gp130 is the common signal transducing receptor subunit for the interleukin-6-type family of cytokines. Its extracellular region (sgp130) is predicted to consist of five fibronectin type III-like domains and an NFL-terminal Ig-like domain. Domains 2 and 3 constitute the cytokine-binding region defined by a set of four conserved cysteines and a WSXWS motif, respectively. Here we determine the disulfide structure of human sgp130 by peptide mapping, in the absence and presence of reducing agent, in combination with Edman degradation and mass spectrometry. Of the 13 cysteines present, 10 form disulfide bonds, two are present as free cysteines (Cys279 and Cys444), and one (Cys397) is modified by a WSXWS motif. Of the 11 potential disulfide bonds, two are present as free cysteines (Cys279 and Cys469), and one (Cys397) is modified by a WSXWS motif. The 11 potential N-glycosylation sites, Asn61, Asn108, Asn132, Asn263, Asn357, Asn361, Asn413, and Asn442 are glycosylated but not Asn289 and Asn430. The disulfide bonds, Cys115-Cys122 and Cys31-460, are consistent with known cytokine-binding region motifs. Unlike granulocyte colony-stimulating factor receptor, the connectivities of the four cysteines in the NH2-terminal domain of gp130 (Cys115-Cys122 and Cys31-460) are consistent with known superfAMILY of Ig-like domains. An eight-residue loop in domain 5 is tethered by Cys436-Cys444. We have created a model predicting that this loop maintains Cys436 in a reduced form, available for ligand-induced intramolecular disulfide bond formation. Furthermore, we postulate that domain 5 may play a role in the disulfide-linked homodimerization and activation process of gp130.
277 amino acids. The amino acid sequence of murine (23) and rat gp130 (24) share an overall sequence identity of 85 and 88%, respectively, with that of human gp130. There are 11 potential N-linked glycosylation sequons in the extracellular domain of the predicted 101-kDa human gp130 (11), suggesting that the mature ~130-kDa gp130 (11) is highly glycosylated. The extracellular region of gp130 has a modular structure consisting of six domains of ~100 amino acids each. The NH₂-terminal domain (domain 1) is predicted to be a member of the immunoglobulin superfamily (Ig-like) (25, 26), which has “Greek key” β-sheet topology (27). In the Ig-like fold, neighboring β-strands form hydrogen bonds in an antiparallel fashion to form a β-sheet, and two β-sheets pack against each other to produce a hydrophobic core. Domains 2–6 of the gp130 are classified as fibronectin type III (FN III)-like modules, a subclass of the β-sandwich fold (26). The topology of these domains is similar to those of the Ig-like modules, the notable exception being the “sheet switching” of β-strand D from the first β-sheet of an Ig-like domain to form β-strand C’ on the second β-sheet of FN III domains. The two membrane distal FN III domains (domains 2 and 3) form the cytokine-binding region (CBR) that is characteristic of class I cytokine receptors (28, 29). CBRs are characterized by two conserved disulfide bonds in their NH₂-terminal domain and a WSXWS motif in their COOH-terminal domain.

The functional anatomy of the extracellular region of gp130 is still poorly understood. A truncated form of gp130 lacking the membrane-proximal FN III modules and the cytoplasmic and transmembrane domains has been shown to bind a complex of either IL-6,IL-6R or LIF,LIF receptor (30, 31). Valγεδ in the BC loop of the COOH-terminal domain of the CBR (domain 3 of gp130) has been implicated in the interaction of gp130 with IL-6,IL-6R (32) and IL-11-IL-11 receptor (33). The location of this residue corresponds to a tryptophan residue in the growth hormone receptor that has been shown to be critical for ligand binding (34), suggesting a conserved mode of ligand binding among the cytokine receptor superfamily. The Ig-like module of gp130 (domain 1) has been shown to be involved in the interaction of gp130 with ligands that induce homodimerization of gp130 (IL-6, IL-6, and IL-11) and essential for the formation of high affinity hexameric complexes (15, 35). This domain is thought to bind sites III in IL-6 and IL-11 (36) and is essential for the formation of high affinity hexameric complexes (15, 35). The three membrane-proximal FN III modules of gp130, although not directly involved in ligand binding, may be required for gp130 dimerization by ligands such as IL-6 (40) and IL-11 (41) or transmembrane signaling events such as stabilization and/or orientation of transmembrane receptor dimers (5).

The structure of the COOH-terminal domain of the gp130 CBR (domain 3) has been determined by NMR (42), and the complete CBR (i.e. domains 2 and 3) by x-ray crystallography (43); however, there is no complete structure for gp130 or any member of the gp130 family of receptors. To elucidate the tertiary structure of the gp130 extracellular region, we have purified human sggp130 using a Chinese hamster ovary (CHO) cell expression system (13, 44). This form of human sggp130 contains 11 potential N-linked glycosylation sites and 13 cysteine residues. Four cysteines are located in the Ig-like domain (domain 1), four are in the NH₂-terminal FN III domain of the CBR (domain 2), one is in the COOH-terminal FN III domain of the CBR (domain 3), one is in domain 4, and three are in domain 5. Previously, we have shown that affinity purified sggp130 bound a binary complex of IL-6,IL-6R with a 2:2:2 stoichiometry (13). Here, reverse phase HPLC peptide mapping under reducing and nonreducing conditions in combination with mass spectrometric and NH₂-terminal sequence analysis was used to determine the cysteine connectivities and carbohydrate attachment sites for sggp130.

**EXPERIMENTAL PROCEDURES**

Materials—4-Vinylpyridine was purchased from Aldrich. Trifluoroacetic acid (HPLC/Spectro grade) was from Pierce. Dithiothreitol (DTT, Ultra grade) was obtained from Calbiochem-Novabiochem. Sequencing grade trypsin (EC 3.4.21.4), chymotrypsin (EC 3.4.21.1), pepsin (EC 3.4.25.1), and neuraminidase (EC 3.2.1.18) were from Roche Molecular Biochemicals. An endoglycosidase preparation from *Bacillus subtilis* (endo F; Sigma) was from Flavobacterium meningosepticum (45) containing three β-N-acetylglucosaminidase F (endo F) activities (F1, F2, and F3) as well as peptide N-glycosidase F (46, 47) was a kind gift from Dr. G. E. Norris (Massey University, New Zealand). HPLC grade solvents were from Mallinkrodt, and all other buffers and reagents (Analar grade) were from BDH. Anti-human gp130 monoclonal antibodies AM64, GPZ55, GPX7, and GPX22 (11, 48) were from Dr. K. Yasukawa (TOSOH, Tokyo, Japan). All buffers and solutions were prepared with deionized water purified and polished by a tandem Milli-RO and Milli-Q system (Millipore).

**Chromatography**—Reverse phase (RP)-HPLC was performed using either a Vydac 5-μm, 300 A octadeyl silica column (inner diameter, 250 × 4.6 mm; Vydac RP-300 T, 300 A octysilica column (inner diameter, 100 × 4.6 mm) (Applied Biosystems), operated at a flow rate of 0.5 ml/min. For capillary RP-HPLC, a fused silica (inner diameter, 50 × 0.2 mm) was packed “in-house” with Brownlee RP-300 7-μm, 300 A octysilica (49). For SEC, an analytical Superose 12 column (inner diameter, 300 × 10 mm) operated at 0.5 ml/min was employed. Rapid desalting and buffer exchange of proteins was performed using a Fast-desaltingTM column (Phenomenex, Torrence, CA). Purification of the Extracellular or Soluble Domain of the Recombinant IL-6 Signaling Receptor gp130 (sgp130)—sgp130 was purified from the conditioned medium of CHO (G16) cells stably transfected with a plasmid (PECEdhfrgp620) that encodes the extracellular domain of gp130 truncated at amino acid 621 (44). Briefly, transfected CHO cells were grown in a large scale fermentation apparatus with a working volume of 1.25 liters (New Brunswick Colligent plus fermenter). Cell conditioned medium was concentrated (20-fold) by ultrafiltration using a Sartoon Mini apparatus (Sartorius, Germany) fitted with a 30,000 molecular weight cut-off filter. sgp130 was purified from concentrated CHO cell medium by binding to a 10-ml column of AM64-Sepharose and eluting sgp130 with 4 M MgCl₂. sgp130 was further purified by preparative SEC (>95% pure, as judged by SDS-PAGE and Western blot analysis), and bst sgp130-containing fractions were pooled. The bst sgp130 was exchanged using a Fast-desaltingTM column prior to storage at ~20 °C. SDS-Polyacrylamide Gel Electrophoresis and Isoelectricfocusing—Samples were analyzed by SDS-PAGE on precast 4–20% polyacrylamide gels (Novex) according to the manufacturer’s instructions. Protein bands were visualized by staining with Coomassie Brilliant Blue as described (50). Glycosylation of sgp130—sgp130 (400 μg, 1 mg/ml) in 0.1 M sodium phosphate buffer, pH 7.4, containing 0.025 M EDTA was treated with 1 unit of neuraminidase and/or endoglycosidase mixture (37 °C, 16 h).

Labeling of Free Cysteine Residues in sgp130—Prior to desulfurization, free cysteine residues in sgp130 (400 μg) were treated with a 5-fold molar excess of 4-vinylpyridine in 0.1 M Tris-HCl, pH 8.4, for 1 h at 25 °C in the dark. The modified protein was buffer exchanged using a Fast-desaltingTM column operated at 1 ml/min. 4-Vinylpyridine was chosen as the alkylating agent because S-β-(4-pyridyl methyl)cysteine (Pec) containing peptides can be identified by their characteristic absorption spectra at 254 nm (51) and are also readily identified during ESI-MS by the presence of an ion of m/z 106 corresponding to collision-induced dissociation (CID), characteristic of the protonated S-pyridylthioly moiety (50).

Peptide Mapping of sgp130—sgp130 (400 μg) in 2 ml of 1% (v/v) ammonium bicarbonate, containing 2 mM calcium chloride, was digested at 37 °C for 16 h with either trypsin or chymotrypsin at an enzyme to substrate ratio of 1:20. For pepsin digestion, sgp130 (400 μg,
~4 nmol) in 5% formic acid was digested for 1 h at 37 °C using an enzyme to substrate ratio of 1:20. Resultant peptides were fractionated by RP-HPLC on either a Vydac C18 column or a Brownlee RP-300 column at 0.5 ml/min at 45 °C. The column eluent was split (~1:180) post-detector, using a stainless steel Tee-union (52), whereas the remainder (99.4%) was collected for further analysis.

For disulfide-containing peptide identification, a portion of the digest (25%) was reduced with 10 mM DTT at 45 °C for 1 h and rechromatographed under identical conditions. Peaks whose retention times shifted upon reduction were subjected to NH₂-terminal sequence analysis.

NH₂-terminal Sequence Analysis—NH₂-terminal sequence analyses were performed using a Hewlett-Packard biphasic NH₂-terminal protein sequencer (model G1005A, Hewlett-Packard) using version 3.0 chemistry as described (50).

Electrospray Mass Spectrometry—On-line MS analysis of peptide fractions was performed using either a Finnigan LCQ quadrupole ion trap mass spectrometer or a TSQ-700 triple quadrupole mass spectrometer, both equipped with an ESI source as described (50, 53). S-Pyridylethylated peptides were identified by parent-ion scanning using the TSQ mass spectrometer by monitoring all peptides for the labile pyridine group (m/z 106) following CID. Source CID/single ion monitoring using the LCQ ion trap mass spectrometer was employed to identify S-pyridylethyl cysteine-containing peptides as described (52). Peptides were identified from their CID product ion spectra using either the Finnigan Xcalibur-Biomass software or Protein Prospector MS-Product algorithm (ProteinProspector Pacific-Rim mirror).

Matrix-assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry—Peptides analyzed by MALDI-TOF mass spectrometry (Compact MALDI-IV fitted with a 337 nm laser, Kratos) were co-crystallized with α-cyano 4-hydroxycinnaminic acid (16 mg of matrix/ml aqueous 60% acetonitrile/0.1% (v/v) trifluoroacetic acid). Matrix (0.5 μl) was deposited onto a clean sample slide immediately followed by 0.5 μl of peptide fraction. Spectra were calibrated using the external standards angiotensin (1297.5 [M+H]⁺) and a matrix-derived ion (173.17 [M+H]⁺).

Homology Modeling—Domain 5 of sgp130 was modeled using the two FN III domains from the structure of the cytoplasmic tail of human integrin α₅β₅ (54), which showed 19 and 16% amino acid sequence identity. These particular FN III domains were chosen as templates to minimize the insertions and deletions in the loop regions. The coordinates for the template were taken from the Protein Data Bank entry 1g3. Domain 5 of sgp130 was manually aligned with the template of the cytoplasmic tail of integrin α₅β₅ crystal structure (54), conserving the hydrophobic and sequence patterns of the FN III β-sheets.

Fig. 1. SDS-PAGE and IEF of recombinant sgp130. A, SDS-PAGE 4–20% acrylamide precast gel. Lane 1, protein molecular weight standards (SeeBlue™, Novex); lane 2, affinity-purified sgp130; lane 3, neuraminidase-treated sgp130; lane 4, sgp130 treated with a mixture of neuraminidase and endoglycosidase. The sample load was ~10 μg. B, IEF-PAGE (pH 3–10). Lane 1, protein IEF standards (Broad pl calibration kit; Amersham Pharmacia Biotech). Lanes 2–4 were the same as for A. Proteins were visualized by Coomassie Blue staining (50).

Fig. 2. Comparative peptide mapping of a tryptic digest of sgp130. A, sgp130 was digested with trypsin as described under “Experimental Procedures,” and generated peptides were fractionated by RP-HPLC using a Vydac octadecyl 5-μm, 300 Å silica column (inner diameter, 250 × 4.6 mm) with a gradient from 0–70% B, where solvent A was 0.1% (v/v) trifluoroacetic acid and solvent B was 0.09% trifluoroacetic acid (v/v) acetonitrile. Flow rate was 0.5 ml/min, and temperature was 45 °C. Peptides were monitored with four different wavelengths (215, 254, 280, and 290 nm), and spectral data were obtained on-the-fly by diode array detection. Cysteine-containing peptide fractions identified by NH₂-terminal sequence analysis and MS analysis are annotated by the letter T (see also Table I). All peptide fractions were analyzed, and the obtained sequence information is in perfect agreement with the published amino acid sequence (11) (data not shown). B, chromatography of a portion (25%) of the tryptic digest of sgp130 from A following reduction with DTT (see “Experimental Procedures”). Chromatographic conditions are identical to A. The arrow indicates the position of T4 when chromatographed under nonreducing conditions (see A). Left inset, diode array spectra of T1 and T2. S-Pyridylethylated peptides (T1 and T2) were identified by their characteristic spectral properties (maxima at A_{254} nm) indicated by an arrow. Right inset, SDS-PAGE analysis of tryptic peptide fraction T5 and T6. Fractions T5-T6 from Fig. 2 (A) were pooled (10% of fraction T5 (1 ml) and T6 (0.9 ml)) and concentrated to 10 μl by centrifugal lyophilization, diluted 3:1 (v/v) with nonreducing SDS sample buffer, and loaded onto a 4–20% polyacrylamide precast gel, as described in Fig. 1. Protein was visualized by silver staining (76). Lane 1, protein molecular weight standards (Mark12™, Novex); lane 2, undigested gp130; lane 3, fraction T5-T6. The sample load was ~100 ng.
### Human Signal Transducer gp130

#### Table I

| Peptide fraction | Sequence (residue number) | Observed mass | Calculated mass | Δ mass |
|------------------|---------------------------|---------------|----------------|-------|
| T1               | CMKEDGK (279–285)         | 914.6 (914.4<sup>a</sup>) | 809.9         | +104.7 |
| T2               | CILITVPVAGDGPSPESIK (469–489) | 2314.6     | 2210.5         | +104.1 |
| T3               | CYSDAVLTPACDFQATIVPMDLKLK (387–409) | 2562.7 (2443.0<sup>b</sup>) | 2443.8         | +118.9 |
| T4               | T4a YILEWCVLSDK (431–441) | 3221.3     | 3222.6         | −1.3   |
| T4b               | PECIDMQEDEGTVHR (442–457) | 5448.1     | 5448.0         | +0.1   |
| T7               | FADCK (147–151)           | 1584.7     | 1584.8         | −0.1   |
| T8/C1            | DTPSCDVTYDSTVFNVNWEAENALGKVTSDHINFDPVYK (155–197) | 2867.9     | 2868.4         | −0.5   |
| T8/C2            | GMDY (32–35)              | 1046.4     | 1046.4         | +0.0   |
| P1               | TDLAGLNLQCTGNLTF (70–85)  | 3936.1     | 3935.6         | +0.5   |
|                  | TISSLPKPEKRKDSSTVNEKGKRMDWGGRETHL (98–132) | 4100.0     | 4100.0         | 0.0    |

<sup>a</sup> Peptide fraction labeled according to the protease digest or subdigestion, T, trypsin; C, chymotrypsin, P, pepsin, TC, tryptic peptide subdigested with chymotrypsin.

<sup>b</sup> Peptide sequences identified from MS and MS/MS spectral data using the programs Xcalibur-Biomass, MS product as well as manual (de novo) interpretation. Underlined sequences were confirmed by Edman sequencing. One-letter abbreviations were used for amino acids. C-Cys, S-cysteynl-cysteine.

<sup>c</sup> Residue numbers denote amino acid positions in the sequence of the mature protein (11).

<sup>d</sup> Numbers refer to the observed average masses of the peptide(s). Numbers in parentheses refer to the observed reduced average masses.

<sup>e</sup> Numbers refer to the sums of the calculated average masses of the deduced peptide(s).

*Tryptic peptide fraction T3 contained two peptide ions with 97% (2562.7 Da) being S-cysteynalted and 3% (2548.4 Da) being S-pyridylethylated.

The MODELLER program (55) was used to generate a model of FN III domain 5. The quality of the model was assessed as described (56) in particular using the ProsaII program (57). A disulfide restraint between Cys<sup>336</sup> and Cys<sup>444</sup> in FN III domain 5 was introduced in accordance with our experimental results. Over 100 models of domain 5 were generated with the final model being chosen on the basis of the quality checks described. The final model of gp130 domain 5 follows integrin α<sub>β</sub><sub>4</sub> domain 1 in the B-C and C-E loops, and follows domain 2 in the C-C' loop and G strand. The ProsaII Z-score of the final model is 4.70 and 6.26, respectively (Table I). The yield of PTH-Pec in peptide T3 further reduced the complexity. Taken together, these data indicate that the heterogeneity of the sgp130 preparation is primarily due to differential N-linked glycosylation.

The immunological integrity of sgp130 was evaluated by Western blot analysis using a panel of anti-sgp130 monoclonal antibodies (AM64, GPX7, GPX22, and GPZ35; Ref. 48). Although monoclonal antibody AM64 recognized sgp130 under both nonreducing and reducing conditions, GPX7, GPX22, and GPZ35 (35) only recognized conformational epitopes of sgp130 under nonreducing conditions (data not shown). The functional integrity of the sgp130 used in this study was assessed by its ability to bind the IL-6-sIL-6R binary complex as well as its ability to form a stable IL-6-sIL-6R-sgp130 ternary complex (13). The mass of the ternary IL-6 receptor complex observed by analytical SEC (~380 kDa) (data not shown) was in good agreement with a value calculated for a hexameric complex of two molecules each of IL-6, IL-6R, and gp130 (13).

**Identification of Unpaired Cysteine(s) in sgp130**—Soluble gp130 was alkylated with 4-vinylpyridine at pH 8.5, digested with trypsin, and then chromatographed by RP-HPLC using a trifluoroacetic acid/acetonitrile solvent system (Fig. 2). Using ESI-MS/MS, the parent-ion scan of the total tryptic digest of the S-pyridylethylated sgp130 identified three Pec containing peptides with molecular masses of 914.6 Da (T1), 2314.6 Da (T2), and 2548.4 Da (T3) (data not shown). Two of these Pec-containing peptides (T1 and T2) were also confirmed by their UV absorbance at A<sub>254 nm</sub> (see tryptic peptide fractions T1 and T2 in Fig. 2A, inset).

Edman degradation identified tryptic peptide T1 as residues Cys<sup>279</sup>–Lys<sup>285</sup>, T2 as residues Cys<sup>469</sup>–Lys<sup>489</sup>, and T3 as residues Ser<sup>387</sup>–Lys<sup>409</sup>. Cysteines were observed as PTH-Pec derivatives at positions 279, 469, and 397 in these peptides, respectively (Table I). The yield of PTH-Pec in peptide T3 (Cys<sup>397</sup>) was significantly lower than that observed for the PTH-Pec derivatives in T1 and T2. Further confirmation of the
identity of these peptides was obtained by rechromatography and subsequent ESI-MS/MS analysis (Table I). Automated CID MS/MS analysis of the doubly charged ion of peptide T2 (m/z 1158.9) localized the difference in mass observed from the calculated mass to Cys469 (Fig. 3).

To evaluate the extent of S-pyridylethylolation of peptide fraction T3, this fraction (from Fig. 2A) was rechromatographed using a capillary RP-HPLC system coupled on-line to ESI-MS/MS (Fig. 4A). Although Edman degradation of T3 revealed only one sequence (residues Ser387–Lys409), examination of the average mass spectrum of the major total ion current peak (Fig. 4B) reveals two peptide ions with molecular masses of 2548.4 and 2562.7 Da that differed by 14 Da. The charge state distribution patterns and relative ion intensities of these ions are very similar. The parent-ion scan of the 2548.4-Da peptide ion (Fig. 4C) revealed that this peptide was S-pyridylethylated (Figs. 4C and 4D). Because the mass difference between S-pyridylethyl and S-cysteinyl moieties is 14 Da, the possibility that Cys397 in T3 may also be cysteinylated was further examined.

To this end, peptide fraction T3 was reduced with DTT and immediately reanalyzed by capillary RP-HPLC/ESI-MS. It can be seen from the total ion current profile (Fig. 5A) that the minor peak (fraction I) corresponds to the 2548.4-Da PeC-modified form of tryptic peptide Ser387–Lys409 (compare Figs. 5B and 4D). The major peak (Fig. 5A, fraction II), which was resolved from the PeC-pectaining fraction, has a mass of 2343.9 Da (Fig. 5C), indicating a loss of 119.7 Da from the parent 2562.7-Da peptide upon reduction (compare Figs. 5C and 4D). Further analysis of the doubly charged ion of T3 (m/z 1282.3) by automated CID-MS/MS identified the peptide as Ser387–Lys409 containing the additional mass of 118.9 Da located at Cys397 (Fig. 6). Taken together, these data indicate that Cys397 is cysteinylethylated (119 Da) and that this cysteine modification is readily removed upon reduction with DTT. A comparison of the two forms of the tryptic peptide fraction T3 (2548.4- and 2562.7-Da peptide ions) in sgp30 indicated that the major cysteine modification was S-cysteinylionation (97%) with ~3% as the free thiol (as revealed by S-pyridylethylation following treatment with 4-vinylpyridine).

**Determination of Disulfide Bonds of sgp30 by Proteolytic Digestion and Peptide Mapping**—A portion (25%) of the tryptic digest of S-pyridylethylated sgp30 was reduced with 10 mM DTT at pH 8.0 and rechromatographed by RP-HPLC using chromatographic conditions identical to those employed for the separation of the nonreduced S-pyridylethylated tryptic digest of sgp30 (Fig. 2). A comparison of panels A and B in Fig. 2 revealed that, upon reduction, only peptide fraction T4, had an altered retention time, suggesting that it is a disulfide-containing peptide. Edman degradation of fraction T4 (Table I) identified two peptides, T4a APXTTGWQEDGTVHR (residues 442–457) and T4b YILEWXVLSDK, (residues 431–441).

Further analysis of peptide fraction T4 by MALDI-TOF MS yielded an average molecular mass of 3220.7 Da (Fig. 7), which is in good agreement with the mass observed by ESI ion trap MS (3222.4 Da). The ion (m/z 1611.9) observed by MALDI-TOF MS is the doubly charged form of the parent peptide ion (m/z 3221.7). In addition, two further peptide ions of (m/z 1856.0) and (m/z 1368.6) were observed (Fig. 7). Presumably, these peptide ions are the pseudomolecular ions corresponding to reduced forms of peptide T4a and T4b (Table I), respectively, that have arisen by prompt fragmentation (58) of the disulfide linkage.

No peptides from the NH2-terminal region of sgp30, including four potential disulfide-containing peptides, were obtained using this approach. Close inspection of the peptide map (Fig. 2A) revealed a shallow peak (peptide fraction T5-T6) eluting late in the gradient. Edman degradation of these pooled fractions yielded two peptide sequences in low yield, one corresponding to the NH2 terminus of sgp30 and the other commencing at residue 63 (data not shown). Nonreducing SDS-PAGE analysis of peptide fraction T5-T6 revealed a single protein band of ~50 kDa (Fig. 2A, inset), suggesting that this fraction contained at least domains 1 and 2 of sgp30, held together by disulfide bonds, consistent with a partially digested NH2-terminal “core” polypeptide.

To establish the cysteine connectivities of the Ig-like domain of the molecule, S-pyridylethylated sgp30 was subjected to extensive digestion (48 h) using a high concentration of trypsin (enzyme to substrate ratio 1:5). Peptide mapping of this trypsin digest (under reducing and nonreducing conditions (data not shown) revealed the identity of two disulfide-containing peptide fractions. One of these (T7) was identified as Cys350–Cys160 (Table I), whereas the other fraction (T8) contained cysteine residues Cys6, Cys26, Cys32, and Cys81 linked by disulfide bonds. To determine the connectivities of Cys6, Cys26, Cys32, and Cys81, tryptic peptide fraction T8 was subdigested with chymotrypsin. Using a combination of Edman degradation...
and MS analysis of the resultant peptides, disulfide bond linkages were established between Cys 6–Cys32 and Cys 26–Cys81 (Table I).

An additional cysteine connectivity in sgp130 was identified following S-pyridylethylolation, deglycosylation (see “Experimental Procedures”), and then digestion with pepsin. Peptide mapping of the resultant peptic digest of S-pyridylethylated sgp130 (400 μg) under reducing and nonreducing conditions (data not shown) revealed the presence of a disulfide-containing peptide P1 (Table I). Edman degradation and MS analysis established the identity of peptide P1 as a linear peptide (residues Ile98–Leu132) containing the disulfide linkage Cys 112–Cys122 (Table I). NH₂-terminal sequence analysis and MS analysis of the peptic peptides of sgp130 reconfirmed the identity of the S-pyridylethylated Cys279 and Cys469 and S-cysteinylated Cys397.

N-linked Glycosylation Sites of sgp130—Of the 11 possible N-glycosylation sites (Asn-Xaa-(Ser/Thr) motifs) in the extracellular domain of gp130, nine were found to be glycosylated as determined by the nonappearance of asparagine residues at positions 21 (NFT), 61 (NRT), 109 (NLT), 135 (NYT), 205 (NLS), 357 (NYT), 361 (NAT), 531 (NYT), and 542 (NET) at the expected cycle during automated Edman degradation of tryptic digest generated peptides (Table II). Further confirmation of N-glycosylated asparagine residues 61, 205, 357, 361, 531, and 542 was provided by Asn/Asp conversion in both tryptic and peptic sgp130 peptides following endoglycosidase F treatment (59) (see Asn/Asp conversion at position 109 in peptide P1 (Table I)). Asparagine residues 224 (NPS) and 368 (NLT) were identified as PTH-asparagine during Edman degradation indicating that these residues are not N-glycosylated.

DISCUSSION

To date there is no complete three-dimensional structure for the extracellular domain of gp130 or any member of the gp130 family of receptors. The results presented here for the cysteine connectivity pattern and carbohydrate attachment sites of sgp130 provide a foundation for the determination of the tertiary structure of this molecule and, ultimately, of the IL-6-IL-6R-gp130 complex. Mass spectrometric and NH₂-terminal sequence analysis of proteolytic digests of sgp130 identified five disulfide linkages (Cys6–Cys32, Cys26–Cys81, Cys112–Cys122, Cys150–Cys160, and Cys436–Cys444), two free cysteines (Cys279 and Cys469), one cysteinylated cysteine (Cys397), and nine N-
linked glycosylation sites (asparagines 21, 61, 109, 135, 205, 357, 361, 531, and 542), whereas the two potential N-glycosylation sites at Asn 224 (NPS) and Asn 368 (NLS) have no carbohydrate moieties attached. Previous studies have shown that NPS is not favorable for N-linked glycosylation, and the NLS is only glycosylated in 40% of the instances studied (60). Fig 8 shows a schematic representation of the disulfide bonding pattern and N-glycosylation sites of gp130.

The two disulfide linkages of the NH2-terminal Ig-like domain of gp130 (Cys 6–Cys 32 and Cys 26–Cys 81) (Fig. 9A) are found between the A strand and B–C loop and between the B and F strands, respectively, with the second disulfide linkage characteristic of a typical I-type Ig-like fold (25). The amino acid sequences of the Ig-like domains of gp130 and GCSFR are homologous, suggestive of an identical fold, yet their disulfide patterns differ (Fig. 9A). The percentage amino acid sequence identity is 19% over the whole Ig-like domain, with 30% identity over the region encompassing the first three cysteines (Fig. 9A).

The two consecutive disulfide linkages in the NH2-terminal FN III domain of the CBR of gp130 (Cys 112–Cys 122 and Cys 150–Cys 160) correspond to those experimentally determined by a peptide mapping approach for the CBR of other class I cytokine receptor families, namely GCSFR (61, 62), growth hormone receptor (63), and leptin receptor (64). These cysteine connectivities have also been confirmed by the x-ray crystal structure analysis of the CBR of gp130 (43), the growth hormone-growth hormone receptor complex (65), the GCSF-GCSFR complex (66), the growth hormone receptor-prolactin receptor complex (67), the erythropoietin receptor-peptide agonist complex (68), the IL-4–IL-4R complex (69), and IL-12R (70). An inspection of the gp130 CBR three-dimensional structure reveals that the N-linked carbohydrate moieties found at Asn 109 (A strand, domain 2), Asn 135 (C strand, domain 2), and Asn 205 (A strand domain 3) are consistent with these asparagines being solvent exposed, whereas the potential N-glycosylation site at Asn 224 would not be expected to be glycosylated because this asparagine residue is partially buried in the structure (43). Cys 279, Cys 397, and Cys 469 located in domains 3, 4, and 5, respectively, do not form disulfide bonds and occur as either a free thiol or in the case of Cys 397, a cysteinylated form (Fig. 9A). Previously, we have reported the presence of both free and cysteinylated cysteine residues in the human IL-6R (52). The occurrence of free cysteines has also been reported for GCSFR
and leptin receptor (64). Cys279, which is present as a free thiol in sgp130, is not conserved among the species variants of gp130 whose sequences have been determined to date (i.e. mouse (23), rat (24), chicken (71), and frog (2)). Analysis of the gp130 CBR crystal structure (43) indicates that Cys 279 is situated on the B strand of domain 3 and is partially buried. Cys397, which is located at the end of the G strand of domain 4 and likely forms part of the hinge region with domain 5, was found to be S-cysteinylated. This finding is consistent with Cys397 being solvent exposed.

Cys469, located in domain 5, is the only free cysteine conserved in all species of gp130 studied to date. As such, Cys469 may be involved in the homodimerization related intracellular signaling of this molecule that is reported to involve disulfide-bond formation (12). The free thiol Cys469 is located at the NH₂ terminus of the F strand flanked by conserved lysine and tyrosine residues and would be predicted to be solvent exposed and hence modified. Our homology modeling data predict that although the free thiol Cys469 is orientated away from the hydrophobic core of domain 5, it is shielded from the solvent by a short eight-residue “protective loop” tethered by the disulfide bond Cys436–Cys444 in strands C and C′ of domain 5 (Fig. 10).

A conformational change in gp130 would be required to reposition this protective loop and expose Cys469 to possible disulfide bond formation with another gp130 molecule. Müller-
Newen and colleagues (73) have recently demonstrated, using a panel of monoclonal antibodies, that the enforced dimerization of gp130 dimerization alone is not sufficient for receptor activation and that an additional conformation requirement is needed. A similar phenomenon has also been reported for the erythropoietin receptor (74) and is also applicable to other types of receptors from both eukaryotic and prokaryotic origins (75), where dimerization alone is not sufficient for receptor activation and an additional conformation requirement is needed. A similar phenomenon has also been reported for the erythropoietin receptor (74) and is also applicable to other types of receptors.

**Acknowledgments—**We thank Dr. H. Ji for assistance in production of sgp130 used in this study, Dr. G. E. Reid, J. Eddes, and D. F. Frecklington for help in tandem MS/MS, and Dr. M. J. Layton and A. R. Cole for critical reading of the manuscript.

**REFERENCES**

1. Simpson, R. J., Hammacher, A., Smith, D. K., Matthews, J. M., and Ward, L. D. (1996) *Protein Sci.* 5, 929–935
2. Taga, T., and Kishimoto, T. (1997) *Annu. Rev. Immunol.* 15, 797–819
3. Heinrich, P. C., Behrmann, I., Muller-Newen, G., Schaper, F., and Graeve, L. (1998) *Biochem. J.* 334, 297–314
4. Elson, G. C., Leclerc, E., Guillet, C., Chevalier, S., Plun-Favreau, H., Froger, J., Suard, I., de Cognac, A. B., Delphes, Y., Bonnefoy, J. Y., Gauchat, J. F., and Gascou, H. (2000) *Nat. Neurosci.* 3, 867–872
5. Bravo, J., and Heath, J. K. (2000) *EMBO J.* 19, 2399–2411
6. Kishimoto, T., Akira, S., Narazaki, M., and Taga, T. (1995) *Blood* 86, 1243–1254
7. Yoshida, K., Taga, T., Saito, M., Sueyoshi, S., Kumanogoh, A., Tanaka, T., Fujisawa, H., Hirata, M., Yamagami, T., Nakahata, T., Hirabayashi, Y., Yoneda, Y., Tanaka, K., Wang, W. Z., Mori, C., Shiota, K., Yoshida, N., and Kishimoto, T. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 407–411
8. Betz, U. A. K., Bloch, W., van den, B. M., Yoshida, K., Taga, T., Kishimoto, T., Addicks, K., Rajewsky, K., and Muller, W. (1998) *J. Exp. Med.* 188, 1955–1965
9. Hirota, H., Chen, J., Betz, U. A., Rajewsky, K., Gu, Y., Ross, J., Jr., Muller, W., and Chien, K. R. (1999) *Cell* 97, 189–198
10. Hirota, H., Yoshida, K., Kishimoto, T., and Taga, T. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 4862–4866
11. Hibii, M., Murakami, M., Saito, M., Hirano, T., Taga, T., and Kishimoto, T.
