Disulfide Structure and *N*-Glycosylation Sites of an Extracellular Domain of Granulocyte-Colony Stimulating Factor Receptor

Mitsuru Haniu,* Thomas Horan, Tsutomu Arakawa, John Le, Viswanathan Katta, Shinichi Hara, and Michael F. Rohde

Departments of Protein Chemistry and Protein Structure, Amgen, Inc., 1840 DeHavilland Drive, Thousand Oaks, California 91320

Received January 29, 1996; Revised Manuscript Received July 30, 1996[⊗]

ABSTRACT: An extracellular domain containing 603 amino acid residues of human granulocyte-colony stimulating factor receptor was expressed in Chinese hamster ovary cells. The affinity-purified material has previously been shown to dimerize when combined with the ligand. In this paper we have characterized the primary structure of this active receptor. Laser desorption mass spectrometry of the purified receptor showed a broad peak at a molecular weight of 84 000, ranging from 77 000 to 91 000. The molecular weight heterogeneity is due to glycosylation. Since the molecular weight based on the amino acid sequence is 67 322, by subtraction the carbohydrate content is approximately 17 000. Disulfide structure of the receptor was determined by peptide mapping in the absence and presence of reducing agent. Sequence and mass spectral analyses of these peptides showed the receptor to contain eight disulfide bonds and three free cysteines. These disulfide bonds are consistent with the known domain motifs of the receptor in that no interdomain disulfides were present. One of the three free cysteines is reactive with alkylating agents, while the others are less reactive, probably being buried in the interior of the molecule. Blocking the free cysteines did not affect the ligand binding. Carbohydrate moieties are somewhat evenly spaced throughout the molecule, at eight different N-glycosylation sites, some of which show heterogeneity in their compositions. Glycosylation seems necessary for stabilizing the molecule against disulfide-linked oligomerization of the receptor, indicating that the free cysteine residues become reactive for oxidation and disulfide exchange upon deglycosylation.

Granulocyte-colony stimulating factor (G-CSF)[1] is a cytokine that plays an essential role in proliferation and differentiation of blood progenitor cells and is used to treat patients suffering from granulopoenia (Lieschke & Burgess, 1992a,b). It functions by binding its specific receptor found on the surface of progenitor cells of neutrophilic granulocyte lineage. Full-length cDNAs of G-CSF receptor have been isolated from both mouse leukemia cells and human placenta (Fukunaga *et al*.,1990a,b) and code for a large extracellular domain, a single transmembrane domain, and a cytoplasmic domain. The extracellular domain of the mature protein consists of an immunoglobulin-like (Ig-like) domain, a cytokine receptor homology (CRH) domain, and fibronectin type III (FN III) repeating domains. The CRH domain contains a cysteine-rich putative ligand-binding domain and is found in many other cytokine receptors (Fukunaga *et al*., 1991). Cysteine residues are largely localized in the extracellular domain, i.e., the human receptor contains nineteen cysteine residues in the extracellular domain, while the cytoplasmic domain has only three cysteines. Furthermore, there is some species difference with respect to the localization of cysteine residues since the murine G-CSF receptor contains only sixteen cysteines in its extracellular domain. It is presumed that the extra cysteine residues in the human receptor exist in the free state. Hiraoka *et al*. (1994) and Anaguchi *et al*. (1995) reported the partial disulfide structure of the ligand-binding domain or CRH domain of murine G-CSF receptor expressed in *Escherichia coli* and from their study predicted the disulfide structure of human CRH domain. The three dimensional structure of the G-CSF receptor has not yet been elucidated. We have previously generated a soluble G-CSF receptor using a CHO-cell expression system and shown that the affinity-purified receptor is fully active for binding its ligand, G-CSF, at a stoichiometry of 2 receptors:2 ligands (Horan *et al*., 1996) We report here the disulfide structure and carbohydrate attachment in this active G-CSF receptor. Hopefully, this information will aid in elucidating the tertiary structure of the extracellular domain of this receptor.

MATERIALS AND METHODS

Materials. Pepsin was obtained from Sigma Chemical Co. (St. Louis, MO), and thermolysin and neuramidase (or sialidase) were from Boehringer/Mannheim (Indianapolis, IN). DTNB, 4-vinylpyridine, and NEM were purchased from Sigma Chemical Co. BNPS-skatole and CHAPS were obtained from Pierce Chemical Co. (Rockford, IL), and 4-HCCA was from Sigma Chemical Co. The *N*- and

* To whom correspondence should be addressed. Tel: (805) 447-3117. FAX: (805) 499-7464.

[⊗] Abstract published in *Advance ACS Abstracts,* September 15, 1996.

[1] Abbreviations: G-CSF, granulocyte-colony stimulating factor; G-CSFR, G-CSF receptor; soluble receptor, extracellular domain of the receptor; CRH, cytokine receptor homology; CHAPS, 3-(3-(cholamidopropyl)dimethylammonio)-1- propanesulfonate; CHO, Chinese hamster ovary; GdHCl, guanidine hydrochloride; DTNB, 5,5′-dithiobis(2-nitrobenzenesulfonate); 5-IAF, 5-iodoacetamidofluorescein; MALDI, matrix-assisted laser desorption ionization; BNPS-skatole, 2-(2′-nitrophenylsulfenyl)-3-methylbromoindoleimine; 4-HCCA, α-cyano-4-hydroxycinnamic acid; 4-VP, 4-vinylpyridine; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; NEM, N-ethylmaleimide.

S0006-2960(96)00213-9 CCC: $12.00

O-glycanases were from Genzyme (Cambridge, MA). Other chemicals are HPLC grade.

Preparation of an Extracellular Domain of G-CSF Receptor. A cDNA construct coding for the extracellular domain (603 amino acid residues) of the G-CSF receptor was expressed in Chinese hamster ovary (CHO) cells as described elsewhere (Horan *et al*., 1996). The receptor was partially purified from CHO-cell conditioned media using batch chromatography with Q-Sepharose and hydroxylapatite and was followed by ligand affinity chromatography and gel filtration. The final purified product showed a single band on SDS-gel electrophoresis.

Labeling of Free Cysteine Residues. For disulfide determination of soluble G-CSF receptor, free cysteine residues of the native protein (100 μg) were treated with a 100-fold molar excess of iodoacetamide, 4-vinylpyridine, or 5-IAF at pH 7.5 in the presence of 6 M GdHCl for 15 h at 25 °C. The modified protein was purified by reversed phase HPLC using a Vydac C18 column (4.6 × 250 mm) using a linear gradient from solvent A (0.1% TFA) to solvent B (0.1%TFA-90% acetonitrile). The number of free sulfhydryl groups was determined with DTNB titration (Ellman, 1959) according to the methods of Habeeb (1972). Briefly, about 0.25 mM of G-CSF receptor was prepared in PBS, pH 7.5, and 2 mg of DTNB/mL stock solution was prepared in 50 mM Tris buffer, pH 8.0. A standard assay mixture of 0.6 mL contained 0.005 mM of the receptor, 0.05 mM or 0.1 mM of DTNB in 0.1M Tris-HCl buffer, pH 8.0, and varying concentrations of urea, i.e., 0, 2, 4, 5, 6, and 8 M. The receptor was mixed with DTNB, and the reaction was monitored at 412 nm for 30 min with a Beckman DU650 UV/vis spectrophotometer (Brea, CA). The number of free sulfhydryl residues in the receptor was calculated using 13.6/mM/cm as absorbance coefficient of TNB. All experiments were carried out at 25 °C using a Peltier temperature controller (Beckman).

Deglycosylation of Soluble Receptor. Soluble G-CSF receptor (0.25 mg/mL) was incubated overnight at 4 °C in either the presence or the absence of 20 mM NEM in 20 mM potassium phosphate, pH 7.0. NEM-treated or untreated receptor (50 μg) was digested with sialidase (10 milliunits), *N*-glycanase (2 units), *O*-glycanase (2 units), or all three enzymes together for 18 h at 37 °C, after adding CHAPS to 5 mM. Mock reactions containing no glycosidases were treated identically to the digested samples. Deglycosylated samples, mock reactions, and untreated receptor samples were analyzed by SDS−PAGE under reducing and non-reducing conditions on 8% polyacrylamide gels (Novex, San Diego, CA), including Novex Mark 12 molecular weight standards (Laemmli, 1970). Protein bands on the gels were visualized by staining with Coomassie Blue.

Enzymatic and Chemical Cleavages. The native receptor (100 mg) was digested with pepsin (2 mg) in 0.02 N HCl at pH 2 for 20 h at 37 °C. Thermolytic digestion was performed in 0.1 M Tris-HCl buffer, pH 7.2, at 37 °C for 20 h with an enzyme−substrate ratio of 1:20. Tryptophan cleavage by BNPS-skatole was performed as described previously (Saviege & Fontana, 1977), followed by three successive washes with ethyl acetate (0.5 mL). Upon each addition of the ethyl acetate, the mixture was resuspended and centrifuged to remove excess reagents.

Peptide Isolation and Sequence Analysis. Peptides were separated with a Vydac C18 reversed-phase HPLC (4.6 × 250 mm) with a three-step linear gradient from 2% to 20% solvent B over 20 min, from 20% to 40% solvent B over 40 min, and finally from 40% to 60% solvent B over 10 min. Sequence analysis of peptides was performed as described (Haniu *et al*., 1994), using Applied Biosystems sequencer models 470A and 477A.

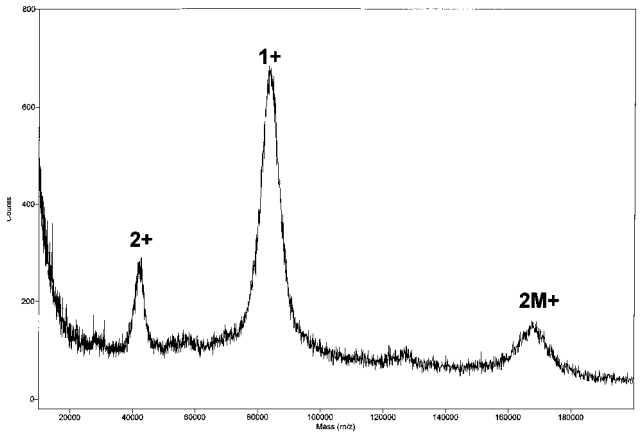


FIGURE 1: MALDI mass spectral analysis of soluble G-CSF receptor. The sample (2−5 μg) was loaded on a slide with 4-HCCA. The signal was obtained using laser desorption mass spectrometry using Voyager (PerSeptives). Single-charged ion (1+) and double-charged ion (2+) are indicated. A signal of $2M^+$ corresponds to a dimeric state.

Mass Spectrometry of the Soluble Domain of G-CSF Receptor and Peptides. Matrix-assisted laser desorption ionization (MALDI) mass spectrometry of proteins was performed using 4-HCCA as a matrix. Samples dissolved in PBS were spotted on the sample cartridge, and then washed with dry ice-cooled water. The analysis was carried out under the appropriate voltage to obtain an optimized signal. Cys-containing peptides were analyzed by Sciex API triple-quadrupole mass spectrometer with an ionspray interface. The dried sample was dissolved in 0.1% TFA−50% acetonitrile and flow-injected into the ion-spray interface using a Michrom BioResources Ultrafast Microprotein Analyzer. The carrier solvent was 0.1% TFA−50% acetonitrile/water with 0.1% TFA flowing at 5 μL/min. The scan range was 300−2400 amu with a step of 0.5 amu. The mass units and standard deviation were calculated using Sciex Hypermass software.

RESULTS AND DISCUSSION

The soluble G-CSF receptor used in this study was prepared from CHO-cell conditioned media from cells transfected with a gene coding for the entire extracellular domain of the receptor. We have shown in the previous paper (Horan *et al*., 1996) that the affinity-purified soluble receptor is fully active, in that it formed reversibly associating complexes with G-CSF at a stoichiometry 2:2. This active material was extensively characterized in disulfide structure and glycosylation as described below.

Protein Analysis of an Extracellular Domain of G-CSF Receptor. Affinity-purified soluble G-CSFR showed a single band on an SDS-gel, exhibiting a molecular weight of approximately 94 000. MALDI mass spectrometry of the soluble G-CSF receptor showed a broad peak at a molecular weight of 84 000, but the molecular weights were highly heterogeneous due to glycosylation, ranging from 77 000 to 91 000 (Figure 1). The N-terminal sequence analysis of the

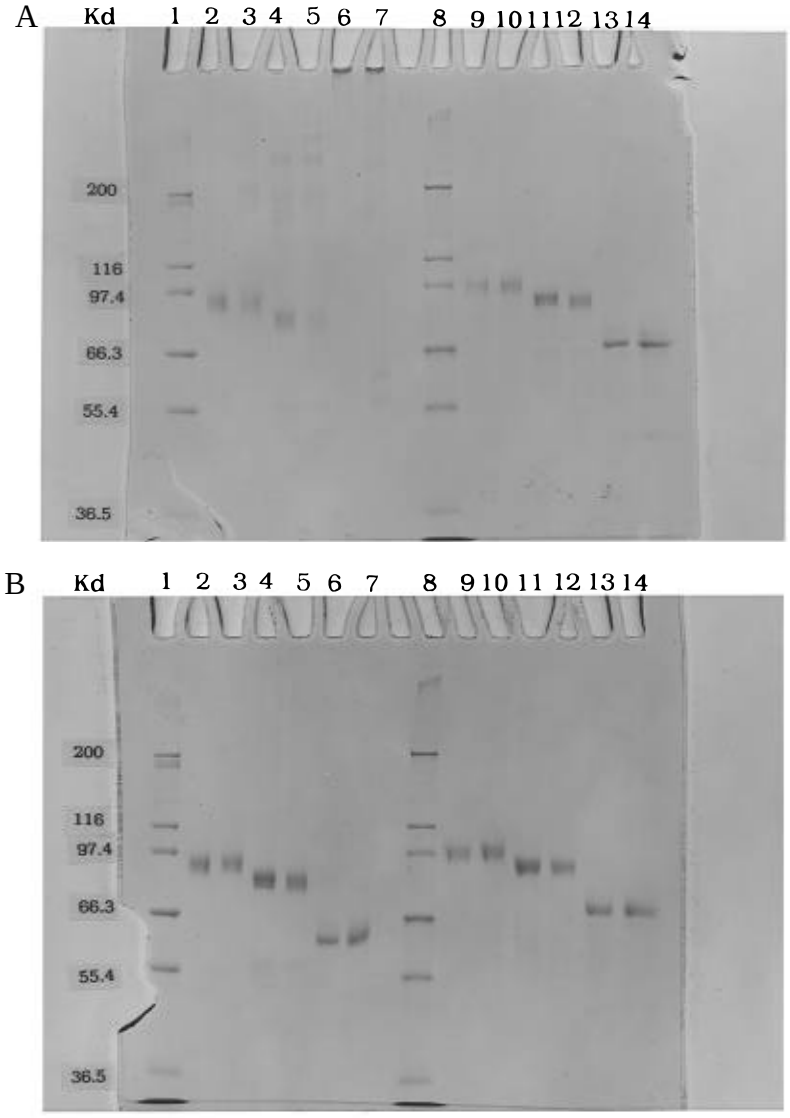


FIGURE 2: SDS−polyacrylamide gel electrophoresis of deglycosylated solubleG-CSF receptor. (A) SDS−PAGE of deglycosylated receptor. Lanes 1 and 8, standard protein mixtures; muscle myosin (200 kDa), β-galactosidase (116 kDa), phosphorylase (97.4 kDa), bovine serum albumin (66.3 kDa), glutamate dehydrogenase (55.4 kDa), lactate dehydrogenase (36.5 kDa) and carbonic anhydrase (31 kDa); lanes 2 and 9, soluble receptor; lanes 3 and 10, receptor incubated at 37 °C in 5 mM CHAPS; lanes 4 and 11, receptor treated with sialidase; lanes 5 and 12, incubated at 4 °C; lanes 6 and 13, receptor treated with *N*-glycanase; lanes 7 and 14, treated with sialidase and both *N*- and *O*-glycanases; lanes 1−7 were run under nonreducing conditions, and lanes 8−14 were run under reducing conditions. (B) SDS−PAGE of NEM-treated deglycosylated receptor. Exact same order as above in A, using soluble G-CSF receptor treated with NEM as described in Materials and Methods.

receptor indicated a correct signal peptide cleavage since the first nine residues were EXGHISVSA (residues 1−9), despite lacking an extra N-terminal Glu (Fukunaga *et al*., 1990b). X denotes an unidentified residue. The C-terminal peptide was isolated from a cyanogen bromide digest and found to have the sequence TLTPEGSELH (residues 594−603). Both N-terminal and C-terminal sequences determined indicate that the protein contains 603 amino acid residues, resulting in the calculated molecular weight of 67 322 and hence leaving a carbohydrate content of about 17 000. The total mass unit of 84 000 is consistent, within experimental error, with that determined by sedimentation equilibrilium measurements (Horan *et al*., 1996).

Labeling of Free Sulfhydryl Residues. Free cysteine residues were modified with SH alkylating agents including iodoacetamide, iodoacetate, DTNB, or 5-IAF under denaturing conditions. The number of free cysteines was determined by DTNB titration with or without denaturation. No free sulfhydryl residues were detected in either the absence of urea or in the presence of 4 M urea. When the receptor was incubated with 4 M urea at 25 °C for 20 min prior to DTNB titration, no free cysteine was detected either. However, when the sample was incubated with 6 or 8 M urea, free cysteines became accessible to DTNB and were estimated to be approximately 1.0 or 1.8 mol per mol of protein, respectively. This suggests that at least two cysteine residues exist in a free state but are buried in the folded structure. Meanwhile, a fluorescent SH reagent, 5-IAF, was not accessible to any free cysteine residues probably due to its bulky fluorophore (data not shown).

Deglycosylation of Soluble G-CSF Receptor. Soluble G-CSF receptor was treated with enzymes to remove both N- and O-linked carbohydrates, and the deglycosylation reactions were analyzed using SDS−PAGE. Several observations can be made from the two SDS−gels which are shown in Figure 2. Both gels were loaded with intact or deglycosylated receptor samples run under reducing or nonreducing conditions. The gel in Figure 2A contains samples

which have not been modified, while those shown in the gel in Figure 2B have been treated with NEM to block reactive, unpaired cysteine residues. In comparing the two nonreduced halves of the gels (lanes 1−7) one may clearly observe disulfide linked aggregation in Figure 2A, but not in Figure 2B. This aggregation is slight in the mock reaction (lane 3) and severe in the *N*-glycanase-treated samples (lanes 6 and 7). Materials in these lanes of Figure 2A barely migrate into the stacking portion of the gel, but upon reduction (lanes 13 and 14) one can assume the expected molecular weight of approximately 67 000. In Figure 2B, the nonreduced *N*-glycanase-treated sample (lane 6) as well as the sample treated with sialidase and *N*- and *O*-glycanase (lane 7) appears to migrate somewhat below the peptide-only molecular weight of 67 322. The reducing half of the gels shows a considerably higher MW band for these samples (lanes 13 and 14), which nearly comigrates with the marker protein bovine serume albumin at 66 300. We routinely observe a slight increase in mobility for other non-reduced proteins compared to their migration after reduction. Although the difference seems rather large in this case, we have found no evidence of proteolytic degradation which in fact may be expected to show a more pronounced effect in the reduced samples. When NEM-treated deglycosylated receptor samples were applied to a gel filtration column to separate the deglycosylated receptor from glycosidases, removed carbohydrates, and any undigested receptor, aggregation of deglycosylated receptor was still observed (data not shown). This finding suggests that in addition to disulfide linked aggregation, a significant non-covalent aggregation upon deglycosylation exists as well. Presumably, these non-covalent interactions are dissolved upon treatment with SDS, as evidenced by lanes 6 and 7 of Figure 2B.

Pepsin-Generated Peptide Map of the Soluble Receptor for Determination of Disulfide Linkages. The purified receptor was digested with pepsin as described in Materials and Methods, and the numerous resulting peptides were purified by reversed-phase HPLC (Figure 3A). Comparison of the peptide map with the DTT-reduced map indicates that many disulfide-containing peptides are present (data not shown). Sequence analyses of all peptides revealed that more than 10 peptides contained disulfide linkages. In summary, the sequence and mass spectral analyses of the pepsin-generated peptides revealed four disulfide linkages, Cys107-Cys118, Cys153-Cys162, Cys224-Cys271, and Cys364-Cys371, respectively, with two free cysteines, Cys163 and Cys228, identified as well.

Table 1 summarizes the sequence and mass spectral analyses of the peptic peptides. Peptide P35.7 was found to contain three cysteine residues in a single peptide chain. Cys153 was not detected, while Cys162 was detected as DiPTH-Cys by sequence analysis. We conclude that this peptide contains a disulfide, of Cys153-Cys162, and a free cysteine, Cys163. This free cysteine was occasionally observed as a glutathione adduct, probably due to modification that occurred during large-scale cell culture of the receptor. The analogous peptic peptide of the IAM-modified receptor showed that Cys162 gave a significant diPTH-Cys, confirming that Cys153 links Cys162 and that Cys163 exists in the free state. The results are consistent with sequence analyses of murine G-CSF receptor (Hiraoka *et al*., 1994), in which the residue corresponding to Cys163 is Ser.

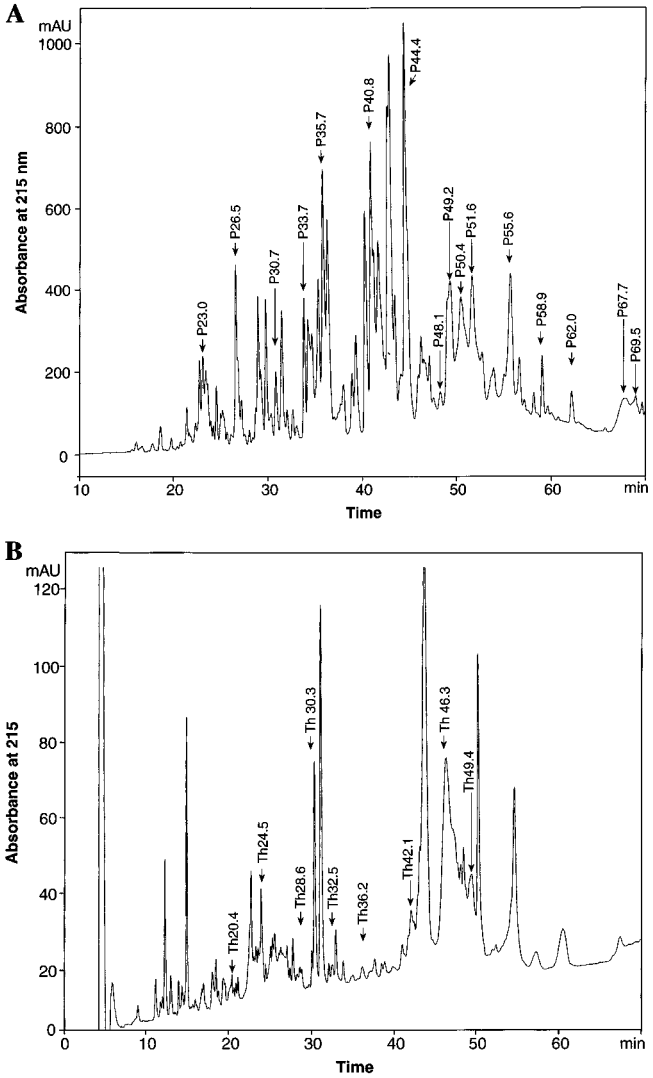


FIGURE 3: (A) HPLC separation of peptic digest of soluble G-CSF receptorThe peptic digest was directly subjected to reverse phase HPLC using a Vydac C18 (4.6 × 250 mm). The peptides were eluted with a linear gradient from solvent A (0.1% TFA) to solvent B (0.1% TFA-90% acetonitrile) as described in Materials and Methods. (B) HPLC purification of thermolytic digest of peptic fragment P-50. See Materials and Methods for experimental procedures.

Sequence analysis of peptide P40.8 showed it to span residues 362−381. No detection of Cys364 and detection of Cys371 as diPTH-Cys indicate the disulfide Cys364-Cys371.

Peptide P55.6 was analyzed by sequencer, indicating that it contained the disulfide linkage Cys107-Cys118 since Cys107 was not observed, while Cys118 was detected as diPTH-Cys. Since this peptide contained two N-glycosylation sites, direct mass spectrometry did not give useful information for the disulfide assignment. When this peptide was deglycosylated with *N*-glycanase and subjected to mass spectral analysis, it gave the molecular mass of 4281 amu, matching the calculated mass for two peptides and confirming the disulfide assignment in these two peptides.

Sequence analysis of another peptide, P58.9, revealed two sequences as shown in Table 1. Since these two sequences each contain a cysteine in the peptide, i.e., Cys224 and Cys271, the coelution of the two peptides suggest the disulfide linkage Cys224-Cys271. Mass spectral analysis of peptide P58.9 gave the corresponding mass units of this

Table 1: Summary of Cysteine- and Cystine-Containing Peptides

no.	peptide	sequence (residue no.)	obsd (calcd) mass
1.	P35.7	D X V P K D G Q S H C X I P R K H L L L (152-171)	2259.8 (2261.0)
2.	P40.8	P L X N T T E L S C T F L P S E A Q E (362-381)	3993 (2502.2)
3.	P48.1	T L K S F K S R G N X Q T Q G D S I L (133-151)	4307 (4303)
		V Q A E N A L G T S M S P Q L X L D PMD (179-199)	
4.	P55.6	R A G Y P P A I P H N L S X L M N L T T S S L I X Q W E P G P E T H L P T S F (94-132)	4281.0 (4281.9)[b]
5.	P58.9	R T M D P S P E A A P P Q A G X L Q L(209-227)	2790.0 (2789.4)
		Y E L X G L L (268-274)	
6.	P62.0	X W E P W Q P G L (228-236)	1113.2 (1115)
7.	P50Th20.4	F K S R G N X Q T Q G D S (137-149)	1661.2 (1660.0)
		L X (193-194)	
8.	P50Th24.5A	H I N Q K X E L R H K P Q R G E A S W (237-255)	2711 (2707)
		I R X (283-285)	
9.	P50Th24.5B	L G D P I T A S X (14-22)	ND
		I I K Q N X S H L D (23-32)	
10.	P50Th28.6	E X G H (1-4)	ND
		F L S C X (74-78)	
11.	P50Th30.3	I N Q K X E L R H K P Q R G E A S W (238-255)	ND
		I R X (283-285)	
12.	P50Th32.5	I L D X V P K D G Q S H C X (150-163)	ND
13.	P50Th42.1	L S C (105-107)	2274 (2274)
		L I C Q W E P G P E T H L P T S F (116-132)	

[a] Numbers in parenthesis show residue nos. based on the mature sequence. "C" denotes cysteine residue detected as a diPTH-Cys. "X" denotes unidentified residue due to Cys. N denotes asparagine residue which was glycosylated. [b] After treatment with *N*-glycanase.

peptide, confirming the existence of the disulfide bond Cys224-Cys271.

Peptide P62.0 showed a single peptide sequence, and the mass spectral result indicates the corresponding mass units, suggesting that Cys228 exists as a free sulfhydryl residue (data not shown). Previously, we reported that Cys228 was modified with a cross-linker, *m-N*-ethylmaleimidobenzoyl-*N*-hydroxysuccinimide, under native conditions (Haniu *et al*., 1995).

Thermolytic Digestion of Fraction P-50. Since peptide fractions P50.4, P51.6, and P67.7 contained similar disulfide-linked peptides, these samples were combined. The combined fraction (renamed P-50) was further digested with thermolysin as described in Materials and Methods. As shown in Figure 3B, many peptides were isolated from this digestion. The peptide map after DTT reduction of this digest was not deduced because of the limitation of sample amount. However, sequence analyses of all peptides revealed several Cys-containing peptides. Peptide fraction P50Th20.4 contained two peptides, LC (residues 193−194) and FK-SRGNCQTQGDS (residues 137−149). Mass spectral data showed this fraction to have a mass of 1661.2, supporting the presence of the disulfide linkage Cys143-Cys194 (Table 1).

Similarly, sequence analysis of peptide P50Th24.5A shows two sequences consisting of IRC (residues 283−285) and HINQKCELRHKPQRGEAW (residues 237−255). MALDI mass spectral results from this peptide are consistent with formation of the disulfide bond Cys242-Cys285 (Figure 4). Peptide P50Th30.3 was identical to the above P50Th24.5A except for lacking the N-terminal histidine.

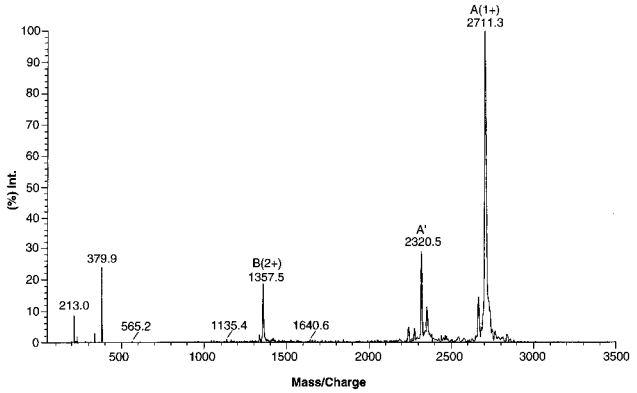


FIGURE 4: Mass spectrometry of Cys-containing peptide P50Th24.5A. MALDI mass spectrometry of the peptide was performed. 1+ and 2+, denote a single-charged molecular ion and its double-charged ion, respectively. The signal of 2,320 amu (A′) corresponds to the mass of one of the Cys-peptide, HINQKCELRHPQRGEAW (residues 237−255). Another signal (379.9 amu) corresponding to peptide IRC (residues 283−285) was derived from the cleavage of disulfide linkage under high laser energy.

Table 2: Summary of N-Glycosylation Sites in Human G-CSF Receptor

no.	consensus sequence (residue nos.)	N-glycosylation[a]	peptides	mouse[b]
1.	NCS (27−29)	G	P67.7	C
2.	NHT (69−71)	G	P69.5	C
3.	NLS (104−106)	G	P50.4A	C
		N	P55.6	
4.	NLT (110−112)	G	P50.4B	N
5.	NTT (365−367)	G	P40.8	C
6.	NKT (450−452)	G	P49.2	C
7.	NQS (547−549)	G	P23.0	N
		N	P26.5	
8.	NAS (555−557)	N	P44.4	C
9.	NST (586−588)	G	P23.0, P30.7, P35.7	C

[a] G, glycosylated; N, non-glycosylated. [b] C, conserved in murine receptor; N, not conserved.

The sequence analysis of peptide P50Th24.5B suggests the presence of disulfide bond Cys22-Cys28 since these two peptides coeluted in a single peak. Mass spectral data is lacking because of limited sample amount.

Peptides P50Th42.1 gave two sequences, LSC (residues 105−107) and LICQWEPGPETHLPTSF (residues 116−132) with cycle three containing significant diPTH-Cys. This result, taken together with the fact that both cysteine residues appear at cycle 3, demonstrates the existence of disulfide linkage between Cys107 and Cys118 (Marti *et al*., 1987; Haniu *et al*., 1994).

Peptide P50Th32.5 contained one peptide sequence, ILDCVPKDGQSHCC (residues 150−163) in which Cys153 is linked to Cys162, because cycle 12 showed diPTH-Cys upon PTH analysis, as previously identified from peptide P35.7.

Sequence analysis of peptide P50Th36.2 showed two Cys-containing peptides, FKSRGNCQTQGDS (residues 137−149) and MSPQLCLDPMD residues 189−199), indicating the existence of the disulfide linkage, Cys143-Cys194.

Carbohydrates in Soluble G-CSF Receptor. Multiple N-glycosylation sites were determined by disappearance of asparagine residues at the expected positions during sequence analysis. Table 2 shows the summary of N-glycosylation sites analyzed in this study. Nine potential N-glycosylation sites are reported from cDNA sequence (Fukunaga *et al*., 1990b), whereas one site, NAS (residues 555−557), was found not to be glycosylated. Two other sites (residues 104−106 and residues 547−549) were heterogeneous since both glycosylated and non-glycosylated peptides were isolated as indicated in Table 2. Meanwhile, the function of carbohydrate moieties was difficult to assess because of the instability of deglycosylated protein. Figure 5 shows the disulfide structure and N-glycosylation sites in the extracellular domain of G-CSF receptor. Although no X-ray crystallographic data of G-CSF receptor have been published yet, this study and other reports (Hiraoka *et al*., 1994; Anaguchi *et al*., 1995) suggest that the CRH domain may retain a complicated structure through disulfide bonding and carbohydrate structure affecting a conformational change of the receptor during ligand binding (Horan *et al*., 1996).

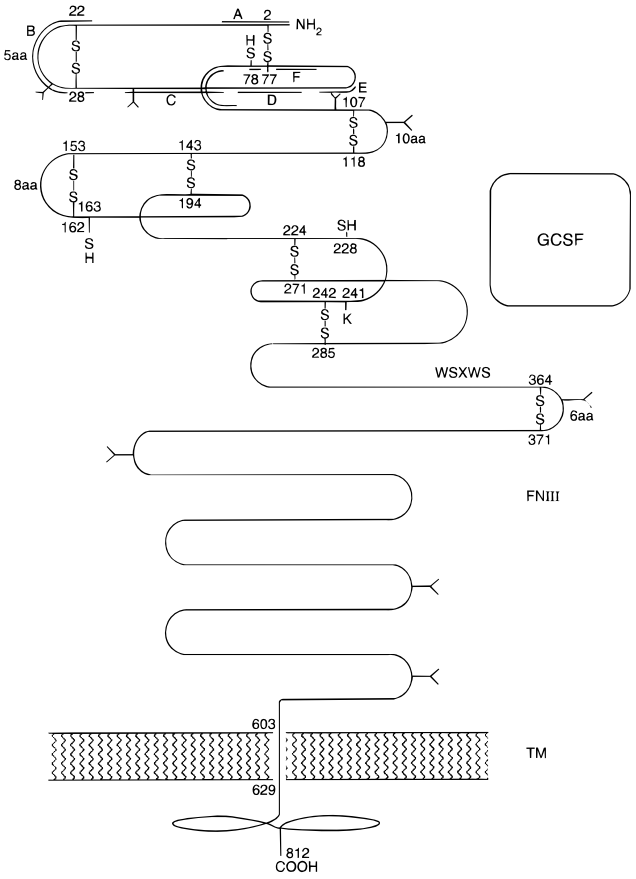
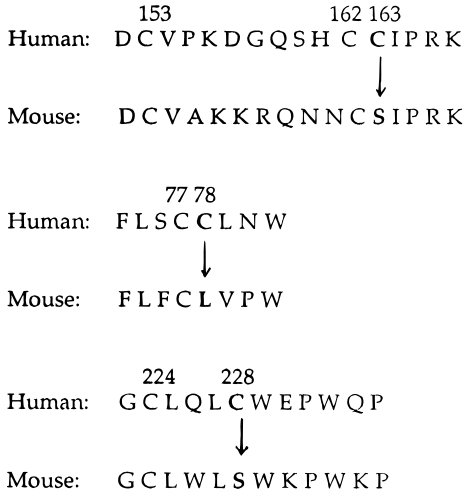


FIGURE 5: Disulfide structure and carbohydrate attachments of soluble G-CSF receptor. Overall schematical drawing is based on the disulfide determination and analysis of N-linked carbohydrates. According to the reports (Hiraoka *et al*., 1994; Haniu *et al*.,1995), G-CSF receptor binds G-CSF at cytokine homology domain (CRH) as indicated. Lines A−F show possible β-strands

Comparison between Human G-CSF Receptor and Other Related Receptors. The extracellular domain of murine G-CSF receptor does not contain any free cysteines, differing from the human receptor (Fukunaga *et al*. 1990b). In our study, these extra cysteine residues (Cys78, Cys163, and Cys228) in the human receptor were determined to be in the free state. However, since these cysteines are not conserved in the murine receptor as shown in Table 3, they may not be important for ligand binding and signal transduction. Cys-228 was able to be alkylated by *N*-ethylmaleimide under native conditions (Haniu *et al*. 1995), whereas two other free cysteines were scarcely reactive.

Table 3: Comparison of Cysteine-Containing Peptides from the Human Receptor and the Murine Receptor



[a] Bold face with arrows show cysteine residues lacking in the murine receptor.

DTNB titration in the absence or presence of denaturant also shows similar results.

Carbohydrate moieties may not directly affect ligand binding, but could contribute to overall protein stability or solubility. We have shown that this may likely be the case for the G-CSF receptor, and that the carbohydrate(s) responsible for this are N-linked. While we have not conclusively ruled out the presence of O-linked carbohydrates, these moieties would have to be very small (200−300 Da) and probably not involved in this stabilization.

The immunoglobulin superfamily structural motif is characterized by seven to nine β-strands, which form two β-sheets stacking each other (Bork *et al*., 1994; Lesk & Chothia, 1982). Assuming that the Ig-domain of G-CSF receptor comprises seven β-strands, an attempt was made to assign them into two stacking β-sheets (Figure 5). Other Ig-like domains such as gp-130, Ig-V light chain, and α-agglutinin (Bork *et al*., 1994; Wojciechowicz *et al*., 1993) can also be modeled similarly. A common structure seen within the Ig superfamily is a disulfide bond connecting β-strands B and F located near each other in the two stacking β-sheets. In the case of G-CSF receptor, Cys2 present in β-strand A form a disulfide bond with Cys77 located in β-strand F, while two cysteine residues present in β-strand C forms a disulfide bond. The observed disulfide structures between β-strands are inconsistent with this criteria for forming an Ig-like structure.

REFERENCES

Anaguchi, H., Hiraoka, O., Yamazaki, K., Naito, S., & Ota, Y. (1995) *J. Biol. Chem. 270*, 27845−27851.
Bork, P., Holm, L., & Sander, C. (1994) *J. Mol. Biol. 242*, 309−320.
Ellman, G. L. (1959) *Arch. Biochem. Biophys. 82*, 70−77.
Fukunaga, R., Ishizaka-Ikeda, E., Seto, Y., & Nagata, S. (1990a) *Cell 61*, 341−350.
Fukunaga, R., Seto, Y., Mizushima, S., & Nagata, S (1990b) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8702−8706.
Fukunaga, R., Ishizaka-Ikeda, E., Pan, C.-X., Seto, Y., & Nagata, S. (1991) The *EMBO J. 10*, 2855−2865.
Habeeb, A. F. S. A. (1972) *Meth*ods *Enzymol 25*, 457−464.
Haniu, M., Ackline, C., Kenney, W. C., & Rohde, M. F. (1994) *Int. J. Protein Peptide Res. 43*, 81−86.
Haniu, M., Horan, T., Arakawa, T., Le, J., Katta, V., & Rohde, M. F. (1995) *Arch. Biochem. Biophys. 324*, 344−356.
Hibi, M., Murakami, M., Saito, M., Hirano, T., Taga, T., & Kishimoto, T. (1990) *Cell 63*, 1149−1157.
Hiraoka, O., Anaguchi, H., Yamazaki, K., Fukunaga, R., Nagata, S., & Ota, Y. (1994) *J. Biol. Chem. 269*, 22412−22419.
Horan, T., Wen, J., Narhi, L. O., Parker, V., Arakawa, T., & Philo, J. (1996) *Biochemistry 35*, 4886−4896.
Laemmli, U. K. (1970) *Nature 227*, 680−685.
Lesk, A. M. & Chothia, C. (1982) *J. Mol. Biol. 160*, 325−342.
Lieschke, G. J., & Burgess, A. W. (1992a) *N. Eng. J. Med. 327*, 28−35.
Lieschke, G. J., & Burgess, A. W. (1992b) *N. Eng. J. Med. 327*, 99−106.
Lipke, P. N., Chen, M.-H., Nobel, H. D., Kurjan, J., & Kahn, P. C. (1995) *Protein Sci. 4*, 2168−2178.
Marti, T., Osselet, S. J., Titani, K., & Walsh, K. A. (1987) *Biochemistry 26*, 8099−8109.
Saviege, W. B., & Fontana, A. (1977) *Methods Enzymol. 47*, 459−469.
Wojciechowicz, D., Lu, C-F., Kurjan, J., & Lipke, P. N. (1993) *Mol. Cell. Biol. 13*, 2554−2563.

BI960213U