X-ray Crystallography and Mass Spectroscopy Reveal that the N-lobe of Human Transferrin Expressed in *Pichia pastoris* Is Folded Correctly but Is Glycosylated on Serine-32[†,‡]

Maria C. Bewley,[§,||] Beatrice M. Tam,[⊥] Jasmine Grewal,[⊥] Shouming He,[#] Steven Shewry,[§,▽] Michael E. P. Murphy,[⊥] Anne B. Mason,[○] Robert C. Woodworth,[○] Edward N. Baker,[§,▽] and Ross T. A. MacGillivray*[,⊥]

Institute of Molecular Biosciences, College of Sciences, Massey University, Private Bag 11222, Palmerston North, New Zealand, Departments of Biochemistry and Molecular Biology and of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada, and Department of Biochemistry, University of Vermont, Burlington, Vermont 05405-0068

Received October 14, 1998; Revised Manuscript Received December 7, 1998

ABSTRACT: The ferric form of the N-lobe of human serum transferrin (Fe(III)-hTF/2N) has been expressed at high levels in *Pichia pastoris*. The Fe(III)-hTF/2N was crystallized in the space group $P4_12_12$, and X-ray crystallography was used to solve the structure of the recombinant protein at 2.5 Å resolution. This represents only the second *P. pastoris*-derived protein structure determined to date, and allows the comparison of the structures of recombinant Fe(III)-hTF/2N expressed in *P. pastoris* and mammalian cells with serum-derived transferrin. The polypeptide folding pattern is essentially identical in all of the three proteins. Mass spectroscopic analyses of *P. pastoris*- hTF/2N and proteolytically derived fragments revealed glycosylation of Ser-32 with a single hexose. This represents the first localization of an O-linked glycan in a *P. pastoris*-derived protein. Because of its distance from the iron-binding site, glycosylation of Ser-32 should not affect the iron-binding properties of hTF/2N expressed in *P. pastoris*, making this an excellent expression system for the production of hTF/2N.

The transferrins are a family of metal-binding proteins that function in the chelation and transport of ferric iron (*1*). Like other transferrins, human serum transferrin (hTF[1]) consists of a single polypeptide chain of M_r 80 000 that is comprised of two homologous lobes (the N-lobe and C-lobe) joined by a short bridging peptide. The amino acid sequence (*2*), cDNA sequence (*3*), and gene structure (*4*) of human transferrin are consistent with the occurrence of a gene duplication event during the evolution of transferrins. Each lobe of transferrin contains a high-affinity binding site for ferric iron and other transition metals. X-ray crystallographic analyses of lactoferrin (*5*) and other transferrins (*6*, *7*) have shown that the iron-binding site consists of four ligands from the protein and two ligands from a synergistically bound anion such as carbonate. In its role in iron metabolism, transferrin carries iron from sites of absorption in the gut to peripheral cells where the Fe(III)−transferrin complex is internalized by transferrin receptor-mediated endocytosis (*8*). After the pH- and receptor-dependent release of iron, both the transferrin and transferrin receptor are recycled to the plasma membrane where each is reused (for reviews, see refs *1*, *9*).

For studying the iron-binding and receptor-binding functions of transferrins more readily, recombinant transferrins have been expressed by using both prokaryotic and eukaryotic systems. Initially, the N-lobe of human transferrin (*10*) and several point mutants (*11*) were expressed in baby hamster kidney cells using the plasmid pNUT as a vector (Table 1). The X-ray crystal structure of BHK-derived Fe(III)-hTF/2N (*12*) showed that the polypeptide chain folded in a way similar to that of the N-lobe of human serum transferrin (*13*) and other transferrins whose structures are known (see ref *12*). The pNUT-BHK cell system has been used to express other transferrins including glycosylated and nonglycosylated hTF (*14*), hTF/2C (*15*), hLF (*16*), hLF/2N

[†] Supported by grants from the U. S. Public Health Service (R01 DK 21739 to R.C.W. and R01 HD 20859 to E.N.B.), and the Medical Research Council of Canada (to R.T.A.M.). During his stay at Massey University, R.T.A.M. was supported in part by a Killam Senior Research Fellowship and by USPHS Grant R01 DK 35533. E.N.B. also acknowledges research support as an International Research Scholar of the Howard Hughes Medical Institute.

[‡] Atomic coordinates of the protein structure reported in this paper has been deposited with the Protein Data Bank at Brookhaven National Laboratories with PDB Code 1b3e.

* Corresponding author. Tel: (604)822.3027. Fax: (604)822.4364. E-mail: macg@unixg.ubc.ca.

[§] Massey University.

[||] Current address: Brookhaven National Laboratory, Upton, NY 11973-5000.

[⊥] Department of Biochemistry and Molecular Biology, University of British Columbia.

[#] Department of Chemistry, University of British Columbia.

[▽] Current address: School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand.

[○] University of Vermont.

[1] Abbreviations: hTF, human serum transferrin; hTF/2N, recombinant N-lobe of human transferrin comprising residues 1−337; hTF/2C, recombinant C-lobe of human transferrin comprising residues 338−681; hLF, human lactoferrin; hLF/2N, recombinant N-lobe of human lactoferrin; oTF, chick ovotransferrin; oTF/2N, recombinant N-lobe of ovotransferrin; bTF, bovine transferrin; BHK cells, baby hamster kidney cells; EDTA, ethylenediamine-tetraacetic acid, rms, root-mean-square; amu, atomic mass units; ESI-MS, electrospray ionization mass spectroscopy; CAM, carboxamidomethyl; TIC, total ion chromatogram.

10.1021/bi9824543 CCC: $18.00
Published on Web 02/02/1999

Table 1: Expression of Recombinant Transferrins

expression system	protein	expression level (mg/L)	reference
E. coli	hTF	15	*19*
E. coli	hTF	not given	*20*
E. coli	hTF/2N, hTF/2C	not given	*21*
E. coli	hTF, hTF/2N, hTF/2C	60[a]	*22, 53*
P. pastoris	hTF/2N	50	*26*
P. pastoris	hTF/2N	240−1000	*27, 28*
A. oryzae	hLF	25	*25*
Baculovirus	hTF, bTF	20	*23*
Baculovirus	hTF	20	*24*
pNUT-BHK cells	hTF/2N	55−120	*10*
pNUT-BHK cells	hTF/2C	12	*15*
pNUT-BHK cells	hTF	125	*14, 15*
pNUT-BHK cells	hTF (non-glycosylated)	25−95	*14, 15*
pNUT-BHK cells	hLF	20	*16*
pNUT-BHK cells	hLF/2N	35	*17*
pNUT-BHK cells	oTF (non-glycosylated)	80	*18*
pNUT-BHK cells	oTF/2N	30−55	*18*

[a] Yield after renaturation was ~3 mg/mL.

(*17*), and nonglycosylated oTF and oTF/2N (*18*). Because the growth of mammalian tissue culture cells is expensive and labor-intensive, several alternative expression systems have been developed, including *Escherichia coli* (*19*−*22*), baculovirus and insect cells (*23*, *24*), *Aspergillus oryzae* (*25*), and *Pichia pastoris* (*26*−*28*) (Table 1).

In this manuscript, we report the crystallization and three-dimensional structure of Fe(III)-hTF/2N expressed in *P. pastoris*. The polypeptide chain folds in a fashion identical to that of Fe(III)-hTF/2N expressed in BHK cells (*12*) and to the N-lobe of serum transferrin (*13*). In contrast to BHK-derived hTF/2N which is nonglycosylated, *P. pastoris* hTF/2N is made in both a glycosylated form and a nonglycosylated form (*27*). By using mass spectroscopy, we have identified the glycosyl moiety as a single hexose (probably mannose) attached to Ser-32.[2]

EXPERIMENTAL PROCEDURES

Materials. DEAE Sephacel and Sephacryl S-100 were obtained from Pharmacia Biotech, Baie D′Urfé, PQ. The 0.22 μm filtration membranes, Amicon stirred cell concentrators, and ultrafiltration (10,000 MW cutoff) membranes were from Millipore, Nepean, ON. Analytical mixed bed resin was from BioRad, Mississauga, ON. Trypsin was obtained from Boehringer Mannheim, Laval, PQ. The solvents used for HPLC were trifluoroacetic acid (from Pierce Chemical Co., Rockford, IL) and acetonitrile (from Fisher Scientific Canada, Nepean, ON). All other reagents were of analytical grade or higher and were obtained from the Sigma Chemical Company or Fisher Scientific Canada.

Expression and Purification of hTF/2N in P. pastoris. hTF/2N was expressed in BHK cells and in *P. pastoris*, and purified from the cell culture media as described previously (*27*, *29*). The purified preparations of hTF/2N migrated as single bands when analyzed by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate, and were of the expected mass distribution when analyzed by ESI-MS (*11*, *27*). Amino-terminal sequence analysis of the purified hTF/2N was performed on a Perkin-Elmer − Applied Biosystems model 476A protein sequencer at the Nucleic Acid and Protein Service facility at the University of British Columbia, and gave only the expected N-terminal sequence of human transferrin (*2*).

[2] All amino acid numbering is taken from the human transferrin sequence (*3*).

Table 2: Data Collection Statistics of *P. pastoris*-Derived Fe(III)-hTF/2N

space group	$P4_12_12$
cell dimensions (Å)	$a = b = 73.5$, $c = 156.7$
resolution (Å)	2.5
R_{merge} on I	0.066 (0.245)[a]
$\langle I \rangle / \langle \sigma(I) \rangle$	17.0 (4.2)
completeness (%)	96.9 (96.0)
unique reflections	15 238
redundancy	3.9

[a] Figures in parentheses refer to the outer shell (2.5−2.6 Å).

Crystallization and Data Collection. *P. pastoris*-derived hTF/2N was dissolved in 150 mM sodium bicarbonate buffer and concentrated to approximately 80 mg/mL using an Amicon Centricon Microconcentrator. The purified protein was crystallized by using the sitting drop method; the well solution contained ethanol (40%, v:v) in 50 mM sodium cacodylate buffer, pH 6.25. After 2 months at 4 °C, deep red rhombohedral crystals appeared. Data were collected on an R-Axis IIC imaging plate mounted on a Rigaku RU-200 rotating anode generator with a graphite monochromator operating at 50 kV and 100 mA. An Oxford Cryostream cooling unit was used to maintain the crystal temperature at 4 °C during data collection. The data collection statistics are summarized in Table 2. The data were processed by using the computer programs DENZO and SCALEPACK (*30*).

Structural Solution and Refinement. Prior to refinement, a random subset of 1069 reflections (8% of the data) was flagged and used only in the calculation of the free *R*-factor (*31*). The coordinates of BHK-derived Fe(III)-hTF/2N (*12*) minus solvent atoms, the Fe(III) atom, and the carbonate anion were used directly to phase the *P. pastoris* Fe(III)-hTF/2N structure. Each domain of Fe(III)-hTF/2N was defined as a separate rigid body structure; rigid body refinement was performed by using the program XPLOR (*32*) for all data between 10 and 2.5 Å resolution. After rigid body refinement, the *R*-factor was 0.339 and the free *R*-factor was 0.332. Electron density maps ($2F_o - F_c$ and $F_o - F_c$) were calculated and the temperature factors reset to 20.0 Å^2. An Fe(III) atom and a carbonate anion were added, and the model was refined with XPLOR using the simulated annealing protocol at 1500 K for all data between 6 and 2.5 Å resolution. Further refinement in XPLOR was punctuated by rounds of manual rebuilding using the program TOM (*33*). Water molecules were added when they formed hydrogen bonds to at least one protein atom and were present in the BHK-derived Fe(III)-hTF/2N structure. Finally, tightly restrained individual temperature factor refinement was carried out. These rounds of refinement and model building reduced the *R*-factor to 0.177 and the free *R*-factor to 0.255. The free *R*-factor decreased at all stages of the refinement, indicating that refining tightly restrained individual temperature factors was appropriate. The final model contains residues 4−333 of Fe(III)-hTF/2N, 64 water molecules, one iron atom, and one carbonate ion. The refinement statistics of the final model are shown in Table 3; the average *B*-values of the *P. pastoris* Fe(III)-hTF/2N model are similar to those

Table 3: Refinement Statistics of *P. pastoris*-Derived Fe(III)-hTF/2N

resolution range (Å)	6.0−2.5
molecules in asymmetric unit	1
working *R*-factor	0.177 (12421 reflections)
free *R*-factor	0.255 (1069 reflections)
rms deviation from ideal geometry	
bond length (Å)	0.010
bond angles (deg)	1.8
average *B*-factor (main chain) (Å^2)	34.79
average *B*-factor (side chains) (Å^2)	38.51
average *B*-factor of solvent (Å^2)	38.46
number of non-H atoms	2627
protein atoms	2558
Fe(III) + CRB	5
number of solvent atoms	64

determined for the BHK-derived orthorhombic crystal form under the same conditions of data collection (*12*). The quality of the final model was assessed by using the program PROCHECK (*34*).

Electrospray Ionization Mass Spectroscopy. ESI-MS of apo-hTF/2N and derived peptides was performed using a PE-Sciex API 300 triple-quadrupole mass spectrometer (PE-Sciex, Thornhill, Ontario) equipped with an ionspray ion source; details of the instrumentation have been reported previously (*35*). To reduce and alkylate the cysteine residues, we dissolved samples of BHK- and *P. pastoris*-derived hTF/2N at a concentration of 10 mg/mL in 0.2 M Tris-HCl buffer, pH 8.6, containing 8 M deionized urea and 1% (v:v) 2-mercaptoethanol. After 3 h at room temperature, the cysteine residues were alkylated by the addition of excess iodoacetamide. The insoluble, carboxamidomethylated hTF/2N polypeptides were dialyzed against distilled water, lyophilized, and dissolved in 30% (v/v) acetic acid. The samples were introduced into the mass spectrometer through a column (1 × 50 mm) of PLRP-S resin equilibrated with 80% solvent A (0.05% trifluoroacetic acid, 2% acetonitrile in water) and 20% solvent B (0.045% triflouroacetic acid, 90% acetonitrile in water). The polypeptides were eluted with a gradient of 20−100% solvent B over 5 min at a flow rate of 50 mL/min. The quadrupole mass analyzer was scanned over an *m*/*z* range of 600−2400 Da.

Proteolytic Digestion of Carboxamidomethylated hTF/2N and Analysis by ESI-MS. Samples (approximately 1 mg/mL) of BHK- and *P. pastoris*-derived hTF/2N were reduced and carboxamidomethylated as described above and dialyzed against tryptic digest buffer (20 mM Tris HCl, pH 8.5, containing 20 mM $CaCl_2$). Trypsin (5% w:w) was added, and digestion was allowed to proceed overnight at room temperature, during which the precipitated protein gradually redissolved. The reaction was terminated by the addition of acetic acid to 30% (v/v). For mass analysis, the tryptic peptides were separated by reverse-phase HPLC using an Ultrafast microprotein analyzer (Michrom BioResources Inc., Auburn, CA) directly interfaced with the mass spectrometer. Peptides were applied to a column (1 × 150 mm) of Delta-Pak C-18 resin equilibrated with solvent A and eluted with a gradient (0−60%) of solvent B over 60 min at a flow rate of 0.7 mL/min. The quadrupole mass analyzer was scanned over a *m*/*z* range of 300−2200 Da. A postcolumn splitter was used to introduce 10% of the eluent into the mass spectrometer while the remaining 90% was collected in a fraction collector for further analysis.

Sequence Analysis of the Glycopeptide and Identification of the Glycosylated Residue. After identification of the tryptic glycopeptide by mass analysis, the corresponding fraction was cleaved with pepsin (*35*) and the resulting peptides were injected into the mass spectrometer via the HPLC column as before. The quadrupole mass analyzer was scanned in the third quadrupole over the range from 50 to 1590 amu. The mass spectrometer conditions were identical to those used previously except that the first quadrupole was set to admit only the single-charged glycosylated species with a *m*/*z* ratio of 1191 into the collision cell. In the second quadrupole, the peptide ions were fragmented by collision with nitrogen under conditions that favor peptide bond cleavage. Fragment ions were observed in the third quadrupole, and the peptide sequences were deduced from the masses of the resulting peptide fragments.

RESULTS

Three-Dimensional Structure of P. pastoris-Derived Fe(III)-hTF/2N. *P. pastoris*-derived Fe(III)-hTF/2N crystallized under similar conditions and in the same space group as BHK-derived Fe(III)-hTF/2N (*12*). The crystals were deep red in color indicating that iron was retained in the crystal form. In contrast to the BHK-derived Fe(III)-hTF/2N that formed large crystals within 7 days, the *P. pastoris* protein crystallized much more slowly and gave smaller crystals. As the *P. pastoris* Fe(III)-hTF/2N is glycosylated heterogeneously (*27*), it is possible that the crystals consist of a minor species of the preparation. The lower concentration of such a minor species would explain the slower rate of crystallization compared to the nonglycosylated BHK Fe(III)-hTF/2N (*12*).

A typical crystal of *P. pastoris* Fe(III)-hTF/2N diffracted to 2.5 Å resolution. The cell dimensions (Table 2) and diffraction data showed that the structure was isomorphous with the tetragonal crystal form of the BHK-derived Fe(III)-hTF/2N (*12*). After rigid body refinement, molecular dynamics refinement, and manual rebuilding, the final model for the *P. pastoris* Fe(III)-hTF/2N has a crystallographic *R*-factor of 0.177 and a free *R*-factor of 0.255 (Table 3). A Ramachandran plot of main torsion angles (*36*) generated by the program PROCHECK (*34*) showed that 88.1% of the residues fall in the most favored regions, with a further 10.5% in allowed regions. Only two residues (Leu-146 and Leu-294) fall outside the permissable regions. Leu-146 lies at the end of a surface loop that is disordered in the BHK-derived hTF/2N structure (*12*) and in hLF/2N (*37*). Leu-294 has ϕ and φ angles of 73° and −45°, respectively, and allows a tight turn that has the conformation of a reverse classical γ-turn (*38*); this γ-turn is found in all of the transferrin structures reported to date (see ref *12*).

As shown in Figure 1, the *P. pastoris* Fe(III)-hTF/2N has a folding pattern identical to that of BHK-derived Fe(III)-hTF/2N and other transferrins whose structures have been determined (see refs *1*, *12*). The polypeptide folds into two domains that are separated by a deep cleft containing the metal-binding site. In Fe(III)-hTF/2N, the ligands to the iron are donated by the side chains of Asp-63, Tyr-95, Tyr-188, and His-249 together with a bidentate carbonate; in turn, the carbonate is stabilized by a hydrogen-bonding network involving the side chains of Arg-124 and Thr-120 and the

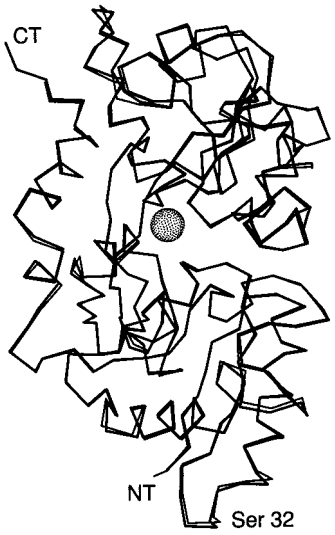

FIGURE 1: Comparison of the structures of Fe(III)-hTF/2N derived from *P. pastoris* and BHK cells (*12*). The α-carbon backbones of the molecules are shown for *P. pastoris* Fe(III)-hTF/2N (thick line) and the tetragonal crystal form of BHK-derived Fe(III)-hTF/2N (thin line). The structures were superimposed using the program Xtal-View (*55*). The N-termini (NT) and C-termini (CT) are labeled. The iron atom is represented by the sphere at the bottom of the cleft between the two domains. The residue that is glycosylated by *P. pastoris* is labeled (Ser-32).

main chain amides of Ala-126 and Gly-127. The stability of the iron-binding site in the *P. pastoris* Fe(III)-hTF/2N is indicated by the low crystallographic *B*-factors for the side chains of the liganding residues (average *B*-factors for the side chain atoms of Asp-63, Tyr-95, Tyr-188, and His-249 are 24.0, 25.6, 16.8, and 22.0, respectively). There are three regions of the polypeptide chain that have high crystallographic *B*-factors; these regions correspond to surface loops involving residues 137−145, 215−222, and 276−281. Similar high temperature factor profiles have been reported for these regions in BHK-derived Fe(III)-hTF/2N and other members of the transferrin family (*12*). The common residues (4−331) of the BHK- and *P. pastoris*-derived Fe(III)-hTF/2N structures superimpose well with an rms deviation of alpha carbons of 0.33 Å, and an rms deviation for all atoms of 0.66 Å.

In contrast to the BHK structure, which displayed disorder at the anion-binding site (*12*), the *P. pastoris* Fe(III)-hTF/2N metal-binding site shows only a single conformation that is identical to the metal-binding sites observed in lactoferrin and other transferrins (*1*). Although it is possible that disorder might not be apparent at the lower resolution of the present analysis (2.5 Å), we believe that the difference may result from the higher pH of crystallization of the *P. pastoris* protein (pH 6.25) compared with the BHK protein (pH ∼6.0) (*12*). The *P. pastoris* protein may be just above the pH at which significant protonation of the carbonate occurs. As reported previously (*12*), the BHK-derived hTF/2N has a folding pattern similar to that found in the N-lobe of human serum transferrin (*13*). The *P. pastoris*-derived hTF/2N maintains this folding pattern.

The preparation of *P. pastoris* Fe(III)-hTF/2N used for crystallization was heterogeneous, consisting of nonglycosylated protein and glycosylated protein with 1−2 hexose units (*27*). Although *P. pastoris* can perform both N- and O-glycosylation (*39*), hTF/2N does not contain any potential N-glycosylation sites of the Asn-X-Ser type (*2*). To determine whether any residues of hTF/2N in the crystal showed evidence of O-glycosylation, we examined the electron density maps for additional electron density in the vicinity of surface-bound serine and threonine residues. Although the electron densities for the atoms for every serine and threonine side chain were visible, no additional density was observed that could account for a hexose residue(s). As the addition of one or two hexose residues would not be expected to cause thermal disorder in the side chain, it is probable that the crystals consist of the nonglycosylated form of Fe(III)-hTF/2N. The low concentration of nonglycosylated protein in the crystallization well is consistent with the lower rate of crystal formation compared to the BHK-derived protein.

ESI-MS of P. pastoris hTF/2N. To determine the glycosylation site of *P. pastoris*-derived hTF/2N, we used an ESI-MS approach. Because the glycosylation pattern of *P. pastoris* hTF/2N varies slightly from batch to batch (unpublished observations), samples of hTF/2N from BHK cells and from *P. pastoris* were reduced and carboxamidomethylated, and analyzed by ESI-MS. Consistent with previous studies (*11*) and taking into account the CAM group mass of 58 amu, BHK-derived CAM-hTF/2N is nonglycosylated and consists of a single species with a mass of 38 083 amu; this is within experimental error of the theoretical mass of 38 079 amu. The *P. pastoris* CAM-hTF/2N consisted of a minor species of a mass of 38 084 amu (corresponding to the nonglycosylated polypeptide) and a major species with a mass of 38 249 amu. This major species corresponds to the polypeptide containing an additional mass of 165 amu and is consistent with the addition of an O-linked single hexose unit (theoretical additional mass of 162 amu). However, this analysis does not differentiate between a single species in which a particular residue is glycosylated and multiple species having different glycosylated residues.

To clarify this point, we digested samples of CAM-hTF/2N from BHK cells and *P. pastoris* with trypsin. The resulting peptides were separated by reverse-phase HPLC and detected by ESI-MS. When the spectrometer was scanned in the normal LC-MS mode, the total ion chromatogram (TIC) of the BHK-derived and *P. pastoris*-derived protein digests displayed similar spectra consisting of a number of peaks (Figure 2a,c). When the peptide masses in each peak in the *P. pastoris* sample were compared with the masses of the peptides from the TIC of the BHK sample, the peptides were present at comparable retention times with the exception of one peptide (labeled * in Figure 2a,c). The *P. pastoris*-hTF/2N digest had a peptide with a mass of 1578 amu (Figure 2d) which was not present in the BHK-hTF/2N digest (Figure 2b). In addition, the BHK-hTF/2N digest had a peptide with a mass of 1416 amu (Figure 2b). The difference in mass of these peptides is 162 amu and accounts for the difference found in the intact polypeptides.

The identity of the glycopeptide was investigated by searching the amino acid sequence of hTF/2N for potential tryptic peptides of mass 1416 amu, but there were no peptides corresponding to this mass. However, a cysteine-containing peptide with a mass of 1359 amu was found; the corresponding CAM-cysteine peptide would have a mass of 1359 + 58 = 1417 amu which is within experimental error of the observed peptide of 1416 amu. The peptide corresponds to residues 28−41 of hTF/2N, and has the following sequence:

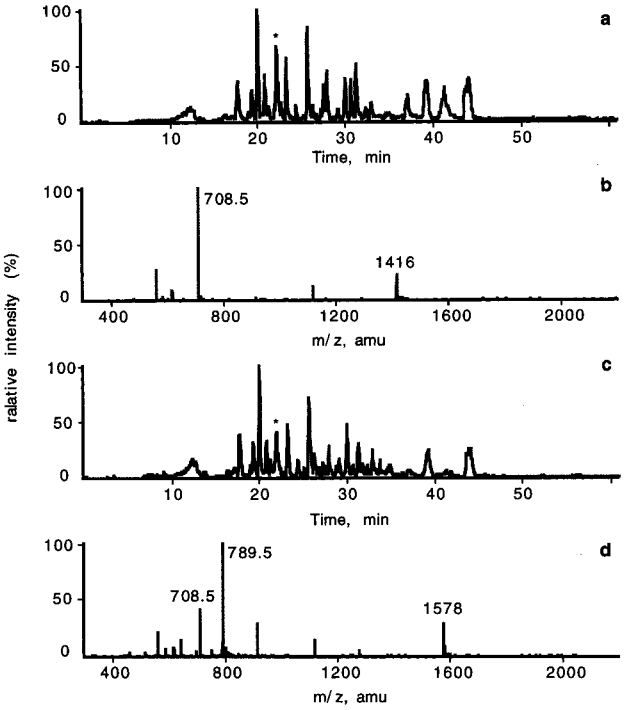

FIGURE 2: Comparative ESI-MS analysis of the tryptic peptides of hTF/2N: (a) ESI-MS TIC of tryptic digest of BHK-derived hTF/2N; (b) ESI-MS/MS TIC of the labeled peak from panel a; (c) ESI-MS TIC of tryptic digest of *P. pastoris*-derived hTF/2N; and (d) ESI-MS/MS TIC of the labeled peak from panel c. The peptides that differ between the two tryptic digests in panels a and c are labeled (*). See text for details.

NH_2-Ser-Val-Ile-Pro-Ser-Asp-Gly-Pro-Ser-Val-Ala-Cys-Val-Lys-COOH

Residue Ser-28 is preceded by Lys-27 allowing trypsin to cleave the Lys-Ser bond and give rise to the peptide. The isolated peptide was reintroduced into the mass spectrometer; fragmentation patterns confirmed that the peptide consisted of residues 28−41 of hTF/2N (data not shown). Taken together, these results show that the glycosylation of *P. pastoris* hTF/2N is confined to a single peptide.

Assuming that *P. pastoris* adds a single hexose to the hydroxyl group of serine or threonine residues, the hTF/2N glycopeptide contains three serine residues that are candidates for the glycosylation site. Inspection of the three-dimensional structure of Fe(III)-hTF/2N showed that the peptide constitutes a surface loop that is far removed from the iron-binding site (the distance from the Ser-32 Cα to the Fe(III) atom is 35 Å) (Figure 1). Residue Ser-28 is unlikely to be available for glycosylation as it points to the interior of the protein molecule (Figure 3). In contrast, Ser-32 and Ser-36 both appear to be likely candidates for glycosylation as they lie on the surface of the outer portion of domain 1 of hTF/2N and project directly out into the external environment (Figure 3). To determine which serine was the glycosylated residue, we introduced the peptide into the mass spectrometer under conditions that favored peptide bond cleavage. Although this peptide fragment analysis eliminated Ser-28 as the glycosylated residue, no peptide bond cleavages were observed between Ser-32 and Ser-36 so that it was not possible to distinguish between these residues (data not shown). To overcome this problem, we digested the peptide with pepsin which cleaved the Ala-CAMCys bond, and resulted in an

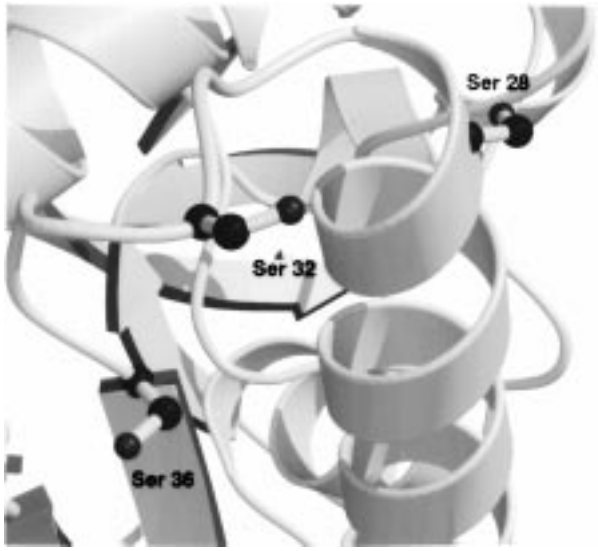

FIGURE 3: Orientation of the serine side chains contained within the tryptic glycopeptide of *P. pastoris*-derived hTF/2N. The image was generated using the programs Molscript (*56*) and Raster3D (*57*).

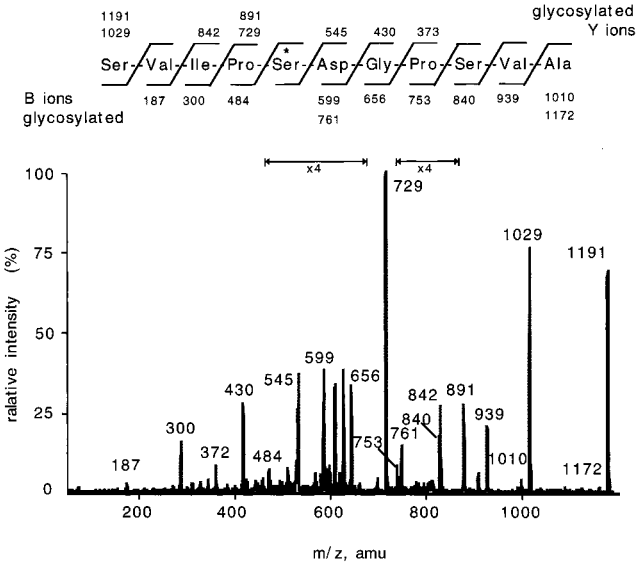

FIGURE 4: ESI-MS/MS of the glycopeptide (residues 28−38) of *P. pastoris*-derived hTF/2N. The intact glycosylated and nonglycosylated peptides with their respective masses of 1191 and 1029 amu are seen on the spectrum. The masses of the Y ions (resulting from a charge at the C-terminus of the peptide) are shown above the peptide sequence; the masses of B ions (resulting from a charge at the N-terminus of the peptide) are given below the peptide sequence. The derivation of peptide fragments is also shown, and includes the masses of glycosylated peptide fragments that are present in the chromatogram. For a region of the spectrum, the scale was expanded by 4-fold as indicated (×4).

11 residue peptide (consisting of residues 28−38 of hTF/2N; Ser-Val-Ile-Pro-Ser-Asp-Gly-Pro-Ser-Val-Ala) and a tripeptide (residues 39−41 of hTF/2N; CAMCys-Val-Lys). The 11 residue peptide was then introduced into the mass spectrometer under conditions that favored peptide bond cleavage. As shown in Figure 4, many fragmentation peptides were detected on the tandem mass spectrum. A signal corresponding to the intact glycosylated peptide with a mass of 1191 amu appears on the chromatogram together with the corresponding deglycosylated peptide (mass 1029 amu). Presumably, some deglycosylation of the peptide occurred

under the experimental conditions used. The spectrum also contained peptides of mass 729 and 891 (Figure 4); these peptides result from cleavage of the Ile-Pro bond and correspond to the glycosylated and nonglycosylated forms of Pro-Ser-Asp-Gly-Pro-Ser-Val-Ala, respectively. As the hexose is present on this peptide, Ser-28 is excluded as the site of hexose addition. The spectrum in Figure 4 also shows peptides resulting from cleavage of the Asp-Gly bond. The peptide Ser-Val-Ile-Pro-Ser-Asp is present in both a glycosylated form (761 amu) and a nonglycosylated form (599 amu). As Ser-28 has been excluded already, these peptide fragments show unambiguously that Ser-32 is the site of hexose attachment. As indicated on the top of Figure 4, each of the other peptide masses can be attributed to peptides resulting from cleavage of other peptide bonds either as Y ions or B ions.

DISCUSSION

Structure of P. pastoris-derived Fe(III)-hTF/2N. Several other *P. pastoris*-derived recombinant proteins have been crystallized and X-ray diffraction studies initiated, including human cathepsin L (*40*), human cathepsin K complexed with an inhibitor (*41*), human procarboxypeptidase A2 (*42*), and *Saccharomyces cerevisiae* class 1 α1,2 mannosidase (*43*). The three-dimensional structure of *P. pastoris*-derived human protein kinase C interacting protein 1 has been determined (*44*); however, the structure of the naturally derived protein has not been reported. Thus, the hTF/2N structure reported here represents the first evidence that a recombinant protein expressed in *P. pastoris* has the same three-dimensional structure as the naturally occurring protein (*13*) and the recombinant protein expressed in mammalian cells (*12*).

Glycosylation by P. pastoris. Many recombinant proteins are glycosylated in heterologous systems. In yeast such as *P. pastoris*, recombinant proteins can contain N- and/or O-linked glycosyl moieties (39, *45*, *46*) but in contrast to N-glycosylation, little is known about O-glycosylation. When the N-linked attachment site for human urokinase was mutated, the recombinant urokinase was still secreted by *P. pastoris* and contained O-linked oligosaccharides (*47*) but their structure(s) and site(s) of attachment were not determined. Similarly, Duman et al. (*46*) found that recombinant human plasminogen (kringle 1−4 domain) contained little N-linked glycosylation but had O-linked (α1,2) mannans containing dimeric, trimeric, tetrameric, and pentameric oligosaccharides with the major amount distributed equally between the dimer and trimer species. Again, the sites of O-linked glycan attachment were not determined. Thus, the glycan in hTF/2N represents the first identification of the site of attachment of a *P. pastoris*-derived O-linked glycan. Although the ESI-MS analysis cannot reveal the identity of the hexose attached to the hTF/2N, the presence of mannose in other *P. pastoris* O-linked glycans suggests that it is probably a single mannose residue.

It is not clear why *P. pastoris* adds a hexose to Ser-32 of hTF/2N, but it may be the result of fortuitous glycosylation as found in other proteins expressed in heterologous systems. Bacterial proteins are not normally glycosylated; however, expression of *Clostridium tetani* tetanus toxin fragment C in *S. cerevisiae* resulted in two glycosylated forms (*48*). Similarly, in activated monocytes, interleukin (IL)-1α and IL-1β are synthesized as precursors that lack a hydrophobic leader sequence. After intracellular processing, IL-1α and IL-1β are secreted by a route that circumvents the normal secretory pathway involving the endoplasmic reticulum and Golgi (*49*, *50*). Thus, despite having a consensus N-glycosylation sequence (Asn-Cys-Thr), monocyte-derived IL-1α and IL-1β are not glycosylated. However, when recombinant IL-1α and IL-1β are expressed in *S. cerevisiae*, these proteins pass through the normal secretion pathway and are N-glycosylated (*51*, *52*). As shown in Figure 3, both Ser-32 and Ser-36 in hTF/2N point away from the polypeptide chain and into the solvent making them accessible to the glycosylating enzyme; from the structure, however, it is not clear why only Ser-32 is glycosylated. To obtain a nonglycosylated preparation of hTF/2N, we intend to mutate Ser-32 to alanine to give S32A-hTF/2N. Given the distance from the iron-binding site, this mutation should not affect the iron-binding properties of the hTF/2N but should ensure that no O-linked glycosylation occurs on this residue. As a result, the S32A-hTF/2N should be expressed at high yield in a single, homogeneous form. In addition, point mutations made on the S32A-hTF/2N backbone should be homogeneous by mass spectroscopy allowing this technique to be used in their characterization.

P. pastoris as an expression system for hTF/2N. When expressing recombinant proteins in a heterologous system, the requirement for correctly folded protein and ease of obtaining high yields of purified protein must be balanced. As summarized in Table 1, several forms of transferrin, ovotransferrin, and lactoferrin have been expressed in a variety of heterologous systems. A prokaryotic system such as *E. coli* allows the easy production of large amounts of biological material, but transferrin can only be produced in an unfolded form having little biological activity (*19*−*22*). A refolding step allows the production of transferrins that can bind iron and have appropriate absorption and EPR spectra, but this step reduces the yield to only ∼3 mg/L (*53*). At the other extreme, mammalian tissue culture systems such as BHK cells offer authentic folding of complex, disulfide-containing proteins. In the case of the pNUT-BHK cells system, the use of a mutant dihydrofolate reductase gene allows immediate selection of transformed cells that have incorporated very high copy numbers of recombinant plasmid thereby reducing the selection time required if gene amplification is required (*10*). However, BHK cells require attachment to a surface for growth which is a disadvantage for large scale fermentation. Both *Aspergillus oryzae* (*25*) and insect cells (*23*, *24*) combine suspension growth in simple growth medium with correct folding of eukaryotic proteins. The lower yield of recombinant proteins expressed in these systems (Table 1) is offset by the less labor-intensive culture conditions. As shown in Table 1, *P. pastoris* offers the advantage of high-density growth in liquid medium with the advantage of a very high yield of secreted protein. Because *P. pastoris* secretes few other proteins, subsequent downstream processing is simplified. For example, a recent study (*54*) utilized *P. pastoris* to express several point mutations of hTF/2N that exhibited reduced rates of iron release. As we have now shown that *P. pastoris*-derived hTF/2N is structurally identical to BHK-derived hTF/2N, *P. pastoris* appears to be an excellent host for the production of this recombinant protein.

ACKNOWLEDGMENT

We thank Drs. Stephen Withers, Mark Brown, and Marta Guarna for helpful suggestions about glycosylation in *P. pastoris*.

REFERENCES

1. Baker, E. N. (1994) *Adv. Inorg. Chem. 41*, 389−463.
2. MacGillivray, R. T. A., Mendez, E., Shewale, J. G., Sinha, S. K., Lineback-Zins, J., and Brew, K. (1983) *J. Biol. Chem. 258*, 3543−3553.
3. Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., van Bragt, P. H., Baldwin, W. D., and Bowman, B. H. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2752−2756.
4. Park, I., Schaeffer, E., Sidoli, A., Baralle, F. E., Cohen, G. N., and Zakin, M. M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3149−3153.
5. Haridas, M., Anderson, B. F., and Baker, E. N. (1995) *Acta Crystallogr., Sect. D 51*, 629−646.
6. Bailey, S., Evans, R. W., Garratt, R. C., Gorinsky, B., Hasnain, S., Horsburgh, C., Jhoti, H., Lindley, P. F., Mydin, A., Sarra, R., and Watson, J. L. (1988) *Biochemistry 27*, 5804−5812.
7. Kurokawa, H., Mikami, B., and Hirose, M. (1995) *J. Mol. Biol. 254*, 196−207.
8. Klausner, R. D., Ashwell, J. V., Van Renswoude, J. B., Harford, J., and Bridges, K. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2263−2267.
9. Ponka, P. (1997) *Blood 89*, 1−25.
10. Funk, W. D., MacGillivray, R. T. A., Mason, A. B., Brown, S. A., and Woodworth, R. C. (1990) *Biochemistry 29*, 1654−1660.
11. Woodworth, R. C., Mason, A. B., Funk, W. D., and MacGillivray, R. T. A. (1991) *Biochemistry 30*, 10824−18029.
12. MacGillivray, R. T. A., Moore, S. A., Chen, J., Anderson, B. F., Baker, H., Luo, Y., Bewley, M., Smith, C. A., Murphy, M. E. P., Wang, Y., Mason, A. B., Woodworth, R. C., Brayer, G. D., and Baker, E. N. (1998) *Biochemistry 37*, 7919−7928.
13. Zuccola, H. J. (1993) Ph.D. Thesis, Georgia Institute of Technology, Atlanta, Georgia.
14. Mason, A. B., Miller, M. K., Funk, W. D., Banfield, D. K., Savage, K. J., Oliver, R. W. A., Green, B. N., MacGillivray, R. T. A., and Woodworth, R. C. (1993) *Biochemistry 32*, 5472−5479.
15. Mason, A. B., Tam, B. M., Woodworth, R. C., Oliver, R. A., Green, B. A., Lin, L.-N., Brandts, J. F., Savage, K. J., Lineback, J. A., and MacGillivray, R. T. A. (1997) *Biochem. J. 326*, 77−85.
16. Stowell, K. M., Rado, T. A., Funk, W. D., and Tweedie, J. W. (1991) *Biochem. J. 276*, 349−355.
17. Day, C. L., Stowell, K. M., Baker, E. N., and Tweedie, J. W. (1992) *J. Biol. Chem. 267* (20), 13857−13862.
18. Mason, A. B., Woodworth, R. C., Oliver, R. W., Green, B. N., Lin, L. N., Brandts, J. F., Savage, K. J., Tam, B. M., and MacGillivray, R. T. (1996) *Biochem. J. 319*, 361−368.
19. Hershberger, C. L., Larson, J. L., Arnold, B., Rosteck, P. R., Williams, P., DeHoff, B., Dunn, P., O'Neal, K. L., Riemen, M. W., Tice, P. A., Crofts, R., and Ivancic, J. (1991) *Ann. N. Y. Acad. Sci. 646*, 140−154.
20. Ikeda, R. A., Bowman, B. H., Yang, F., and Lokey, L. K. (1992) *Gene 117*, 265−269.
21. Steinlein, L. M., and Ikeda, R. A. (1993) *Enzymol. Microbiol. Technol. 15*, 193−199.
22. de Smit, D. H., Hoefkens, P., de Jong, G., van Duin, J., van Knippenberg, P. H., and van Eijk, H. G. (1995) *Int. J. Biochem. Cell Biol. 27*, 839−850.
23. Retzer, M. D., Kabani, A., Button, L. L., Yu, R. H., and Schryvers, A. B. (1996) *J. Biol. Chem. 271*, 1166−1173.
24. Ali, S. A., Joao, H. C., Csonga, R., Hammerschmid, F., and Steinkasserer, A. (1996) *Biochem. J. 319*, 191−195.
25. Ward, P. P., Lo, J.-Y., Duke, M., May, G. S., Headon, D. R., and Conneely, O. M. (1992) *Bio/Technology 10*, 784−789.
26. Steinlein, L. M., Graf, T. N., and Ikeda, R. A. (1995) *Protein Exp. Purif. 6*, 619−624.
27. Mason, A. B., Woodworth, R. C., Oliver, R. W. A., Green, B. N., Lin, L. N., Brandts, J. F., Tam, B. M., Maxwell, A., and MacGillivray, R. T. A. (1996) *Protein Exp. Purif. 8*, 119−25.
28. Guarna, M., Lesnicki, G. L., Tam, B., Robinson, J., Radziminski, C. Z., Hasenwinkle, D., Boraston, A., Jervis, E., MacGillivray, R. T. A., Turner, R. F. B., and Kilburn, D. G. (1997) *Biotechnol. Bioeng. 56*, 279−286.
29. Mason, A. B., Funk, W. D., MacGillivray, R. T. A., and Woodworth, R. C. (1991) *Protein Exp. Purif. 2*, 214−220.
30. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
31. Brünger, A. T. (1992) *Nature 335*, 472−474.
32. Brünger, A. T. (1992) XPLOR Version 3.1; a system for X-ray crystallography and NMR, Yale University Press, New Haven, CT.
33. Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268−272.
34. Laskowski, R. A., MacArthur, M. W., Moss D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
35. Tull, D., Burgoyne, D. L., Chow, D. T., Withers, S. G., and Aebersold, R. (1996) *Anal. Biochem. 234*, 119−125.
36. Ramakrishnan, C., and Ramachandran, G. N. (1965) *Biophys. J. 5*, 909−933.
37. Day, C. L., Anderson, B. F., Tweedie, J. W., and Baker, E. N. (1993) *J. Mol. Biol. 232*, 1084−1100.
38. Matthews, B. W. (1972) *Macromolecules 5*, 818−819.
39. Heimo, H., Palmu, K., and Suominen, I. (1997) *Protein Exp. Purif. 10*, 70−79.
40. Coulombe, R., Li, Y., Takebe, S., Menard, R., Mason, P., Mort, J. S., and Cygler, M. (1996) *Proteins 25*, 398−400.
41. Linnevers, C. J., McGrath, M. E., Armstrong, R., Mistry, F. R., Barnes, M. G., Klaus, J. L., Palmer, J. T., Katz, B. A., and Bromme, D. (1997) *Protein Sci. 6*, 919−921.
42. Reverter, D., Garcia-Saez, I., Catasus, L., Vendrell, J., Coll, M., and Aviles, F. X. (1997) *FEBS Lett. 420*, 7−10.
43. Dole, K., Lipari, F., Herscovics, A., and Howell, P. L. (1997) *J. Struct. Biol. 120*, 69−72.
44. Lima, C. D., Klein, M. G., Weinstein, I. B., and Hendrickson, W. A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5357−5362.
45. Romanos, M. A., Scorer, C. A., and Clare, J. J. (1992) *Yeast 8*, 423−488.
46. Duman, J. G., Miele, R. G., Liang, H., Grella, D. K., Sim, K. L., Castellino, F. J., and Bretthauer, R. K. (1998) *Biotechnol. Appl. Biochem. 28*, 39−45.
47. Tsujikawa, M., Okabayashi, K., Morita, M., and Tanabe, T. (1996) *Yeast 12*, 541−553.
48. Romanos, M., Makoff, A. J., Fairweather, N. F., Beestley, K. M., Slater, D. E., Rayment, F. B., Payne, M. M., and Clare, J. J. (1991) *Nucleic Acids Res. 19*, 1461−1467.
49. Hazuda, D. J., Lee, J. C., and Young, P. R. (1988) *J. Biol. Chem. 263*, 8473−8479.
50. Young, P. R., Hazuda, D. J., and Simon, P. L. (1988) *J. Cell Biol. 107*, 447−456.
51. Livi, G. P., Ferrara, A., Roskin, R., Simon, P. L., and Young, P. R. (1990) *Gene 88*, 297−301.
52. Livi, G. P., Lillquist, J. S., Miles, L. M., Ferrara, A., Sathe, G. M., Simon, P. L., Meyers, C. A., Gorman, J. A., and Young, P. R. (1991) *J. Biol. Chem. 266*, 15348−15355.
53. Hoefkens, P., de Smit, M. H., de Jeu-Jaspars, N. M. H., Huijskes-Heins, M. I. E., de Jong, G., and van Eijk, H. G. (1996) *Int. J. Biochem. Cell Biol. 28*, 975−982.
54. Steinlein, L. M., Ligman, C. M., Kessler, S., and Ikeda, R. A. (1998) *Biochemistry 37*, 13696−13703.
55. McRee, D. E. (1992) *J. Mol. Graphics 10*, 44−46.
56. Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
57. Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr., Sect. D 50*, 869−873.

BI9824543