J. Biochem. **127**, 65–72 (2000)

Characterization of the *N*-Oligosaccharides Attached to the Atypical Asn-X-Cys Sequence of Recombinant Human Epidermal Growth Factor Receptor

Chiaki Sato,* Jae-Hoon Kim,† Yoshito Abe,* Kazuki Saito,† Shigeyuki Yokoyama,† and Daisuke Kohda*,[1]

**Department of Structural Biology, Biomolecular Engineering Research Institute, Furuedai 6-chome, Suita, Osaka 565–0874; and* †*Yokoyama CytoLogic Project, ERATO, Japan Science and Technology Corporation, c/o Tsukuba Research Consortium, Tokodai, Tsukuba, Ibaraki 300–2635*

Received September 3, 1999; accepted October 14, 1999

The extracellular domain of human EGF receptor (sEGFR) produced by CHO cells has been used in various biophysical studies to elucidate the molecular mechanism of EGF-induced receptor activation. We have found that the CHO sEGFR contains one oligosaccharide chain attached to an atypical *N*-glycosylation consensus sequence, Asn^{32}-X^{33}-Cys^{34}. The oligosaccharide structure at Asn^{32} is a mixture of the monosialo and asialo forms of a core fucosylated biantennary complex-type oligosaccharide. Deletion of this atypical glycosylation site by replacement of Asn^{32} with lysine changed neither the expression nor function of the full length EGFR in CHO cells. The glycosylation at Asn^{32} in CHO sEGFR was incomplete: 20% of Asn^{32} remained unmodified. Thus, CHO sEGFR itself is heterogeneous with respect to the glycosylation at Asn^{32}, which may cause problems in biophysical studies. An attempt to remove the oligosaccharide at Asn^{32} enzymatically did not succeed under nondenaturing conditions. Therefore, sEGFR with the mutation of $Asn^{32} \rightarrow Lys^{32}$ is useful for biophysical and biochemical studies, and, particularly, for X-ray crystallography.

Key words: Asn-X-Cys, CHO cells, epidermal growth factor receptor, EGF, glycosylation.

Epidermal growth factor receptor (EGFR) is the cell membrane receptor for epidermal growth factor (EGF) and its relatives (*1*). These ligands have a common sequence motif, called EGF-like, and share EGF-like folding (*2*). Ligand binding to the extracellular domain induces the dimerization of EGFR (*3*). This triggers the tyrosine phosphorylation of EGFR itself through an intermolecular phosphorylation mechanism (*4*). To understand the molecular mechanism of EGFR activation *at the structural level*, a large amount (mg–g) of the receptor protein is necessary. The extracellular domain of human EGFR (sEGFR) can be prepared in large quantities from the conditioned media of recombinant Chinese hamster ovary (CHO) cells (*5*), insect cells infected by recombinant baculovirus (*6, 7*), and the A431 human tumor cell line (*8*). The extracellular domains secreted by CHO cells and insect cells are 621–624-residue fragments, whereas that secreted by A431 cells is a 616-residue fragment plus 18 C-terminal amino acids of unknown origin. These sEGFR preparation have been used in many biophysical studies, for example, the single particle analysis of electron microscopic images (*5*), as well as circular dichroism and fluorescence (*7*), laser light- (*9*) and X-ray scattering (*10*), and isothermal titration calorimetry (*10*) analyses. The crystallization of the A431 sEGFR in complex with EGF has been published (*11*), but the structure has not yet been obtained (*12*).

The full length EGFR produced by various human cell lines is heavily *N*-glycosylated, but not *O*-glycosylated (*13, 14*). The mature 170-kDa EGFR is composed of a single polypeptide chain of 1,186 amino acid residues and a substantial amount (*ca.* 40 kDa) of *N*-linked oligosaccharide. The sEGFR preparations from various sources are all 95–115 kDa glycoproteins consisting of a 68–70 kDa polypeptide; thus 30–40 kDa are derived from oligosaccharide (*5–8*). The amino acid sequence of human sEGFR contains eleven *N*-glycosylation consensus sequences, Asn-X-Ser/Thr, where X is any amino acid except Pro. The glycosylation sites of sEGFR have been investigated experimentally using the CHO material, but ambiguity remains at some sites (*15*).

Previously, we determined the connections of the 25 disulfide bonds in the CHO sEGFR (*16*). In that study, we found modification on the Asn^{32} residue, but its characterization was not carried out. Here we show that the previously unidentified modification is actually *N*-glycosylation linked to the first Asn residue in the sequence Asn-Asn-Cys, rather than the usual *N*-glycosylation motif, Asn-X-Ser/Thr sequence. We report a basic characterization of the structure and function of the oligosaccharide attached to this atypical site. The information presented here will be

[1]To whom correspondence should be addressed. Tel: +81-6-6872-8218, Fax: +81-6-6872-8219, E-mail: kohda@beri.co.jp

Abbreviations: CHO, Chinese hamster ovary; EGF, epidermal growth factor; EGFR, EGF receptor; PA-, pyridylamino; MALDI-MS, matrix assisted laser desorption/ionization mass spectrometry; PCR, polymerase chain reaction; sEGFR, the extracellular domain of EGFR; MAP kinase, mitogen-activated protein kinase.

useful in designing new constructs of sEGFR that are suitable for biophysical studies.

MATERIALS AND METHODS

Preparation of the "peak C4" Fraction after Cyanogen Bromide Cleavage of sEGFR—sEGFR (residues 1–621) was produced by overexpression in CHO cells (*5, 17*), and was purified from the conditioned medium by Affi-Gel Blue (Bio-Rad) chromatography (*16*). The fragments resulting from the cyanogen bromide cleavage of 5 mg sEGFR were separated by reversed-phase HPLC, and "peak C4", containing a peptide fragment with oligosaccharides attached to Asn^{32}, was collected and lyophilized (*16*).

V8 Protease Digestion of the Fragment C4—The purified cyanogen bromide fragment (fragment C4) was dissolved in 100 µl of 50 mM potassium phosphate buffer, pH 6.0, and digested with 5 µl of V8 protease (2 mg/ml, Wako) at 37°C for 20 h. The generated fragments were separated by reversed-phase HPLC on a Pegasil ODS column (4.6 mm × 150 mm, Senshu Scientific) at a flow rate of 0.5 ml/min with a 1–40% linear gradient of acetonitrile in 0.1% trifluoroacetic acid over 120 min.

Preparation of Pyridylaminated Oligosaccharides—The lyophilized fragment C4 was dissolved in 50 µl of 50 mM potassium phosphate buffer, pH 8.6, and mixed with 5 units of *N*-glycosidase F (recombinant, alias PNGase F, Boehringer Mannheim). The reaction mixture was incubated overnight at 37°C to release the *N*-linked oligosaccharide chains. Salts were removed with tandemly-connected columns containing 0.5 ml of Dowex AG50W-X8 (H^+) resin and AG1-X8 (OH^-) equilibrated with water, and the eluant containing the oligosaccharide was dried under vacuum. Pyridylamination of the oligosaccharide was carried out according to the published method (*18*). Briefly, 20 µl of coupling reagent [552 mg of 2-aminopyridine (Wako) dissolved in 200 µl of acetic acid, and stored at –20°C] was added, and the mixture was heated at 90°C for 60 min in a sealed tube. Seventy microliters of reducing reagent [100 mg of borane-dimethylamine complex (Wako) dissolved in a mixture of 40 µl of acetic acid and 20 µl of water, freshly prepared] was added to the solution, which was heated at 80°C for 35 min in the resealed tube. Aqueous ammonia (300 µl) and chloroform (300 µl) were added to the reaction mixture. After vortexing, the recovered aqueous phase was dried, and dissolved in 100 µl of water. The solution was loaded onto a Sephadex G-15 column (30 ml, Pharmacia) equilibrated with 10 mM ammonium hydrogen carbonate, pH 8.0. The fraction containing pyridylaminated (PA-) oligosaccharides was collected and concentrated under vacuum.

Exoglycosidase Digestions—Digestions of PA-labeled oligosaccharides were carried out in a volume of 50 µl at 37°C for 24 h, using the following exoglycosidases and buffers: Sialidase (alias neuraminidase, from *Arthrobacter ureafaciens*) at 1 U/ml in 100 mM sodium acetate, pH 5.5; bovine kidney α-L-fucosidase at 0.6 U/ml in 200 mM sodium citrate/phosphate, pH 5.5; β-galactosidase from an *E. coli* overproducer at 1,600 U/ml in 200 mM sodium citrate/phosphate, pH 7.0; or *N*-acetyl-β-D-glucosaminidase from *Diplococcus pneumoniae* at 0.6 U/ml in 100 mM sodium citrate/phosphate, pH 5.5. All of the exoglycosidases were purchased from Boehringer Mannheim.

HPLC Separation of PA-Oligosaccharides—PA-oligosaccharides were analyzed by reversed-phase HPLC on a Shim Pack CLC-ODS column (100 Å, 6 mm × 150 mm, Shimadzu) using a LC-10A system and a RF-10AXL fluorescence detector (Shimadzu). The excitation and emission wavelengths were 320 and 400 nm, respectively. The column temperature was maintained at 40°C. A 5–55% linear gradient was applied with solvent A (100 mM acetic acid-triethylamine, pH 4.0) and solvent B (solvent A containing 0.5% *n*-butanol) at a flow rate of 1 ml/min over 100 min. The retention times are expressed in glucose units using a mixture of PA-glucose oligomers (DP = 3–22, Takara) as the standard.

Analytical Methods—N-terminal sequencing of peptides was performed using an Applied Biosystems model 492 Protein Sequencer. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometric (MALDI-TOF-MS) analysis of the (glycosylated) peptides was carried out using a Voyager Elite spectrometer (PerSeptive Biosystems) as described (*16*). For the MALDI-MS analysis of oligosaccharides, PA-oligosaccharide samples (*ca.* 10 pmol dissolved in 0.1% trifluoroacetic acid containing 60% acetonitrile) were mixed with an equal volume of matrix solution (2,5-dihydroxybenzoic acid saturated in 10% ethanol). The data were acquired in the positive ion linear mode.

Deglycosylation of sEGFR in Native and Denatured States—Complete deglycosylation of sEGFR was carried out with *N*-glycosidase F under denaturing conditions. sEGFR (16 µg) was dissolved in 50 µl of 200 mM potassium phosphate buffer, pH 8.6. Then, 1 µl of 10% SDS and 0.2 µl of β-mercaptoethanol were added, and the mixture was heated at 90°C for 3 min. Next, one unit of *N*-glycosidase F and 5 µl of 15% NP-40 were added, and the mixture was incubated at 37°C for 18 h. For the deglycosylation of sEGFR under nondenaturing conditions, the additions of SDS and β-mercaptoethanol, and the heat step were omitted. Deglycosylated proteins were analyzed by SDS-PAGE under reducing conditions on 7.5% polyacrylamide gels using a PhastSystem (Pharmacia). Proteins in the gels were stained with PhastGel Blue R (Pharmacia).

Isolation of Human EGFR cDNA and Construction of the Expression Plasmid for EGFR—The full-length cDNA encoding EGFR was isolated by PCR from a human placenta Quick-Clone cDNA (Clontech). Two primers (5′-CGGGGAGCTAGCATGCGACCCTCCGGGACGGCCGGG-3′ and 5′-TATCCTGGTACCTCATGCTCCAATAAATTCAC-TGCTTTGTGG-3′) were designed to add *Nhe*I and *Kpn*I recognition sequences at the 5′- and 3′-ends, respectively. The amplified DNA was cloned into the *Nhe*I–*Kpn*I sites of the mammalian expression plasmid pcDNA3.1/Zeo(+) (Invitrogen) to create pcDNA-EGFR. The introduction of the N32K mutation, in which Asn^{32} of EGFR is replaced by Lys, was performed by PCR. The mutation in the EGFR sequence was confirmed by DNA sequence analysis.

Cell Culture and Establishment of EGFR-Expressing Cell Lines—CHO cells were cultured in Ham's F-12 medium supplemented with 10% fetal bovine serum. The cells were transfected with an expression vector carrying the wild type or mutant receptor by the lipofectamine method according to the recommendations of the manufacturer (Gibco-BRL). Stable transfectants were selected with zeocin (Invitrogen) at a concentration of 0.1 mg/ml. Independent colonies were isolated, and then EGFR expression was con-

firmed with an anti-EGFR antibody (1005, Santa Cruz Biotechnology).

Stimulation, Lysate Preparation, Immunoprecipitation, and Western Blotting—CHO cells were serum-starved in Ham's F-12 medium containing 1 mg/ml bovine serum albumin for 24 h, and then human EGF was added for the times indicated to stimulate the cells. The cells were washed in ice-cold PBS and were lysed in 30 mM Tris buffer, pH 7.4, containing 150 mM NaCl, 5 mM EDTA, 40 mM β-glycerophosphate, 10% glycerol, 1% Triton X-100, 1 mM phenylmethylsulfonylfluoride, 1 mM sodium orthovanadate, 10 μg/ml aprotinin, and 10 μg/ml leupeptin. The lysate was cleared of cell-debris by centrifugation, and the protein content was determined with a Protein Assay Kit (BioRad).

For the detection of autophosphorylation, the EGFRs were immunoprecipitated in the lysate with the anti-EGFR antibody and protein A–agarose (Boehringer Mannheim), and analyzed in 7.5% SDS–polyacrylamide gels. The fractionated proteins in the gels were then transferred to a nitrocellulose membrane for immunoblotting with the anti-EGFR antibody or an anti-phosphotyrosine antibody (PY20, Santa Cruz Biotechnology). Protein bands were detected by the ECL system (Amersham Pharmacia Biotech).

For the detection of MAP kinase activation, 20 μg of lysates were separated in a 10% SDS–polyacrylamide gel. The proteins were transferred to a nitrocellulose membrane, and immunoblotted with an anti–phospho-p44/42 MAP kinase (Thr^{202}/Tyr^{204}) antibody (New England Biolabs). Additionally, the blotted membrane was stripped and reprobed with an anti–MAP kinase antibody (C-14, Santa Cruz Biotechnology) to determine the protein amounts.

RESULTS

Our previous study provided information on the attachment of oligosaccharide chains in the recombinant extracellular domain of the human EGF receptor (sEGFR) produced in CHO cells (*16*). We found four sites (Asn^{328}, Asn^{337}, Asn^{544}, Asn^{579}) were actually glycosylated, but two sites (Asn^{172}, Asn^{599}) remained unmodified. The data from an-

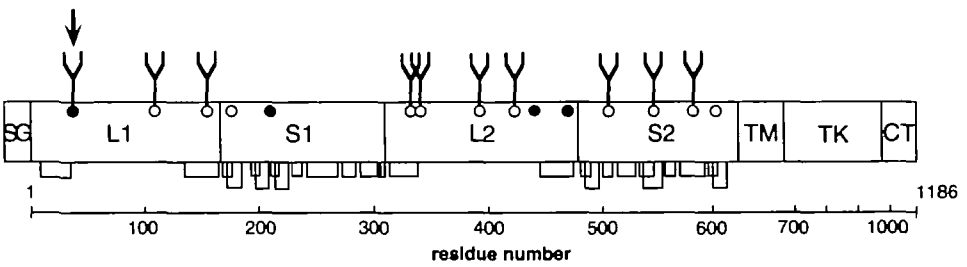

Fig. 1. **Summary of *N*-glycosylation sites of recombinant human EGF receptor produced by CHO cells.** This figure was made on the basis of the data from our previous study (*16*), and that from Smith *et al.* (*15*). Open circles (○) show the positions of the typical *N*-glycosylation consensus, Asn-X-Ser/Thr, whereas closed circles (●) show the positions of the atypical *N*-glycosylation sequence, Asn-X-Cys. X in the triplet sequences represents any amino acid other than Pro. Nine out of eleven typical *N*-glycosylation sequences are actually glycosylated. One of the four atypical *N*-glycosylation motifs is glycosylated, indicated by a vertical arrow. Below the schematic drawing of the human EGF receptor protein, the locations of the disulfide bonds are shown (*16*). SG, signal sequence; L1 and L2, large domains; S1 and S2, small cysteine-rich domains; TM, transmembrane region; TK, tyrosine kinase domain; CT, C-terminal region.

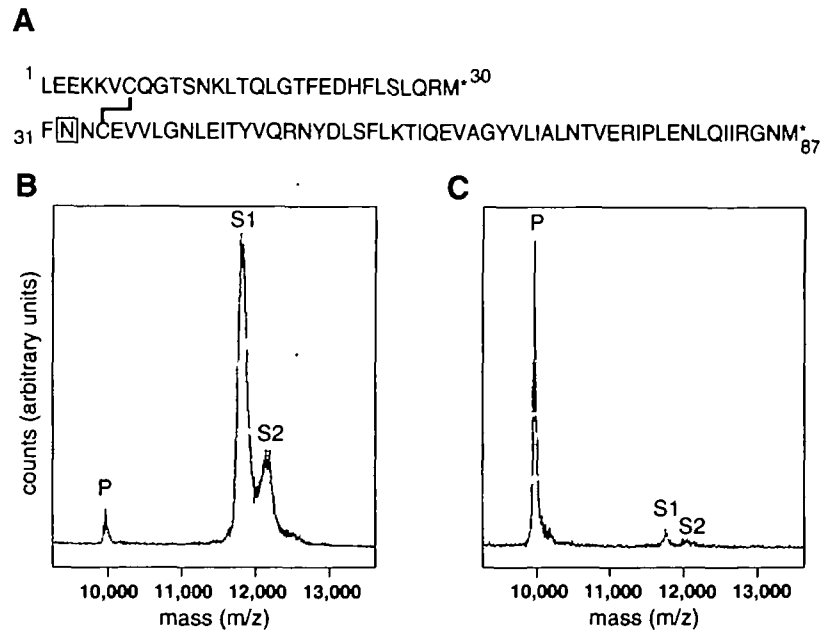

Fig. 2. **MALDI-MS analyses of the cyanogen bromide fragment containing the atypical glycosylation site.** A: Amino acid sequence and disulfide structure of the cyanogen bromide fragment in the 'peak C4' fraction (fragment C4). The modified Asn at residue 32 is boxed. M* denotes a homoserine residue formed by cyanogen bromide treatment. B: MALDI-MS spectra of fragment C4 before *N*-glycosidase F treatment. Three signals, P, S1, and S2, were observed. The *m/z* value of peak P (9,949 *m/z*), coincides with the theoretical mass of unmodified fragment C4 (9,942, 9,960, 9,978). Note that each signal splits into three peaks due to homoserine lactone formation, –18 *m/z* (one lactone) and –36 *m/z* (two lactones). C: the same as B, but after *N*-glycosidase F treatment.

other group were also incomplete, but supplement our data to reveal that the remaining five sites (Asn^{104}, Asn^{151}, Asn^{389}, Asn^{420}, Asn^{504}) are all glycosylated (*15*). Thus, nine out of the eleven *N*-glycosylation consensus sequences are *N*-glycosylated in the CHO sEGFR (Fig. 1).

We also noticed the modification of Asn^{32}, which is the first residue in the sequence Asn-Asn-Cys. We inferred that the modification of Asn^{32} is asparagine-linked glycosylation, but this was not confirmed experimentally in the previous study (*16*). To prepare a peptide fragment containing the modified Asn residue, the CHO sEGFR was cleaved by cyanogen bromide without reduction or deglycosylation. The resultant (glycosylated) peptide fragments were separated by reversed-phase HPLC to give four peaks, C1–C4 (see Fig. 1A of Ref. *16*). The last peak, C4, contained a single species of fragment (fragment C4) consisting of two peptides, Leu^{1}–Met^{*30} and Phe^{31}–Met^{*87}, connected by a disulfide bond, Cys^{7}–Cys^{34} (Fig. 2A). Met* denotes a homoserine residue derived from a methionine residue during cyanogen bromide treatment. The amino acid sequence of one peptide was confirmed correctly as Leu^{1}-Glu^{2}-Glu^{3}-Lys^{4}-Lys^{5}- by Edman degradation, but the sequence of the other peptide was read as Phe^{31}-X^{32}-Asn^{33}-X^{34}-Glu^{35}-. By comparison with the amino acid sequence of human EGFR, X^{32} is Asn, and X^{34} is Cys.

Structure of the Oligosaccharide Chains Attached to Asn^{32}—Matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS) analysis revealed three molecular ion peaks, P, S1, and S2, from fragment C4 (Fig. 2B). The S1 and S2 peaks disappeared after *N*-glycosidase F treatment, suggesting that the modification of Asn^{32} is *N*-glycosylation (Fig. 2C). The oligosaccharides attached at Asn^{32} were released by *N*-glycosidase F, and labeled by the pyridylamination (PA) reaction. The PA-oligosaccharides were fractionated by reversed-phase HPLC (Fig. 3A). The molecular masses (*m/z*, H^+ form) of the collected peaks were measured by MALDI-MS analysis. The two major peaks were identified as PA-oligosaccharides, corresponding to the S1 and S2 peaks in Fig. 2B. Sialidase digestion (Siaα2-↓→x) of recovered peak S2 resulted in the conversion of S2 to S1 (Fig. 3, B and C). Considering the mass difference between the two peaks, 291.3 Da, peak S2 was identified as peak S1 with one sialic acid (*N*-acetylneuraminic acid) residue attached. Defucosylation of S1 was carried out with bovine kidney fucosidase (Fucα1-↓→x) (Fig. 3D). The mass difference, 146.3 Da, indicates the removal of one fucose residue from S1. Further digestions with *E. coli* β-galactosidase (Galβ1-↓→x) and *Diplococcus* *N*-acetyl-β-D-glucosaminidase (GlcNAcβ1-↓→2) of the defucosylated S1 generated a trimannosyl core structure (Fig. 3, E and F). The fucose residue is attached to the trimannosyl core structure because digestions with the β-galactosidase and *N*-acetyl-β-D-glucosaminidase of S1 without fucosidase treatment generated a PA-oligosaccharide with the molecular mass of the trimannosyl core plus a fucose residue (data not shown).

Strictly, lectin or methylation analysis is necessary to determine the linkage of monosaccharide units in addition to exoglycosidase digestion and mass analysis. We can, however, estimate the oligosaccharide structures by comparing the retention times from the reversed-phase HPLC column with those of structure-known standards. The retention times of S1 and its defucosylated form were found to be identical to those of the authentic PA-oligosaccharide

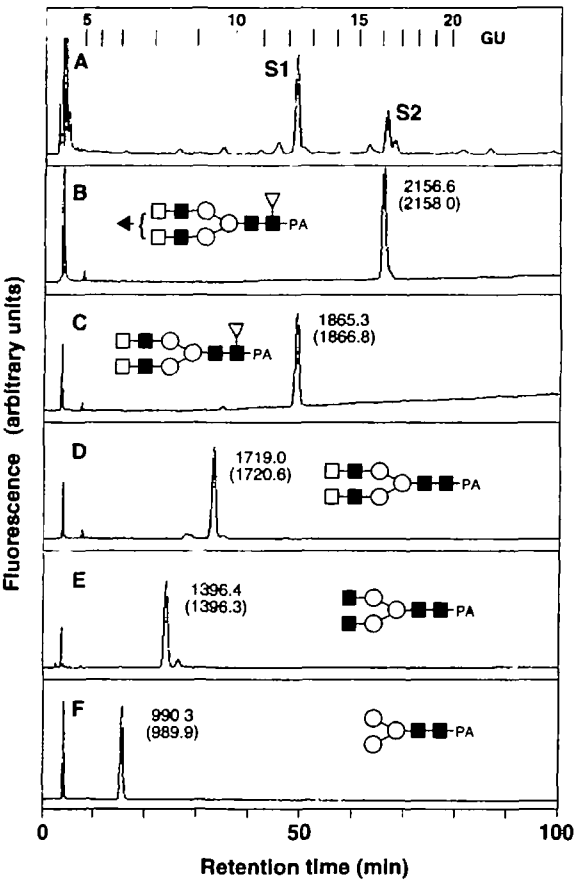

Fig. 3. **Reversed-phase HPLC analysis of the sequential exoglycosidase digests of oligosaccharides attached to Asn^{32}.** A: Separation of oligosaccharides after fluorescent labeling with 2-aminopyridine. The integers at the top indicate the positions of the standard PA-glucose oligomers. B and C: separation before and after digestion with *A. ureafaciens* sialidase. D: after digestion with bovine kidney α-L-fucosidase. E: after digestion with *E. coli* β-galactosidase. F: after digestion with *D. pneumoniae* *N*-acetyl-β-D-glucosaminidase. Numbers near the peaks are observed *m/z* values in MALDI-MS analyses; theoretical mass values are in parentheses. Symbols used for monosaccharide units are: ▲, sialic acid; □, galactose; ■, GlcNAc; ○, mannose; ▽, fucose.

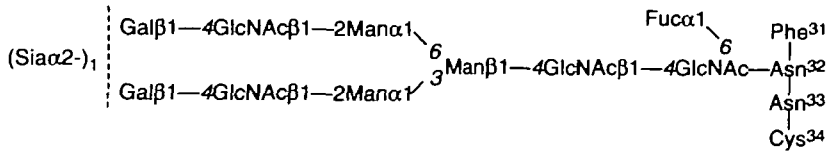

Fig. 4. **Deduced oligosaccharide structures attached to Asn^{32} of the CHO sEGFR.** The asialo form is the major form (S1), and the monosialo form is the minor form (S2). The linkage positions that were not determined experimentally, but were estimated from comparison with the retention times of standards on the reversed-phase column are italicized.

standards, PA-Sugar Chain009 and Chain001 (Takara), respectively. The Chain001 standard is a biantennary complex type oligosaccharide with the most common linkages, and the Chain009 is the same structure with a fucose α1-6 linked to the reducing terminal GlcNAc. Thus, the oligosaccharide attached to Asn[32] of the sEGFR produced by CHO cells was a mixture of the monosialo and asialo forms of a 6-linked core fucosylated biantennary complex type (Fig. 4).

The Percentage of Glycosylation at Asn[32]—The glycosylation at Asn[32] was partial, as clearly shown by the mass peak labeled 'P', which corresponds to the unglycosylated form (Fig. 2B). Accurate determination of the extent of glycosylation from the mass spectrum is difficult due to the different efficiencies of the ionization process. Thus, the cyanogen bromide fragment containing the oligosaccharide at Asn[32] was further digested with V8 protease, and then subjected to reversed-phase HPLC (Fig. 5). Individual peaks were recovered and analyzed by MALDI-MS and *N*-terminal sequencing. Peak 7 and peak 9 correspond to the glycosylated and unglycosylated forms of a fragment consisting of two peptides, Lys[4]–Glu[21] and Phe[31]–Glu[35], connected by a disulfide bond (inset of Fig. 5). The comparison of the peak intensities reveals that the percentage of glycosylation at Asn[32] is about 80%[2]. In contrast, glycosylation at the usual *N*-glycosylation sites is almost 100%, because neither we nor the other group found any evidence of incomplete glycosylation (*15, 16*).

The partial glycosylation at Asn[32] may cause problems related to sample homogeneity when CHO-produced sEGFR is used in various biophysical studies. Thus, it is important to know the sensitivity of the oligosaccharide at Asn[32] to *N*-glycosidase F treatment under nondenaturing conditions. All of the oligosaccharide chains were easily removed by *N*-glycosidase F treatment after heat treatment in the presence of SDS and β-mercaptoethanol (Fig. 6, 95→70 kDa), but only half of the oligosaccharides were released without the SDS/heat treatment (95→81 kDa). This partially deglycosylated sEGFR was digested with cyanogen bromide, and fragment C4 was prepared. The mass spectrum obtained was identical to that in Fig. 2B, indicating that the oligosaccharide at Asn[32] is *resistant* to enzymatic digestion under nondenaturing conditions (data not shown). Therefore, the reduction of the heterogeneity at Asn[32] of the CHO sEGFR demands harsh treatment, such as unfold-refolding or disulfide reduction–oxidation.

Activity of N32K-EGFR Mutant Expressed in CHO Cells—To examine the effects of the oligosaccharide chain attached to the atypical glycosylation site on the receptor activities, a mutant of the full length EGFR with an Asn[32]→Lys[32] substitution (N32K-EGFR) was expressed in CHO cells. After serum starvation, the transfected cells were stimulated with EGF, lysed, and immunoprecipitated with the anti-EGFR antibody. Endogenous EGFR is not expressed at a detectable level in control CHO cells (Fig. 7). The recombinant wild type human EGFR (WT-EGFR) was expressed in the transfected cells. N32K-EGFR was also expressed at a slightly higher level in the cells.

EGF stimulation is known to trigger tyrosine autophosphorylation of EGFR through receptor dimerization. In transfected CHO cells, N32K-EGFR is autophosphorylated in response to the ligand to an extent similer to that of WT-EGFR (Fig. 7). Thus, the oligosaccharide chain at Asn[32] does not participate in the ligand-induced dimerization of the receptor. Phosphorylation of mitogen-activated protein kinase (MAP kinase) is also known to occur in various EGF-stimulated cells, and this can be used as a major signaling probe for the downstream pathway of EGFR involving several signal-transducing components, such as the Shc, Grb2, Sos, and Ras proteins (*19, 20*). Activation of MAP kinase was assayed by immunoblotting the cell lysate with an antibody that specifically recognizes the phosphorylated MAP kinase. In Fig. 8, a 1-min stimulation by EGF at a concentration of 10 ng/ml activated MAP kinase in both the WT-EGFR and N32K-EGFR transfectants, whereas no activation of the kinase occurred in control host

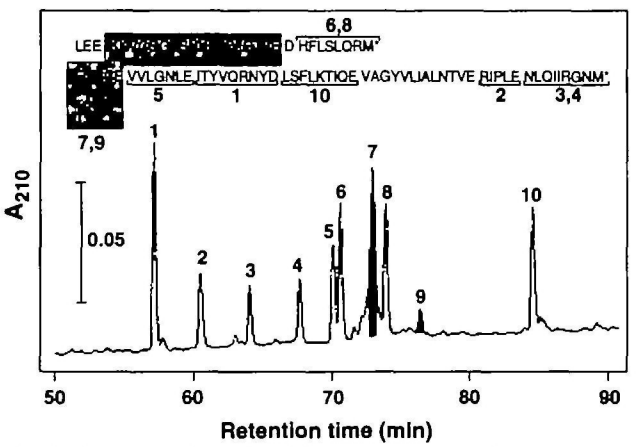

Fig. 5. **Reversed-phase HPLC separation of V8 protease digest of the fragment C4.** The inset shows the structure of the fragment C4 with the V8 protease cleavage sites. All of the expected fragments were recovered and assigned, except for the very short fragments (<4 residues), and Val[61]–Glu[73]. The latter fragment was not recovered, probably due to its strong hydrophobicity. Each fragment derived from the two C-termini of fragment C4 gave two peaks due to the formation of homoserine lactone.

[2] Simple comparison of the peak intensities gives a ratio of 85:15 (peak 7:peak 9). The double bonds in the oligosaccharide moiety in the peak 7 contribute to the absorbance at 210 nm. Assuming that the double bonds in the peptide and oligosaccharide have the same absorbance coefficients, the ratio is 75:15 after correction, thus about 80% is glycosylated. Measurement of the absorbance at 280 nm failed due to the low absorbance of the peptides.

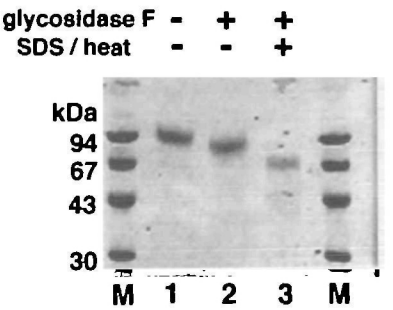

Fig. 6. **SDS-PAGE of deglycosylated sEGFR in native and denatured states.** Lane M, molecular weight markers (Pharmacia); lane 1, untreated sEGFR; lane 2, deglycosylated sEGFR in the native state; lane 3, deglycosylated sEGFR in the denatured state.

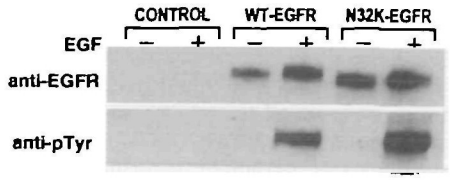

Fig. 7. **Expression and autophosphorylation of WT-EGFR and N32K-EGFR in CHO cells.** Serum-starved WT-EGFR and N32K-EGFR transfectants were treated with or without human EGF (10 ng/ml) at 37°C for 5 min. Cell lysates were immunoprecipitated with an anti-EGFR antibody and electrophoresed, and then transferred to a nitrocellulose membrane. The membrane was blotted with the anti-EGFR antibody (upper panel) or with an anti-phosphotyrosine antibody (lower panel).

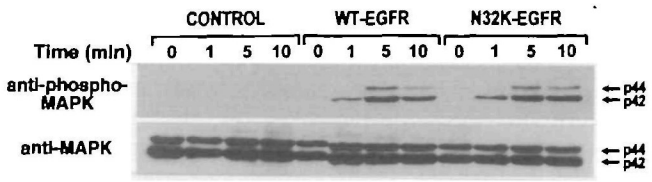

Fig. 8. **EGF-induced activation of p44/42 MAP kinase.** Serum-starved WT-EGFR and N32K-EGFR transfectants were treated with human EGF (10 ng/ml) for the times indicated. Cell lysates with equivalent amounts of total protein were electrophoresed and transferred to a nitrocellulose membrane. The membrane was blotted with an anti-phospho-MAP kinase antibody (upper panel). After detection of the protein bands, the membrane was stripped and reprobed with an anti–MAP kinase antibody to confirm the equivalent amounts of protein in each lane (lower panel).

cells. Activation was maximal at 5 min and slightly decreased at 10 min in both cell types due to down-regulation of the MAP kinase pathway. In summary, the behavior of the N32K mutant was similar to that of the wild type receptor *in vivo*. Thus, the oligosaccharide chain attached at Asn^{32} has little influence on the expression and signal-transducing function of EGFR.

DISCUSSION

The experiments in this study have demonstrated that the extracellular domain of the recombinant EGF receptor produced by CHO cells has an oligosaccharide attached at Asn^{32} (Fig. 1). This is unexpected, because Asn^{32} is not a part of the usual *N*-glycosylation consensus, Asn-X-Ser/Thr, but actually is within an Asn-X-Cys sequence. There are four Asn-X-Cys motifs in the human EGFR sequence, but no glycosylation was found at the other three sites.

Asparagine-linked glycosylation begins with the transfer of a high mannose core oligosaccharide, $Glc_3Man_9GlcNAc_2$, from a dolichol-linked donor onto asparagine residues in nascent proteins (*21*). This reaction is catalyzed by oligosaccharyltransferase. Then, the structure of the high-mannose type oligosaccharide is modified by other enzymes in the Golgi. The mature structures of oligosaccharides vary considerably, but glycosylation at a given asparagine residue is determined by the specificity of the oligosaccharyltransferase. It has been known for a long time that the consensus sequence, Asn-X-Ser/Thr, where X cannot be Pro, is prerequisite, but in itself insufficient for *N*-glycosylation to occur (*22*). The local folding around the potential *N*-glycosylation sites affects the efficiency of glycosylation (*23*). Using an *in vitro* model system, the asparagine residue of a hexapeptide, Tyr-Asn-Gly-Cys-Ser-Val, was recognized and glycosylated by the oligosaccharyltransferase in calf liver microsomal fractions at a lower, but significant relative rate (*24*). Thus, the glycosylation at Asn-X-Cys sequences should not be a rare event, but a very limited number of reports of such atypical glycosylation sites have been published thus far. To our knowledge, sEGFR is just the ninth example. Table I summarizes the proteins with atypical *N*-glycosylation at Asn-X-Cys motifs (*25–32*).

The proteins in Table I are mainly secreted soluble proteins, but two proteins, CD69 and the EGF receptor, are membrane proteins with a single transmembrane region. Although the sEGFR used mainly in this study is a truncated form of the EGF receptor secreted by cultured CHO cells, a slightly faster migration of the N32K-EGFR mutant than the wild type receptor in the SDS–polyacrylamide gel (Fig. 7) suggests that the full length EGFR expressed in CHO cells also contains the atypical oligosaccharide chain at residue 32. It is interesting to know whether sEGFR obtained from insect cells or A431 human cells, or intact EGF receptors from natural sources contain such atypical glycosylation. Although we have no experimental evidence, we expect that they are *N*-glycosylated at Asn^{32}. The extent of the glycosylation and its oligosaccharide structure may differ. The finding that human protein C derived from a cultured cell line also possesses the atypical *N*-glycosylation as well as protein C from human plasma supports this notion (*30*). Table I also demonstrates that there are no obvious consensus amino acid residues other than the asparagine and cysteine residues, and the percentage of atypical glycosylation varies greatly from 1 to 100%. In the case of the CHO sEGFR, the glycosylation is 80%. The cysteine residues are all involved in disulfide bonds except for the CD69 protein, suggesting that the disulfide bond formation occurs after the oligosaccharyltransferase-catalyzed glycosyl transfer. The structures of the oligosaccharides at the Asn-X-Cys motif have been reported for some proteins. They are the usual complex-type oligosaccharides, indicating that these oligosaccharides at the Asn-X-Cys motif are processed by the same Golgi maturation enzymes as the oligosaccharides at the Asn-X-Ser/Thr motif.

N-linked oligosaccharides can have effects on the function of the EGF receptor in some cases. The EGF receptors expressed in the human cell line, U373 MG, contain a bisecting GlcNAc oligosaccharide structure. An increase in the content of the bisecting oligosaccharide structure by genetic manipulation of U373 MG cells led to a decrease in both EGF binding and EGF receptor autophosphorylation (*33*). The *N*-glycosylation of sEGFR at Asn^{32}, however, does not seem to be biologically significant, because the deletion of the glycosylation at Asn^{32} by amino acid substitution affected neither its biosynthesis nor its signal transduction activity (Figs. 7 and 8). In fact, the Asn^{32} residue is only conserved in the human, murine, and avian EGF receptors, and not in EGF receptors from fish, fly, or nematode, or in the human EGFR-related ErbB family proteins.

From a practical point of view, characterization of the glycosylation is necessary because the CHO-produced

TABLE I. **Glycosylated Asn-X-Cys sites on proteins.**

Protein	Source	Sequence[a]	Glycosylation	*N*-Oligosaccharide structure[b]	Reference[c]
Protein C	Bovine plasma	VPY NAC VHA (1)	100%	n.d.	*25* (1982)
von Willebrand factor	Human plasma	WRY NSC APA (7)	100%	n.d.	*26* (1986)
Protein C	Human plasma	VPH NEC SEV (1)	70%	n.d.	*27* (1990)
Fetal antigen 1	Human amniotic fluid	FSG NFC EIV (4)	Partial	n.d.	*28* (1994)
		IVA NSC TPN	Partial	n.d.	
Fetal antigen 1	Murine amniotic fluid	FSG NFC EIV (4)	5%	Sialo fucosylbi- and triantennary complex (mass)	*29* (1997)
		AAT NSC TPN	100%	Sialo fucosylbiantennary complex (mass)	
CD69	Human Jurkat T cells	SAQ NAC SEH (1)	30%	Complex type (endoH-resistant)	*30* (1997)
α-Lactalbumin	Human milk	QSR NIC DIS (1)	1%	(1,886 and 1,740 Da by mass)	*31* (1997)
α1T-glycoprotein	Human plasma	LLR NCC NTE (2)	100%	Sialo biantennary complex (chromatography)	*32* (1998)
Human EGF receptor	Hamster CHO cells	RMF NNC EVV (4)	80%	Sialo fucosylbiantennary complex (mass)	This study

[a]Amino acid sequence near the glycosylated Asn-X-Cys site. The fetal antigen 1 proteins from both human and mouse have two glycosylated Asn-X-Cys sites. The number of occurrences of the Asn-X-Cys motif (X is a non-proline residue) in the sequence is shown in parentheses. All of the cysteine residues in the glycosylated Asn-X-Cys sites form disulfide bonds, except for CD69. [b]The methods for the determination of the *N*-glycan structures are shown in parentheses. mass, mass spectrometry; n.d., not determined. [c]The year of publication is in parentheses.

sEGFR has been widely used in various biophysical studies. Protein homogeneity is of primary importance for proper data interpretation. The heterogeneity of the oligosaccharide chains due to terminal sialic acid and galactose residues is a common problem to glycoproteins, but the heterogeneity due to the partial glycosylation at Asn^{32} is unique to the CHO sEGFR preparation. The former heterogeneity can be reduced by using a variant of CHO cells. The glycosidase-defective Lec8 cell line produces a protein with *N*-linked oligosaccharides lacking the terminal sialic acid and galactose residues (*34*), and has been used in the successful crystallization of the insulin-like growth factor–1 receptor (*35, 36*). On the other hand, the latter heterogeneity cannot be removed either by using the Lec8 cells or glycosidase digestion under nondenaturing conditions (Fig. 6). Glycosidase digestions under (partial) denaturing conditions may be required, but does not guarantee the maintenance of the biological conformation. Therefore, the N32K-sEGFR preparation, in which the partial glycosylation is removed by the amino acid replacement of Asn^{32}, produced by the CHO Lec8 cells is the most suitable for future biophysical studies. The heterogeneity of oligosaccharide moieties is a particular nuisance in preparing crystals of glycoproteins for structure determination. In this sense, the Lec8-produced N32K sEGFR may give good crystals for X-ray crystallography.

Finally, we emphasize that *N*-glycosylation at the Asn-X-Cys motif is a more general event than we thought, although it occurs less frequently than *N*-glycosylation at the Asn-X-Ser/Thr motif. Owing to recent progress in mass spectrometry, we expect that not a few glycoproteins will be shown to have oligosaccharide chains at Asn-X-Cys motifs. The heterogeneity of glycoproteins should be discussed and examined, considering the possibility of Asn-X-Cys glycosylation.

We thank Drs. Irit Lax and Joseph Schlessinger (New York University, New York) for providing us with CHO cells that overexpress sEGFR, and Dr. Masafumi Odaka (RIKEN, Saitama) and Dr. Fuyuhiko Inagaki (Hokkaido University, Sapporo) for the large-scale preparation of sEGFR. We also thank Dr. Ikuo Fujii (BERI) for access to the Shimadzu HPLC system equipped with a fluorescence Acknowl detector.

REFERENCES

1. Carpenter, G. and Cohen, S. (1990) Epidermal growth factor. *J. Biol. Chem.* **265**, 7709–7712
2. Campbell, I.D., Cooke, R.M. , Baron, M., Harvey, T.S., and Tappin, M.J. (1989) The solution structures of epidermal growth factor and transforming growth factor alpha. *Prog. Growth Factor Res.* **1**, 13–22
3. Ullrich, A. and Schlessinger, J. (1990) Signal transduction by receptors with tyrosine kinase activity. *Cell* **61**, 203–212
4. Honegger, A.M., Schmidt, A., Ullrich, A., and Schlessinger, J. (1990) Evidence for epidermal growth factor (EGF)-induced intermolecular autophosphorylation of the EGF receptors in living cells. *Mol. Cell Biol.* **10**, 4035–4044
5. Lax, I., Mitra, A.K., Ravera, C., Hurwitz, D.R., Rubinstein, M., Ullrich, A., Stroud, R.M., and Schlessinger, J. (1991) Epidermal growth factor (EGF) induces oligomerization of soluble, extracellular, ligand-binding domain of EGF receptor. *J. Biol. Chem.* **266**, 13828–13833
6. Hurwitz, D.R., Emanuel, S.L., Nathan, M.H., Sarver, N., Ullrich, A., Felder, S., Lax, I., and Schlessinger, J. (1991) EGF induces increased ligand binding affinity and dimerization of soluble epidermal growth factor (EGF) receptor extracellular domain. *J. Biol. Chem.* **266**, 22035–22043
7. Greenfield, C., Hiles, I., Waterfield, M.D., Federwisch, M., Wollmer, A., Blundell, T.L., and McDonald, N. (1989) Epidermal growth factor binding induces a conformational change in the external domain of its receptor. *EMBO J.* **8**, 4115–4123
8. Ullrich, A., Coussens, L., Hayflick, J.S., Dull, T.J., Gray, A., Tam, A.W., Lee, J., Yarden, Y., Libermann, T.A., Schlessinger, J., Downward, J., Mayes, E.L.V., Whittle, N., Waterfield, M.D., and Seeburg, P.H. (1984) Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* **309**, 418–425
9. Odaka, M., Kohda, D., Lax, I., Schlessinger, J., and Inagaki, F. (1997) Ligand-binding enhances the affinity of dimerization of the extracellular domain of the epidermal growth factor receptor. *J. Biochem.* **122**, 116–121
10. Lemmon, M.A., Bu, Z., Ladbury, J.E., Zhou, M., Pinchasi, D., Lax, I., Engelman, D.M., and Schlessinger, J. (1997) Two EGF molecules contribute additively to stabilization of the EGFR dimer. *EMBO J.* **16**, 281–294
11. Günther, N., Betzel, C., and Weber, W. (1990) The secreted form of the epidermal growth factor receptor. *J. Biol. Chem.* **265**, 22082–22085
12. Degenhardt, M., Weber, W., Eschenburg, S., Dierks, K., Funari, S.S., Rapp, G., and Betzel, C. (1998) Crystallization and preliminary X-ray crystallographic analysis of the EGF receptor ectodomain. *Acta Cryt.* D**54**, 999–1001

13. Soderquist, A.M. and Carpenter, G. (1984) Glycosylation of the epidermal growth factor receptor in A-431 cells. *J. Biol. Chem.* **259**, 12586–12594
14. Gamou S. and Shimizu, N. (1988) Glycosylation of the epidermal growth factor receptor and its relationship to membrane transport and ligand binding. *J. Biochem.* **104**, 388–396
15. Smith, K.D., Davies. M.J., Bailey, D., Renouf, D.V., and Hounsell, E.F. (1996) Analysis of the glycosylation patterns of the extracellular domain of the epidermal growth factor receptor expressed in Chinese hamster ovary fibroblasts. *Growth Factors* **13**, 121–132
16. Abe, Y., Odaka, M., Inagaki, F., Lax, I., Schlessinger, J., and Kohda, D. (1998) Disulfide bond structure of human epidermal growth factor receptor. *J. Biol. Chem.* **273**, 11150–11157
17. Kohda, D., Odaka, M., Lax, I., Kawasaki, H., Suzuki, K., Ullrich, A., Schlessinger, J., and Inagaki, F. (1993) A 40-kDa epidermal growth factor/transforming growth factor α-binding domain produced by limited proteolysis of the extracellular domain of the epidermal growth factor receptor. *J. Biol. Chem.* **268**, 1976–1981
18. Hase, S. (1994) High-performance liquid chromatography of pyridylaminated saccharides. *Methods Enzymol.* **230**, 225–237
19. de Vries-Smits, A.M., Burgering, B.M., Leevers, S.J., Marshall, C.J., and Bos, J.L. (1992) Involvement of p21ras in activation of extracellular signal-regulated kinase 2. *Nature* **357**, 602–604
20. Buday, L. and Downward, J. (1993) Epidermal growth factor regulates p21ras through the formation of a complex of receptor, Grb2 adaptor protein, and Sos nucleotide exchange factor. *Cell* **73**, 611–620
21. Kornfeld, R. and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664
22. Marshall, R.D. (1972) Glycoproteins. *Annu. Rev. Biochem.* **41**, 673–702
23. Pless, D.D. and Lennarz, W.J. (1977) Enzymatic conversion of proteins to glycoproteins. *Proc. Natl. Acad. Sci. USA* **74**, 134–138
24. Bause, E. and Legler, G. (1981) The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the *N*-glycosylation step during glycoprotein biosynthesis. *Biochem. J.* **195**, 639–644
25. Stenflo, J. and Fernlund, P. (1982) Amino acid sequence of the heavy chain of bovine protein C. *J. Biol. Chem.* **257**, 12180–12190
26. Titani, K., Kumar, S., Takio, K., Ericsson, L.H., Wade, R.D., Ashida, K., Walsh, K.A., Chopek, M.W., Sadler, J.E., and Fujikawa, K. (1986) Amino acid sequence of human von Willebrand factor. *Biochemistry* **25**, 3171–3184
27. Miletich, J.P. and Broze, G.J., Jr. (1990) β Protein C is not glycosylated at asparagine 329. *J. Biol. Chem.* **265**, 11397–11404
28. Jensen, C.H., Krogh, T.N., Højrup, P., Clausen, P.P., Skjødt, K., Larsson, L.-I., Enghild, J.J., and Teisner, B. (1994) Protein structure of fetal antigen 1 (FA1). *Eur. J. Biochem.* **225**, 83–92
29. Krogh, T.N., Bachmann, E., Teisner, B., Skjødt, K., and Højrup, P. (1997) Glycosylation analysis and protein structure determination of murine fetal antigen 1 (mFA1). *Eur. J. Biochem.* **244**, 334–342
30. Vance, B.A., Wu, W., Ribaudo, R.K., Segal, D.M., and Kearse, K.P. (1997) Multiple dimeric forms of human CD69 result from differential addition of *N*-glycans to typical (Asn-X-Ser/Thr) and atypical (Asn-X-Cys) glycosylation motifs. *J. Biol. Chem.* **272**, 23117–23122
31. Giuffrida, M.G., Cavaletto, M., Giunta, C., Neuteboom, B., Cantisani, A., Napolitano, L., Calderone, V., Godovac-Zimmermann, J., and Conti, A. (1997) The unusual amino acid triplet Asn-Ile-Cys is a glycosylation consensus site in human α-lactalbumin. *J. Protein Chem.* **16**, 747–753
32. Araki, T., Haupt, H., Hermentin, P., Schwick, H.G., Kimura, Y., Schmid, K., and Torikata, T. (1998) Preparation and partial structural characterization of α1T-glycoprotein form normal human plasma. *Arch. Biochem. Biophys.* **351**, 250–256
33. Rebbaa, A., Yamamoto, H., Saito, T., Meuillet, E., Kim, P., Kersey, D.S., Bremer, E.G., Taniguchi, N., and Moskal, J.R. (1997) Gene transfection-mediated overexpression of β1,4-*N*-acetylglucosamine bisecting oligosaccharides in glioma cell line U373 MG inhibits epidermal growth factor receptor function. *J. Biol. Chem.* **272**, 9275–9279
34. Stanley, P. (1989) Chinese hamster overy cell mutants with multiple glycosylation defects for the production of glycoproteins with minimal carbohydrate heterogeneity. *Mol. Cell. Biol.* **9**, 377–383
35. McKern, N.M., Lou, M., Frenkel, M.J., Verkuylen, A., Bentley, J.D., Lovrecz, G.O., Ivancic, N., Elleman, T.C., Garrett, T.P.J., Cosgrove, L.J., and Ward, C.W. (1997) Crystallization of the first three domains of the human insulin-like growth factor-1 receptor. *Protein Sci.* **6**, 2663–2666
36. Garrett, T.P.J., McKern, N.M., Lou, M., Frenkel, M.J., Bentley, J.D., Lovrecz, G.O., Elleman, T.C., Cosgrove, L.J., and Ward, C.W. (1998) Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. *Nature* **394**, 395–399