Peptide, Disulfide, and Glycosylation Mapping of Recombinant Human Thrombopoietin from Ser1 to Arg246

Ross C. Hoffman,* Henrik Andersen, Kathleen Walker, Jonathan D. Krakover, Sunil Patel, Michael R. Stamm, and Sherri G. Osborn

Department of Biological Structure, ZymoGenetics, Inc., Seattle, Washington 98102

Received May 6, 1996; Revised Manuscript Received August 26, 1996[⊗]

ABSTRACT: Thrombopoietin (TPO) is a hematopoietic factor involved in the regulation of megakaryocytopoiesis. Full length recombinant human TPO (332 residues) has been expressed in BHK cells and purified to homogeneity using conventional means. Peptide, disulfide, and glycosylation mapping of human TPO from residues 1 to 246 has been carried out using liquid chromatography−electrospray mass spectrometry (LC−ESMS). A modification of the ramped orifice method of Carr and co-workers [Carr et al. (1993) *Protein Sci. 2*, 183−196] is employed, providing additional information for assignment of the LC−ESMS chromatograms. With the modification, b- and y-series peptide ions are produced via front-end CID which confirms the mass-based assignments. The results of our analysis of TPO indicate that the amino acid sequence of TPO 1−246 is as expected from the transfected cDNA with complete cleavage of the signal peptide. Two unique disulfides are formed between the four cysteines in the cytokine domain of TPO: Cys7−Cys151 and Cys29−Cys85. The glycosylation map indicates the position, occupancy, and structures of the *N*- and *O*-glycans in TPO 1−246. In addition, site specific structural characterization of the PNGase F-liberated *N*-glycans has been performed following purification by high-pH anionic exchange chromatography with pulsed amperometric detection (HPAEC-PAD); the results corroborate the LC−ESMS data. The *N*-glycans are of the complex type with the core-fucosylated disialylated biantennary and trisialylated triantennary structures predominating. The *O*-glycans are of the mucin type with the monosialylated and disialylated GalGalNAc-S/T structures predominating. Furthermore, we propose that the C-terminal domain of TPO be further divided into two domains on the basis of sequence homology among the cloned sequences and glycosylation/structural features: an *N*-glycan domain (154−246) and an *O*-glycan domain (247−332).

Thrombopoietin (TPO)[1] is the primary humoral regulator of blood platelet levels. Binding of TPO to c-mpl, a member of the cytokine receptor superfamily (Bazan, 1990a; Vignon et al., 1992), stimulates megakaryocytopoiesis by promoting both the proliferation and maturation of megakaryocytes [for reviews, see Lok and Foster (1994), Kaushansky (1995a,b), and Ellis et al. (1995)]. As such, this hematopoietic factor is a potential therapeutic agent for the treatment of thrombocytopenia (Metcalf, 1994). The murine TPO cDNA was cloned using a c-mpl ligand-dependent cell line for screening an expression cDNA library; the human, canine, and rat TPO cDNAs have been isolated by standard molecular cloning techniques (Lok et al., 1994; de Sauvage et al., 1994; Bartley et al., 1994; Ogami et al., 1995). The polypeptide sequences deduced from the cDNAs are highly homologous with an average of ∼70% amino acid identity.

TPO has initially been described as possessing a two-domain structure: a cytokine domain comprising the first 153 amino acids and a larger C-terminal domain following it (de Sauvage et al., 1994; Lok et al., 1994). The cytokine domain of TPO is highly conserved among the cloned sequences with 84% amino acid identity between human and murine TPO. Significant homology to erythropoietin (EPO) has been noted in the cytokine domain with 23% amino acid identity and 50% amino acid similarity between the human sequences (de Sauvage et al., 1994). Studies with recombinant forms of human and murine TPO truncated at residue 153 (de Sauvage et al., 1994; Wada et al., 1996) indicate that the cytokine domain is responsible for c-mpl binding and is sufficient for the in vitro stimulatory activity of the molecule. Four conserved cysteine residues are found in the cytokine domain, and three of these are positionally conserved with reference to human EPO (de Sauvage et al., 1994; Lok et al., 1994). It is anticipated that the cytokine domain of TPO will have a structure similar to that proposed for its sequence homolog EPO (Boissel et al., 1993); i.e. it will be a member of the four-helix bundle family of growth factors and cytokines (Bazan, 1990b). The cytokine domain is thus defined both functionally and structurally.

* Author to whom correspondence should be addressed.

[⊗] Abstract published in *Advance ACS Abstracts*, November 1, 1996.

[1] Abbreviations: ACN, acetonitrile; BHK, baby hamster kidney; BSA, bovine serum albumin; CHO, Chinese hamster ovary; CID, collision-induced dissociation; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid disodium salt; EPO, erythropoietin; FACE, fluorophore-assisted carbohydrate electrophoresis; Fuc, fucose; Gal, galactose; GalNAc, *N*-acetylgalactosamine; G-CSF, granulocyte colony-stimulating factor; GlcNAc, *N*-acetylglucosamine; GM-CSF, granulocyte/macrophage colony-stimulating factor; Hex, hexose; HexNAc, *N*-acetylhexosamine; HPAEC-PAD, high-pH anionic exchange chromatography with pulsed amperometric detection; IL-, interleukin-; LC−ESMS, liquid chromatography−electrospray mass spectrometry; LacNAc, *N*-acetyllactosamine; LIF, leukemia inhibitory factor; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; Man, mannose; NeuNAc, *N*-acetylneuraminic acid; Onco M, oncostatin M; PE-TPO, pyridylethylated thrombopoietin; TFA, trifluoroacetic acid; TIC, total ion current; TPCK, L-1-chloro-3-(4-tosylamido)-4-phenyl-2-butanone; TPO, thrombopoietin; Tris, tris(hydroxymethyl)aminomethane; XIC, extracted ion current.

S0006-2960(96)01075-6 CCC: $12.00

TPO is unique among the cytokines identified to date in that it possesses a multidomain structure. The largely undefined C-terminal domain is separated from the cytokine domain by a conserved dibasic site (R153-R154) which may indicate that proteolytic processing plays some role in the maturation of TPO in vivo (de Sauvage et al., 1994; Lok et al., 1994). This domain is less well conserved than the cytokine domain in terms of both amino acid identity and overall length among the cloned sequences (de Sauvage et al., 1994; Lok et al., 1994; Bartley et al., 1994). Although human and murine TPO have 84% amino acid identity in the cytokine domain, they have only 62% identity in the "carbohydrate" domain. Furthermore, murine TPO is the longest (335 amino acids) and rat TPO is the shortest (305 amino acids) of the cloned sequences; the 30-amino acid deletion occurs in the C-terminal domain (Ogami et al., 1995). It has been observed that the C-terminal domain possesses four conserved N-glycosylation sequons and a region rich in serine, threonine, and proline residues, suggesting that the domain is a candidate for modification by both N- and O-glycosylation. However, neither a structural model nor a functional role has been hypothesized for the C-terminal domain.

The amino acid sequence of a protein deduced from a cloned cDNA can provide some useful structural information based on sequence homologies to related proteins or to recognition motifs for modifying enzymes. However, there are a staggering variety of co- and post-translational events that impact proteins, including proteolytic processing, amino acid modification, and disulfide bond formation. One of the most difficult protein modifications to predict accurately is glycosylation. Glycosylation may be dependent on the structural context in which the acceptor site is presented; e.g. the N-glycosylation sequon, Asn-X-Ser/Thr where X is not proline, is necessary for N-glycosylation, but it is not sufficient (Hunt & Dayhoff, 1970; Sinohara & Maruyama, 1973; Gavel & von Heijne, 1990). Furthermore, cell line-dependent variations in the available heterologous expression systems (Goto et al., 1988; Kagawa et al., 1988; Conradt et al., 1989; James et al., 1995) and bioprocess factors can affect the glycosylation state of proteins (Goochee & Monica, 1990; Goochee, 1992; Watson et al., 1994; Gawlitzek et al., 1995). It is evident from the variety and complexity of post-translational events that simple sequence analysis will fail to provide adequate information for characterizing recombinant proteins. A thorough structural characterization of a recombinant protein produced in a given expression system under a defined set of bioprocess factors is necessary, especially in the case of biotherapeutics where it may be essential that the protein be correctly modified in order to maintain its full biological activity [for review, see Cumming (1991)].

We have expressed full length human TPO (hTPO) in BHK cells and purified the secreted protein to homogeneity using conventional means. The average mass of hTPO calculated from its amino acid sequence is 35.5 kDa; however, BHK-recombinant hTPO has an average mass of 57.5 kDa as measured by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. The 62% increase in mass suggests extensive post-translational modification with carbohydrate. We provide herein an analysis of the amino acid sequence, disulfide bonds, and glycosylation state of BHK-recombinant hTPO from Ser1 to Arg246. Also presented is a modification of the ramped or stepped orifice method developed by Carr and co-workers (1993) for facile identification of glycopeptides via LC−ESMS. Our modification provides additional information to aid in the assignment of LC−ESMS chromatograms of digested proteins.

EXPERIMENTAL PROCEDURES

Expression and Purification. Full length recombinant human TPO was expressed in BHK cells (BHK 570; ATCC CRL 10314) using a stirred tank method and was purified by conventional chromatographic steps, including immobilized dye affinity, ion exchange, hydrophobic interaction, and size exclusion chromatography (R. Senderoff, personal communication). All chemicals used were of reagent or analytical grade. Purity was ≥95% as judged by SDS−PAGE, amino acid analysis, and N-terminal sequencing.

Reduction and Alkylation. TPO (5.0 mg) at a concentration of 0.4 mg/mL was reduced in 6.0 M guanidine hydrochloride, 130 mM Tris-HCl, and 263 μM EDTA at pH 7.6 containing a 25-fold molar excess of dithiothreitol (DTT) over the number of cysteine residues. The reduction reaction was incubated at 37 °C for 1 h. Following reduction, a 5-fold molar excess of 4-vinylpyridine over DTT was added and the alkylation reaction was incubated in the dark at room temperature for 2 h.

The pyridylethylated TPO (PE-TPO) was immediately desalted using reverse-phase HPLC on a 7.5 × 150 mm, 8 μm, 1000 Å PLRP-S column (Polymer Laboratories, Amherst, MA) equilibrated in 10% ACN + 0.1% TFA. Following injection, the column was washed with 10% ACN + 0.1% TFA until the guanidine hydrochloride had completely eluted, and then a 10 to 80% ACN gradient over 20 min was initiated. PE-TPO routinely eluted from 17.0 to 20.0 min; the peak was collected and lyophilized. Lyophilized PE-TPO was stored at −20 °C.

PNGase F Digestion. PE-TPO (1.0 mg) was treated with recombinant *F. meningosepticum* PNGase F (Oxford GlycoSystems, Rosedale, NY) under the following conditions: 1.0 mg/mL protein in 20 mM sodium phosphate and 50 mM EDTA at pH 7.5, 94 u/mL PNGase F, and 37 °C for 21 h. The PNGase F-digested PE-TPO was stored at −20 °C and thawed as needed.

Trypsin Digestion. Samples (0.5 mg) of native TPO, PE-TPO, and PNGase F-treated PE-TPO were digested with sequencing grade trypsin (Boehringer-Mannheim, Indianapolis, IN) under the following reaction conditions: 1.0 mg/mL protein in 100 mM Tris, 100 μM TPCK, and 10% ACN at pH 8.5; 1:50 trypsin:protein; and 37 °C for 48 h. Trypsin was reconstituted in 1 mM HCl at 1 mg/mL just prior to initiating the digest. Digests were stored at −20 °C and thawed as needed.

N-Terminal Amino Acid Sequence Analysis. N-Terminal amino acid sequence analyses were performed on a model 476A protein sequencer system from Applied Biosystems (Foster City, CA). Data analysis was performed with Applied Biosystems model 610A data analysis system for protein sequencing, v.1.2.2. All reagents used were from Applied Biosystems. HPLC-purified peptides and glycopeptides were loaded onto Biobrene-treated precycled glass fiber filters for sequencing.

HPAEC-PAD. *N*-Glycans were removed from PE-TPO and reverse-phase HPLC-purified *N*-glycopeptides using PNGase F as detailed above. After incubation with the enzyme, protein or peptides were precipitated with ethanol and the supernatant was retained for chromatography. Chromatography was performed on a DX500 HPLC with an ED40 electrochemical detector, a GP40 pump, and a CarboPac-PA100 column (Dionex, Sunnydale, CA). Following equilibration and sample injection, the column was washed with 20 mM sodium acetate/100 mM NaOH for 5 min followed by a gradient to 200 mM sodium acetate/100 mM NaOH over 55 min. Eluant from the detector was passed through an anion self-regenerating suppressor (Dionex) in order to desalt the purified *N*-glycans.

The concentration of *N*-glycan in each of the six major HPAEC-PAD peaks was determined by monosaccharide composition analysis of fucose with the assumption that the molar ratio of fucose to *N*-glycan was 1:1. Multiple injections of a known amount of an *N*-glycan permitted calculation of its PAD response factor (PAD-RF); different *N*-glycan structures do not give a simple linear PAD response (Townsend et al., 1989). The relative abundance of all *N*-glycan structures present in an HPAEC-PAD chromatogram could then be calculated by adjusting the relative peak areas by their respective PAD-RFs. Since monosaccharide composition analysis was not possible for the minor peaks present in the HPAEC-PAD chromatogram, they were assigned a PAD-RF that was an average of the PAD-RFs for the six major peaks.

Monosaccharide Composition. Monosaccharide composition analysis was performed on HPAEC-PAD-purified *N*-glycans using a DX500 HPLC with an ED40 electrochemical detector, a GP40 pump, and a CarboPac-PA100 column (Dionex).

For sialic acid analysis, 1−2 mg of *N*-glycan was vacuum centrifuged to dryness, resuspended in 400 μL of 0.1 N TFA, and incubated at 80 °C for 1 h. Following vacuum centrifugation to dryness, the hydrolyzed sample was resuspended in H_2O and injected onto the Dionex system equilibrated in 50 mM sodium acetate/100 mM NaOH. A gradient to 180 mM sodium acetate/100 mM NaOH over 25 min was used.

For neutral monosaccharide analysis, the hydrolysis was carried out in 2.0 N TFA at 100 °C for 4 h. Separation was effected using a 25 min isocratic gradient of 14 mM NaOH.

In both types of analyses, Dionex monosaccharide standards were used to calibrate the instrument. The glycoprotein fetuin was used as a positive control to judge the performance of the instrument (Sigma, St. Louis, MO).

Fluorophore-Assisted Carbohydrate Electrophoresis (FACE). Several of the HPAEC-PAD-purified *N*-glycans were sequenced using the FACE system (Glyko, Novata, CA). *N*-Glycans were vacuum centrifuged to dryness and resuspended in 5 μL of 0.15 M 8-aminonaphthalene-1,3,6-trisulfonic acid in 15% acetic acid and 5 μL of 1.0 M sodium cyanoborohydride in dimethyl sulfoxide. After incubation at 37 °C for 16 h, the derivatized *N*-glycans were again vacuum centrifuged to dryness.

Glycosidase digests of the derivatized *N*-glycans were carried with 20−40 pmol of *N*-glycan resuspended in 100 mM sodium citrate/phosphate at pH 6.0 and 37 °C for 16 h. Following digestion, the *N*-glycans were loaded into the lanes of an oligosaccharide-profiling gel per the procedure in the FACE kit (Glyko). The six lanes contained undigested *N*-glycan, *N*-glycan + sialidase, *N*-glycan + sialidase + β-galactosidase, *N*-glycan + sialidase + β-galactosidase + β-*N*-acetylhexosaminidase, *N*-glycan + sialidase + β-galactosidase + β-*N*-acetylhexosaminidase + α-mannosidase, and *N*-glycan + sialidase + β-galactosidase + β-*N*-acetylhexosaminidase + α-mannosidase + α-fucosidase. The sources of the exoglycosidases are as follows: sialidase from Newcastle Disease Virus and *Clostridium perfringens*, β-galactosidase from *Streptococcus pneumonia*, β-*N*-acetylhexosaminidase from Jack bean, α-mannosidase from Jack bean, and α-fucosidase from almond meal. After electrophoresis at a constant current of 20 mA, the gel was imaged in the SE1000 FACE imager and analyzed using FACE software v2.3 (Glyko). The bands were assigned a DP (degree of polymerization) value on the basis of their position relative to oligosaccharide ladder standards included in the kit. Increases or decreases in DP are related to changes in mobility of the *N*-glycan upon digestion with the exoglycosidase library.

MALDI-TOF Mass Spectrometry. MALDI time-of-flight mass spectra were measured with a TofSpec SE mass spectrometer (MicroMass, Beverly, MA) equipped with a 337 nm UV laser. Full scale laser output was measured to be 180 μJ/shot with a 4 ns pulse width; the laser was attenuated using the OPUS software system. The laser coarse adjustment was set at 20% transmittance, and the fine adjustment was varied between 70 and 100% transmittance. Protein and *N*-glycan spectra were recorded in the linear TOF mode (1.5 m flight path). The target was held at 25 kV with an extraction voltage of 8.3 kV and a focus voltage of 23 kV; the detector was held at 3.8 kV. Mass spectra were generated from the average of 30 shots at five different positions within a sample spot on three different sample spots, thus resulting in 450 shots total. A software-determined smoothing was applied, allowing a peak detection of the internal standards at 80% baseline with 0.250% peak slope. The $(M + H)^+$ and $(M + 2H)^{2+}$ peaks of bovine serum albumin (BSA) were used as an internal standard for the protein spectra; *N*-glycan spectra were externally calibrated to the spectrum of a core-fucosylated disialylated biantennary *N*-glycan standard (Oxford GlycoSystems). Proteins were spotted by premixing 1 μL of 2 pmol/μL BSA and ~0.5 μg of sample with 1 μL of sinapinic acid in 30:70:0.1 ACN/H_2O/TFA, and then 2 μL was spotted on the target and allowed to air-dry at room temperature; *N*-glycans were spotted using a similar procedure in 2,5-dihydroxybenzoic acid.

LC−ESMS. Tryptic digests of TPO were separated by reverse-phase HPLC on a Michrom Ultrafast Microprotein Analyzer with a 1.0 × 150 mm, 5 μm, 300 Å Vydac C4 column (Michrom BioResources Inc., Auburn, CA). The following standard gradient was employed for all LC−ESMS runs: 1 to 90% B over 180 min at 50 μL/min (A, 2% ACN + 0.1% TFA; B, 90% ACN + 0.095% TFA). Injections of 40 μg of digested TPO (~1 nmol) were used to collect the three LC−ESMS data sets: native TPO, PE-TPO, and PNGase F-treated PE-TPO. Thus, all the LC−ESMS mapping data was obtained from 120 μg of protein. Eluant flow was split with a T-splitter to give a flow rate of ~15 μL/min into the electrospray mass spectrometer.

Mass spectra were recorded on a PE-Sciex API-III$^+$ triple-quadrupole mass spectrometer equipped with an articulated,

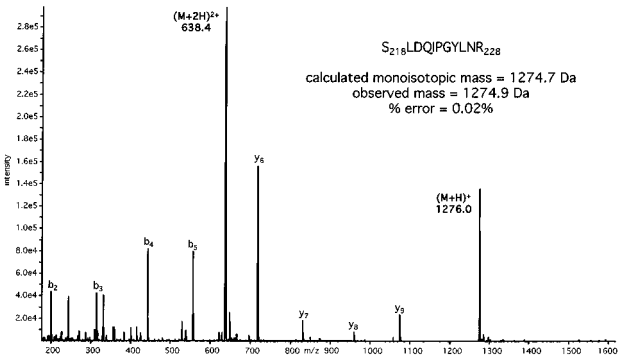

FIGURE 1: Mass spectrum (180−1620 *m/z*) of the native TPO tryptic peptide eluting at 36.5 min (sum of 20 scans). The singly and doubly charged peptide ions are shown in boldface, and the positions of the b and y ions are indicated. The assignment of the peak to TPO 218−228 and the percent error in mass associated with that assignment are shown in the inset.

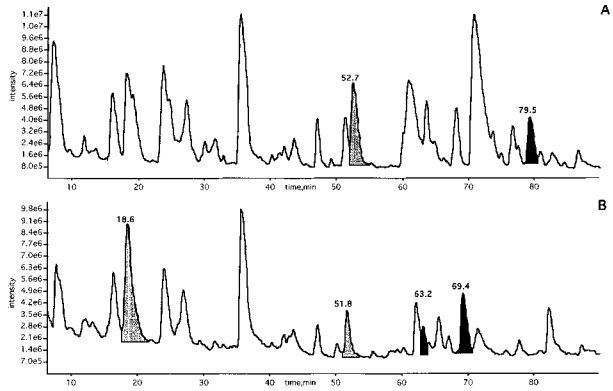

FIGURE 2: Comparison of the LC−ESMS TIC chromatograms (150−2200 *m/z*) for the trypsin digests of (A) native TPO and (B) PE-TPO. The peaks for the disulfide-bonded tryptic peptides are highlighted in panel A, and their elution times are shown: 1−10 + 141−153 at 52.7 min (gray) and 26−52 + 79−98 at 79.5 min (black). The peaks for the four pyridylethylated tryptic peptides are highlighted in panel B, and their elution times are shown: 1−10 at 18.6 min (gray), 141−153 at 51.8 min (gray), 79−98 at 63.2 min (black), and 26−52 at 69.4 min (black).

pneumatically assisted nebulization probe and an atmospheric pressure ionization source (PE-Sciex Instruments, Norwalk, CT). Standard instrument operating conditions were N_2 (99.995%) curtain gas at 1.2 L/min and "zero grade" compressed air nebulizer gas at 50 psi. The instrument was tuned and calibrated using an equimolar solution of poly(propylene glycol) (PPG) 425, 1000, and 2000 in 50:50:0.1:0.1 H_2O/methanol/ACN/formic acid (v/v/v/v) and 2 mM NH_4OAc.

LC−ESMS chromatograms were acquired using an ion spray voltage of 4.8 kV. Quadrupole one of the spectrometer was scanned from 150 to 2200 *m/z* using a step size of 0.2 *m/z* and a dwell time of 0.4 ms; the orifice voltage was ramped from 140 to 80 V over the 150−500 *m/z* range and was held constant at 80 V over the 500−2200 *m/z* range (4.8 scans/s).

RESULTS

Peptide Mapping. Peptide mapping of recombinant TPO was carried out in order to confirm the amino acid sequence of the heterologously expressed protein. Following digestion of the protein with trypsin, the peptides were separated by reverse-phase HPLC and eluted directly into the source of a triple-quadrupole electrospray mass spectrometer (LC−ESMS). The masses of the tryptic peptides were then compared with the peptide masses expected from a trypsin digest of the cDNA-translated sequence in order to assign the chromatogram by mass matching.

For example, the peptide eluting at 36.5 min in the chromatogram of the trypsin digest of native TPO gives a mass spectrum with a singly charged ion at 1276.0 *m/z* and a doubly charged ion at 638.4 *m/z*. Thus, the peptide has a mass of 1274.9 Da (Figure 1). Only one tryptic peptide derived from the hTPO sequence is expected to have a mass in this range; TPO 218−228 has a calculated monoisotopic mass of 1274.7 Da. The error in this assignment based on mass is 0.02% and is typical of the error associated with all mass-matching assignments.

The assignment of the peptide at 36.5 min is aided by the observation of four b ions (b_{2-5}) and four y ions (y_{6-9}) which confirm the presence of the three-amino acid sequence D-Q/K-I/L in the peptide and which are found at the masses consistent for fragmentation of the undecapeptide SLDQIPGYLNR. Therefore, the peak in the chromatogram at 36.5 min can be unambiguously assigned to TPO 218−228 on the basis of its mass and fragmentation pattern.

Glycopeptides produced sugar oxonium ions using the ramped orifice method but did not fragment along the polypeptide backbone. Yet for nearly every tryptic glycopeptide in the digests, the nonglycosylated species was also observed (indicating partial occupancy). Fragmentation of the nonglycosylated species could then be used to confirm the sequence and mass-based assignments. However, there were three tryptic glycopeptides for which the nonglycosylated species were not observed: two *N*-glycopeptides (TPO 178−191 and TPO 229−237) and one *O*-glycopeptide (TPO 154−164). The chromatogram of PNGase F-treated PE-TPO contained the *N*-deglycosylated TPO 178−191 and TPO 229−237 which did undergo fragmentation.

Using this method, we were able to confirm that the amino acid sequence of recombinant hTPO is identical to that predicted from the published human cDNA sequence from residue S1 to R246 (de Sauvage et al., 1994). TPO 1−153 and TPO 165−245 were confirmed by mass matching and the observation of at least three b and/or y ions per peptide. Mass matching alone accounted for TPO 154−164 and TPO 246. Fragmentation was not observed for the *O*-glycopeptide TPO 154−164 or the peptide due to incomplete trypsinolysis at R245, TPO 238−246 (present in small amounts relative to TPO 238−245). The elution times, masses, fragment ions observed, and assignments for all peptides in the chromatograms of the trypsin digests of recombinant hTPO can be found in Table 1 of the Supporting Information.

Disulfide Mapping. The disulfide pairing of the four cysteine residues present in the cytokine domain of hTPO was determined by comparison of the chromatograms of trypsinized native and PE-TPO. Fortuitously, there was at least one trypsin cleavage site between each of the four cysteine residues, obviating the need for a secondary digest. A comparison of the total ion current chromatograms (TICs) for native and PE-TPO is shown in Figure 2.

Two peptides with masses of 2420.5 and 4986.4 Da were observed in the digest of native hTPO, eluting at 52.7 and 79.5 min, respectively (Figure 2, upper panel), and were absent in the digest of PE-TPO (lower panel). These masses

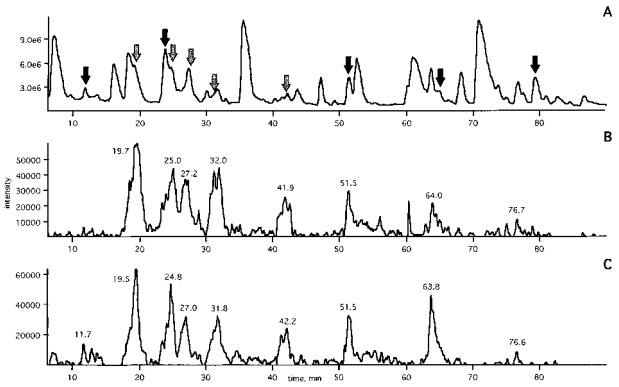

FIGURE 3: (A) LC−ESMS TIC chromatogram (150−2200 *m/z*) for the trypsin digest of native TPO. (B) XIC chromatogram for 204.2 *m/z*, HexNAc. (C) XIC chromatogram for 292.2 *m/z*, NeuNAc. The elution positions of the *O*-glycopeptide and *N*-glycopeptide peaks are indicated by the black and gray arrows, respectively, in panel A.

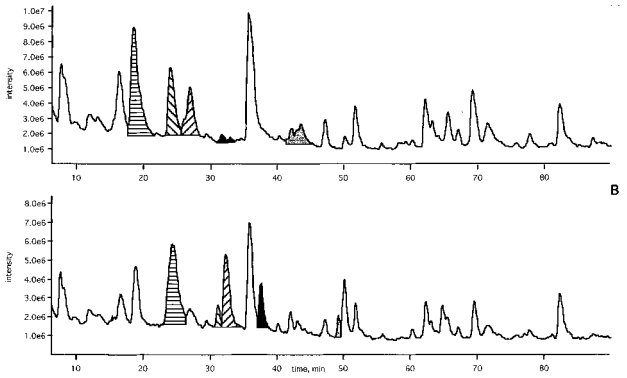

FIGURE 4: Comparison of the LC−ESMS TIC chromatograms (150−2200 *m/z*) for the trypsin digests of (A) PE-TPO and (B) PNGase F-treated PE-TPO. The peaks for the five *N*-glycopeptides are highlighted in panel A: TPO 229−237 (horizontal stripes), TPO 208−217 (left diagonal stripes), TPO 206−217 (right diagonal stripes), TPO 178−191 (black), and TPO 165−177 (gray). The peaks for the five *N*-deglycosylated peptides are highlighted in panel B with the same patterns to indicate the peak shifts after PNGase F treatment.

match the expected masses for the TPO tryptic fragments 1−10 + 141−153 (2421.0 Da) and 26−52 + 79−98 (4986.9 Da). Four tryptic peptides not observed in the digest of native hTPO were produced from the trypsin digest of PE-TPO. The peaks at 18.6, 51.8, 63.2, and 69.4 min have masses of 1132.0, 1501.4, 2120.3, and 3079.5 Da, respectively (Figure 2, lower panel). These masses have been assigned to the TPO tryptic peptides 1−10 (1132.3 Da), 141−153 (1501.9 Da), 79−98 (2120.5 Da), and 26−52 (3079.5 Da) with one pyridylethylcysteine residue per peptide. Thus, the two disulfide bridges in the cytokine domain of hTPO are Cys7−Cys151 and Cys29−Cys85.

Glycosylation Mapping. The large difference between the molecular mass expected from the amino acid sequence of hTPO (35 469 Da) and the mass of the recombinant protein determined by MALDI-TOF mass spectrometry (57 539 Da) clearly indicates that the protein has been modified by the addition of carbohydrate. The sequence of hTPO contains six N-glycosylation sequons and an unknown number of O-glycosylation sites. Glycosylation mapping of hTPO 1−246 was used to determine the position, occupancy, and structure of the *N*- and *O*-glycans.

N- and *O*-glycopeptides present in the trypsin digests of hTPO were visualized in the LC−ESMS chromatograms using the ramped orifice method described in Experimental Procedures. The TIC chromatogram of trypsinized native TPO is shown in Figure 3A. Panels B and C show the extracted ion current chromatograms (XICs) for 204.2 and 292.2 *m/z*, respectively. These masses correspond to the singly charged oxonium ions for *N*-acetylhexosamines (e.g. GalNAc and GlcNAc) and *N*-acetylneuraminic acid (NeuNAc), respectively. HexNAc and NeuNAc ion peaks in the XIC chromatograms reveal the positions of 10 glycopeptides in the trypsin digest of native TPO. This technique does not distinguish between *N*- and *O*-glycopeptides, however, as both *N*-glycans and *O*-glycans can contain HexNAc and NeuNAc monosaccharide residues. The following sections will describe the methods used to differentiate between the two types of glycans and the characterization of those glycans.

Mapping of N-Glycans by LC−ESMS. hTPO contains six N-glycosylation sequons which are candidates for modification by addition of carbohydrate: N_{176}RT, N_{185}FT, N_{213}QT, N_{234}GT, N_{319}TS, and N_{327}LS. The choice of trypsin as the protease allowed us to characterize the first four sites via LC−ESMS. Distinguishing between the *N*- and *O*-glycopeptides observed in the XICs was carried out by comparing the trypsin digests of PE-TPO and PNGase F-treated PE-TPO. Figure 4 shows the TIC chromatogram from pyridylethylated TPO in the upper panel and the TIC chromatogram from PNGase F-treated PE-TPO in the lower panel. Peak shifts were observed for five tryptic peptides: TPO 165−177, TPO 178−191, TPO 206−217, TPO 208−217, and TPO 229−237. The peak shifts occur due to loss of the hydrophilic glycan and conversion of the modified Asn to an Asp residue. This conversion results in longer retention times and allows for easy identification of the *N*-deglycosylated peptides by mass matching and fragmentation. Note that incomplete cleavage after K207 resulted in two peptides containing N213 (TPO 206−217 and TPO 208−217).

PNGase F treatment revealed that the first four N-glycosylation sequons were all modifed. Two of these sites, N176 and N213, were only partially occupied by *N*-glycans, as we observed both the glycosylated and nonglycosylated peptide species; the nonglycosylated species eluted with a retention time intermediate between that of the glycosylated and N-deglycosylated peptides. For example, N-glycosylated TPO 206−217 elutes from ∼26.2 to 27.1 min, nonglycosylated TPO 206−217 at 31.8 min, and *N*-deglycosylated TPO 206−217 at 32.4 min. Nonglycosylated forms of peptides TPO 178−191 and TPO 229−237 were not observed, suggesting 100% occupancy at sites N185 and N234.

The five tryptic *N*-glycopeptide peaks all had broad elution profiles due to microheterogeneity of the attached glycan; peptides with larger *N*-glycans eluted before those with smaller *N*-glycans. Summation of the mass spectra over the broad elution peak for TPO 229−237 revealed the $(M + 2H)^{2+}$ and $(M + 3H)^{3+}$ ions for glycopeptide masses ranging from ∼3100 to 5100 Da. This is illustrated in Figure 5. Subtraction of the base mass for peptide 229−237 from the reconstructed masses permitted identification of the *N*-glycan structures through mass matching; the *N*-glycans range in mass from 2061 to 4031 Da. The various species observed in the reconstructed mass spectrum (Figure 5, bottom panel) differed in mass by units of HexNAcHex (366 Da) and NeuNAc (291 Da) as one would expect for successive

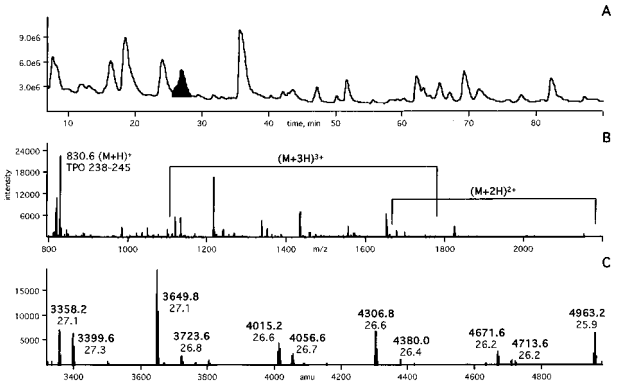

FIGURE 5: (A) LC−ESMS TIC chromatogram (150−2200 *m/z*) of the trypsin digest of PE-TPO. (B) Summed mass spectra from 27.0 to 29.0 min from 800 to 2200 *m/z*. The singly charged ion of the coeluting peptide TPO 238−246 is indicated. The brackets show the ranges for the triply and doubly charged ions of the heterogeneous *N*-glycopeptide TPO 206−217. (C) Reconstructed mass spectrum from 3500 to 5000 amu of the original mass spectrum in panel B. The masses of the different species of *N*-glycopeptide TPO 206−217 are shown in boldface. The elution times of the various species are indicated below their masses.

additions of antennae in complex *N*-glycans. The elution times for the different *N*-glycopeptide species are also indicated above the $(M + 3H)^{3+}$ ions.

The masses of the *N*-glycans at all four of the N-terminal sites in hTPO are consistent with complex *N*-glycans. At each of these sites, the mass of the smallest *N*-glycan is consistent with a monosialylated biantennary structure and the mass of the largest *N*-glycan is consistent with either a tetrasialylated pentaantennary structure or a tetrasialylated tetraantennary structure with one *N*-acetyllactosamine (LacNAc) repeat. The masses of all the *N*-glycan species are consistent with core fucosylation. In addition to these structures, we also observe masses consistent with structures in which a single Hex has been substituted by a HexNAc. *N*-Glycans with this single substitution in the bi-, tri-, and tetraantennary structures were observed.

The relative abundance of the various *N*-glycan structures on each of the five *N*-glycopeptides was estimated by comparison of the $(M + 3H)^{3+}$ ion intensities for each species. In the case of TPO 206−217 and TPO 208−217, intensities for each species in the two mass spectra were summed to calculate the *N*-glycan populations at N213. This method was used instead of comparing reconstructed mass peak intensities because the $(M + 2H)^{2+}$ ions were only observable (i.e. below 2200 *m/z*) for the lower-mass species. Since differences in carbohydrate structures on a glycopeptide can affect its efficiency of ionization, these population calculations based on ion intensity must be viewed as rough estimates of the actual solution populations. Table 1 gives the relative abundance of each *N*-glycan structure at N176, N185, N213, and N234. Note that the degree of branching decreases as one proceeds from the first to the fourth site.

Mapping of N-Glycans by HPAEC-PAD. In order to confirm the LC−ESMS structure assignments and to provide more reliable figures for relative abundance of the various oligosaccharides, analysis of the *N*-glycans liberated from the whole protein and from the individual tryptic *N*-glycopeptides by PNGase F was carried out. Following chromatographic separation via HPAEC-PAD, the purified *N*-glycans were analyzed by a variety of techniques, including MALDI-TOF mass spectrometry, monosaccharide composition, and FACE sequencing.

PE-TPO was treated with PNGase F to release the intact *N*-glycans from the protein, and following precipitation of the protein, the *N*-glycans were separated via HPAEC-PAD. Twenty peaks were observed in the HPAEC-PAD chromatogram (Figure 6). Four of the peaks (8, 14, 18, and 20) were identified as products of epimerization of four major peaks (7, 13, 17, and 19). Reinjection of each of the purified peaks 7, 13, 17, and 19 gave rise to peaks in the position of peaks 8, 14, 18, and 20, respectively. The purified epimers also had the same mass as their counterparts as determined by MALDI-TOF mass spectrometry. This chromatogram indicated the relative abundance of the different *N*-glycans averaged over all occupied sites in hTPO.

Analysis of the *N*-glycans released from the individual tryptic *N*-glycopeptides permitted site specific characterization of the carbohydrate microheterogeneity. A tryptic digest of PE-TPO was separated under the reverse-phase HPLC conditions described in Experimental Procedures, and the individual peptide peaks were collected. Using the LC−ESMS chromatogram as a guide, fractions close to the time of elution for a specific *N*-glycopeptide were subjected to N-terminal sequencing and those fractions containing any

Table 1: Relative Abundance of the *N*-Glycan Structures on Whole Recombinant Human TPO and at the Four N-Terminal N-Glycosylation Sites as Determined by LC−ESMS and HPAEC-PAD

	relative abundance by LC−ESMS, %				relative abundance by HPAEC-PAD, %				
N-glycan structure	N176	N185	N213	N234	N176	N185	N213	N234	TPO
monosialylated biantennary	4.4	4.4	9.9	10.4	2.8	5.9	10.6	12.7	8.7
monosialylated biantennary + Gal≫GalNAc[a]	6.9	8.4	9.9	9.9	1.8	3.7	8.1	7.9	6.7
disialylated biantennary	18.0	24.2	**26.5**	**32.2**	17.8	20.7	19.1	**30.7**	23.9
monosialylated triantennary	1.7	3.9	2.7	1.7	2.5	2.5	2.8	1.6	2.1
disialylated triantennary	6.2	6.8	7.5	5.2	5.7	4.7	5.7	5.8	6.8
disialylated triantennary + Gal≫GalNAc	3.2	4.1	7.2	3.6	8.0	4.5	6.6	3.9	3.6
trisialylated triantennary	**25.4**[b]	**28.0**	16.8	22.9	**35.4**	**30.2**	**26.1**	24.1	**27.3**
disialylated tetraantennary	2.2	2.4	1.6	1.4	nd[c]	nd	nd	nd	nd
trisialylated tetraantennary	7.6	3.0	5.8	4.4	4.9	4.7	3.4	2.3	2.6
trisialylated tetraantennary + Gal≫GalNAc	2.7	1.5	1.1	1.0	1.7	0.9	1.0	nd	0.6
tetrasialylated tetraantennary	20.0	12.4	10.9	6.7	19.5	22.2	16.6	10.9	17.6
tetrasialylated pentaantennary or tetrasialylated tetraantennary + LacNAc	1.7	0.9	0.1	0.6	nd	nd	nd	nd	nd
average number of antennae	3.03	2.82	2.73	2.61	3.04	2.98	2.83	2.62	2.81

[a] Gal≫GalNAc indicates a substitution of GalNAc for Gal in one of the antennae. [b] The most abundant *N*-glycan structure at each of the N-glycosylation sites appears in boldface. *c*None detected.

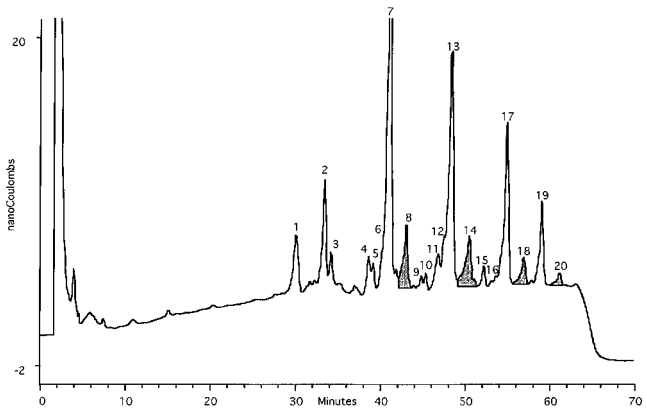

FIGURE 6: HPAEC-PAD chromatogram of the *N*-glycans released from whole PE-TPO by PNGase F. The 20 *N*-glycan peaks collected for carbohydrate analysis are indicated. The four peaks due to epimerization of major peaks preceding them have been highlighted.

of that *N*-glycopeptide sequence were pooled for analysis. Tryptic *N*-glycopeptides TPO 206−217 and TPO 208−217 were pooled together for characterization of site N213. After PNGase F digestion, the deglycosylated peptides were separated from the *N*-glycans and the four *N*-glycan pools were subjected to HPAEC-PAD. As with the whole protein, each of the four pools produced HPAEC-PAD chromatograms with 20 peaks present; i.e. there was no *N*-glycan structure that was unique to a single site.

It must be noted that a small amount of the incompletely cleaved tryptic *N*-glycopeptide 165−191 was discovered by N-terminal sequencing. Tryptic peptide 165−191 was not observed in the LC−ESMS chromatogram. It is likely that the microheterogeneity of two N-glycosylation sites made the elution of this minor peptide so broad it passed undetected. Another factor contributing to the indetectability of TPO 165−191 is its mass-to-charge ratio. The mass of TPO 165−191 with two disialylated biantennary complexes attached would be 7623.6 Da, and only the $(M + 4H)^{4+}$ ion of such a mass would be below 2200 *m/z*. The presence of this incomplete cleavage product containing sites N176 and N185 may skew the *N*-glycan populations calculated for N176. This effect would occur if the size of the *N*-glycan at N176 is affecting the cleavage at R177 through steric hindrance of trypsin. If cleavage by trypsin after R177 were affected by the size of the *N*-glycan at N176, one would expect an enrichment in the population of smaller *N*-glycans for glycopeptide TPO 165−177. This is not the case as the profile of this site is actually enriched in the larger *N*-glycans and qualitatively resembles the profiles at the other sites. Thus, the *N*-glycan populations calculated for site N176 appear to be truly representative and not skewed by the small amount of incomplete trypsinolysis at R177.

Structural assignment of the majority of the purified *N*-glycans was carried out through mass matching. MALDI-TOF mass spectrometry was used to determine the masses of the purified *N*-glycans. Three of the *N*-glycans were assigned structures solely on the basis of their elution position in the HPAEC-PAD chromatogram. Finally, monosaccharide composition analysis and FACE sequencing were used for a more in-depth characterization of the six major *N*-glycan species. Monosaccharide composition analysis was used to determine the relative amounts of fucose, mannose, galactose, GalNAc, GlcNAc, and NeuNAc present in the six major *N*-glycans. FACE sequencing with specific exoglycosidases was used to provide information on the oligosaccharide linkages in the six major *N*-glycans. The structural assignments of all twenty *N*-glycan peaks present in the HPAEC-PAD chromatograms are found in Table 2. The results of the latter two analyses are shown in Table 2 of the Supporting Information. The results of these analyses confirm the structures proposed in Figure 7. Note that the underlined linkages have been experimentally determined, while the other linkages were proposed by analogy to previously published *N*-glycan structures. There were some limitations in the structural characterization.

In the structure proposed for peak 1, it is not possible from our data to determine whether the GalNAc for Gal substitution occurs on the α1−6 or α1−3 antenna, nor is it possible to determine whether the sialic acid is attached α2−3 to the terminal Gal or GalNAc residue. With respect to the structure proposed for peak 2, it was not possible to determine to which antenna the sialic acid is α2−3 linked. Likewise, with our data, we were unable to differentiate between the two isomeric triantennary structures proposed for peak 13, β1−2 antenna on the α1−6 arm with β1−2 and β1−4 antennae on the α1−3 arm and β1−2 and β1−6 antennae on the α1−6 arm with β1−2 antenna on the α1−3 arm. Since these two *N*-glycan structures coelute in HPAEC, it is entirely possible the fraction is a mixture of the two isomers (Hermentin et al., 1992).

The relative abundance of each of the *N*-glycans in whole hTPO and at sites N176, N185, N213, and N234 is shown in Table 1. These populations are calculated from the relative peak areas in the HPAEC-PAD chromatograms. Note that the degree of branching decreases as one proceeds from the first to the fourth site in hTPO.

Comparison of N-Glycan Mapping Methods. The relative abundance of the different *N*-glycan structures at each of the first four N-glycosylation sites in hTPO was determined by two independent methods: comparison of *N*-glycopeptide $(M + 3H)^{3+}$ ion intensities obtained by LC−ESMS and comparison of HPAEC-PAD chromatogram peak areas for PNGase F-released *N*-glycans. Though relative ion intensities are not necessarily an accurate representation of actual solution populations, the populations calculated from the HPAEC-PAD data largely corroborate the LC−ESMS estimates. Table 1 shows the relative abundance of biantennary, triantennary, and tetraantennary structures at each of these four sites for the LC−ESMS and HPAEC-PAD methods.

Both methods show the biantennary structures predominate at site N234 and the triantennary structures predominate at sites N176 and N185. However, the methods differ at site N213, where the biantennary structures predominate in the LC−ESMS data and the triantennary structures predominate in the HPAEC-PAD data. The largest absolute difference between the two data sets is for the relative abundance of trisialylated triantennary *N*-glycans at site N176: 25.4% (LC−ESMS) vs 35.4% (HPAEC-PAD), a 28% error in the LC−ESMS figure.

Despite these differences, both methods show a clear trend of diminishing *N*-glycan branching as one proceeds from the first to the fourth *N*-glycosylation site in hTPO. The average number of antennary arms at each of the four sites is shown in Table 1. These values are strikingly similar for the two methods, with the largest percent difference of only 5% at site N185.

Table 2: Structural Assignment of HPAEC-PAD Peaks

HPAEC-PAD peak	elution time, min	MALDI-TOF MS mass, Da	expected mass, Da	structure assignment
1[a]	30.2	2163.0	2164.0 (2 Na^+)[b]	**monosialylated biantennary + Gal≫GalNAc**
2	33.4	2123.8	2122.9 (2 Na^+)	**monosialylated biantennary**
3	34.3	2486.7	2488.2 (2 Na^+)	monosialylated triantennary
4	38.7	2840.4	2842.5 (3 Na^+)	disialylated triantennary + Gal≫GalNAc
5	39.2	2848.8	2842.5 (3 Na^+)	disialylated triantennary + Gal≫GalNAc
6	40.4	nd	−	disialylated triantennary
7	40.9	2434.2	2436.2 (3 Na^+)	**disialylated biantennary**
8	43.1	2436.5	2436.2 (3 Na^+)	epimer of peak 7c
9	44.8	nd	−	trisialylated tetraantennary + Gal≫GalNAc
10	45.4	nd	−	trisialylated triantennary
11	46.8	3476.6	3480.0 (4 Na^+)	trisialylated tetraantennary
12	47.4	3477.4	3480.0 (4 Na^+)	trisialylated tetraantennary
13	48.2	3110.1	3114.7 (4 Na^+)	**trisialylated triantennary**
14	50.4	3111.5	3114.7 (4 Na^+)	epimer of peak 13[c]
15	52.0	3120.6	3114.7 (4 Na^+)	trisialylated triantennary
16	53.4	3802.6	3793.5 (5 Na^+)	tetrasialylated tetraantennary
17	54.6	3795.0	3793.3 (5 Na^+)	**tetrasialylated tetraantennary**
18	56.7	3796.5	3793.3 (5 Na^+)	epimer of peak 17[c]
19	58.8	3793.0		**tetrasialylated tetraantennary**
20	61.0	3795.0	3793.3 (5 Na^+)	epimer of peak 19[c]

[a] The six largest peaks are shown in boldface. [b] The *N*-glycans were observed as sodium adducts with one more sodium than the number of sialic acids present. [c] This peak is due to epimerization of the *N*-glycan immediately preceding it.

Mapping of O-Glycans by LC−ESMS. The presence of glycopeptides that were unaffected by treatment with PNGase F indicated that there were *O*-glycopeptides in hTPO. PNGase F treatment indicated that five of the glycopeptide signals observed in Figure 3 were due to *N*-glycopeptides, suggesting there are five *O*-glycopeptides in the trypsin digest of native rhTPO. In fact, there are six tryptic *O*-glycopeptides in TPO, but one of them contains multiple glycosylation sites. This *O*-glycopeptide does not ionize very efficiently and thus was missed by the ramped orifice method.

The six tryptic *O*-glycopeptides all had broad elution profiles, indicating microheterogeneity of the attached *O*-glycans. For example, summation of the mass spectra across the peak for TPO 192−199 revealed the $(M + H)^+$ and $(M + 2H)^{2+}$ ions for peptide masses ranging from ∼979 to 1725 Da. This is illustrated in Figure 8. Subtraction of the base mass for TPO 192−199 from the reconstructed masses allowed identification of the *O*-glycan structures through mass matching; the *O*-glycans range in mass from 203 to 948 Da. The various species observed in the reconstructed mass spectrum differed in mass by units of Hex (162 Da) and NeuNAc (291 Da). The *O*-glycan masses for all of the O-glycosylation sites in hTPO 1−246 are consistent with mucin-type *O*-glycans. At each of these sites, the smallest *O*-glycan is a simple HexNAc and the largest *O*-glycan is the disialylated HexHexNAc.

O-Glycans were detected on the following tryptic peptides: TPO 1−10, TPO 26−52, TPO 99−117, TPO 154−164, TPO 192−199, and TPO 238−245. Partial occupancy was observed for all the tryptic *O*-glycopeptides except 154−164 which was not observed as a nonglycosylated species. *O*-Glycopeptide TPO 154−164 was observed both alone and as part of the incompletely digested peptide TPO 154−177 with no *N*-glycan present. The sequence TPO 154−164 was observed with one to three O-glycans attached, indicating multiple sites in the peptide.

In the case of tryptic *O*-glycopeptides TPO 1−10 and TPO 238−245, only one possible attachment site was present in each peptide, S1 and S244, respectively. Three sites were present in TPO 154−164, and the peptide was observed with up to three *O*-glycans attached, indicating that all three possible sites were capable of being glycosylated: T158, T159, and S163. The remaining three *O*-glycopeptides had multiple Ser and/or Thr residues present and only a single *O*-glycan. Thus, the methods used in this study were insufficient for determining unambiguously the site of attachment.

Elhammer and co-workers have written an algorithm to calculate the probability of glycosylation for a Ser or Thr residue at the center of a sliding nine-residue window (Elhammer et al., 1993). The probability of glycosylation for the Ser or Thr residue is based on a summation of the amino acid probabilities of appearing in an O-glycosylation site for each of the nine residues. Individual amino acid probabilities were calculated from a data base of 196 O-glycosylated peptide segments. When applied to the sequences of the three *O*-glycopeptides with multiple sites, one Ser or Thr residue in each peptide had a much larger probability than the others. The probabilities of the three sites present in TPO 26−52 are as follows: S27, $p = 0.03$; T37, $p = 0.25$; and S47, $p < 0.01$. The probabilities of the two sites present in TPO 99−117 are as follows: T110, $p = 0.33$; and S106, $p = 0.01$. The probabilites for the three sites in TPO 192−199 are as follows: T192, $p = 0.60$; T193, $p = 0.14$; and S195, $p = 0.14$. The results of the algorithm suggest that the O-glycosylation sites in the three ambiguous *O*-glycopeptides are T37, T110, and T192.

The linkages for the *O*-glycans present in hTPO 1−246 were not determined experimentally; however, linkages for these types of *O*-glycans have been reported previously. The structures of the four *O*-glycans consistent with the masses we observed are GalNAc-S/T, Galβ1−3GalNAc-S/T, NeuNAcα2−3Galβ1−3GalNAc-S/T, and NeuNAcα2−3Galβ1−3(NeuNAcα2−6)GalNAc-S/T (Vliegenthart et al., 1983). The relative abundance of the four *O*-glycan structures at each of the sites was estimated by comparison of the reconstructed mass spectra peak intensities (Table 3).

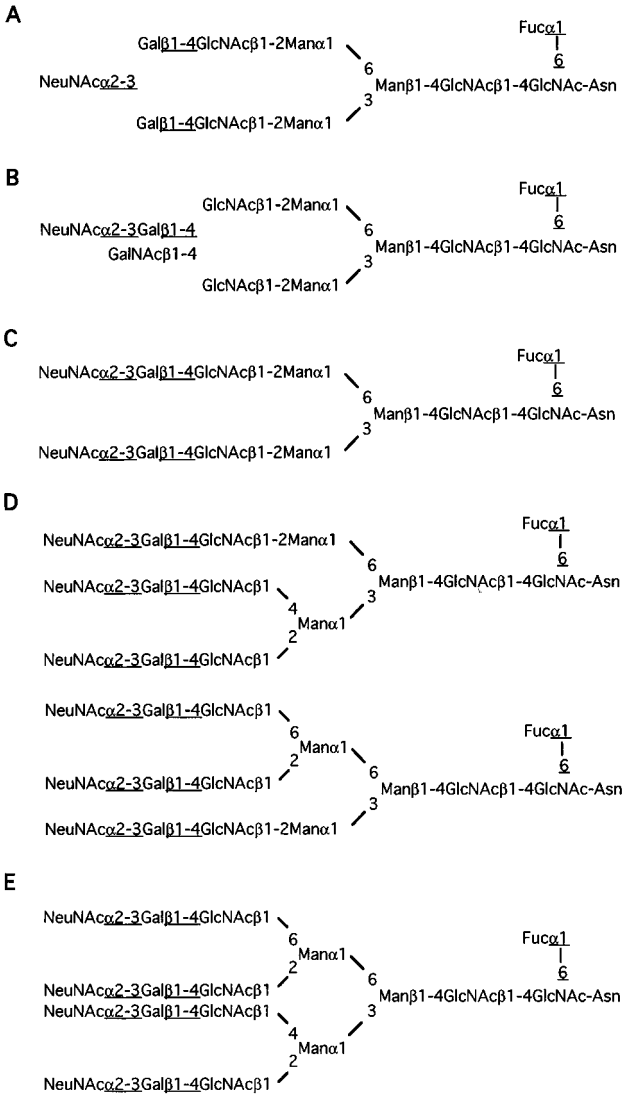

FIGURE 7: Proposed structures for the five major *N*-glycans observed in recombinant human TPO: (A) monosialylated biantennary, (B) monosialylated biantennary + one GalNAc for Gal substitution, (C) disialylated biantennary, (D) two isomers of trisialylated triantennary, and (E) tetrasialylated tetraantennary. All the *N*-glycan structures are core fucosylated. The underlined linkages were experimentally determined by FACE sequencing; the other linkages are proposed by analogy to previously reported *N*-glycan structures.

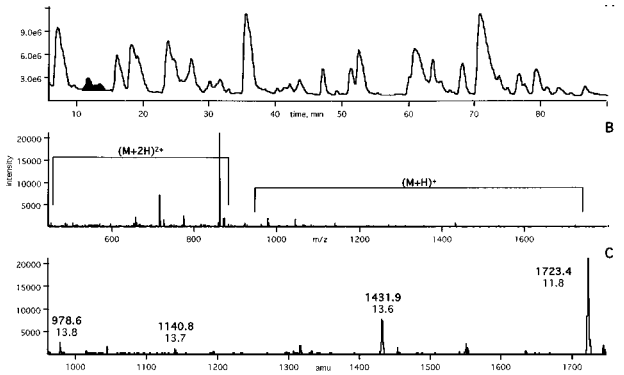

FIGURE 8: (A) LC−ESMS TIC chromatogram (150−2200 *m/z*) of the trypsin digest of PE-TPO. (B) Summed mass spectra from 11.2 to 12.0 min from 400 to 1800 *m/z*. The brackets show the ranges for the doubly and singly charged ions for the heterogeneous *O*-glycopeptide TPO 192−199. (C) Reconstructed mass spectrum from 900 to 1800 amu of the original mass spectrum in panel B. The masses of the different species of *O*-glycopeptide TPO 192−199 are shown in boldface. The elution times of the various species are indicated below their masses.

DISCUSSION

The ramped or stepped orifice method was developed by Carr and co-workers (1993) in order to facilitate identification of glycopeptides in LC−ESMS chromatograms of protease-digested glycoproteins. When a glycopeptide elutes and is ionized, carbohydrate oxonium ions are produced from the attached glycans by the high orifice voltage via front-end CID and are detected in the low-mass range. Performing single-ion extractions of the characteristic carbohydrate oxonium ion masses from the total ion current (TIC) chromatogram clearly reveals the elution positions of any glycopeptides present in the digest (Carr et al., 1993).

We found that this method could be enhanced to yield not only sugar fragment ions from glycopeptides but also b and y ions from the nonglycosylated peptides present in the trypsin digests. This procedural modification simply required slightly higher orifice voltages than those recommended by Carr and co-workers in order to promote front-end CID of the peptide ions. Observation of the sequence diagnostic b and y ions permitted direct confirmation of the peptide assignments made by mass matching. It must be noted that glycopeptides which eluted over long periods of time due to microheterogeneity and ionized less efficiently than nonglycosylated peptides produced carbohydrate ions but not peptide fragment ions under these conditions. Our method can be used on both single- and triple-quadrupole electrospray instruments and represents an easy way to confirm mass-based assignments in the LC−ESMS chromatograms of peptides.

The complete peptide, disulfide, and glycosylation map of BHK-expressed recombinant hTPO 1−246 is shown in Figure 9. The peptide map indicates that the amino acid sequence of the cytokine domain (1−153) and half of the C-terminal domain (154−246) are as expected from translation of the human cDNA sequence. The signal peptide was completely cleaved, and there was no evidence of any aminopeptidase activity. Nor was there evidence for amino acid modifications (e.g. hydroxylation, methylation, or deamidation) other than glycosylation.

The four cysteine residues of hTPO are found in the cytokine domain and form two disulfides: Cys7−Cys151 and Cys29−Cys85. Homology to EPO suggests that the cytokine domain of TPO is a member of the four-helix bundle family of growth and hematopoietic factors. Three-dimensional structures of porcine and human growth hormone (Abdel-Meguid et al., 1987; de Vos et al., 1992) and a growing number of cytokines in this family have been elucidated by NMR spectroscopy and/or X-ray crystallography, including human interleukins-2, -4, and -5, human granulocyte/macrophage and granulocyte colony-stimulating factors, and murine leukemia inhibitory factor (Brandhuber et al., 1987; Bazan, 1992; McKay, 1992; Powers et al., 1992; Smith et al., 1992; Walter et al., 1992a,b; Wlodawer et al., 1992; Milburn et al., 1993; Diederichs et al., 1991; Hill et al., 1993; Robinson et al., 1994). Recently, a secondary-structure analysis of one of the largest members of this family, human oncostatin M (195 residues), was carried out using multidimensional NMR spectroscopy and indicated the positions of the four helices in the sequence (Hoffman et al., 1996). Although there is weak sequence homology

Table 3: Relative Abundance of the *O*-Glycan Structures at Each of the O-Glycosylation Sites as Determined by LC−ESMS

O-glycan structure	relative abundance estimated by LC−ESMS, %					
	S1	S27, **T37**, or S47[a]	S106 or **T110**	154−164[b]	**T192**, T193, or S195	S244
GalNAc	18.5	28.0	nd[c]	nd	7.6	10.8
Galβ1−3GalNAc	11.0	3.0	nd	nd	3.6	26.8
NeuNAcα2−3Galβ1−3GalNAc	**56.0**[d]	**56.9**	10.3	38.6	23.6	57.2
NeuNAcα2−3Galβ1−3(NeuNAcα2−6)GalNAc	14.5	12.1	**89.7**	**61.4**	**65.2**	5.3

[a] *O*-Glycopeptides with multiple sites are shown with the most probable site of attachment in boldface. [b] Tryptic peptide 154−164 was observed with one to three *O*-glycans attached, indicating that all three sites present, T158, T159, and S163, are capable of being glycosylated. These are the relative amounts of the *O*-glycan structures on the *O*-glycopeptide when all three sites are occupied. [c] None detected. [d] The most abundant *O*-glycan structure at each of the O-glycosylation sites appears in boldface.

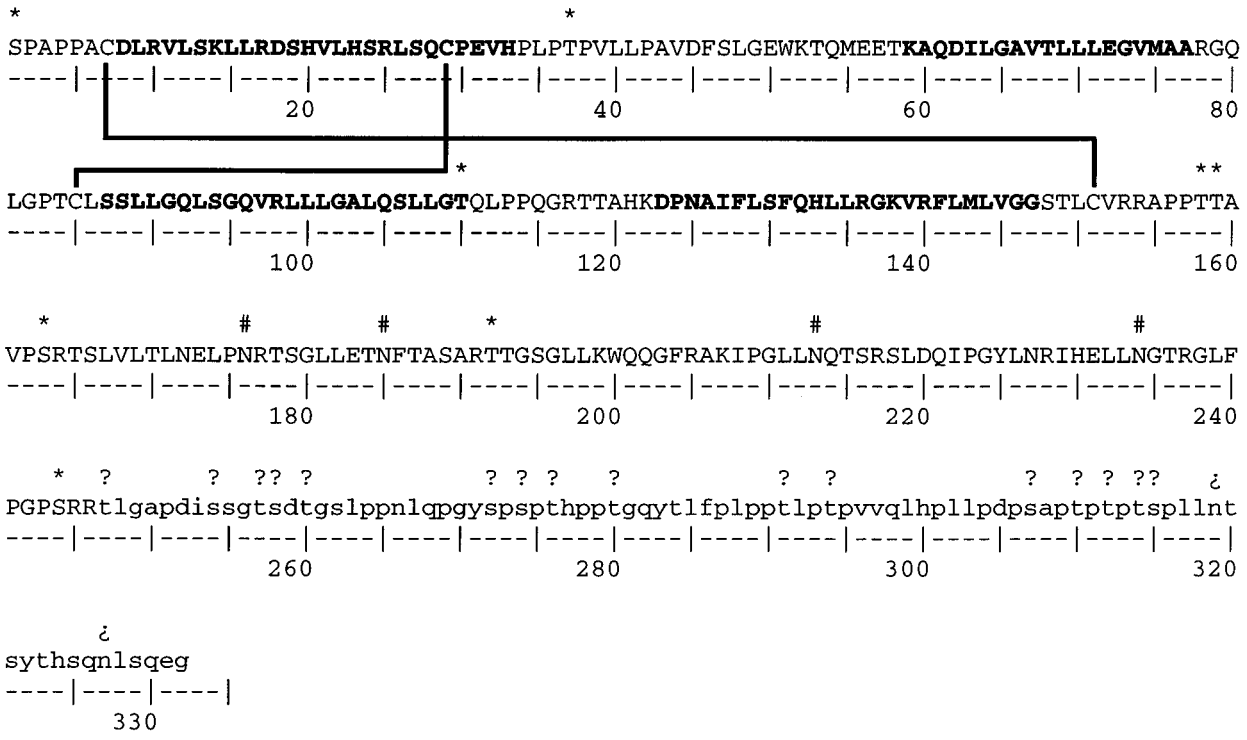

FIGURE 9: Peptide, disulfide, and glycosylation map of the cytokine and *N*-glycan domains of BHK-recombinant hTPO. The sequence of hTPO is shown with the cytokine and *N*-glycan domains in capitals and the unmapped *O*-glycan domain in small letters. The two disulfides in the cytokine domain are shown by connecting lines. The experimentally determined glycosylation sites are marked by asterisks (O-glycosylation) and by pound signs (N-glycosylation). α-Helices proposed by S. Presnell (personal communication) are shown in boldface. The two N-glycosylation sequons in the *O*-glycan domain are marked by upside-down question marks, and the possible O-glycosylation sites calculated to have a p of >0.19 by the prediction algorithm of Elhammer and co-workers (1993) are marked by question marks.

among these family members, all possess a core structure composed of left-handed four-helix bundle with unique up−up−down−down topology. In addition, some disulfides are structurally conserved among family members; e.g. a disulfide bridge between a cysteine in helix B and a cysteine in the overhand CD loop is observed in the structures of IL-2, IL-4, and GM-CSF (Wlodawer et al., 1993).

A preliminary model of the cytokine domain of hTPO has been proposed on the basis of (1) secondary-structure prediction algorithms and (2) an alignment of the hTPO sequence with the structures of IL-2, IL-4, and GM-CSF and the model proposed for EPO (Brandhuber et al., 1987; Bazan, 1992; McKay, 1992; Powers et al., 1992; Smith et al., 1992; Walter et al., 1992a,b; Wlodawer et al., 1992; Diederichs et al., 1991; Boissel et al., 1993). This approach suggests that the four major helical segments in hTPO are Asp8−His33, Lys59−Ala77, Ser87−Thr110, and Asp123−Gly147 (S. Presnell, personal communication). The Cys7−Cys151 bridge in hTPO is consistent with linking the cysteine amino terminal to helix A with the cysteine carboxy terminal to helix D. This particular disulfide linking the termini of the cytokine domain is analogous to one observed in the structure of hIL-4, Cys3−Cys127, and predicted in the model of hEPO, Cys7−Cys161 (Powers et al., 1992; Smith et al., 1992; Walter et al., 1992a; Wlodawer et al., 1992; Boissel et al., 1993). The Cys29−Cys85 bridge in hTPO is also analogous to one observed in the hIL-4 structure, Cys24−Cys65. In hIL-4, this disulfide links the cysteine at the start of the overhand AB loop to the cysteine in the short BC loop (Powers et al., 1992; Smith et al., 1992; Walter et al., 1992a; Wlodawer et al., 1992). Thus, the disulfide bonds in hTPO are consistent with the helices proposed and membership in the four-helix bundle cytokine family.

Analysis of the *N*-glycans present in recombinant hTPO by both LC−ESMS and HPAEC-PAD indicates a heterogeneous population of complex *N*-glycans. The five major species observed are monosialylated biantennary, monosialylated biantennary with one GalNAcGlcNAc-R antenna, disialylated biantennary, trisialylated triantennary, and tetrasialylated tetraantennary. All of the *N*-glycans are core-fucosylated. These *N*-glycans are consistent with asparagine-linked carbohydrate structures previously observed on

recombinant proteins expressed in mammalian cell lines such as BHK and CHO [e.g. Sasaki et al. (1987), Tsuda et al. (1988), Hironaka et al. (1992), Bergwerff et al. (1993), and Legrand et al. (1995)]. The GalNAcGlcNAc-R antennae have been observed previously on nonrecombinant human urokinase, CHO-recombinant mt-PA, and BHK-recombinant FVIIa, but unlike with urokinase and mt-PA, we saw no evidence for sulfation of the terminal GalNAc residue in hTPO (Bergwerff et al., 1995; Chan et al., 1991; Klausen & Bayne, 1995). Furthermore, in the *N*-glycan structures found on BHK-recombinant hTPO, there was no evidence for fucosylation of GlcNAc residues in the antennae as observed in BHK-recombinant human FVIII (Hironaka et al., 1992), for sulfation of galactose as observed in C127-recombinant human t-PA (Pfeiffer et al., 1992), or for O-acetylation of the *N*-acetylneuraminic acid residues as observed in CHO-recombinant human K2tu-PA and EPO (Bergwerff et al., 1993; Rush et al., 1995). The LacNAc repeats observed on the antennae of complex *N*-glycans in CHO-recombinant human IL-6 and EPO were present in a very small amount (<2%) if at all (Orita et al., 1994; Rush et al., 1993; Hokke et al., 1995).

The first four N-glycosylation sequons in hTPO—N_{176}RT, N_{185}FT, N_{213}QT, and N_{234}GT—are all glycosylated. The LC−ESMS data suggest 100% occupancy for sites N185 and N234 and partial occupancy for sites N176 and N213. The initial transfer of dolichol pyrophosphate oligosaccharide to a candidate asparagine occurs cotranslationally in the rough endoplasmic reticulum [for review, see Kornfeld and Kornfeld (1985)]. Statistical analysis of N-glycosylated proteins and a circular dichroism study of synthetic peptide substrates have suggested that the preference of the oligosaccharyltransferase for one site over another is dependent on the ability of the amino acid sequence at the acceptor site to adopt a turn-like conformation (Beeley, 1977; Aubert et al., 1981). A favorable conformation may promote catalysis via a hydrogen bond between the side chain amide of the Asn and the side chain oxygen of the Ser/Thr or may be stabilized by a hydrogen bond between the backbone amide of the Ser/Thr and the side chain carboxyl of the Asn (Bause & Legler, 1981; Bause, 1983; Imperiali & Shannon, 1991). It is possible that decreased efficiency of N-glycosylation at N176 and N213 is a result of local interactions among the adjacent amino acids that hinder the formation of peptide conformations at the acceptor site favorable to the oligosaccharyltransferase. N176 has a prolyl residue at the N-1 position, and N213 has a prolyl residue at the N-4 position. These prolyl residues may restrict conformational freedom at these sites. Previous studies have noted that a Pro at position N + 1 or N + 3 inhibits or prevents glycosylation of acceptor peptides, though a Pro at N − 1 apparently has no effect (Bause, 1983; Roitsch & Lehle, 1989).

In addition to the disparity in occupancy, there is a definite trend toward decreased structural complexity of the *N*-glycans as one proceeds from the first site to the fourth site. The trisialylated triantennary structure is the most common at N176 and N185, while the disialylated biantennary structure is the most common at N234. Likewise, the relative abundance of tetraantennary structures decreases across the four sites. The number of antennae on complex *N*-glycans is determined by the sequential action of GlcNAc transferases I−VI in the medial cisternae of the Golgi apparatus (Schachter, 1986; Brockhausen et al., 1988; Schachter, 1991). A previous study suggests that the position of an oligosaccharide with respect to the N terminus of the protein may affect the level of processing (Pollack & Atkinson, 1983). Pollack and Atkinson found in a survey of *N*-glycoproteins that *N*-glycans closer to the amino terminus were more likely to be complex while those farther away were more likely to be high mannose. Our results suggest that the effect they observed may apply to multiple complex *N*-glycans as well. The more branched *N*-glycans in hTPO may simply be more accessible to the later GlcNAc transferases by virtue of the conformation of the folded polypeptide such that the degree of accessibility fortuitously coincides with the distance from the amino terminus. Several theories have been advanced to explain the observation of site specific *N*-glycan processing, including site-directed processing, differential accessibility, and differential kinetic rates [for an excellent review, see Cumming (1991)].

O-Glycans of the mucin type are observed in hTPO 1−246. There is microheterogeneity at each of the O-glycosylation sites, and the four species of *O*-glycans observed are all derived from a common synthetic pathway: the Tn antigen, GalNAc-S/T; the T antigen, GalGalNAc-S/T; and the mono- and disialylated GalGalNAc-S/T. Mucin-type *O*-glycans have been previously reported on recombinant proteins such as human EPO and glycophorin A expressed in CHO and BHK cells (Sasaki et al., 1987; Nimtz et al., 1993; Påhlsson et al., 1994).

It is not possible to predict a priori where O-glycosylation will occur as there is no strict consensus sequence for UDP-GalNAc−polypeptide *N*-acetylgalactosaminyltransferase which initiates O-glycosylation post-translationally in the Golgi apparatus [for review, see Carraway and Hull (1989)]. There is often a positionally independent preponderance of Ser, Thr, Pro, Ala, and Gly residues near a glycosylated Ser or Thr residue (Elhammer et al., 1993). When the O-glycosylation prediction algorithm of Elhammer and co-workers was applied to the sequence for hTPO 1−246, there are nine sites with a p of >0.19: S1, T37, T110, T158, T159, S163, T165, S189, and T192. The hTPO glycosylation map indicates that the program is fairly accurate, only overpredicting glycosylation at two sites, T165 ($p = 0.32$) and S189 ($p = 0.40$), and underpredicting glycosylation at one site, S244 ($p = 0.08$).

Mucin-type *O*-glycans have been observed at structurally homologous positions in several recombinant cytokines. For example, the overhand loop connecting the C and D helices has been shown to be O-glycosylated in recombinant human EPO (S126), G-CSF (T133), IL-6 (T139), and oncostatin M (T160, T162, and/or S165) (Takeuchi & Kobata, 1991; Clogston et al., 1993; Orita et al., 1994; Hoffman et al., 1996). Also, recombinant hIL-2 is O-glycosylated amino-terminal to helix A (T3) (Robb et al., 1984). Consistent with the predicted helices in the hTPO model, the cytokine domain of hTPO is O-glycosylated in similar locations: amino-terminal to helix A (S1) and in both the overhand loops, AB (T37) and CD (T110). The preference for O-glycosylation in the solvent-exposed loops of these proteins highlights the importance of structural context as well as of neighboring sequence in predicting O-glycosylation events. In addition to the O-glycosylation sites in the cytokine domain, hTPO has five sites in 153−246: T157, T158, S163, T192, and S244. On the basis of sequence analysis of the cloned TPO homologs and our structural characterization of

recombinant hTPO, we propose a division of the loosely defined C-terminal domain into two domains: an *N*-glycan domain (154−246) and an *O*-glycan domain (247−332). The *N*-glycan domain directly follows the cytokine domain and contains the four conserved N-glycosylation sites. The rat and murine *N*-glycan domains have one additional N-glycosylation sequon, and the canine has two additional sequons that may be glycosylated. In the human sequence, the *N*-glycan domain is 93 amino acids long and separated from the *O*-glycan domain by an additional dibasic site (R245-R246). The murine and human *N*-glycan domain sequences have 69% amino acid identity. The pattern of O-glycosyation in the *N*-glycan domain suggests that the two termini and the central portion of this domain are not involved in secondary-structure elements. This observation does not preclude the possibility that the N domain adopts a unique folded structure.

The *O*-glycan domain is the last domain, and the structural feature that characterizes it is the high percentage of proline, serine, and threonine residues. The human *O*-glycan domain is 86 amino acids long with 24% proline and 28% serine + threonine, indicating that this domain is a prime candidate for extensive O-glycosylation. The prediction algorithm of Elhammer and co-workers reveals 16 Ser and Thr residues with a p of >0.19 in the *O*-glycan domain. Not surprisingly, MALDI-TOF mass spectrometry indicates that hTPO 247−332 has a mass of 16.3 kDa while the polypeptide accounts for 8.9 kDa, only 55% of the measured mass. It has the least degree of sequence homology among the three proposed domains, 56% amino acid identity between the human and murine sequences. There are two N-glycosylation sequons in the last 21 amino acids of this domain in the human, murine, and canine sequences; they have been deleted in the truncated rat sequence. The amino acid composition and potentially heavy O-glycosylation of the *O*-glycan domain of hTPO is reminiscent of the mucins, and it is likely that this domain shares a characteristic structural feature of the mucins, i.e. an extended filamentous conformation (Shogren et al., 1989; Jentoft, 1990). To date, a lack of cleavage sites for specific proteases (e.g. trypsin, endoproteinase Glu-C, and endoproteinase Asp-N) has greatly hampered efforts to provide a peptide and glycosylation map of the *O*-glycan domain of hTPO.

We have provided a peptide map of full length BHK-recombinant hTPO from S1 to R246 and shown that the amino acid sequence is identical to that of the translated cDNA with no amino acid modifications other than glycosylation. The two disulfide bridges in the cytokine domain are consistent with this domain of TPO being a member of the four-helix bundle family of cytokines. The glycosylation map shows that the four N-glycosylation sequons in the *N*-glycan domain of TPO are occupied with complex *N*-glycans predominantly of the biantennary and triantennary type. The pattern of O-glycosylation in the cytokine domain is consistent with the structural model proposed, and O-glycosylation in the *N*-glycan domain suggests that the termini and central portion of this domain are not involved in secondary-structure elements. Finally, the molecular mass of the *O*-glycan domain, the high preponderance of prolyl residues, and the fact that it has 16 predicted O-glycosylation sites suggest that it does not form a compact globular domain but may have a filamentous structure characteristic of the mucins. Three-dimensional structural analysis of hTPO and functional analysis of the two C-terminal domains will greatly aid our understanding of this unique cytokine.

ACKNOWLEDGMENT

The authors thank Scott R. Presnell for information concerning his structural model of the cytokine domain of human TPO and Åke Elhammer for providing us with the O-glycosylation prediction algorithm. We also thank Karen DeJongh and Niels Kristian Klausen for a critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Table 1 with the elution time, masses, fragment ions observed, and assignments of all the peaks in the tryptic maps of recombinant human TPO and Table 2 with the structural characterization data and assignments for the 20 *N*-glycan HPAEC-PAD peaks from recombinant human TPO (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Abdel-Meguid, S. S., Shiel, H.-S., Smith, W. W., Dayringer, H. E., Violand, B. N., & Bentle, L. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6434−6437.

Aubert, J. P., Helbecque, N., & Loucheux-Lefebvre, M. H. (1981) *Arch. Biochem. Biophys. 208*, 20−29.

Bartley, T. D., Bogenberger, J., Hunt, P., Li, Y.-S., Lu, H. S., Martin, F., Chang, M.-S., Samal, B., Nichol, J. L., Swift, S., Johnson, M. J., Hsu, R.-Y., Parker, V. P., Suggs, S., Skrine, J. D., Merewether, L. A., Clogston, C., Hsu, E., Hokom, M. M., Hornkohl, A., Choi, E., Pangelinan, M., Sun, Y., Mar, V., McNinch, J., Simonet, L., Jacobsen, F., Xie, C., Shutter, J., Chute, H., Basu, R., Selander, L., Trollinger, D., Sieu, L., Padilla, D., Trail, G., Elliott, G., Izumi, R., Covey, T., Crouse, J., Garcia, A., Xu, W., Del Castillo, J., Biron, J., Cole, S., Hu, M. C.-T., Pacifici, R., Ponting, I., Saris, C., Wen, D., Yung, Y. P., Lin, H., & Bosselman, R. A. (1994) *Cell 77*, 1117−1124.

Bause, E. (1983) *Biochem. J. 209*, 331−336.

Bause, E., & Legler, G. (1981) *Biochem. J. 195*, 639−644.

Bazan, J. F. (1990a) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6934−6938.

Bazan, J. F. (1990b) *Immunol. Today 11*, 350−354.

Bazan, J. F. (1992) *Science 257*, 410−412.

Beeley, J. G. (1977) *Biochem. Biophys. Res. Commun. 76*, 1051−1055.

Bergwerff, A. A., van Oostrum, J., Asselbergs, F. A., Burgi, R., Hokke, C. H., Kamerling, J. P., & Vliegenthart, J. F. (1993) *Eur. J. Biochem. 212*, 639−656.

Bergwerff, A. A., van Oostrum, J., Kamerling, J. P., & Vliegenthart, J. F. (1995) *Eur. J. Biochem. 228*, 1009−1019.

Boissel, J.-P., Lee, W.-R., Presnell, S. R., Cohen, F. F., & Bunn, H. F. (1993) *J. Biol. Chem. 268*, 15983−15993.

Brandhuber, B. J., Boone, T., Kenney, W. C., & McKay, D. B. (1987) *Science 238*, 1707−1709.

Brockhausen, I., Carver, J. P., & Schachter, H. (1988) *Biochem. Cell Biol. 66*, 1134−1151.

Carr, S. A., Huddleston, M. J., & Bean, M. F. (1993) *Protein Sci. 2*, 183−196.

Carraway, K. L., & Hull, S. R. (1989) *BioEssays 10*, 117−121.

Chan, A. L., Morris, H. R., Panico, M., Etienne, A. T., Rogers, M. E., Gaffney, P., Creighton-Kempsford, L., & Dell, A. (1991) *Glycobiology 1*, 173−185.

Clogston, C., Hu, S., Boone, T., & Lu, H. (1993) *J. Chromatogr. 637*, 55−62.

Conradt, H. S., Nimtz, M., Dittmar, K. E. J., Lindenmaier, W., Hoppe, J., & Hauser, H. (1989) *J. Biol. Chem. 264*, 17368−17373.

Cumming, D. A. (1991) *Glycobiology 1*, 115−130.

de Sauvage, F. J., Hass, P. E., Spencer, S. D., Malloy, B. E., Gurney, A. L., Spencer, S. A., Darbonne, W. C., Henzel, W. J., Wong, S. C., Kuang, W.-J., Oles, K. J., Hultgren, B., Solberg, L. A., Jr., Goeddel, D. V., & Eaton, D. L. (1994) *Nature 369*, 533−538.

de Vos, A. M., Ultsch, M., & Kossiakoff, A. A. (1992) *Science 255*, 306−312.
Diederichs, K., Boone, T., & Karplus, P. A. (1991) *Science 254*, 1779−1782.
Elhammer, Å. P., Poorman, R. A., Brown, E., Maggiora, L. L., Hoogerheide, J. G., & Kézdy, F. J. (1993) *J. Biol. Chem. 268*, 10029−10038.
Ellis, M. H., Avraham, H., & Groopman, J. E. (1995) *Blood Rev. 9*, 1−6.
Gavel, Y., & von Heijne, G. (1990) *Protein Eng. 3*, 433−442.
Gawlitzek, M., Valley, U., Nimtz, M., Wagner, R., & Conradt, H. S. (1995) *J. Biotechnol. 42*, 117−131.
Gelpi, E. (1995) *J. Chromatogr. A 703*, 59−80.
Goochee, C. F. (1992) *Dev. Biol. Stand. 76*, 95−104.
Goochee, C. F., & Monica, T. (1990) *Bio/Technology 8*, 421−427.
Goto, M., Akai, K., Murakami, A., Hasimoto, C., Tsuda, E., Ueda, M., Kawanishi, G., Takahashi, N., Ishimoto, A., Chiba, H., & Saski, R. (1988) *Bio/Technology 6*, 67−71.
Hermentin, P., Witzel, R., Vliegenhart, J. F. G., Kamerling, J. P., Nimtz, M., & Conradt, H. S. (1992) *Anal. Biochem. 203*, 281−289.
Hill, C. P., Osslund, T. D., & Eisenberg, D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5167−5171.
Hironaka, T., Furakawa, K., Esmon, P. C., Fournel, M. A., Sawada, S., Kato, M., Minaga, T., & Kobata, A. (1992) *J. Biol. Chem. 267*, 8012−8020.
Hoffman, R. C., Moy, F. J., Price, V., Richardson, J., Kaubisch, D., Frieden, E. A., Krakover, J. D., Castner, B. J., King, J., March, C. J., & Powers, R. (1996) *J. Biomol. NMR 7*, 273−282.
Hokke, C. H., Bergwerff, A. A., Van Dedem, G. W., Kamerling, J. P., & Vliegenthart, J. F. (1995) *Eur. J. Biochem. 228*, 981−1008.
Hunt, L. T., & Dayhoff, M. O. (1970) *Biochem. Biophys. Res. Commun. 39*, 757−765.
Imperiali, B., & Shannon, K. L. (1991) *Biochemistry 30*, 4374−4380.
James, D. C., Freedman, R. B., Hoare, M., Ogonah, O. W., Rooney, B. C., Larionov, O. A., Dobrovolsky, V. N., Lagutin, O. V., & Jenkins, N. (1995) *Bio/Technology 13*, 592−596.
Jentoft, N. (1990) *Trends Biochem. Sci. 15*, 291−294.
Kagawa, Y., Takasaki, S., Utsumi, J., Hosoi, K., Shimizu, H., Kachibe, N., & Kobata, A. (1988) *J. Biol. Chem. 263*, 17508−17515.
Kaushansky, K. (1995a) *Int. J. Hematol. 62*, 7−15.
Kaushansky, K. (1995b) *Blood 86*, 419−431.
Klausen, N. K., & Bayne, S. (1995) *Glycoconjugate J. 12*, 409.
Kornfeld, R., & Kornfeld, S. (1985) *Annu. Rev. Biochem. 54*, 631−664.
Legrand, D., Salmon, V., Coddeville, B., Benaissa, M., Plancke, Y., & Spik, G. (1995) *FEBS Lett. 365*, 57−60.
Lok, S., & Foster, D. C. (1994) *Stem Cells (Dayton) 12*, 586−598.
Lok, S., Kaushansky, K., Holly, R. D., Kuijper, J. L., Lofton-Day, C. E., Oort, P. J., Heipel, M. D., Burkhead, S. K., Kramer, J. M., Bell, L. A., Sprecher, C. A., Blumberg, H., Johnson, R., Prunkard, D., Ching, A. F. T., Mathewes, S. L., Bailey, M. C., Forstrom, J. W., Buddle, M. M., Osborn, S. G., Evans, S. J., Sheppard, P. O., Presnell, S. R., O'Hara, P. J., Hagen, F. S., Roth, G. J., & Foster, D. C. (1994) *Nature 369*, 565−568.
McKay, D. B. (1992) *Science 257*, 412−413.
Metcalf, D. (1994) *Nature 369*, 519−520.
Milburn, M. V., Hassell, A. M., Lambert, M. H., Jordan, S. R., Proudfoot, A. E. I., Graber, P., & Wells, T. N. C. (1993) *Nature 363*, 172−176.
Nimtz, M., Martin, W., Wray, V., Klöppel, K.-D., Augustin, J., & Conradt, H. S. (1993) *Eur. J. Biochem. 213*, 39−56.
Ogami, K., Shimada, Y., Sohma, Y., Akahori, H., Kato, T., Kawamura, K., & Miyazaki, H. (1995) *Gene 158*, 309−310.
Orita, T., Oh-eda, M., Hasegawa, M., Kuboniwa, H., Esaki, K., & Ochi, N. (1994) *J. Biochem. 115*, 345−350.
Påhlsson, P., Blackall, D. P., Ugorski, M., Czerwinski, M., & Spitalnik, S. L. (1994) *Glycoconjugate J. 11*, 43−50.
Pfeiffer, G., Stirm, S., Geyer, R., Strube, K. H., Bergwerff, A. A., Kamerling, J. P., & Vliegenthart, J. F. (1992) *Glycobiology 2*, 411−418.
Pollack, L., & Atkinson, P. H. (1983) *J. Cell Biol. 97*, 293−300.
Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1992) *Science 256*, 1673−1677.
Robb, R., Kutny, R., Panico, M., Morris, H., & Chowdry, V. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6486−6490.
Robinson, R. C., Grey, L. M., Staunton, D., Vankelecom, H., Vernallis, A. B., Moreau, J. F., Stuart, D. I., Heath, J. K., & Jones, E. Y. (1994) *Cell 77*, 1101−1116.
Roitsch, T., & Lehle, L. (1989) *Eur. J. Biochem. 181*, 525−529.
Rush, R. S., Derby, P. L., Strickland, T. W., & Rohde, M. F. (1993) *Anal. Chem. 65*, 1834−1842.
Rush, R. S., Derby, P. L., Smith, D. M., Merry, C., Rogers, G., Rohde, M. F., & Katta, V. (1995) *Anal. Chem. 67*, 1442−1452.
Sasaki, H., Bothner, B., Dell, A., & Fukuda, M. (1987) *J. Biol. Chem. 262*, 12059−12076.
Schachter, H. (1986) *Biochem. Cell Biol. 64*, 163−181.
Schachter, H. (1991) *Glycobiology 1*, 453−461.
Shogren, R., Gerken, T. A., & Jentoft, N. (1989) *Biochemistry 28*, 5525−5536.
Sinohara, H., & Maruyama, T. (1973) *J. Mol. Evol. 2*, 117−122.
Smith, L. J., Redfield, C., Boyd, J., Lawrence, G. M. P., Edwards, R. G., Smith, R. A. G., & Dobson, C. M. (1992) *J. Mol. Biol. 224*, 900−904.
Takeuchi, M., & Kobata, A. (1991) *Glycobiology 1*, 337−346.
Townsend, R. R., Hardy, M. R., Cumming, D. A., Carver, J. P., & Bendiak, B. (1989) *Anal. Biochem. 182*, 1−8.
Tsuda, E., Goto, M., Murakami, A., Akai, K., Ueda, M., Kawanishi, G., Takahashi, N., Sasaki, R., Chiba, H., Ishihara, H., Mori, M., Tejima, S., Endo, S., & Arata, Y. (1988) *Biochemistry 27*, 5646−5654.
Vignon, I., Mornon, J.-P., Cocault, L., Mitjavila, M.-T., Tambourin, P., Gisselbrecht, S., & Souryi, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5640−5644.
Vliegenthart, J. F., Dorland, L., & van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem. 41*, 209−374.
Wada, T., Nagata, Y., Nagahisa, H., Okutomi, K., Ha, S. H., Ohnuki, T., Kanaya, T., Masumura, M., & Todokoro, K. (1995) *Biochem. Biophys. Res. Commun. 213*, 1091−1098.
Walter, M. R., Cook, W. J., Ealick, S. E., Nagabhushan, T. L., Trotta, P. P., & Bugg, C. E. (1992a) *J. Mol. Biol. 224*, 1075−1085.
Walter, M. R., Cook, W. J., Zhao, B. G., Cameron, R. P., Ealick, S. E., Walter, R. L., Reichert, P., Nagabhushan, T. L., Trotta, P. P., & Bugg, C. E. (1992b) *J. Biol. Chem. 267*, 20371−20376.
Watson, E., Shah, B., Liederman, L., Hsu, Y.-R., Karkare, S., Lu, H. S., & Lin, F.-K. (1994) *Biotechnol. Prog. 10*, 39−44.
Wlodawer, A., Pavlovsky, A., & Gustchina, A. (1992) *FEBS Lett. 309*, 59−64.
Wlodawer, A., Pavlovsky, A., & Gustchina, A. (1993) *Protein Sci. 2*, 1373−1382.

BI961075B