# Assignment of the Inter- and Intramolecular Disulfide Linkages in Recombinant Human Macrophage Colony Stimulating Factor Using Fast Atom Bombardment Mass Spectrometry†

Michael O. Glocker,‡ Brian Arbogast,‡ Jolanda Schreurs,§ and Max L. Deinzer*,‡

*Department of Agricultural Chemistry, Oregon State University, Corvallis, Oregon 97331, and Chiron Corporation, 4560 Horton Street, Emeryville, California 94608*

*Received August 7, 1992; Revised Manuscript Received November 6, 1992*

ABSTRACT: The disulfide bridges in recombinant human macrophage colony stimulating factor (rhM-CSF), a 49-kDa homodimeric protein, were assigned. The 18 cysteines in the dimer form three intermolecular and two sets of three intramolecular disulfide bonds. The intermolecular disulfide bridges hold the dimer together and form symmetric bonds in which Cys31 and Cys157/Cys159 from one monomer unit are linked to the corresponding cysteines of the second monomer. The intramolecular disulfide bonds are located between Cys7–Cys90, Cys48–Cys139, and Cys102–Cys146, respectively. The resistance of native M-CSF to proteolytic cleavage was overcome by an initial chemical cleavage reaction using BrCN. The close proximity of four cysteines (Cys139, Cys146, Cys157, and Cys159) results in a tight core complex that makes the protein undigestable for most proteases. Digestion using endoprotease Asp-N resulted in cleavage at Asp156 near the C-terminal end of this region, thereby opening the complex structure.

Macrophage colony stimulating factor (M-CSF)[1] is a cytokine of tremendous pharmaceutical interest (Aggarwal & Pocsik, 1992; Aukerman et al., 1991). M-CSF is a hematopoietic growth factor that stimulates the survival, proliferation, and differentiation of mononuclear phagocytes (Aukerman et al., 1991; Clark & Kamen, 1987; Stanley et al., 1983; Das & Stanley, 1982). Native M-CSF has been isolated from human urine and shown to be secreted as a homodimeric glycoprotein (Ralph et al., 1988; Stanley & Heard, 1977). Three human M-CSF cDNA clones, each encoding an M-CSF polypeptide of a different length ($\alpha$, 256 aa; $\beta$, 554 aa; and $\gamma$, 438 aa), have been isolated (Ceretti et al., 1988; Wong et al., 1987; Kawasaki et al., 1985). The $\alpha$, $\beta$, and $\tau$ polypeptides all end with a 23 amino acid hydrophobic region followed by 36 residues of a "cytoplasmic" tail.

The subject of our studies was recombinant human M-CSF, 218 amino acids in length (Figure 1), that was expressed from a truncated form of the longest cDNA clone (M-CSF $\beta$) (Ralph et al., 1986, 1988). To improve the homogeneity of the recombinant product, the first three amino acids of the mature native M-CSF sequence were deleted such that the N-terminal amino acid is serine. This material was recovered from *Escherichia coli* as insoluble inclusion bodies and refolded under renaturaturing conditions (Halenbeck et al., 1989). M-CSF is biologically active only as a correctly folded dimer

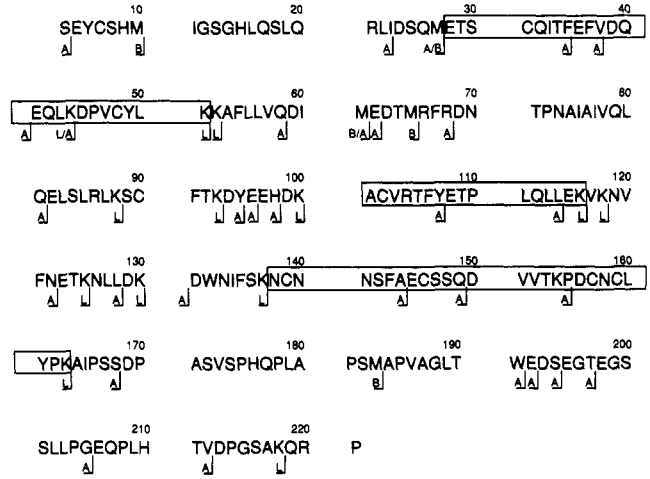


FIGURE 1: rhM-CSF amino acid sequence. Numbering of the sequence starts with Ser4 since the first three amino acids from the native M-CSF were deleted (Halenbeck et al., 1989). The boxed peptides are linked together via disulfide bridges and elute on the HPLC in one peak (fraction 3m in Figure 4) after BrCN and Lys-C digest. The cleavage sites for Asp-N, A; BrCN, B; and Lys-C, L are indicated.

(Das & Stanley, 1982). The two monomers are linked by disulfide bridges, resulting in a 49-kDa homodimer with 18 cysteines. Proper formation of disulfide bonds is crucial for attaining the correct three-dimensional structure of proteins; therefore, it was important to determine the locations of all disulfide bonds in the molecule.

Mass spectrometric methods have successfully been used for the study of disulfide bridges of peptides and proteins (Sorensen et al., 1990; Yazdanparast et al., 1987; Morris & Pucci, 1985; Buko & Fraser, 1985; Takao et al., 1984), and several comprehensive reviews are available on this subject (Carr et al., 1991; Smith & Zhou, 1990; Carr et al., 1990; Morris & Greer, 1988). In this report we describe the detection of the inter- and intramolecular disulfide bonds of rhM-CSF. Chemical and proteolytic cleavage techniques, followed by high-performance liquid chromatographic (HPLC) separation,

† This work was supported in part by the National Institutes of Health (Grant NIEHS ES00210). This paper is Oregon Agricultural Experiment Station Publication No. 9952.

* To whom correspondence should be addressed.

‡ Oregon State University.

§ Chiron Corp.

[1] Abbreviations: aa, amino acid; BrCN, cyanogen bromide; BSA, bovine serum albumin; calcd, calculated; DTT, dithiothreitol; FAB-MS, fast atom bombardment mass spectrometry; m, homoserine; m′, homoserine lactone; MALDI-TOF-MS, matrix assisted laser desorption ionization time of flight mass spectrometry; $MH^+$, protonated molecular ion; MW, molecular weight; nd, not determined; Nd-YAG, neodynium–yttrium aluminum garnet; obsd, observed; SAP-V8, *Staphylococcus aureus* protease V8; rhM-CSF, recombinant human macrophage colony stimulating factor; RP, reversed phase; TFA, trifluoroacetic acid; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/93/0432-482$04.00/0 

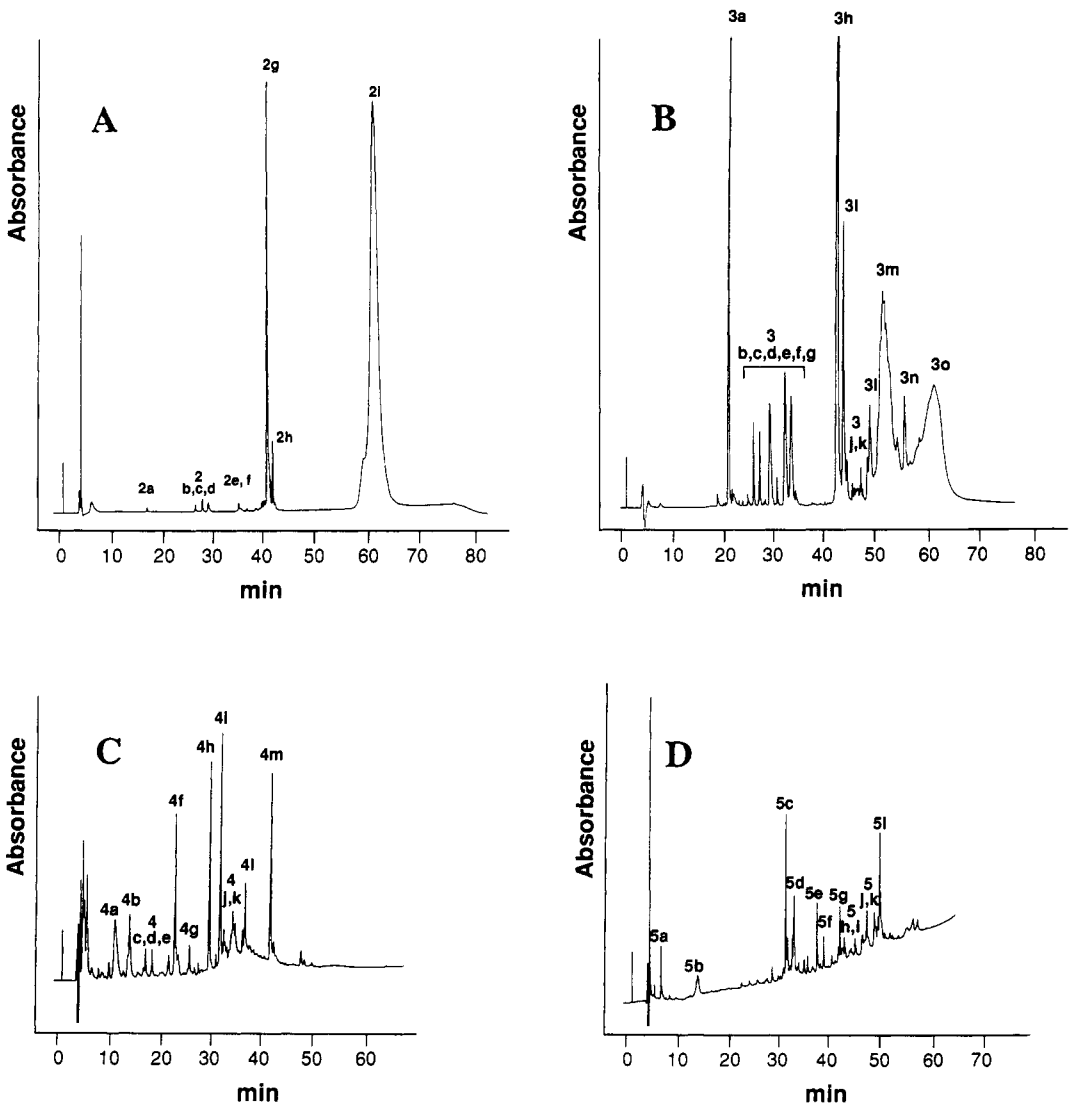


FIGURE 2: (A) HPLC chromatogram of BrCN treated rhM-CSF. Peak numbers indicate the individually collected fractions. For mass assignments see Table II. (B) HPLC chromatogram of rhM-CSF treated with both Lys-C and BrCN. For mass assignments see Table III. Disulfide-linked peptides were found in fractions 3b, 3c, 3e, 3k, and 3m. (C) HPLC chromatogram of thermolysin, Lys-C, and BrCN cleaved rhM-CSF. For mass assignments see Table IV. Fraction 4g contains a disulfide-linked dipeptide. (D) HPLC chromatogram of rhM-CSF treated with Asp-N, Lys-C, and BrCN. For mass assignment see Table V. Disulfide-linked peptides were present in fractions 5c, 5h, and 5j.

and fast atom bombardment mass spectrometric (FAB-MS) analyses were used for the rapid identification of the disulfide bridges.

## MATERIALS AND METHODS

*HPLC Purification and Separations.* The HPLC system consisted of an ALTEX model 322 gradient liquid chromatograph equipped with a Waters 486 tunable absorbance detector set at 220 nm. Linear gradients were applied using 0.1% TFA in $H_2O$ as solvent A and 0.08% TFA in $CH_3CN$ solvent B at a flow rate of 1 mL/min. All solvents were HPLC grade, and water was purified by passage through a Waters Millipore Milli-Q purification system. Protein purification was performed on a VYDAC C4 RP column (30 nm, 10 μm, 4.6 × 250 mm); peptide separations were carried out on a VYDAC C18 RP column (30 nm, 5 μm, 4.6 × 250 mm).

*Purification and Buffer Exchange for rhM-CSF.* Highly purified recombinant human macrophage colony stimulating factor (rhM-CSF) was supplied by Chiron Corp. (Emeryville, CA). To exchange buffers, the protein was purified by HPLC using the following gradient: initial concentration of solvent B was 10%. After 5 min the concentration was increased to 60% B over a period of 60 min. This solvent composition was held for an additional 10 min. Protein-containing fractions (major fraction 1a and minor fractions 1b and 1c) were collected manually, pooled, and lyophilized on a speedvac concentrator (Savant). The protein fraction (1a, >90%) was redissolved in 500 μL of a mixture of 0.1% TFA/$H_2O$ and 0.08% TFA/$CH_3CN$ (60:40), giving a protein concentration of 4.5 mg/mL (Bio-Rad protein assay kit using BSA as standard).

*BrCN Cleavage of Purified M-CSF.* A 5 M BrCN solution (10 μL) dissolved in $CH_3CN$ (approximately 2000-fold molar excess relative to the amount of Met residues present) was added to 500 μL of protein solution (4.5 mg/mL, 46 nmol) in 0.1% TFA/$H_2O$:0.08% TFA/$CH_3CN$ (60:40). After 20 h at 25 °C in the dark, 500 μL of $H_2O$ was added, and excess BrCN was removed using the speedvac concentrator. The volume was adjusted to 200 μL, and the sample was subjected to HPLC separation. After 1 min at 10% B the concentration was raised to 60% B in 65 min. Fractions were collected (Figure 2A) and used for FAB-MS experiments after freeze drying.

*Buffer Exchange after BrCN Reaction.* After BrCN cleavage the major protein fragment (Figure 2A, fraction 2i) was concentrated using the speedvac concentrator while

carefully avoiding precipitation. After evaporation of most of the $CH_3CN$, the sample was transferred to a microconcentrator device (Amicon centricon; MW cutoff, 10 000). The buffer was exchanged by adding $NH_4HCO_3$ and subsequent ultrafiltration at 4 °C; 300 μL of 50 mM $NH_4HCO_3$ (pH 8.5) was added and reduced to a volume of approximately 100 μL. This procedure was repeated three times. The final volume was adjusted to 300 μL, producing a protein concentration of 4 mg/mL.

*Lys-C Digestion of BrCN Treated rhM-CSF.* Lys-C (lysyl endopeptidase, EC 3.4.21.50, Wako Chemicals) solution (10 μL, 1 mg/mL) dissolved in 50 mM $NH_4HCO_3$ (pH 8.5) was added to 300 μL of protein solution (4 mg/mL) in 50 mM $NH_4HCO_3$ (pH 8.5) (E:S = 1:120, w/w). The mixture was incubated for 3 h at 37 °C. Peptides were separated using the HPLC conditions described above. Fractions were collected (Figure 2B) and used for FAB-MS experiments after freeze drying. Fraction 3m was redissolved in 150 μL of 50 mM $NH_4HCO_3$ (pH 8.5) giving a protein concentration of 6 mg/mL.

*Thermolysin Digestion of Lys-C Digested and BrCN Treated rhM-CSF.* Thermolysin (EC 3.4.24.4, Sigma) solution (2.5 μL, 1 mg/mL) dissolved in 50 mM $NH_4HCO_3$ (pH 8.5) was added to 30 μL of protein solution (6 mg/mL) in 50 mM $NH_4HCO_3$ (pH 8.5) (E:S = 1:72, w/w). The mixture was maintained at 37 °C for 3 h and then subjected to HPLC separation. The initial solvent composition, 10% B, was maintained for a period of 5 min and then changed to 45% B over a period of 55 min. Fractions were collected (Figure 2C), concentrated to dryness, and used for FAB-MS analysis.

*Asp-N Digestion of Lys-C Digested and BrCN Treated rhM-CSF.* Asp-N (Boehringer) solution (25 μL, 0.04 mg/mL) dissolved in 10 mM Tris-HCl (pH 7.5) was added to 30 μL of protein solution (6 mg/mL) in 50 mM $NH_4HCO_3$ (pH 8.5) (E:S = 1:180, w/w). $CH_3CN$ (5 μL) was added, and the mixture was incubated at 37 °C for 3 h. The sample was subjected to HPLC separation using the same conditions as described above. Fractions were collected (Figure 2D), and after lyophilization the samples were used for FAB-MS analysis.

*Reduction of rhM-CSF and Disulfide-Linked Peptides.* Aliquots (2.5 μL) of each of the HPLC-purified and redissolved samples containing disulfide-linked peptides or purified protein were lyophilized. Samples were redissolved in 30 μL of 50 mM $NH_4HCO_3$ (pH 8.5) containing 10 mg/mL DTT. After 3 h at 37 °C, the reaction was stopped by freeze drying.

*Xe-FAB-MS Analysis of Peptides before and after Reduction.* HPLC purified and lyophilized samples were redissolved in 5 μL of 0.1% TFA/$H_2O$ solution. Aliquots (2 μL) of each fraction were mixed on the probe with 2 μL of matrix consisting of 3-nitrobenzylalcohol (3-NBA) or glycerol/toluenesulfonic acid (TSA) or dithiothreitol–dithioerythritol (5:1, w/w, DTT–DTE). FAB-MS analysis was carried out on a Kratos MS-50 double-focusing mass spectrometer operated at a resolution of 1000 for raw data collection. Narrow scan medium resolution experiments were carried out at a resolution of 1600. Xenon was used to generate the primary ionizing beam from an Ion-Tech gun operated at 7–8 keV. Ions were accelerated from the ion source at 8 keV. The Kratos postacceleration detector was modified to allow operation at 25 keV. The scan rate was 30 s per decade. Molecular weight calculations and reconstructions of isotope distribution patterns were carried out using the Kratos DS90 software.

Table I: Predicted and Experimental Masses by MALDI-TOF-MS for rhM-CSF

| HPLC peak | $MH^{n+}_{calcd}$ [a] | $MH^{n+}_{obsd}$ [b] |
|---|---|---|
| native 1a | | |
| $MH^+$ | 49 047 | 49 073 [c] |
| $MH_2^{2+}$ | 24 524.5 | 24 535 [c] |
| reduced 1a | | |
| $MH^+$ | 24 525 | 24 554 [d] |
| $MH_2^{2+}$ | 12 263.5 | 12 291 [d] |

[a] Average mass. [b] Error range: ≤0.1%. [c] Internal mass calibration with cytochrome *c* and insulin. [d] External mass calibration with cytochrome *c*.

Table II: Predicted and Experimental Masses of Isolated Peptides[a] Derived from rhM-CSF after BrCN Cleavage

| HPLC peak | $MH^+_{calcd}$ [b] | $MH^+_{obsd}$ [c] | peptide (aa) |
|---|---|---|---|
| 2a | 840.5 | 841 | (214–221) |
| 2b | 1388.7 | 1389 | (170–183m) |
| | 1370.7 | 1371 | (170–183m′) |
| 2c | 1370.7 | 1371 | (170–183m′) |
| 2d | 1388.7 | 1389 | (170–183m) |
| 2e | 2314.2 | 2315 [e] | *f* |
| 2f | 3916.3 [d] | 3916 | (184–221) |
| 2g | 3916.3 [d] | 3916 | (184–221) |
| 2h | 1834.9 | 1836 [e] | (11–27m′) |
| | 3094.3 [d] | 3094 | (184–213) |
| 2i | – | nd | |

[a] See Figure 2A. [b] Monoisotopic mass. [c] Rounded-off values. [d] Average mass. [e] Due to the resolution chosen, the observed masses are higher than the calculated monoisotopic values. [f] Not assigned.

*MALDI-TOF-MS Analysis of Native and Reduced rhM-CSF.* C4 RP HPLC purified rhM-CSF (1 μL) was diluted 1:10 with 9 μL of sinapinic acid (5 mg/mL) dissolved in a 1:1 mixture of 0.1% TFA/$H_2O$ and $CH_3CN$. From this mixture, 1 μL (4.6 pmol of rhM-CSF) was used for the experiment. Cytochrome *c* was used as an external mass standard. For internal mass standards, a mixture of cytochrome *c* (0.5 pmol) and insulin (0.5 pmol) was used. Samples were desorbed using a Nd-YAG laser (355 nm) and 20 kV acceleration voltage.

*Amino Acid Sequence Analysis.* Peptides were subjected to Edman degradation using a model 470A gas-phase protein sequencer (Applied Biosystems).

## RESULTS

Peptide mapping of the reduced and iodoacetamide-alkylated rhM-CSF confirmed the amino acid sequence (Figure 1) and the correct C-terminal end (KQRP) of the protein (data not shown). MALDI-TOF-MS experiments comparing the native with the reduced rhM-CSF further proved the existence of a covalently linked homodimer (see Table I for mass assignments). No additional posttranslational modifications of the *E. coli*-derived rhM-CSF were observed.

Treatment with BrCN cleaved the C-terminus (aa 184–221) as well as the N-terminal peptide aa 11–27. The HPLC separated peptides (Figure 2A) were identified using FAB-MS experiments (Table II). Under the conditions used, some mild acid hydrolysis at Asp–Pro bonds was observed, giving peptides aa 170–183m′, aa 184–213, and aa 214–221. Cleavage at Met61 or Met65 was not complete, and oxidation occurred at the methionines (see Table III).

Lys-C digestion of the BrCN treated rhM-CSF produced a mixture of peptides that was separated by HPLC (Figure 2B). FAB-MS experiments were carried out on all fractions in order to identify the peptides (Table III). The HPLC trace indicates that digestion was not complete, because fraction 3o elutes with the same retention time as did the starting material (Figure 2A, fraction 2i). The HPLC fractions that contained

Table III: Predicted and Experimental Masses of Isolated Peptides[a] Derived from rhM-CSF after BrCN and Lys-C Cleavages, before and after Reduction

| HPLC peak | $MH^+_{calcd}$[b] | $MH^+_{obsd}$[c] | peptide (aa) |
|---|---|---|---|
| 3a | 935.4 | 935 | (94–100) |
| 3b | 585.3 | 585 | (89–93)[f] |
| | 808.2 | 808 | (4–10m′)[f] |
| | 826.2 | 826 | (4–10m)[f] |
| | 1390.5 | 1391[d] | (4–10m′) + (89–93)[g] |
| | 1408.5 | 1409[d] | (4–10m) + (89–93)[g] |
| 3c | 585.3 | 585 | (89–93)[f] |
| | 808.2 | 808 | (4–10m′)[f] |
| | 1390.5 | 1391[d] | (4–10m′) + (89–93)[g] |
| 3d | 602.4 | 602 | (126–130) |
| | 851.4 | 851 | (119–125) |
| 3e | 585.3 | 585 | (89–93)[f] |
| | 1167.6 | 1168[d] | (89–93) + (89–93)[g] |
| 3f | 1940.9 | 1942[d] | (164–183m′) |
| 3g | 1940.9 | 1942[d] | (164–183m′) |
| 3h | 909.5 | 909 | (131–137) |
| 3i | 1129.6 | 1130[d] | (52–61m′) |
| 3j | 1001.5 | 1001 | (53–61m′) |
| | 1019.5 | 1019 | (53–61m) |
| | 1492.8 | 1493[d] | (126–137) |
| 3k | 585.3 | 585 | (89–93)[f] |
| | 1911.0 | 1912 | (101–116)[f] |
| | 2493.3 | 2494[d] | (89–93) + (101–116)[g] |
| 3l | 2695.5 | 2696[d] | (66–88) |
| 3m | 2879.2 | 2880[d] | (138–163)[f] |
| | 2863.3 | 2865[d] | (28–51)[f] |
| | 1911.0 | 1912[d] | (101–116)[f] |
| | 837.4 | 837 | (45–51)[f] |
| 3n | 4234.2 | 4234[e] | (53–88 + 2′O′) |
| 3o | see 2i | nd | |

[a] See Figure 2B. [b] Monoisotopic mass. [c] Rounded-off values. [d] Due to the resolution chosen the observed masses are higher than the calculated monoisotopic values. [e] Confirmed by gas-phase Edman amino acid sequencing. [f] Reduced. [g] Nonreduced.

disulfide-linked peptides (3b, 3c, 3e, 3k, and 3m) were reduced separately in $NH_4HCO_3$ solution containing DTT before being subjected to additional FAB-MS analyses. To determine whether or not an observed mass peak in the recorded spectrum was due to disulfide linkages, complete reduction had to be obtained. Only completely eliminated peaks by reduction could be addressed unambiguously as disulfide-linked dipeptides. Reduction on the FAB-MS probe using DTT–DTE as matrix was incomplete and, therefore, not conclusive. Complete reduction was only achieved after treating the corresponding samples with excess DTT for an extended period of time.

Partial reduction of fraction 3c (Figure 2B) on the FAB target showed three relevant peaks in the spectrum (Table III). The peak with $m/z$ 585 corresponds to a Lys-C derived fragment (aa 89–93) bearing Cys90. Mass peak 808 was determined to be a BrCN-produced fragment (aa 4–10) in which Met10 was transformed to a homoserine lactone (m′). This peptide contains Cys7. The peak with $m/z$ 1391 represents the disulfide-linked dipeptide in which Cys7 is connected to Cys90. After complete reduction, the peak at $m/z$ 1391 disappeared, the ion peak with $m/z$ 585 remained, and the peak with $m/z$ 808 shifted to $m/z$ 826 because of the formation of homoserine (m) by $H_2O$ addition to the lactone during the reduction (Table III). Peak 3b (Table III) contained the same peptides; however, the homoserine lactone (m′) was already partially hydrolyzed to homoserine (m) (Shimizu et al., 1992), resulting in the formation of peaks with $m/z$ 808 and $m/z$ 826, as well as $m/z$ 1391 and $m/z$ 1409, respectively (data not shown).

The HPLC fractions 3b and 3c contain the majority of Cys90 bearing peptide aa 89–93 according to the absorbance intensities. In addition, two small HPLC peaks (3e and 3k) were detected that also show disulfide-linked dipeptides in which Cys90 was involved. In partially reduced fraction 3e (Figure 2B), a homodimer of peptide aa 89–93 was present as indicated by the peak with $m/z$ 1168 and the reduction product peak with $m/z$ 585 (Figure 2A). The peak with $m/z$ 1168 disappeared upon complete reduction (Figure 2B). Fraction 3k revealed mass peaks with $m/z$ 585 (aa 89–93), $m/z$ 1912 (aa 101–116), and $m/z$ 2494 (Table III). The latter is the disulfide-linked dipeptide in which Cys90 is connected to Cys102. This peak also disappears after reduction.

Table IV: Predicted and Experimental Masses of Isolated Peptides[a] Derived from rhM-CSF after BrCN, Lys-C, and Thermolysin Cleavages, before and after Reduction

| HPLC peak | $MH^+_{calcd}$[b] | $MH^+_{obsd}$[c] | peptide (aa) |
|---|---|---|---|
| 4a | 502.3 | 502 | (113–116) |
| | 520.3 | 520 | (160–163) |
| 4b | – | 798 | *e* |
| 4c | 1090.5 | 1091[d] | (66–74) |
| 4d | – | 1144 | *e* |
| 4e | 529.3 | 529 | (19–22) |
| 4f | 587.3 | 587 | (108–112) |
| | 656.3 | 656 | (106–110) |
| | 1006.5 | 1006 | (37–44) |
| 4g | 551.6 | *f* | (138–142)[g] |
| | 837.4 | 837 | (45–51)[g] |
| | 1385.6 | 1386[d] | (138–142) + (45–51)[h] |
| 4h | 743.5 | 743 | (111–117) |
| | 897.4 | 897 | (107–112) |
| 4i | 1582.8 | 1583[d] | (102–114) |
| | – | 2132 | *e* |
| 4j | – | 929 | *e* |
| 4k | 913.4 | 913 | (104–110) |
| | 1069.5 | 1069 | (62–69) |
| 4l | 863.5 | 863 | (107–113) |
| | 1824.9 | 1825[d] | (37–51) |
| | – | 2373 | *e* |
| 4m | 1010.5 | 1011[d] | (106–113) |

[a] See Figure 2C. [b] Monoisotopic mass. [c] Rounded-off values. [d] Due to the resolution chosen, the observed masses are higher than the calculated monoisotopic values. [e] Not assigned. [f] Confirmed by gas-phase Edman amino acid sequencing. [g] Reduced. [h] Nonreduced.

The major components in fraction 3m were peptides with $m/z$ 2880 (aa 138–163), $m/z$ 2865 (aa 28–51), $m/z$ 1912 (aa 101–116), and $m/z$ 837 (aa 45–51) as shown by FAB-MS analysis of the reduced sample (Table III). The peptide aa 138–163 contains four cysteines. All other peptides in this fraction were connected to this oligomer via disulfide bonds before reduction. This fraction was used for further studies.

By thermolysin digestion of rebuffered fraction 3m, partial cleavage at Phe143 was accomplished. Thermolysin cleaves peptide bonds formed by the amino groups of hydrophobic residues (Wilkinson, 1986). The HPLC-isolated fraction 4g (Figure 2C) showed mass peaks for the disulfide-linked dipeptide with $m/z$ 1386 and for the partially reduced product (aa 45–51) with $m/z$ 837 (Table IV). A peak corresponding to peptide aa 138–142 ($m/z$ 552) was not observed because of suppression of the ion. However, the mass difference of 549 (3 mass units lower than calculated due to three additional protons added by reduction and protonation) between the peak with $m/z$ 1386 for the disulfide-linked dipeptide and the peak with $m/z$ 837 for peptide aa 45–51 indicated the existence of the peptide aa 138–142. Amino acid sequencing of fraction 4g also proved the existence of peptide aa 138–142 (data not shown). After complete reduction the ion peak with $m/z$ 1386 disappeared proving that it was indeed a disulfide-linked dipeptide connecting Cys48 to Cys139. The masses of all other HPLC-fractionated peptides were assigned (Table IV).

Asp-N digestion after Lys-C and BrCN cleavages of rhM-CSF produced several fragments that could be separated by

Table V: Predicted and Experimental Masses of Isolated Peptides[a] Derived from rhM-CSF after BrCN, Lys-C, and Asp-N Cleavages, before and after Reduction

| HPLC peak | $MH^+_{calcd}$[b] | $MH^+_{obsd}$[c] | peptide (aa) |
|---|---|---|---|
| 5a | 760.4 | 760 | (39–44) |
| 5b | 637.3 | 638 | (36–40) |
| | 658.4 | 658 | (150–155) |
| 5c | 955.4 | 955 | (156–163)[g] |
| | 1905.8 | 1906[d] | (156–163) + (156–163)[h] |
| 5d | 729.5 | 729 | (18–23) |
| | 858.4 | 858 | (82–88) |
| 5e | – | – | [e] |
| 5f | 1421.8 | 1422[d] | (11–23) |
| | 690.4 | 690 | (53–58) |
| 5g | 813.4 | 813 | (108–114) |
| | 837.4 | 837 | (45–51) |
| | 1463.8 | 1464[d] | (41–52) |
| 5h | 553.2 | [f] | (145–149)[g] |
| | 813.4 | 813 | (108–114) |
| | 965.5 | 965 | (45–52) |
| | 1463.8 | 1464[d] | (41–52) |
| | 1911.0 | 1912[d] | (101–116)[g] |
| | 2461.2 | 2462[d] | (101–116) + (145–149)[h] |
| 5i | 553.2 | [f] | (145–149) |
| | 837.4 | 837 | (45–51) |
| | 965.5 | 965 | (45–52) |
| | 1653.9 | 1654[d] | (101–114) |
| | 1911.0 | 1912[d] | (101–116) |
| 5j | 837.4 | 837 | (45–51)[g] |
| | 1303.5 | 1304[d,f] | (138–149)[g] |
| | 1653.9 | 1654[d] | (101–114)[g] |
| | 2204.1 | 2204[d] | (101–114) + (145–149)[h] |
| | 3788.0 | 3788[d] | (138–149) + (101–114) + (45–51)[h] |
| 5k | 1653.9 | 1654[d] | (101–114) |
| 5l | – | – | [e] |

[a] See Figure 2D. [b] Monoisotopic mass. [c] Rounded-off values. [d] Due to the resolution chosen, the observed masses are higher than the calculated monoisotopic value. [e] Not assigned. [f] Confirmed by gas-phase Edman amino acid sequencing. [g] Reduced. [h] Nonreduced.

HPLC (Figure 2D), and some of them (5c, 5h, and 5j) were identified as disulfide-linked peptides. Asp-N cleaves at the N-terminal side of aspartyl residues. Nonspecific cleavages at glutamyl, threonyl, and seryl bonds have also been observed (Drapeau, 1980; Noreau & Drapeau, 1979; Ronk et al., 1991).

Fraction 5c showed a peak with $m/z$ 1906 and its potassium adducts. This peak is due to a disulfide-linked dipeptide, aa 156–163 (Table V). In order to distinguish whether the observed mass was due to one or two disulfide linkages in the dipeptide, higher resolution was needed. A mass difference of 2 Da had to be resolved. Narrow scan (2400–1700 Da) medium-resolution FAB-MS experiments of this dipeptide made a more accurate mass determination possible. The mass of the peptide was determined to be 1905.8 Da (Figure 3A). Comparison with the calculated and predicted isotopic pattern for the assumed elemental composition of this peptide ($C_{80}H_{121}N_{20}O_{26}S_4$) matched the observed $MH^+$ cluster ion pattern (Figure 3B). Thus, both Cys157 and Cys159 in this peptide must form intermolecular disulfide bonds to the corresponding cysteines of the second peptide, and these linkages hold the protein dimer together. After complete reduction, the peak with $m/z$ 1906 disappeared, and a peak with $m/z$ 955 corresponding to the peptide aa 156–163 was present (Table V). Using 3-NBA, a nonreducing matrix, had the disadvantage of poorer signal to noise ratios. Also, in 3-NBA the formation of potassium adducts is favored. The potassium adducts were suppressed by adding toluenesulfonic acid (TSA) to the matrix.

Under the conditions for the digestion, Asp-N cleaved partially at Glu which gave rise to the disulfide-linked peptides in fraction 5h and 5j. In the partially reduced peptide, 5h,

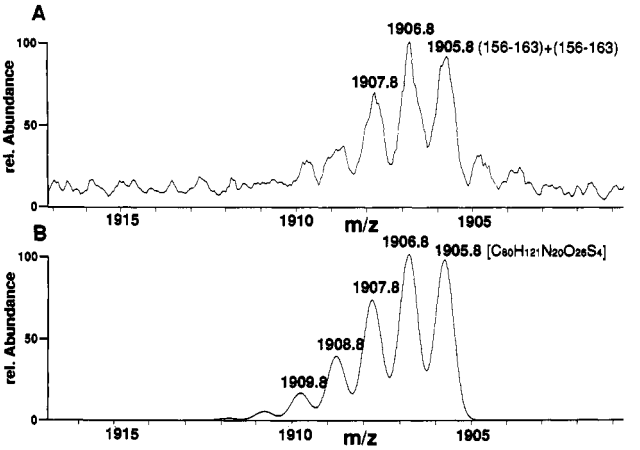


FIGURE 3: Medium-resolution FAB-MS of fraction 5c and reconstructed $MH^+$ isotope distribution. (A) The accurate mass of the disulfide-linked dipeptide was determined to be 1905.8 Da. Glycerol/TSA was used as matrix. (B) Comparison to the reconstructed isotopic pattern of the corresponding elemental composition matched the doubly disulfide-linked dipeptide, where all four half-cysteines are involved in intermolecular disulfide bridges. For experimental details, see Materials and Methods.

mass peaks were found with $m/z$ 1912 for the Lys-C-derived peptide aa 101–116 and with $m/z$ 2462 for the disulfide-linked dipeptide (aa 101–116 + aa 145–149) (and potassium adducts) in which Cys102 was connected to Cys146 (Table V). The Asp-N-derived peptide with $m/z$ 553 (aa 145–149) was not observed due to suppression of the ion. The peak with $m/z$ 2462 disappeared after complete reduction. In fraction 5j, a disulfide-linked dipeptide with $m/z$ 2204 was detected, and partial reduction gave a peak with $m/z$ 1654 (aa 101–114) (Table V). As demonstrated for fraction 5h, the disulfide-bonded dipeptide is the one formed by the connection between Cys102 and Cys146. The Asp-N-derived peptide aa 145–149 again was not detected. After full reduction, the peak with $m/z$ 2204 disappeared, and the only important peaks that remained were those with $m/z$ 1654 and $m/z$ 837 (aa 45–51).

In a second FAB-MS experiment, screening of the high mass range of sample 5j revealed a peak with $m/z$ 3788 (Table V). This peak is the disulfide-linked tripeptide in which peptides aa 45–51 ($m/z$ 837) and aa 101–114 ($m/z$ 1654) are connected to an Asp-N-derived peptide, aa 138–149 ($m/z$ 1304) containing two cysteines (Cys139 and Cys146). Upon complete reduction, the peak with $m/z$ 3788 disappeared, and mass peaks with $m/z$ 1654 (bearing Cys102) and $m/z$ 1304 (aa 138–149) were observed (Figure 3B). Simultaneous formation of mass peak 837 was observed.

## DISCUSSION

Native rhM-CSF (Figure 1) is very resistant to proteolytic cleavage, and only fragments from its C-terminus (aa 163–221) could be obtained, regardless of the protease used. Chemical cleavage was necessary to make rhM-CSF susceptible to proteolytic digestion. The release of the N-terminal peptide aa 11–27 after BrCN treatment changed the tertiary structure of rhM-CSF and made it digestable by Lys-C.

Special care had to be taken in the exchange of one buffer for another after reaction with BrCN. Lyophilization removed the more volatile components of the solvent mixture first and caused precipitation of the protein. The material obtained was then insoluble in $NH_4HCO_3$. Commercially available microconcentrators (AMICON) can only handle small amounts of organic solvents, but they avoid a change in the solvent composition during microfiltration of the protein.

Precipitation, thus, can be avoided. The initial solvent removal using a speedvac concentrator was carried out carefully to avoid precipitation of the protein. Addition of $NH_4HCO_3$ buffer to the protein solution and repeated ultrafiltration resulted in an $NH_4HCO_3$ buffered protein solution that subsequently could be used for further proteolysis.

The reaction times for proteolytic digests were limited to 3 h in order to minimize disulfide bond scrambling under the basic conditions used. Previous studies showed that during the first 3 h of Lys-C digestion scrambling could not be detected (data not shown). Using limited digestion times resulted in incomplete digestion of the starting material and, therefore, in mixtures of peptides with partial cleavages. By HPLC, similar peptides and disulfide-linked diepeptides were separated into different fractions.

After Lys-C digestion only disulfide bonds involving Cys90 could be assigned. The disulfide-linked peptide connecting Cys7 to Cys90 is the major one according to the UV absorbance intensities (Figure 2B, fractions 3b and 3c). Two minor HPLC fractions (Figure 2B, fractions 3e and 3k) also showed the presence of disulfide-linked dipeptides where Cys90 is connected through a disulfide bond to either Cys102 or to a second Cys90. These two disulfide bonds are assumed to derive from minor heterogenous species in the starting material, since varying the digestion conditions did not vary the amount of the observed HPLC peaks (data not shown). Scrambling probably does not account for the constant appearance of these disulfide-linked peptides, as one would expect variable amounts of these alternately linked dipeptides. Two minor impurities (1b, 1c) were detected by RP HPLC of the starting material.

The HPLC trace indicates that fraction 3m (Figure 2B) was heterogenous. The major components in fraction 3m were peptides with $m/z$ 2880 (aa 138–163), $m/z$ 2865 (aa 28–51), $m/z$ 1912 (aa 101–116), and $m/z$ 837 (aa 45–51) containing one or more cysteines, as shown by FAB-MS analysis of the reduced sample (Table III). The peptide aa 138–163 contains four cysteines. All other peptides in this fraction were connected to this oligomer via disulfide bonds before reduction (Table III). Contamination by the neighboring peak 3l ($m/z$ 2696) also was detected. Possible minor components that coelute with this fraction could be due to incomplete cleavage by both BrCN and/or Lys-C. Partial oxidation of the methionines could also contribute to the heterogeneity of this fraction. This fraction was used for further studies.

It was difficult to cleave in the protein region aa 138–163. Although there are cleavage sites, e.g., for SAP-V8 at Glu145 and for chymotrypsin at Phe143, no proteolytic cleavages were observed with these enzymes. This can be rationalized on steric grounds. The close proximity of four cysteines (Cys139, Cys146, Cys157, and Cys159), all of which are connected through disulfide bridges, shield this protein region from enzymatic attack by forming a core structure that exposes only the peripheral chains of the binding partners. Enzymes of broader specificity such as pepsin or proteinase K produced many fragments as judged from the HPLC trace, but identification of the peptides was not feasible because of their great number and nonspecific cleavages (data not shown). Thermolysin achieved only partial cleavage at Phe143 (Figure 2C, fraction 4g). At the N-terminal end of the peptide (aa 138–163) almost no cleavage was achieved regardless of the protease used. Asp-N, however, cleaves the Lys-C-derived peptide (aa 138–163) at Asp156, thus opening up the complex structure from the C-terminal end. This proved to be a very useful enzyme as it cleaved a region of the peptide that contains two cysteines (Cys157 and Cys159) suspected very early on of being involved in the formation of the homodimer. This assumption was substantiated by FAB-MS studies in our laboratory on individually trapped rhM-CSF intermediates obtained during a time course of protein refolding (data not shown).

The Asp-N released disulfide-linked dipeptide (Figure 2D, fraction 5c) with $m/z$ 1906 (2× aa 156–163) (Table V) contains two cysteines per monomer. The existence of two disulfide bonds in this peptide was shown by medium-resolution, narrow scan FAB-MS experiments (Figure 3A). The calculated mass for a dipeptide with two disulfide bonds (elemental composition $C_{80}H_{121}N_{20}O_{26}S_4$) matched the observed mass of 1905.8 Da and the reconstructed isotopic pattern (Figure 3B). It could not be determined whether the disulfide bridges in this dipeptide were symmetric (Cys157–Cys157-linked and Cys159–Cys159-linked) or asymmetric (Cys157–Cys159-linked and vice versa). This question possibly could be answered by subjecting the peptide 5c to two manual Edman degradation cycles and mass spectral analysis of the products. However, peptides produced from the protein after several chemical and proteolytic cleavages followed by HPLC isolation always is attended with considerable losses, and combining manual Edman degradation with FAB-MS would have required fairly high amounts of the peptide. The amount of peptide needed for these experiments simply was not available. Moreover, experience in this laboratory with other proteins shows that sequencing through disulfide bonds is incomplete and contributes to sample loss.

Not all monomer peptides involved in disulfide linkages could be detected by FAB-MS. In each of three cases (Figure 2C, fraction 4g; Figure 2D, fractions 5h and 5j), the spectra after partial reduction on the FAB-MS probe showed the mass peaks for the disulfide-linked dipeptides and only one of the monomers. The missing peptides in each case were deduced by the mass differences between the observed peaks for the dipeptides and the monomers. In fraction 4g, for example, a peak corresponding to peptide aa 138–142 ($m/z$ 552) was not observed, presumably because of suppression of the ion (Bull and Breese index: 3450). The mass difference of 549 (3 mass units lower than calculated due to three additional protons added by reduction and protonation) between the peak with $m/z$ 1386 for the disulfide-linked dipeptide and the peak with $m/z$ 837 for peptide aa 45–51 was indicative of the existence of this peptide (Table IV). Amino acid sequencing of this fraction also proved the existence of the peptide aa 138–142 (data not shown).

The same is true for the mass difference of 550 between the two mass peaks $m/z$ 2462 and $m/z$ 1912 (aa 101–116) in fraction 5h. These data suggest that the disulfide-linked partner was the Asp-N-derived peptide aa 145–149 ($m/z$ 553) with Cys146 connected to Cys102 (Table V). This mass difference of 550 also was observed between the two peaks $m/z$ 2204 and $m/z$ 1654 (aa 101–114) in fraction 5j, further indicating that Cys102 was connected to Cys142 from peptide aa 145–149 (Table V). With amino acid sequencing of these fractions the existence of peptide aa 145–149 was confirmed (data not shown). A calculated Bull and Breese index of 2680 for this peptide is the third worst of all possible combinations, making its detection by FAB-MS unlikely. Peptides with a highly positive Bull and Breese index are difficult to detect by FAB-MS due to suppression, whereas peptides with highly negative values usually show strong peaks. Peptides bearing many polar amino acids result in highly positive Bull and Breese indices. This for instance is the case with all peptides in the region aa 138–149.

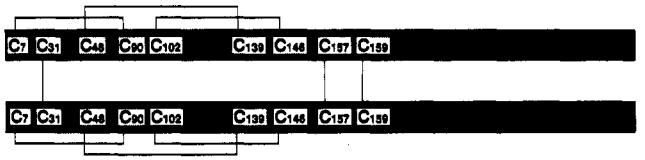


FIGURE 4: Schematic representation of the rhM-CSF dimer. All major inter- and intramolecular disulfide linkages are shown. The intermolecular bond connecting Cys31 with Cys31 was deduced as the only possible combination after all the other bonds were assigned.

The appearance of the peak $m/z$ 837 upon reduction of fraction 5j suggested the presence of another disulfide-linked peptide in this sample. A high mass peak with $m/z$ 3788 was observed which corresponded to a disulfide-linked tripeptide consisting of the peptides aa 45–51 ($m/z$ 837), aa 101–114 ($m/z$ 1654), and aa 138–149 ($m/z$ 13040) (Table V). All these peptides were found upon reduction, but the peptide aa 138–149 gave only a weak signal at $m/z$ 1304. The calculated Bull and Breese value of 5220 is the worst of all calculated possibilities for trying to observe this peptide by FAB-MS.

Three intramolecular disulfide bonds in each monomer and two intermolecular bonds between two monomers were detected. A possible disulfide bond between Cys31 in one monomer and Cys31 in the other remains the only one for which direct experimental evidence was not obtained. The expected Asp-N peptide aa 28–35 consists mainly of polar amino acids. This peptide was not observed presumably due to ion suppression. However, since there are no other free cysteines, Cys31 can only form an intermolecular disulfide bridge with the corresponding cysteine in the partner. The existence of the third intermolecular bond is further substantiated by the finding that in native rhM-CSF no free sulfhydryl groups were found (Boosman, personal communication), and, therefore, no alkylation of possible free sulfhydryl groups in the native rhM-CSF could be achieved (data not shown). Further evidence comes from studies of a truncated version of rhM-CSF, a mutant protein which lacks the last two cysteines, Cys157 and Cys159. This protein also forms a fully biologically active disulfide-linked homodimer, and it was proposed that in this particular protein the remaining seven cysteines in each monomer form the same disulfide bonds as in the rhM-CSF we have studied (Halenbeck et al., 1989).

Two of the now assigned disulfide bonds (Cys7–Cys90 and Cys48–Cys139) had been predicted previously (Bazan, 1991). The data presented here confirm Bazan's hypothesis. Moreover, Bazan's suggestion that Cys31 is probably involved in an intermolecular disulfide bond is also shown to be correct by the data presented here. According to Bazan's computations, the receptor binding moiety of M-CSF is encoded within the first 148 aa. This domain reveals some structural similarity to the stem cell factor (SCF). The disulfide structure of SCF is known (Lu et al., 1991) and matches our results.

A schematic drawing of the assigned disulfide bridges is shown in Figure 4. In summary, it has been shown that intramolecular disulfide bonds exist between the following pairs of cysteines: Cys7–Cys90, Cys48–Cys139, and Cys102–Cys146, and intermolecular disulfide linkages give Cys31–Cys31 and Cys157/159–Cys157/159 pairs.

## ACKNOWLEDGMENT

We thank Ole N. Jensen for help with the MALDI-TOF-MS analyses. We thank A. Boosman for helpful discussions and disclosure of unpublished results. We also thank A. Boosman, K. Koths, and C. Cowgill for critically reading the manuscript.

## REFERENCES

Aggarwal, B. B., & Pocsik, E. (1992) *Arch. Biochem. Biophys. 292*, 335–359.

Aukerman, S. L., Middleton, S., Sampson-Johannes, A., & Ralph, P. (1991) in *Hematopoietic Growth Factors and Interleukins: Biology and Clinical Applications* (Symann, M., Quesenberry, P., & Morstyn, G., Eds.) Gardirer-Caldwell Communications.

Bazan, J. F. (1991) *Cell 65*, 9–10.

Buko, A. M., & Fraser, B. A. (1985) *Biomed. Mass Spectrom. 12*, 577–585.

Carr, S. A., Bean, M. F., Hemling, M. E., & Roberts, G. D. (1990) in *Biological Mass Spectrometry* (Burlingame, A. L., & McCloskey, J. A., Eds.) pp 621–652, Elsevier Science Publishers B. V., Amsterdam.

Carr, S. A., Hemling, M. E., Bean, M. F., & Roberts, G. D. (1991) *Anal. Chem. 63*, 2802–2824.

Ceretti, D. P., Wignall, J., Anderson, D., Tushinski, R. J., Gallis, B. M., Stya, M., Gillis, S., Urdal, D. L., & Cosman, D. (1988) *Mol. Immunol. 25*, 761–770.

Clark, S. C., & Kamen, R. (1987) *Science 236*, 1229–1237.

Das, S. K., & Stanley, E. R. (1982) *J. Biol. Chem. 257*, 13679–13684.

Drapeau, G. R. (1980) *J. Biol. Chem. 255*, 839–840.

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

Kawasaki, E. S., Ladner, M. B., Wang, A. M., Van Arsdell, J., Warren, M. K., Coyne, M. Y., Schweikart, V. L., Lee, M.-T., Wilson, K. J., Boosman, A., Stanley, E. R., Ralph, P., & Mark, D. F. (1985) *Science 230*, 291–296.

Lu, H. S., Clogston, C. L., Wypych, J., Fausset, P. R., Lauren, S., Mendiaz, E. A., Zsebo, K. M., & Langley, K. E. (1991) *J. Biol. Chem. 266*, 8102–8107.

Morris, H. R., & Pucci, P. (1985) *Biochem. Biophys. Res. Commun. 126*, 1122–1128.

Morris, H. R., & Greer, F. M. (1988) *Trends Biotechnol. 6*, 140–147.

Noreau, J., & Drapeau, G. R. (1979) *J. Bacteriol. 140*, 911–916.

Ralph, P., Warren, M. K., Ladner, M. B., Kawasaki, E. S., Boosman, A., & White, T. J. (1986) *Cold Spring Harbor Symp. Quant. Biol. 51*, 679–683.

Ralph, P., Lee, M.-T., & Nakoinz, I. (1988) in *Macrophage-Derived Cell Regulatory Factors. Cytokines* (Sorg, C., Ed.) pp 1–18, Karger, Basel.

Ronk, M., Shively, J. E., Shute, E. A., & Blake, R. C. (1991) *Biochemistry 30*, 9435–9442.

Shimizu, A., Nagel, G. M., & Doolittle, R. F. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2888–2892.

Smith, D. L., & Zhou, Z. (1990) *Methods Enzymol. 193*, 374–389.

Sørensen, H. H., Thomsen, J., Bayne, S., Højrup, P., & Roepstorff, P. (1990) *Biomed. Mass Spectrom. 19*, 713–720.

Stanley, E. R., & Heard, P. M. (1977) *J. Biol. Chem. 252*, 4305–4312.

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

Takao, T., Yoshida, M., Hong, Y.-M., Aimoto, S., & Shimonishi, Y. (1984) *Biomed. Mass Spectrom. 11*, 549–556.

Wilkinson, J. M. (1986) in *Practical Protein Chemistry—A Handbook* (Darbre, A., Ed.) pp 131–148, John Wiley & Sons Ltd., New York.

Wong, G. G., Temple, P. A., Leary, A. C., Witek-Giannotti, J. S., Yang, Y.-C., Ciarletta, A. B., Chung, M., Murtha, P., Kriz, R., Kaufman, R. J., Ferenz, C. R., Sibley, B. S., Turner, K. J., Hewick, R. M., Clark, S. C., Yanai, N., Yokota, H., Yamada, M., Saito, M., Motoyoshi, K., & Takaku, F. (1987) *Science 235*, 1504–1508.

Yazdanparast, R., Andrews, P. C., Smith, D. L., & Dixon, J. E. (1987) *J. Biol. Chem. 262*, 2507–2513.