully digested peptide E-4 and its partially digested counterpart E-(3+4). As is evident from those results, the same extent of norleucine incorporation was found at both methionine positions. Similarly, the peptides representing methionines 61 and 65 (labeled as E-7 and E-8, respectively) were identified by N-terminal sequencing, as shown by their elution positions (retention time) in Table 5 and Figure 5.

II. Reversed-Phase HPLC/ESMS Analysis of Proteolytic Digest. Application of narrow-bore LC/ESMS allowed for the fast and sensitive screening of proteolytic digestion mixtures for the Nle-containing fragments. Samples 6, 9, and 10 were examined. All Glu-C-generated Nle-containing peptides were separated from their Met counterparts, and their presence was evident even in the case of sample 6, which contains only 3% total norleucine (Figure 5). Table 5 combines the chromatographic and mass spectrometric data obtained for all Glu-C fragments that cover methionine-containing regions of M-CSF (4–153). Figure 6 shows sample ES mass spectra of the peptides E-4 and E-(3+4), each carrying one Nle residue, derived from samples 6 and 10 (2.8% Nle, top, and 20.6% Nle, bottom, respectively). Ion current, produced by molecular ions derived from the Nle and Met counterparts and represented by the peak areas in their reconstructed selective ion chromatograms, was used as a measure of the relative abundance of the corresponding peptides. The results obtained for positions 10 and 27, both within peptides E-4 and E-(3+4), for position 61 (peptide E-7), and for position 65 (peptide E-8) were averaged to give the total Nle incorporation value (Figure 1); the low standard deviation of those measurements points to the uniform distribution of norleucine within all methionine positions in the rM-CSF molecule. The total extent of norleucine incorporation thus derived was consistent with the data obtained by AAA, AAS, and ESMS of intact protein (Figure 1).

Figure 7 shows an example of selected reconstructed ion chromatograms for the triply charged molecular ions produced by peptides E-4 and E-(3+4) with 0, 1, and 2 Nle residues (sample 10, 20.6% Nle). The distribution of peptides E-4 and E-(3+4) with 0, 1, and 2 Nle residues was also examined. It was slightly biased toward the doubly substituted species, when compared with binomial probabilities calculated for two points of substitution. LC/MS analysis of the proteolytic digest mixture revealed the presence of fragments whose masses corresponded to normal Glu-C peptides modified by the addition of either +16 Da (oxygen) or +57 Da (CAM). This finding corroborated a tentative assignment of the higher mass species detected by ESMS of intact protein as products of rM-CSF oxidation or additional carboxyamidomethylation. In a chromatographic peak eluting in between the Met- and Nle-containing E-7 peptides, an ion corresponding to the N-terminal pyroglutamyl analog of E-7 was detected. The latter modification, which brings in a −17-Da mass difference from normal E-7, is likely to confuse detection of the Met→Nle mutation (−18 Da), should the two peptides coelute.

III. Bioactivity of Nle-Containing M-CSF. As shown in Figure 8, the bioactivity value for M-CSF sample 10 (with 20% Nle substituted at each of the four methionine residues) is similar to that obtained for M-CSF sample 6 (with 3% Nle) and samples 1 and 2 (with no detectable Nle). From the bioactivity profile of seven samples with varying Nle amounts, it is clear that there is no correlation between the bioactivity and the Nle content.

DISCUSSION

The requirement of very high purity and homogeneity for the recombinant proteins that are used in therapy and experimental studies presents the scientific community with the analytical challenge of proving the integrity of their final product. Despite the fact that our ability to determine the amount and the amino acid sequence of unknown proteins has become very accurate even at low picomole levels (Ericsson et al., 1991; Niece et al., 1991), to date all of the published reports provide only qualitative measurement of Nle in recombinant protein derived from *E. coli* (Bagosian et al., 1989). Nle has served as an internal standard for over 25 years in amino acid analysis when using an ion-exchange column method (Peiz, 1968); however, the precolumn derivatization and reversed-phase HPLC system lack the resolution of the Nle derivative (Bogosian et al., 1989; Violand et al., 1989).

We have developed a method that resolves PTC-Nle from other components, and we show that the Nle (%) value calculated from the combined value of Met + Nle is unreliable due to the loss of methionine. We have shown that the % Nle value derived from using the Met + Nle value is, on the average, 20% higher by the amino acid analysis method and about 10%

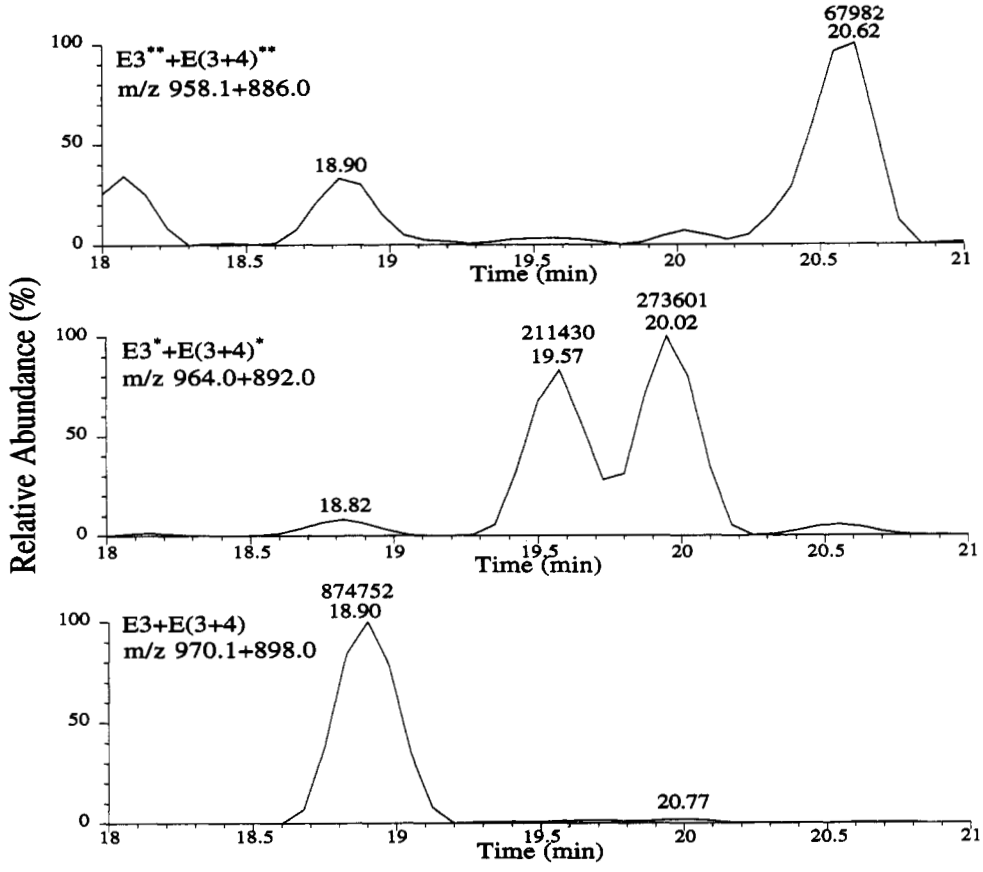

FIGURE 7: Reconstructed selective ion chromatograms of Glu-C-derived peptides E-3 and E-(3+4) from sample 10. Ion current generated by triply charged molecular ions produced from the counterparts containing 2, 1, and 0 Nle residues is shown in the upper, middle, and bottom panels, respectively. Numbers above the retention times represent the peak areas corresponding to the above-described ions, which are specified on the left-hand side of each panel. There is a ~0.4-min delay between the UV and mass spectrometer detectors.

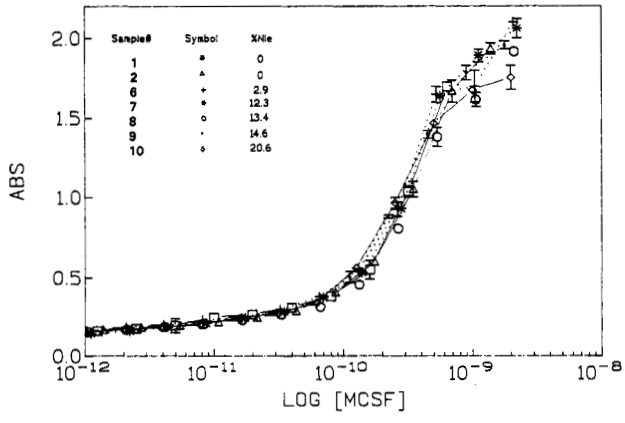

FIGURE 8: Biological activity of various M-CSF samples. The sample number and amount of Nle shown in the insert represents Nle (%) taken from amino acid composition data (Table 3).

higher by automated N-terminal sequencing analysis than expected. This increase reflects a partial oxidation of Met in the intact protein, but it is mainly due to the destruction of methionine during these analyses. Although the destruction of methionine can be minimized by the addition of β-mercaptoethanol as a reducing agent and by avoiding the presence of oxygen during *in vacuo* acid hydrolysis, the precise measurement of methionine is unfortunately very difficult to achieve consistently, even under the best of conditions. However, since Nle is extremely stable under these conditions, we propose here that the calculation of the % incorporation of Nle at the position of Met can be made accurately by using an average residue value of any of the acid-stable residues or any average of several stable residues. Since our amino acid sequencing and mass spectrometry data have also confirmed that the Nle incorporation is uniform at all methionine positions in *E. coli*-synthesized M-CSF, our approach to using automated N-terminal sequencing analysis for the quantitation of Nle should facilitate its accurate determination in any protein with one or more methionine residues in the amino-terminal region (1–25 residues).

Mass spectrometry provides an augmentation of classical protein chemistry methods in confirming a fidelity of protein expression or detecting any deviations from the expected structure. ESMS is especially well-suited to serve as a quality control in recombinant protein technology since it can handle mixtures of species and measures molecular masses of intact proteins with high resolution and accuracy (Covey et al., 1988; Fenn et al., 1989; Smith et al., 1990). In its most sophisticated form, when electrospray ionization is combined with the Fourier transform mass spectrometry, it offers a resolution of 50000–80000, while providing a molecular measurement for a 17-kDa protein with <6 ppm error (Henry et al., 1991). However, even when relatively inexpensive low-resolution quadrupole detectors are used, the accuracy of molecular mass measurements is at the level of 0.01%.

A further noteworthy improvement in apparent resolution is achieved when electrospray spectra are processed using a MaxEnt algorithm, as was demonstrated by Ferrige et al. in resolving two ~16-kDa protein species that were 6 Da apart (Ferrige et al., 1992). Without the aid of MaxEnt data analysis, 12 Da is a practical limit of resolution for a ~16-kDa protein (Witkowska et al., 1991), and we would expect that very similar limitations apply in the case of a slightly heavier CAM-rM-CSF monomer (17.5 kDa). Therefore, all

mutations or modifications resulting in a more than 12-Da change in molecular mass theoretically should be detectable in mixtures with normal counterparts, and the masses of the main species should be measured accurately within 2 Da. Our results fully meet these predictions: we were able to detect the multiple Met→Nle replacements, each resulting in −18-Da mass difference, and the accuracy of all molecular mass measurements was within 1–2 Da. An attempt was made to use ES mass spectra for the quantitation of relative abundances of species. In general, the responses of different analytes are structure-specific rather than uniform, but it is reasonable to expect that very closely related species (e.g., Met and Nle forms of rM-CSF) will ionize and be detected with the same efficiency.

The question of sustained linearity of the ion current response in the whole concentration range was tackled by Van Dorsselaer et al. (1990), who showed very good agreement between the UV absorbance and ESMS quantitation results regarding the relative abundances of full-length (predominant) and truncated forms (minor components) of recombinant protein. ESMS-based calculations of norleucine incorporation generally showed higher levels than those obtained by other techniques (Figure 1). This discrepancy results from a lack of resolution between different rM-CSF species and their various adducts/chemical modifications, which disabled accurate quantitation of whole populations of the Nle- and Met-containing species. Nevertheless, the method proved capable of providing total % Nle generally within 20% of levels found by AAA and AAS. ESMS of intact protein was also successful in detecting low levels of Nle incorporation (sample 6, 2.8% Nle). It must be stressed, however, that the Met→Nle substitution detection limit would depend not only on the total incorporation level but also on the number of mutation sites within the molecule: the higher number of possible sites results in the higher binomial probability of a singly substituted variant (e.g., in sample 6 at 2.8% total Nle, 10.3% of a single Met→Nle species is expected, given four substitution sites).

We anticipate that an even higher sensitivity of Nle detection can be achieved by performing LC/ESMS of proteolytic fragment mixtures, especially when monitoring only selected molecular ions. An advantage of this approach is that molecular ions of interest, in this case the Nle-containing peptides, can be detected and quantitated even when they coelute with other species, albeit with different mass. Within the limitation discussed above, ESMS also allowed for the measurement of a distribution of species carrying different numbers (from 0 to 4) of Nle residues in the molecule. We were unable to confirm those results using an alternative technique: we attempted on-line reversed-phase HPLC/ESMS, but were unsuccessful in achieving full separation of rM-CSF molecules with different degrees of methionine replacement (data not shown). The ESMS results showed an unanticipated disparity between actual distributions and expected binomial probabilities. This would suggest that, in our expression system, the availability of norleucine and, consequently, the frequency of substitution vary during the time of rM-CSF synthesis. However, in accordance with previously reported findings, the incorporation of norleucine was found to be uniform for all methionine positions.

It has been suggested that in those proteins where methionine residues play purely a structural role, such as S. nuclease (Anfinsen & Corley, 1969), RNAse-S-peptide (Rocchi et al., 1969), β-lactogalactosidase (Naider et al., 1972), and adenylate kinase (Gilles et al., 1988), substitution of methionine with Nle does not affect the catalytic activity of the enzyme. Similarly, no loss of bioactivity was reported for bovine somatotropin (Bogosian et al., 1989), ACTH (Boissonas et al., 1966), IL-2 (Tsai et al., 1988), and IGF-1 (Forsberg et al., 1990). The results of this study confirm the absence of any role for any of the eight methionine residues in the bioactivity of M-CSF dimer. However, it is possible that substitution of Nle could have an effect on other assays, specifically marrow cell colony formation on the functional activation of mature monocytes.

ACKNOWLEDGMENT

We acknowledge Dennis Brown, Maurice Guertin, Karen Palumbo, Tom Rohrer, Shih-Queen Lee-Lin, and Yoshimasa Isakari for their technical support and Michael Seidman for helpful discussions throughout the entire work. The work contributed by Masayuki Takahashi is also acknowledged with thanks. The authors are grateful to Brian Green of VG BioTech/Fisons for performing MaxEnt data analysis and to Stuart Smith, Richard T. Jones, Alan J. Smith, Reed J. Harris, and Ronald L. Niece for their critical reading of the manuscript. We also thank Yasuo Iriye for his support and encouragement throughout this work.

REFERENCES

Anfinsen, C. B., & Corley, L. G. (1969) *J. Biol. Chem. 244*, 5149–5152.

Bitsch, F., Witkowska, H. E., Nugent, K., King, D., & Shackleton, C. H. L. (1992) Assessing structural alterations in natural and recombinant proteins by RP-LC/ESI MS. *Proceedings of the 40th ASMS Conference on Mass Spectrometry and Allied Topics*, May 31–June 5, 1992, Washington, D.C., pp 428–9, ASMS, Washington, D.C.

Bitsch, F., Ma, W.-W., Macdonald, F., Nieder, M., & Shackleton, C. H. L. (1993) *J. Chromatogr. 615*, 273–280.

Bogosian, G., Violand, B. N., Dorward-King, E. J., Workman, W. E., Jung, P. E., & Kane J. F. (1989) *J. Biol. Chem. 264*, 531–539.

Boissonas, R. A., Guttman, S., & Pless, J. (1966) *Experientia 22*, 526.

Covey, T. R., Bonner, R. F., Shushan, B. I., & Henion, J. (1988) *Rapid Commun. Mass Spectrom. 2*, 249.

Cowie, D. B., Cohen, G. N., Bolton, E. T., & Robi-Chon-Szulmajster, D. (1959) *Biochim. Biophys. Acta 34*, 39–36.

Ericsson, L. H., Atherton, D., Kutny, R., Smith, A. J., & Crabb, J. W. (1991) in *Methods in Protein Sequence Analysis* (Jornvall, H., Hoog, J.-O., & Gustavsson, M., Eds.) pp 1–9, Birkhauser Verlag, Basel, Switzerland.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989) *Science 246*, 64–71.

Ferrige, A. G., Seddon, M. J., Green, B. N., Jarvis, S. A., & Skilling, J. (1992) *Rapid Commun. Mass Spectrom. 6*, 707–711.

Forsberg, J., Palm, G., Ekebacke, A., Josephson, S., & Hartmanis, M. (1990) *Biochem. J. 271*, 357–363.

Gilles, A.-M., Marliere, P., Rose, T., Sarfati, R., Longin, R., Meier, A., Fermamdjian, S., Monnot, M., Cohen, G. N., & Barzu, O. (1988) *J. Biol. Chem. 263*, 8204–8209.

Halenbeck, R., Kawasaki, E., Wrin, J., & Koths, K. (1989) *Bio/Technology 7*, 710–715.

Henry, K. D., & MacLafferty, F. W. (1991) *J. Am. Chem. Soc. 113*, 5447–5449.

Kerwar, S. S., & Weissback, H. (1970) *Arch. Biochem. Biophys. 141*, 525–532.

Naider, F., Bohak, Z., & Yariv, J. (1972) *Biochemistry 11*, 3202–3208.

Niece, R. L., Ericsson, L. H., Fowler, A. V., Smith, A. J., Speicher, D. W., Crabb, J. W., & Williams, K. R. (1991) in *Methods in Protein Sequence Analysis* (Jornvall, H., Hoog, J.-O., & Gustavsson, M., Eds.) pp 1–9, Birkhauser Verlag, Basel, Switzerland.

Pandit, J., Bohm, A., Jancarik, J., Halenbeck, R., Koths, K., & Kim, S.-H. (1992) *Science 258*, 1358–1362.

Piez, K. A. (1968) Cross-linking of collagen and elastin. *Annu. Rev. Biochem. 37*, 547.

Randhawa, Z. I., Brown., D., Cone, J., Wilkins, J., Witkowska, H. E., Yamanishi, K., Takahashi, M., & Green, B. N. (1991) Structural Charcaterization of the Amino and Carboxy Terminal Heterogeneity in Recombinant Human M-CSF by Peptide Mapping Analysis and Electrospray Ionization Mass Spectrometry. 2nd Interlaken Conference on Advances in Purification of Recombinant Proteins, Interlaken, Switzerland, March 19–22, 1991.

Randhawa, Z. I., Brown, D., Yasuda, S., Yamaninshi, K., Hughes, P., Arthur, P., Green, B., & Witkowska, H. E. (1992) L'incorporation de la norleucine aux positions normalment occuppees par la methionine dans les monochaines de macrophage colony stimulating factor produites par l'E.coli par l'analyse du contenu d'acide amines, la sequence N-terminale, ainsi que l'electrospray ionization mass spectrometry. 5eme Symposium: Technologies D'etude et de purification des proteines, Grenoble, France, March 17–19, 1992.

Rocchi, R., Scatturin, A., Moroder,, L., Marchiori, F., Tamburro, A. M., & Scoffone, E. (1969) *J. Am. Chem. Soc. 91*, 492.

Savige, W. E., & Fontana, A. (1977) in *Methods in Enzymology*: *Interconversion of methionine and methionine sulfoxide*, pp 453–459, Academic Press, Inc., San Diego, CA.

Schecter, Y., Burstein, Y., & Patcornik, A. (1975) *Biochemistry 14*, 4497–4503.

Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J., & Udseth, H. R. (1990) *Anal. Chem. 62*, 882–899.

Stanley, E. R., Guilbert, L. J., Tushinski, R. J., & Bartelmez, S. H. (1983) *J. Cell Biochem. 21*, 151–159.

Trupin, J., Dickerman, H., Nirenberg, M., & Weissback, H. (1966) *Biochem. Biophys. Res. Commun. 24*, 50–55.

Tsai, L. B., Lu, H. S., Kenney, W. C., Curless, C. C., Klein, M. L., Lai, P.-M., Fenton, M., Altrock, B. W., & Mann, M. B. (1988) *Biochem. Biophys. Res. Commun. 156*, 733–739.

Van Dorsselaer, A., Bitsch, F., Green, B., Jarvis, S., Lepage, P., Bischoff, R., Kolbe, H. V. J., & Roitsch, C. (1990) *Biomed. Environ. Mass Spectrom. 19*, 692–704.

Violand, B. N., Siegel, N. R., Bogosian, G., Workman, W. E., & Kane, J. F. (1989) in *Techniques in Protein Chemistry* (Hugli, T. E., Ed.) pp 315–326, Academic Press, San Diego, CA.

Wilkins, J. A., Cone, J., Randhawa, Z. I., Wood, D., Warren, K., & Witkowska, H. E. (1993) *Protein Sci. 2*, 244–254.

Witkowska, H. E., Lubin, B. H., Beuzard, Y., Baruchel, S., Esseltine, D. W., Vichinsky, E. P., Kleman, K. M., Bardakdjian-Michau, J., Pinkoski, L., Cahn, S., Roitman, E., Green, B. N., Falick, A. M., & Shackleton, C. H. L. (1991) Sickle cell disease in a patient with sickle cell trait and compound heterozygosity for hemoglobin S and hemoglobin Quebec-Chori. *New Engl. J. Med. 325*, 1150–1154.

Yamanishi, K., Takahasi, M., Nishida, T., Ohmoto, Y., Takano, M., Nakai, S., & Hirai, Y. (1991) *J. Biochem. 109*, 404–409.

Yamanishi, K., Randhawa, Z. I., Brown, D., Takahashi, M., Masui, Y., Yasuda., S., & Adachi, M. (1993) *J. Biochem. 113*, 81–87.