Analytical Biochemistry **267,** 125–134 (1999)
Article ID abio.1998.3003, available online at http://www.idealibrary.com on IDEAL®

Peptide Mapping and Disulfide Bond Analysis of Myeloid Progenitor Inhibitory Chemokine and Keratinocyte Growth Factor by Matrix-Assisted Laser Desorption Ionization Mass Spectrometry

Vivek Navale,* Parveen Kaushal,† Susan Hunt,† Ileana Burducea,† Reiner Gentz,† Fazal Khan,† and Akos Vertes*,[1]

**Department of Chemistry, George Washington University, Washington, DC 20052; and †Human Genome Sciences Inc., 9410 Key West Avenue, Rockville, Maryland 20850*

Received July 21, 1998

Peptide mapping and matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) were conducted to characterize the human genome-based recombinant proteins. Accurate mass values for the deleted forms of the myeloid progenitor inhibitory factor chemokine (MPIF-1d23), and the keratinocyte growth factor (KGF-2d33) were measured as 8848.55 ± 0.25 and 16,175.87 ± 0.89 Da, respectively. The mass accuracy of delayed ion extraction MALDI-MS measurements was within 20 ppm of the cDNA predicted value. Reduction and alkylation of the chemokine showed the presence of six cysteine residues and three disulfide bonds. Additional confirmation of disulfide bonding among the cysteine residues of the chemokine was demonstrated by identifying the RP-HPLC separated tryptic and endoprotease Glu-C peptides. Three methionine residues of the chemokine were identified by MALDI-MS of its cyanogen bromide (CNBr) cleavage products. The KGF-2d33 growth factor, however, lacked the disulfide bonding between the two-cysteine residues and contained two free sulfhydryl groups. Direct analysis of the growth factor CNBr digest showed 7542.99, 4993.4, and 3107.7 Da peptides, which identified the methionine residues. Peptide mapping mass spectrometry indicated that host-specific posttranslational modifications had not influenced the gene expression work. © 1999 Academic Press

Chemokines, also known as intercrine cytokines are a special group of proteins that belong to a large class of intracellular cytokines. These molecules have an unique ability to elicit chemotactic migration of distinct white blood cells (monocytes, neutrophils, lymphocytes, basophils, and fibroblasts) to active sites of infection (1). The intercrine cytokines are also involved in histamine and leukotriene release, lysosomal enzyme activity, and increased adherence of target immune cells to endothelial cells (2). Although many chemokines may be involved during an inflammatory reaction, certain of these molecules can exhibit specific activities. For example, the macrophage inflammatory protein (MIP-1)[2] is able to suppress hematopoietic stem proliferation, platelet factor-4 is a potent inhibitor of endothelial cell growth. Interleukin-8 promotes proliferation of keratinocytes, and GRO is an autocrine growth factor for melanoma cells.

Structurally the polypeptides of the chemokines have a certain degree of homology at the amino acid level. On the basis of four conserved cysteine residues, these molecules can be differentiated into two subfamilies: the α and β family. In the α subfamily, the cysteine (C) residues are separated by one amino acid (X) and are designated as C–X–C, whereas in the β subfamily, the Cs are proximal to each other (C–C) (3). Recently, a new class of chemokines with the CX_3C motif has been identified from certain non-hematopoi-

[1] To whom correspondence should be addressed. E-mail: vertes@gwu.edu.

[2] Abbreviations used: MIP-1, macrophage inflammatory protein; KGF, keratinocyte growth factor; MPIF, myeloid progenitor inhibitory factor; hMIP-1β, human MIP-1β; FAB, fast atom bombardment; MALDI, matrix-assisted laser desorption ionization; ESI, electrospray ionization; TOF, time-of-flight; DE, delayed extraction; SA, sinapinic acid; DHB, 2,5-dihydroxybenzoic acid; CHCA, α-cyano-4-hydroxycinnamic acid; ACN, acetonitrile; TFA, trifluoroacetic acid; DTT, dithiothreitol; IAA, iodoacetic acid.

0003-2697/99 $30.00

etic cells (4). The molecular mass of chemokines varies from 7 to 15 kDa.

Three chemokines, Rantes, MIP-1α, and MIP-1β, have been reported to act as effective HIV-suppressive factors, preventing the replication of the AIDS virus in infected humans (5). Elevated levels of chemokines are also associated with acute rheumatoid arthritis conditions, chronic pulmonary, and cardiovascular diseases (6). Therefore, the development of specific chemokine inhibitors capable of blocking excessive inflammation is important for prevention and remedy of diseases in humans.

The nonendocrinal growth factors are another group of polypeptides released by a specific cell type that act upon and influence the proliferation of other cell types. The keratinocyte growth factor (KGF) is found to be specific to epithelial cells, identified in a human embryonic lung fibroblast cell line, and consists of a single polypeptide chain of approximately 28 kDa (7). The identification of growth factors derived from epithelial tissues is important for understanding the vast majority of human malignancies related to epithelial cells (8,9). Thus, growth factors and their receptor-mediated signal transduction pathways can provide insights into mechanisms of both normal and malignant cell growth.

The present study focuses on the chemokine myeloid progenitor inhibitory factor (MPIF-1) that was derived from a human aortic endothelium tissue (10). The mature recombinant MPIF-1 protein consists of 99 amino acids and bears significant homology to the human MIP-1α, showing 51% identity. It displays chemotactic activity on resting T lymphocytes and monocytes, a minimal but significant activity on neutrophils, and is negative on activated T lymphocytes. The inhibitory activity studies suggest that MPIF-1 may have the potential to protect the hematopoietic progenitors from the cytotoxic effects of the antiblastic drugs used during cancer therapy (11).

The other polypeptide investigated in this work was obtained from a randomly primed human fetal lung cDNA library of the fibroblast growth factor, the KGF-2 (12a, 12b). *In vivo* studies of KGF-2 have demonstrated efficacy in wound healing. *In vitro* the KGF-2 can cause proliferation of keratinocytes. The isolated KGF-2 polypeptide contains 208 amino acids and shares significant homology (69% similar and 48.5% identical) to KGF-1 (13).

The NMR analysis of human macrophage inflammatory protein-1β (hMIP-1β) has elucidated the three dimensional structure of the molecule (14). However, only relative mass values (in the range of 16 kDa) were reported for the protein. Within a limited mass range (<8800 Da), FAB-MS has been shown to be a sensitive method for molecular mass measurement and disulfide bond analysis of peptides (15). Other methods such as the matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) and electrospray ionization mass spectrometry (ESI-MS) have extended the accessible mass range, with capabilities to identify large proteins and their complexes (>100 kDa) (16). MALDI-MS has also been shown to verify the identity and homogeneity of genetically engineered products (17), to determine disulfide bond assignment in proteins (18), and identify co- and posttranslational modifications (19, 20). By combining enzymatic proteolysis with the MALDI-MS method, conformational changes in proteins have also been identified (21).

The advantages of the MALDI-MS over FAB-MS include its capacity to tolerate salts and buffers present in biological samples and to detect proteins and peptides at subpicomolar concentrations. The mass accuracy of continuous mode MALDI with a linear time-of-flight mass spectrometer (TOF-MS) can vary from 0.01 to 0.5% (22). Recently, significant improvements in resolution and mass accuracy of MALDI-MS systems have been shown by application of the delayed ion extraction (DE) method (23–25). Also, in-source fragmentation of protein ions for protein sequencing has been shown by DE–MALDI-MS (26).

With advancements in human genome research, and the advent of "proteomics," many novel proteins have been produced. An important part of these experiments is to verify the fidelity of gene expression in the host systems and to be able to identify modifications during the protein expression work. The possible errors in the gene translation can be manifested by the formation of proteins with different amino acid sequence than predicted by the cDNA. In all of these experiments the nature of disulfide bonding between the cysteine residues is usually unknown. The disulfide bonds play an important role in the folding and the stabilization of three-dimensional structure of proteins. Molecules with incorrect disulfide linkages may have much lower activity than the desired product. Therefore, the characterization of disulfide linkages in the recombinant DNA products is essential. To date, there has been no molecular analysis of MPIF-1 chemokine and the KGF-2 recombinant proteins. In the present study we demonstrate the application of MALDI-MS for determination of disulfide bonds and free sulfhydryl groups in related human genome-based recombinant proteins. The work focuses on developing an online enzymatic and reversed-phase high-performance liquid chromatography (RP-HPLC) method with MALDI-MS for the analysis of recombinant proteins. Also, a comparative analysis of mass accuracy by continuous and DE-MALDI-MS was conducted for peptide mapping of the molecules.

MATERIALS AND METHODS

Matrix-Assisted Laser Desorption Ionization Experiments

Instrumentation. The configuration of the continuous mode MALDI with a linear TOF-MS system was described previously (27). A nitrogen laser (VSL-337 ND, Laser Science Inc., Newton, MA), emitting at the wavelength of 337 nm, was attenuated and focused onto a stainless-steel probe tip, which was placed in the ion source of the mass spectrometer. Subsequent to the laser desorption and ionization process, the positive ions were accelerated into the 210-cm flight tube by 30 kV accelerating voltage. Typically, spectra generated by 25 to 50 laser shots were averaged to improve the signal-to-noise ratio.

In contrast to the continuous ionization mode (where ions formed at high electric field are immediately accelerated into the field free region), ions in the delayed extraction mode are initially subjected to the same potential between the repeller (sample plate) and the first grid. After a delay of few hundred nanoseconds, a higher potential is applied to the repeller, resulting in a pulsed voltage. This method of DE can compensate for the initial kinetic energy spread of the ions in the ion source of the TOF-MS. By applying suitable pulse voltage and delay time, both the resolution and the mass accuracy of TOF-MS systems can be improved.

DE–MALDI–TOF-MS experiments were conducted by using the Perseptive Biosystems instrument (Voyager RPS, Perseptive Biosystems, Framingham, MA). The instrument can be operated either in the continuous or the DE mode. In the present work, the DE mode was used by initially maintaining the sample plate and the variable-voltage grid at 21 kV. After a delay time of 250 ns, the sample plate voltage was increased to 25 kV, and the ions were accelerated into the flight tube that had ion path length of 1.3 m. The ions were detected with a dual microchannel plate, the signal was recorded by a 500 MHZ digitizer board, and the data were acquired by the standard Voyager RPS software.

Sample preparation. The polynucleotides for the MPIF-1d23 chemokine and the KGF-2d33 growth factor expressed in an *Escherichia coli* host system and the recombinant proteins were provided by Human Genome Sciences Inc. (HGS, Rockville, MD). The procedure for protein expression and purification has been described previously (13). In light of the biological efficacy of the molecules under investigation, the leader peptide (21 amino acids) and an additional 23 amino acids were deleted from the full-length MPIF-1, to form MPIF-1d23. Similarly, the leader peptide (35 amino acids) and additional 33 amino acids were removed from the KGF-2, to form the KGF-2d33.

Three MALDI matrices, sinapinic acid (SA), 2,5-dihydroxybenzoic acid (DHB), α-cyano-4-hydroxycinnamic acid (CHCA) were purchased from Sigma (Sigma Inc., St. Louis, MO), and saturated solutions were prepared in a 7:3 v/v mixture of acetonitrile (ACN) and water. It was observed that for the recombinant proteins studied in this work, the MALDI-MS response with the CHCA matrix was superior to the signal obtained with the SA or DHB matrices. Therefore, CHCA was selected as the MALDI matrix for the analysis of these proteins. The sample preparation was based on 10 μl of the native protein solution (1 mg/ml), diluted by 100 μl of 0.1% trifluoroacetic acid (TFA). Reduced and alkylated samples were diluted by only 5 μl of 0.1% TFA. Bovine insulin, lysozyme, and myoglobin (Sigma Inc.) were added to MPIF-1d23 and KGF-2d33, respectively, as internal calibration compounds. The protein solution containing 5 μl of the internal standard (at the concentration of 10^{-5} M) was mixed with 35 to 40 μl of CHCA solution. Approximately 10 μl of this mixture was deposited on the stainless-steel insertion probe. A slightly modified procedure was adopted during sample preparation for peptide mapping: 5 μl of the sample was directly deposited with 2 μl of angiotensin (average mass of 1297.51 Da) and ACTH (18–39 peptides with average mass of 2466.72 Da) internal standards (Sequazyme C-Peptide Sequencing kit, Perseptive Biosystems, Framingham, MA) and with 5 μl of CHCA (matrix diluent 50% ACN and 0.3% in TFA) onto the probe tip. Direct deposition of the sample on the probe tip may minimize possible peptide losses during pipetting, transfer, and storage, a factor that becomes significant during the analysis of subpicomolar quantities of peptides (28). All samples were air-dried for several minutes prior to introduction into the ion source of the TOF-MS.

Reduction and alkylation. The derivatization experiments were performed using the autosampler of the microanalytical system (Integral, Perseptive Biosystems, Framingham, MA) for rapid reduction and alkylation of the proteins. The method is as follows: 39 μl of 8 M urea, 50 mM Tris buffer, and 10 mM $CaCl_2$ solution was added at pH 8 to 47.5 μl of the protein solution (initial concentration of MPIF-1d23 was 2.2 mg/ml and of KGF-2d33 it was 1.00 mg/ml). The reagents were mixed with 50 μl of 4.5 mM dithiothreitol (DTT) and heated to 50°C for a period of 15 min. Sixty-five microliters of the reduced protein was then mixed with 13 μl of 100 mM iodoacetic acid (IAA). During the second set of experiments only alkylation was carried out for both proteins by replacing the DTT with an identical volume of deionized water. The reduced and alkylated protein samples were analyzed by MALDI-MS within an interval of a few hours.

For some of the peptide mapping experiments the polypeptides were chemically digested by cyanogen bromide treatment. Approximately 1 ml of the MPIF-1d23 (3.15 mg/ml) and KGF-2d33 (2.0 mg/ml) samples was mixed with 1 ml of 0.1 N hydrochloric acid. A small

(unweighed) crystal of cyanogen bromide was added to the sample vials. The mixture was vortexed and subsequently thermostated for 3 h at 37°C. The reaction was quenched by adding 3 ml deionized water and the products were kept frozen until MALDI-MS analysis was performed.

Peptide Preparation and Analysis

Enzymatic digestion procedure. An automated online enzymatic digestion procedure was developed for the proteins with corresponding RP-HPLC method for the separation of the resulting digestion products. The proteolytic digestion was carried out on immobilized trypsin and endoproteinase Glu-C (*Staphylococcus aureus*, V8) columns, where the enzymes were covalently linked to a hydrophilic polystyrene–divinyl benzene support (POROS media, Perseptive Biosystems, Framingham, MA). The immobilized trypsin and V8 cartridges (2.1 mm in diameter and 30 mm in length) were coupled to a C-18 RP-HPLC column (4.6 mm i.d., 250 mm length) packed with 5-μm particle size 300 Å pore octadecyl-coated silica particles (Perseptive Biosystems). A digestion buffer was prepared by dissolving 9 g of Tris (hydroxymethyl) aminomethane acetate in 800 ml of deionized water. The pH was adjusted to 7.8 for the V8 column, and 8.0 for the trypsin column by adding 2 M of solution of HCl or NaOH as required. This was followed by dilution to a total buffer volume of 1 L. The enzymatic columns were washed with the digestion buffer, and equilibrated at 37°C.

RP-HPLC separation of peptides. The RP-HPLC system (INTEGRAL, Perseptive Biosystems) consisted of two small displacement volume (10 μl) dual-piston HPLC pumps, a temperature-controlled buffer cabinet, robotic autosampler, automated column switching valves, a dual syringe pump, injection valves, a multiple column capacity subsystem, and a programmable UV/VIS detector.

A gradient-elution method was developed for the peptide-mapping experiments (29). The starting composition (in percentage) of the mobile phase was 95:5 v/v water and acetonitrile. During a 30-min analysis time the mobile-phase composition was changed to 30:70 (water:acetonitrile) at a flow rate of 1 ml/min. The injection volume was 10.0 μl and the concentration of the samples was in the range of 1.0 to 2 mg/ml. The spectra were obtained by means of a UV detector with the wavelength set at 215 nm. The separated peptides were collected by means of a fraction collector. To account for the dead volume between the detector and the fraction collector, a delay time was set that enabled collection of fractions from the apex of the RP-HPLC peak.

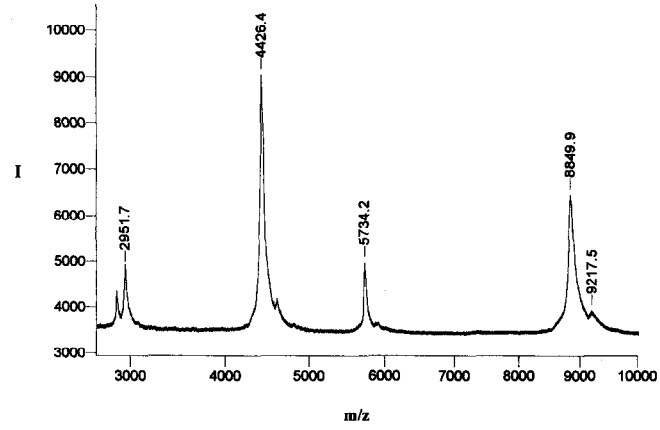

FIG. 1. MALDI-MS spectrum of MPIF-1d23 chemokine from CHCA matrix exhibits singly, doubly and triply charged ions at $[M + H]^+$ 8849.9, $[M + 2H]^{2+}$ 4426.4, and $[M + 3H]^{3+}$ 2951.7, respectively. Insulin is the internal standard with *m/z* 5734.2.

RESULTS AND DISCUSSION

Molecular mass measurements. The MALDI-MS (continuous mode) spectrum of MPIF-1d23 chemokine (concentration, 1 μM) from CHCA matrix is shown in Fig. 1. The predominant peaks correspond to the protonated molecules of the protein in singly, doubly, and triply charged forms at *m/z* 8849.1 ± 3.5, 4426.0 ± 1.7, and 2952.4 ± 1.2 Da, respectively. The insulin internal standard peaks are shown at *m/z* 5734.5 and 2868.7 Da. The molecular mass of MPIF-1d23 was determined to be 8848.1 ± 3.5 Da with a mass accuracy of +100 parts per million (ppm). Several measurements made by the DE–MALD-MS indicated the mass value to be 8848.55 ± 0.25 Da. The mass accuracy was within +18 ppm of the cDNA value. Also, compared to a resolution value of 500 and less by the continuous mode, the DE analysis resulted in a resolution of 1400.

At higher protein concentrations (0.1 mM), a minor component (adjacent to the $[M + H]^+$ peak) was also observed with a *m/z* of 9217.5. This feature may represent a component similar to the MPIF-1d23 protein with an additional mass of 367.6 Da. The spectrum also showed weak traces of the dimer $[2M + H]^+$ at *m/z* 17698, a likely indication of the association of protein molecules during the MALDI process.

The MALDI analysis (continuous mode) of the KGF-2d33 protein from CHCA matrix showed the singly, doubly, and triply charged species, identified as $[M + H]^+$ at 16179 ± 4, $[M + 2H]^{2+}$ at 8090.5 ± 2.8, and $[M + 3H]^{3+}$ at 5395.7 ± 1.4 Da, respectively. The internal calibration compound lysozyme with $[M + H]^+$ at 14,305, $[M + 2H]^{2+}$ at 7154.2, and $[M + 3H]^{3+}$ at 4771.2 Da resulted in the determination of the KGF-2d33 molecular mass as 16,178 ± 4 Da. At higher analyte concentrations several cluster ions of the protein were observed: $[3M + 2H]^{2+}$ at 24,334, $[2M + H]^+$ at 32,359, and $[3M + H]^+$ at 48,429. In all likelihood

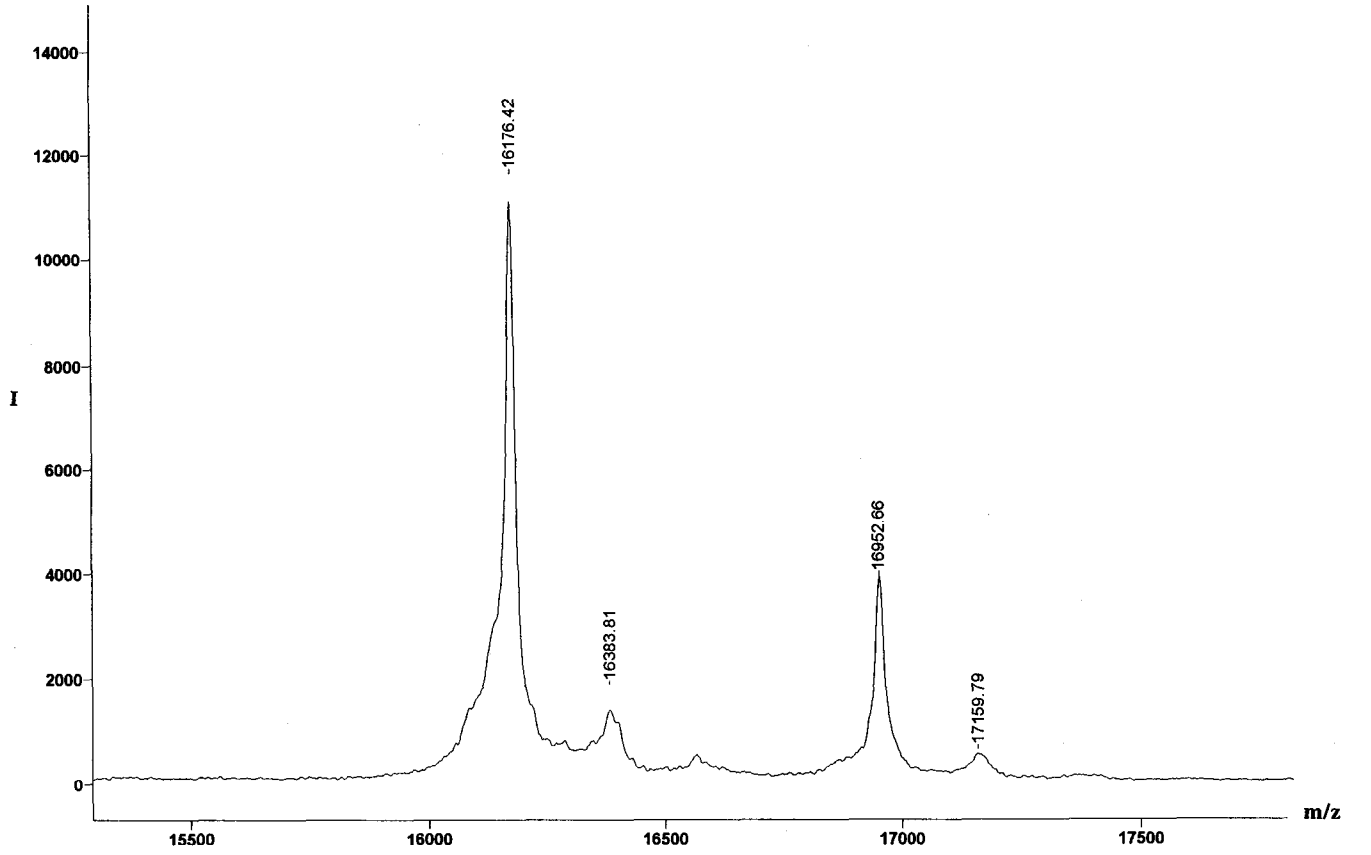

FIG. 2. DE–MALDI-MS spectrum of the KGF-2d33 protein in SA matrix, showing the $[M + H]^+$ at 16,176.42 and the matrix adduct at *m/z* 16,383.81. The internal calibration compound myoglobin is indicated by the $[M + H]^+$ at 16,952.66, and its matrix adduct is shown by *m/z* 17,159.79.

the observed clusters are produced via association during the MALDI process.

Figure 2 illustrates the DE-MALDI-MS spectrum of KGF-2d33 from SA matrix, indicating a mass value of 16,175.87 ± 0.89 Da. The protonated ion for the internal standard myoglobin is shown by *m/z* 16,952.56. Also present in the same spectrum are the SA matrix adducts of the KGF-2d33 and the myoglobin at *m/z* 16,383.82 and 17,159.79, respectively. The mass accuracy for KGF-2d33 by the DE mode was 13 ppm and the resolution was in the range of 1400. Compared to the DE values, mass accuracy and resolution by the continuous mode were less than 150 ppm and 100, respectively. Overall, one order of magnitude improvement in the mass accuracy and resolution was achieved by application of the DE–MALDI-MS method.

Disulfide bond determination. The conventional procedures for disulfide bond determination in proteins usually involve multiple steps. These procedures require that the protein be cleaved by enzymes or chemical reagents between half cystinyl residues. The cystinyl-containing peptides can be separated by chromatographic methods and sequenced. However, isolation and identification of the cysteine-containing fragments in complex mixtures can be tedious, time-consuming, and hindered by contamination. The present study shows a rapid MALDI-MS method for the quantitative analysis of disulfide bonds in the chemokine and the growth factor. This method required mass values of native protein (M_N), the reduced and alkylated species (M_{R+A}), and the alkylated derivative (M_A) of the protein molecule. The mass spectra for the reduced alkylated MPIF-1d23 protein showed the singly, doubly, and triply charged ions $[M + H]^+$ at 9212.1 Da, $[M + 2H]^{2+}$ at 4606.2 Da, and $[M + 3H]^{3+}$ at 3077.1 Da. The M_A values were $[M + H]^+$ at 8851.9 Da and $[M + 2H]^{2+}$ at 4433.3 Da. The M_{R+A} of the KGF-2 protein was calculated based on $[M + H]^+$ at 16,300 Da, $[M + 2H]^{2+}$ at 8150.1 Da, and $[M + 3H]^{3+}$ at 5436.9 Da. The M_{R+A} and M_A of the KGF-2d33 were found to be identical. Table 1 compares the MALDI-MS measurements for M_N, M_{R+A}, and M_A for the chemokine and the growth factor obtained by continuous and delayed extraction MALDI-MS method. The cDNA predicted values for the two proteins are also shown. Because the exact nature of disulfide bonding from the cDNA sequence was not known, the theoretical mass values in Table 1 assume that cysteine residues are bridged together.

TABLE 1

Determination of Number of Cysteines, Free Sulfhydryl Groups, Disulfide Bonds in MPIF-1d23, and KGF-2d33 Proteins by Continuous and DE–MALDI-MS Measurements

Symbol	MPIF-1d23 (cDNA value)[a]	MPIF-1d23 (continuous MALDI-MS)	MPIF-1d23 (DE-mode MALDI-MS)	KGF-2d33 (cDNA predicted value)[a]	KGF-2d33 (MALDI-MS continuous)	KGF-2d33 (MALDI-MS DE-mode)
M_N	8848.4	8848.1 ± 3.5	8848.55 ± 0.25	16,175.6	16,178 ± 4	16,175.87 ± 0.89
M_{R+A}	N/A	9216.5 ± 4.2	9203.17 ± 0.31	N/A	16,298 ± 7	16,291.16 ± 0.25
M_A	N/A	8849.1 ± 2.5	8848.51 ± 0.80	N/A	16,302 ± 8	16,292.37 ± 0.21
N_{cys}	6	6	6	2	2	2
N_{S-H}	N/A	0	0	N/A	2	2
N_{S-S}	N/A	3	3	N/A	0	0

Note. The mass values predicted from the cDNA sequence of the proteins are shown for comparison. The number of cysteines (N_{cys}), number of free sulfhydryl groups (N_{S-H}), and number of disulfide bonds (N_{S-S}) were calculated using the formulas indicated below the table.

[a] The theoretical mass values calculated from the cDNA sequence assumes that disulfide bonds are present between the cysteine residues.

$$N_{cys} = \frac{M_{R+A} - M_N}{59}, \quad N_{S-H} = \frac{M_A - M_N}{(59 - 1)}, \quad N_{S-S} = \frac{N_{cys} - N_{S-H}}{2}$$

CH_3COO^- = 59 (from IAA).

The difference between M_{R+A} and the mass of the intact protein, M_N, divided by the mass of the alkylating agent (in IAA 59 for CH_3COO^-) provided the total number of cysteines. Similarly the mass difference between the M_A and M_N divided by the mass of alkylating agent less the mass of hydrogen (59 − 1) provided the number of the free sulfhydryl groups. The number of the disulfide bonds was determined by the difference between the number of cysteine residues and the number of the free sulfhydryl groups divided by 2. MALDI-MS measurements verified the presence of six cysteines in the MPIF-1d23 molecule and reported three disulfide bonds. Similar measurements on the KGF-2d33 revealed the presence of two free sulfhydryl groups and clearly showed the absence of disulfide linkage between them.

Characterization of peptides. Table 2 provides the cDNA predicted amino acid sequence for the MPIF-1d23 and KGF-2d33 polypeptides. To characterize the proteins expressed in *E. coli*, peptide mapping was performed using chemical and enzymatic digestion procedures in combination with MALDI-MS.

Identification of the MPIF-1d23 peptide (mass 6913.4 Da) formed as a result of cyanogen bromide reaction can be related to the amino acid sequence DRFHATSADCCISYTPRSIPCSLLESYFETNSECSKPGVIFLTKKGRRFCANPSDKQVQVCM. Also, observed in the same MALDI spectrum was a 7044.15-Da peptide that resulted from a missed cleavage between the C-terminus of the first methionine (M) residue and N-terminus of the first aspartic (D) residue (shown in the sequence above, also see Table 2A). The 1485.63-Da peptide was assigned the sequence LKLDTRIKTRKN. The smallest CNBr decomposition product of MPIF-1d23 can have a mass 257.4 Da. This fragment was not identified because the CHCA matrix signal was predominant in the mass region of the spectrum, which may have contained the peptide. The mass accuracy for the CNBr products by continuous mode MALDI-MS was several thousand parts per million, whereas the DE–MALDI-MS provided mass accuracy within 10 to 100 ppm (see Table 3).

The enzymatic action of trypsin on the C-terminal of arginine and lysine can result in the formation of 15 peptides. Table 2B lists the theoretical sequence of the cleavage products, generated with the aid of a software package (Peptide Companion, Window Chem. Software Inc., Fairfield, CA). To minimize the total number of peptide fragments produced as a result tryptic digestion, the MPIF-1d23 molecules were analyzed without the cleavage of the disulfide bonds. This approach enabled us to identify peptides containing the disulfide bridging among the six cysteine residues.

Figure 3 illustrates the RP-HPLC separated tryptic map of the chemokine. Analysis of fractions indicated as B, C, D, and E on the peptide map showed the presence of several disulfide-bonded peptides. The fraction B contained three peptides with *m/z* values 6405.52, 6535.08, and 6941.7. The amino acid sequence for the 6404.52-Da peptide is shown in Table 3. The 6534.4-Da peptide represents a missed cleavage between the two K residues of the SIPCSLLESYFETNSECSKPGVIFLTKK fragment. The mass difference between the 6534.4- and the 6944.7-Da peptide may be accounted by the three amino acid residues, MDR (resulting from incomplete digestion between the C-terminal of R and N-terminal of the F residue).

MALDI-MS analysis of the fraction labeled C (shown

TABLE 2

The cDNA Predicted Amino Acid Sequence of MPIF-1d23 and KGF-2d33

MPIF-1d23

MDRFHATSADCCISYTPRSIPCSLLESYFETNSECSKPGVIFLTKKGRRFCANPSDKQVQVCMRMLKLDTRIKTRKN

A. Cyanogen bromide digest products
M, DRFHATSADCCISYTPRSIPCSLLESYFETNSECSKPGIFLTKKGRRFCANPSDKQVQVCM, RM, LKLDTRIKTRKN

B. Trypsin digest products
MDR, FHAT, SIPCSLLESYFETNSECSK, PGVIFLTK, K, GR, R, FCANPSDK, QVQVCMR, MLK, LDTR, IK, TR, K, N.

C. V8 digestion products
MD, RFHATSAD, CCISYTPRSIPCSLLE, SYFE, TNSE, CSKPGVIFLTKKGRRFCANPSD, KQVQCMRMLKLD, TRIKTRN

KGF-2d33

SYNHLQGDVRWRKLFSFTKYFLKIEKNGKVSGTKKENCPYSILEITSVEIGVVAVKAINSNYYLAMNKKGKLYGSK-
EFNNDCKLKERIEENGYNTYASFNWQHNGRQMYVALNGKGAPRRGQKTRRKNTSAHFLPMVVHS

D. Cyanogen bromide digest products
SYNHLQGDVRWRKLFSFTKYFLKIEKNGKVSGTKKENCPYSILEITSVEIGVVAVKAINSNYYLAM,
NKKGKLYGSKEFNNDCKLERIEENGYNTYASFNWQHNGRQM, YVALNGKGAPRRGQKTRRKNTSAHFLPM, VVHS

E. V8 digestion products
SYNHLQGD, VRWRKLFSFTKYFLKIE, KNGKVSGTKKE, NCPYSILE, ITSVE,
IGVVAVKAINSNYYLAMNKKGKLYGSKE, FNND, CKLKE, RIE, E,
NGYNTYASFNWQHNGRQMYALNGKGAPRRGQKTRRKNTSAHFLPMVVHS

Note. Also, indicated are the peptides resulting from cyanogen bromide, trypsin and V8 enzymatic digestion.

in Fig. 3) resulted in the identification of three peptides with *m/z* 7248.4, 7704.3, and 8115.2 Da. The amino acid sequence for the 7703.34-Da peptide is shown in Table 3. The 7247.4-Da peptide may have resulted from cleavage between the C-terminus of the R and the N-terminus of the M residue. However, the lack of cleavage between the C-terminus of the R and the N-terminus of the F residue can result in a peptide mass of 8114.2 Da (refer Table 2).

The fraction D on the map consisted of peptides with *m/z* values 6559.08, 6687.57, and 6670.9 Da. The amino acid sequence for 6558.08 and 6686.57 peptides is shown in Table 3. The two peptides differ by a K residue, which may have resulted from the missed cleavage between the two K residues shown in the fragment RFCANPSDKK (see Table 3 for the complete sequence).

The fraction E, labeled on the peptide map, contained 6538.6, 6738.7, and 6941.6 Da peptides. The two fraction E peptides (mass 6538 and 6941 Da) appear to possess identical mass identical to the fraction B peptides (mass values 6535 and 6944 Da). It is possible that these two of pairs of peptides are isoforms of each other. Table 3 provides accurate mass measurements of the trypsin-digested peptides that confirm the disulfide linkage between the cysteine residues of the MPIF-1d23 molecules. The specific assignment of disulfide bonding between the cysteine residues was not possible because two cysteine residues were adjoining each other.

The cleavage at the C-terminal of aspartic and glutamic acid residues of the MPIF-1d23 was brought about by the endoproteinase Glu-C (V8) enzyme. The V8 enzymatic digestion of the chemokine can possibly result in eight different peptides if the disulfide linkages are not present (theoretical sequences are shown in Table 2C). Similar to the trypsin digestion experiment, the V8 digestion was not preceded by the reduction experiment. This approach provided a means to verify the disulfide-bonded tryptic peptides. Analysis of the V8 peptides showed the presence of peptides at *m/z* 6929.83, 6682.52, and 6229.25. Table 3 provides the amino acid sequence for the V8 peptides that correspond to masses 6928.83, 6681.52, and 6229.25 Da. The 6928.83 and 6681.52 peptides differ by MD, representing missed cleavage between the C-terminal of D and N-terminal of the R residues (see Table 3). The sequence for the 6229.25 is also shown in Table 3. Analysis of the tryptic and the V8 peptides by DE–MALDI-MS showed that mass accuracy by the DE mode was consistently in the range of 100 to 200 ppm, whereas for the continuous mode it varied from several hundred to 3000 ppm.

In the absence of disulfide bonds, the V8 digestion of KGF-2d33 can result in the formation of eleven peptides (shown in Table 2E). The possible amino acid sequence for the 3060.65-, 1175.36-, 1167-, and 1110.26-Da peptides that were identified by MALDI-MS is shown in Table 3. The accuracy of mass determination for these peptides obtained by the DE mode was in the range of 10 to 100 ppm. Direct MALDI-MS of the V8 fraction indicated the presence of the largest 5748.52-Da peptide (possible sequence shown Table 2E). MALDI-MS analysis of the CNBr products showed the presence of the 7542.99-Da peptide. The 4993.4-Da peptide possibly represents the fragment between the two methionines (numbers 66 and 108), shown by the NKKGKLYGSKEFNNDCK-

TABLE 3

Digestion Products of MPIF-1d23 and KGF-2d33 Proteins Indicating the Possible Amino Acid Sequence of the Peptides

Amino acid sequence	cDNA[a]-based value (Da)	Continuous MALDI-MS and mass accuracy Da (ppm)	DE–MALDI-MS and mass accuracy Da (ppm)
MPIF-1d23 cleavage products			
FHATSADCCISYTPRSIPCSLLESYFETNSECSKPGVIFLTKKGRRFCANPSDKQVQVCMRMLKLDTR	7,704.9[a]	7,702.5 (−311)	7,703.34 (−202)
MDRFHATSADCCISYTPRSIPCSLLESYFETNSECSKPGVIFLTKKGRRFCANPSDKQVQVCM	7,045.2[c]	7,050.0 (+600)	7,044.15 (−149)
DRFHATSADCCISYTPRSIPCSLLESYFETNSECSKPGVIFLTKKGRRFCANPSDKQVQVCM	6,914.0[c]	6,908.2 (−800)	6,913.14 (−124)
FHATSADCCISYTPR SIPCSLLESYFETNSECSKPGVIFLTK RFCANPSDK QVQVCMR	6,559.5[a]	6,557.6 (−289)	6,558.67 (−216)
FHATSADCCISYTPR SIPCSLLESYFETNSECSKPGVIFLTKK RFCANPSDK QVQVCMR	6,687.73[a]	6,669.9 (−2666)	6,686.57 (−173)
FHATSADCCISYTPR SIPCSLLESYFETNSECSKPGVIFLTK FCANPSDK QVQVCMR	6,403.37[a]	6,391.6 (−1838)	6,404.52 (+179)
MDRFHATSADCCISYTPRSIPCSLLE CSKPGVIFLTKKGRRFCANPSD KQVQVCMRMLKLD	6,928.29[b]	6,935.01 (+969)	6,928.83 (+77)
RFHATSADCCISYTPRSIPCSLLE CSKPGVIFLTKKGRRFCANPSD KQVQVCMRMLKLD	6,681.03[b]	6,683.29 (+339)	6,681.52 (+73)
CCISYTPRSIPCSLLE TNSECSKPGVIFLTKKGRRFCANPSD KQVQVCMRMLKLD	6,227.45[b]	6,223.38 (−586)	6,228.25 (+195)
LKLDTRIKTRKN	1,485.81[c]	1,492.0 (+4000)	1,485.63 (−121)
KGF-2d33 cleavage products			
SYNHLQGDVRWRKLFSFTKYFLKIEKNGKVSGTKKENCPYSILEITSVEIGVVAVKAINSNYYLA-MNKKGKLYGSKEFNNDCKLKERIEENGYNTYASFNWQHNGRQM	12,565.5[c]	12,568.4 (+72)	12,563.79 (−295)
IGVVAVKAINSNYYLAMNKKGKLYGSKE	3,059.65[b]	3,065.0 (+1748)	3,060.68 (+336)
KNGKVSGTKKE	1,175.36[b]	1,175.8 (+374)	1,175.35 (−8.5)
FNNDCKLKERIEE	1,636.84[b]	1,639.9 (+1258)	1,637.58 (−158)
FNNDCKLKE	1,109.26[b]	1,111.1 (+756)	1,110.56 (+270)
NCPYSILE	938.07[b]	937.53 (−576)	938.48 (+437)
CKLKERIEE	1,147.36[b]	1,145.98 (−1203)	1,147.69 (+287)
SYNHLQGDVRWRKLFSFTKYFLKIEKNGKVSGTKKENCPYSILEITSVEIGVVAVKAINSNYYLAM	7,542.78[c]	7,567.6 (+3157)	7,542.99 (−104)
NKKGKLYGSKEFNNDCKLKERIEENGYNTYASFNWQHNGRQM	4,992.51[c]	4,993.4 (−22)	4,992.30 (−242)
YVALNGKGAPRRGQKTRRKNTSAHFLPM	3,107.61[c]	3,118.6 (+3536)	3,107.51 (−32)

Note. The experimentally determined mass values (Da) by MALDI-MS were compared with the theoretical (cDNA sequence predicted) mass for the peptides (cleavage method a, trypsin, b, *S. aureus* V8, c, cyanogen bromide). The mass accuracy is expressed in ppm and indicated in parentheses.

[a] The calculated theoretical value (from the cDNA predicted sequence) shown in the table assumes that cysteine residues are linked by disulfide bonds.

LKERIEENGYNTYASFNWQHNGRQM portion of the sequence. Also observed in the same spectrum was the 12,568-Da peptide, which represented the undigested portion of the sequence (i.e., when cleavage between the methionine (M) and the asparagine (N) residues had not occurred). The peptide with the amino acid sequence: YUALNGKGAPRRGQKTRRKNTSAHFLPM accounted for the mass 3107.51 Da. The smallest possible peptide fragment (see Table 2D) resulting from CNBr with mass 392.5 Da was masked by the matrix signal in the MALDI-MS spectrum. A comparative analysis of cDNA predicted mass values with the measured values by both continuous and DE–MALDI-MS is shown in Table 3. The results indicate that significant improvement in mass accuracy occurs when analyzing smaller peptides by DE–MALDI-MS. For example, the mass accuracy of 3107.7 Da by the DE–MALDI-MS was 32 ppm, whereas in the continuous MALDI-MS mode it was 300 ppm. However, for the larger CNBr peptides (e.g., 12,567.5 Da) the mass ac-

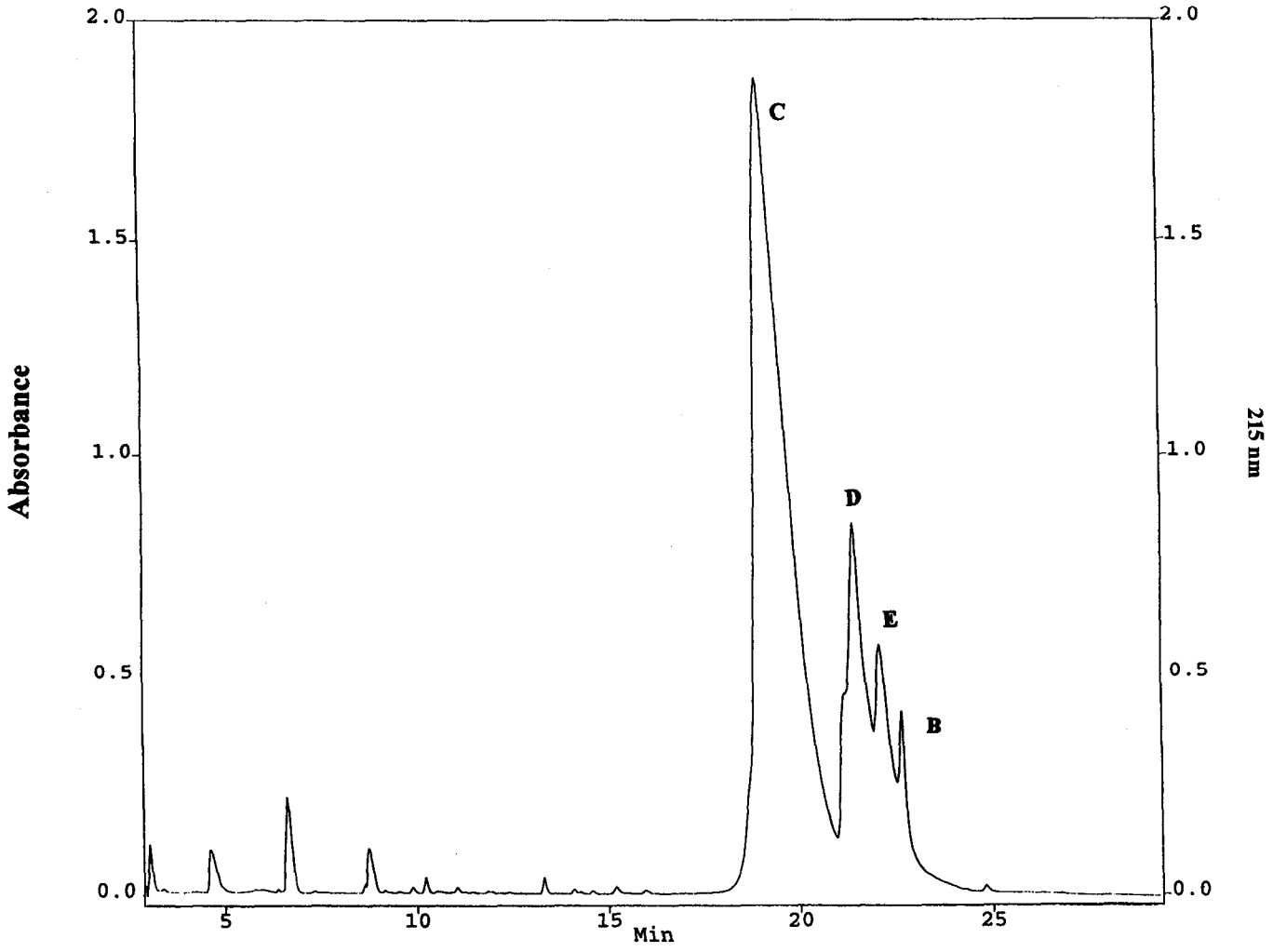

FIG. 3. Tryptic digest of MPIF-1d23 separated by RP-HPLC. The fractions labeled B to E were mass analyzed by MALDI-MS (B, 6404.5, 6534.1, 6940.7; C, 7247.4, 7703.5, 8114.2; D, 6558.1, 6687.6, 6669.9; and E, 6538.6, 6738.7, 6941.6 Da peptides).

curacy provided by both continuous and DE–MALDI-MS were comparable.

In conclusion MALDI-MS provided quantitative analysis of disulfide bonds in MPIF-1d23 and free sulfhydryl groups in KGF-2d33 molecules. Also, an efficient RP-HPLC and off-line MALDI-MS method for peptide mapping of chemokine and growth factor proteins was developed. Unlike direct LC–MS analysis where salts and buffers present in the sample can effect the results, this method was found to be independent of compositional variability in samples requiring almost no sample purification. The targeting of specific peptide fractions for mass measurements and rapid identification of several disulfide-bonded cleavage products were two of the other advantages of this method. The study also showed that DE–MALDI-MS mode is effective in improving the accuracy of mass determination for peptides that have molecular mass less than 10,000 Da. Several partially digested peptide fragments in the RP-HPLC separated fractions of the chemokine and the growth factor were identified by DE–MALDI-MS with a mass accuracy of 10 to 100 ppm. These measurements enabled us to infer that certain posttranslational modifications (e.g., sulfation, phosphorylation, glycosylation, N-terminal acetylation, formylation, myristylation, and palmitylation) had not influenced the expression of the MPIF-1d23 and KGF-2d33 genes in the host system. It is likely that further analysis of the peptides by tandem mass spectrometry can provide insightful information on the primary structure of these molecules.

ACKNOWLEDGMENTS

The recombinant proteins provided with the material transfer agreement (MTA 303) between the George Washington University, Washington, DC, and the Human Genome Sciences Inc. (Rockville, MD) are acknowledged.

REFERENCES

1. Broxmeyer, H. E. (1995) *in* Human Cytokines: Their Role in Disease and Therapy (Aggarwal, B. B., and Puri, R. K., Eds.), pp. 27–36, Blackwell Science, Cambridge, MA.
2. Burd, P. R. (1995) *in* Human Cytokines: Their Role in Disease and Therapy (Aggarwal, B. B., and Puri, R. K., Eds.), pp. 87–100, Blackwell Science, Cambridge, MA.
3. Zack Howard, O. M., Ben-Baruch, A., and Oppenheim, J. J. (1996) *J. Trends BioTechnol.* **14**, 46–51.
4. Bazan, J. F., Bacon, K. B., Hardiman, G., Wang, W., Soo, K., Rossi, D., Greaves, D. R., Zlotnik, A., and Schall, T. J. (1997) *Nature* **385**, 640–644.
5. Cocchi, F., DeVico, A. L., Garrzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995) *Science* **270**, 1811–1815.
6. Litwin, M. S., Gamble, J. R., and Vadas, M. A. (1995) *in* Human Cytokines: Their Role in Disease and Therapy (Aggarwal, B. B., and Puri, R. K., Eds.), pp. 101–131, Blackwell Science, Cambridge, MA.
7. Finch. P. W. H. E., Rubin, J. S., Miki, T., Ron, D., and Aaronson, S. A. (1989) *Science* **245**, 752–755.
8. Rubin, J. S. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 802–806.
9. Chen, B. L., and Arakawa, T. (1996) *J. Pharmacol. Sci.* **85**, 419–422.
10. Human Genome Science Inc. (1997) U.S. Patent No. 5,504,003.
11. Patel, V. P., Kreider, B. L., Li, Y., Li, H., Leung, K., Salcedo, T., Nardelli, B., Pippalla, V., Gentz, S., Thotakura, R., Parmelee, D., Geintz, R., and Garotta, G. (1997) *J. Exp. Med.* **185**, 1163–1172.

12a. Human Genome Science Inc. (1996) U.S. Patent No. WO96/25422.

12b. Human Genome Science Inc. (1996) Project Report (HGS03400).

13. Huddleston, K., Cherry, S., Lucero, S., Chopra, A., Keleti, D., Unsworth, E., Gaulding, S., Rohrer, T., Gentz, R., and Coleman, T. (1997) in Proceedings of the Protein Expression and Production Meeting, San Diego, CA.
14. Lodi, P. J., Garrett, D. S., Kuszewski, J., Tsang, M. L. S., Weatherbee, J. A., Leonard, W. J., Gronenborn, A. G., and Clore, G. M. (1994) *Science* **263**, 1762–1767.
15. Glocker, M. O., Arbogast, B., and Deinzer, M. L. (1995) *J. Am. Soc. Mass Spectrom.* **6**, 638–643.
16. Perera, I., Kantartzoglou, S., and Dyer, P. E. (1996) *Int. J. Mass Spectrom. Ion Processes* **156**, 151–172.
17. Biemann, K. (1994) in Biological Mass Spectrometry: Present and Future (Matsuo, T., Caproli, R. M., Gross, M. L., and Seyama, Y., Eds.), pp. 275–297. Wiley, New York.
18. Hirayama, K., and Akashi, S. (1994) in Biological Mass Spectrometry: Present and Future (Matsuo, T., Caproli, R. M., Gross, M. L., and Seyama, Y., Eds.), pp. 299–312. Wiley, New York.
19. Martin, S. A., Vath, J. E., Wen, Y., and Scoble, H. A. (1994) in Biological Mass Spectrometry: Present and Future; (Matsuo, T., Caproli, R. M., Gross, M. L., and Seyama, Y., Eds.), pp. 313–330, Wiley, New York.
20. Burlingame, A. L., Boyd, R. K., and Gaskell, S. J. (1996) *Anal. Chem.* **68**, 599R–65R.
21. Yang, H. H., Xiaoling, C. L., Amft, M., and Grotemeyer, J. (1998) *Anal. Biochem.* **258**, 118–126

22. Kaufman, R. (1995) *J. Biotechnol.* **41**, 155–175.
23. Vestal, M. P., Juhasz, P., and Martin, S. A. (1995) *Rapid Commun. Mass Spectrom.* **9**, 1044–1050.
24. Juhasz, P., Roskey, M. T., Smirnov, I. P., Haff, L. A., Vestal, M. L., and Martin, S. A. (1996) *Anal. Chem.* **68**, 941–946.
25. Edmondson, R. D., and Russel, D. H. (1996) *J. Am. Soc. Mass Spectrom.* **7**, 995–1001.
26. Katta, V., Chow, D. T., and Rohde, M. F. (1998) *Anal. Chem.* **70**, 4410–4416
27. Olumee, Z., Sadeghi, M., Tang, X., and Vertes, A. (1995) *Rapid Commun. Mass Spectrom.* **9**, 744–752.
28. Tempst, P., Bromage, H. E., Casteels, P., Geromanos, S., Lui, M., Powell, M., and Nelson, R. W. (1996) in Mass Spectrometry in the Biological Sciences (Burlingame, A. L., and Carr, S., Eds.), pp. 105–133, Humana Press, Rahway, NJ.
29. Kaushal, P., Mahadev, N., Hunt, S., Hesselberg, M., Virjee, M., Parmelee, D., and Gentz, R. (1997) in Proceedings of the International Symposium on Proteins, Peptides and Polynucleotides (ISPPP), p. 18, Rockville, MD.