Structural Requirements for Addition of O-Linked Carbohydrate to Recombinant Erythropoietin

Steve Elliott,* Tim Bartley, Evelyne Delorme, Pat Derby, Rob Hunt, Tony Lorenzini, Vann Parker, Michael F. Rohde, and Ken Stoney

Amgen, Amgen Center, Thousand Oaks, California 91320

Received April 19, 1994; Revised Manuscript Received July 7, 1994®

ABSTRACT: To define the structural requirements for addition of O-linked glycosylation *in vivo*, recombinant erythropoietin (rEPO) variants were constructed. Thirty-three independent Ser or Thr substitutions were constructed and examined to see which were subject to O-linked carbohydrate addition. Variants with Thr mutations at positions 123 and 125, but not elsewhere, contained additional carbohydrate, which suggests that several positions around the existing O-linked glycosylation site (Ser126), but not elsewhere, contain the necessary information for O-linked carbohydrate addition. Two forms of the Thr125 variant were identified. One form was glycosylated only at residue 125, and a second form was glycosylated at both Thr125 and Ser126, the normal O-glycosylation site. We have also found that glycosylation is less efficient when rEPO is improperly folded and that prolines at −1 and +1 relative to the O-glycosylation site enhance glycosylation.

Mammalian cells add carbohydrate via N-linkages to Asn residues or via O-linkages to Ser, Thr, hydroxylysine, or hydroxyproline residues. Unlike the N-linked glycosylation machinery which recognizes the sequence Asn-X-Ser/Thr (Pless & Lennarz, 1977), the amino acid sequence and structural requirements for O-linked glycosylation are not completely understood. There is a high degree of specificity for addition of O-linked carbohydrate: certain Ser/Thr residues in proteins are glycosylated with high efficiency, while others are not. This indicates that there are specific sequence or conformational requirements for carbohydrate addition. It is generally assumed that the glycosylated amino acid must be on the protein surface. In addition, there is some evidence that glycosylated residues are in reverse turns of the peptide backbone (Aubert et al., 1976; Fiat et al., 1980). A comparison of glycosylated and nonglycosylated amino acids indicates that prolines are often found at the −1, −3, −6, and +3 positions when glycosylation occurs (O'Connell et al., 1991; Wilson et al., 1991). A Ser or Thr with charged residues at these nearby positions was glycosylated less often. In yeast, *in vitro* mannosyl transfer assays on synthetic peptides indicate that transfer is more efficient when there are nearby prolines and a charged residue at +1 and that a Gly at −1 inhibits glycosyl transfer (Lehle & Bause, 1984). Studies with purified enzymes from bovine colostrum (O'Connell et al., 1992) or porcine submaxillary glands (Wang et al., 1992, 1993) and synthetic peptides indicate that adjacent sequence as well as the length of the peptide influences the efficiency of O-glycosylation. This information led to the conclusion that prolines are important for O-linked carbohydrate addition and that their role may be to induce β-turns. However, this has been disputed (Strahl-Bolsinger & Tanner, 1991).

Human erythropoietin (EPO)[1] is a circulating glycoprotein that contains three N-linked oligosaccharide chains and one O-linked oligosaccharide chain at Ser126 (Lai et al., 1986; Lin et al., 1985). The O-linked site has prolines at −5, −4, and +3 relative to the Ser126, but there is not a Pro at −1 or +1, and the region is not predicted to be in a β-turn. In addition there are charged residues at the −3 and +5 positions. The human EPO gene has been cloned, and recombinant glycoproteins (rEPO) have been produced in mammalian cells. Several studies indicate that the sequence around the Ser126 can affect the efficiency of O-glycosylation (O'Connell et al., 1992; Wang et al, 1993). However, these studies were done with synthetic peptides, and thus it is difficult to dissociate effects on peptide stability from primary sequence effects from effects on secondary structure. Indeed the lack of an identifiable sequon for O-glycosylation points to a secondary or tertiary structure as the substrate and not a primary amino acid sequence. Thus effects of changes in peptide sequence or length may affect conformational stability, which indirectly affects glycosylation. The fact that the length of the peptide influences formation of glycosylated peptides (Wang et al., 1993) suggests that lengthening the peptide stabilizes some structure and thus affects glycosylation. For these reasons it is desirable to examine glycosylation in an intact molecule. Furthermore, it is also important to understand how O-glycosylation is controlled *in vivo*.

We have developed EPO as a model system to examine sequence and structural requirements for O-glycosylation *in vivo*. We have previously reported that a Thr at position 126 in rEPO is glycosylated more efficiently than a Ser (Delorme et al., 1992) and that the efficiency can be determined *in vivo*. Thus, rEPO is a good model system for determining sequence requirements for O-glycosylation *in vivo*. Using *in vitro* mutagenesis, we have found that only certain regions in rEPO contain the necessary information for O-linked glycosylation *in vivo*. The efficiency of glycosylation can be altered by

* Corresponding author.

® Abstract published in *Advance ACS Abstracts,* September 1, 1994.

[1] Abbreviations: IEF, isoelectric focusing; LC/MS, liquid chromatography/mass spectroscopy; PNGaseF, peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase; *O*-glycanase, α-*N*-acetylgalactosaminidase, E.C. 3.2.1.49; PAGE, polyacrylamide gel electrophoresis; EPO, erythropoietin; rEPO, recombinant erythropoietin; rpHPLC, reverse-phase high-performance liquid chromatography; SDS, sodium dodecyl sulfate.

0006-2960/94/0433-11237$04.50/0

mutations adjacent to the glycosylation site. In addition, distant mutations that affect protein folding also affect O-glycosylation. Therefore, primary, secondary, and tertiary protein structure are important for efficient O-glycosylation.

EXPERIMENTAL PROCEDURES

EPO cDNA Clone. A human EPO cDNA clone (Law et al., 1986) derived from a human EPO genomic clone (Lin et al., 1985) was used in all experiments. The coding region was cloned into an SV40 vector as described (Delorme et al., 1992).

Construction of rEPO Variants. An 810-bp fragment containing the EPO coding sequences was transferred to M13mp18 (Yanisch-Perron et al., 1985), and single-stranded DNA was recovered from the supernatant of RZ1032 (*dut ung*) *Escherichia coli* cells transfected with the vector as described (Kunkel et al., 1987). Synthetic oligonucleotides containing the desired mutations were annealed to the vector and extended with Klenow polymerase in the presence of deoxynucleotide triphosphates, ATP, and T4 DNA ligase at 14 °C overnight. The mixture was then transfected into wild-type *E. coli* (JM109; Yanisch-Perron et al., 1985). Plaques containing DNA with the desired mutation(s) were identified by differential hybridization with the mutagenic oligonucleotide(s). The presence of each mutation was confirmed by sequence analysis. The DNA segment was subsequently transferred back into the original vector for expression in COS-1 and CHO cells.

Expression of EPO in Mammalian Cells (COS-1 and CHO). COS-1 and CHO D-cells were transfected by the calcium phosphate method as described previously (Wigler et al., 1978). In some cases COS-1 cells were transfected by electroporation (Delorme et al., 1992). Conditioned medium from COS-1 cells (5-10 mL) was recovered after 3–5 days of growth. Stable CHO cell transformants were selected in medium lacking hypoxanthine and thymidine. Individual colonies were picked, and the higher expressing clones were identified by Elisa assay of conditioned medium. Conditioned medium from the highest expressing clones was recovered for purification.

Analysis of Glycosylation by Western Immunoblotting. rEPO in conditioned medium from COS-1 cells (approximately 20 units) was immunoprecipitated with a rabbit anti-EPO polyclonal antibody and protein A–Sepharose. The EPO-containing pellets were resuspended in 0.2% SDS and boiled for 5 min. NP-40 and $NaPO_4$ (pH 8.5) were then added bringing their final concentrations to 2% and 70 mM, respectively. One microliter PNGaseF (peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase) (Genzyme) was then added to remove N-linked carbohydrate. In some cases the supernatant was treated with neuraminidase and O-glycanase. Neuraminidase treatment was done on PNGaseF-treated samples by adjusting the pH of the sample with 1/16 vol of 1 M $NaPO_4$, pH 6.1, and 1/40 vol of 1 N HCl. One microliter (0.01 units) neuraminidase (*Clostridium perfringens* type X; Sigma) was then added. Samples were treated for 1 h at 37 °C. Subsequent treatment with O-glycanase (Genzyme) was done by adding 1 μL of enzyme followed by an overnight digestion at 37 °C. EPO proteins were separated by SDS–PAGE (15%). Following transfer to 0.2-μm nitrocellulose, the rEPO proteins were visualized with an anti-EPO monoclonal antibody, 9G8A (obtained from D. Chang, Amgen), and an anti-mouse IgG Vectastain kit (Vector Laboratories, Inc., Burlingame, CA).

rEPO Purification. Medium (385 mL) conditioned by CHO cells expressing Thr125 EPO was concentrated and diafiltered against 10 mM Tris-HCl, pH 8.6, using a stirred cell (YM-10 membrane, Amicon, Danvers, MA). The concentrate (17 mL) was applied to an ion-exchange column (2.2 × 1.4 cm, Q-Sepharose fast flow, Pharmacia, Piscataway, NJ) and eluted with a 250-mL linear gradient of 0–250 mM NaCl at a flow rate of 0.4 mL/min. Upon analysis by SDS–PAGE and IEF, five pools were made according to the molecular weight and charge of the rEPO species in each fraction. rEPO in each pool was purified further by rp C4 chromatography with a 20–94% ethanol gradient, in 10 mM Tris-HCl, pH 7.0. The ethanol was then removed by binding the material to Q-Sepharose and eluting with 20 mM sodium citrate, and 250 mM NaCl. Recovered material was judged to be pure according to rpHPLC and SDS–PAGE.

Digestion of Purified rEPO and Sequence Analysis of Peptides. Tryptic digests of purified rEPO were done with sequencing grade trypsin (Boehringer Mannhein) at a trypsin: protein ratio of 1:50. Samples were incubated at 37 °C for 6 h. The peptides were separated by reverse-phase HPLC on a Hewlett-Packard 1090M chromatograph using a Vydac C4 column and a 0–65% (0.06% trifluoroacetic acid/80% acetonitrile) gradient over 90 min. The peptide in each peak was collected, dried in a Speed-Vac, and resuspended in 30 μL of 0.1% trifluoroacetic acid. Sequencing of the peptides was done using either an ABI 470A or 473A protein sequencer.

Mass Spectrometry. Molecular weights of glycosylated rEPO peptides were determined using a triple-quadrapole mass spectrometer (SCIEX API III, Toronto, Canada). LC/MS analysis was performed using a Vydac C4 column, 1 × 150 mm, at 50 μL/min flow. A sample of 50 μg of rEPO tryptic digest in 20 μL was injected onto the column and eluted with a 30-min linear gradient of 0.1% trifluoroacetic acid in water from 0 to 54% acetonitrile. The effluent of the reverse-phase HPLC separation was directed into the ionspray inlet. Scans were accumulated, and their spectra were averaged for various time segments and analyzed using MacSpec version 3.2 software (SCIEX, Toronto, Canada).

RESULTS

(a) Potential Sites for O-Linked Glycosylation in rEPO. Human EPO contains 10 serine and 11 threonine residues, yet only one of these, Ser126, contains O-linked carbohydrate (Delorme et al., 1992; Lai et al., 1986). *In vitro* mutagenesis was used to determine whether Ser or Thr residues introduced at other locations within the rEPO molecule could be glycosylated. Thirty-three sites were chosen across the length of the rEPO coding sequence. The locations of the mutations are shown in Figure 1. Each variant was expressed in COS-1 cells, and conditioned medium was recovered. The rEPO was immunoprecipitated with an anti-EPO polyclonal antibody, treated with *N*-glycanase to remove N-linked carbohydrate, and then subjected to Western immunoblotting.

Two main bands are apparent in rEPO treated with *N*-glycanase (Figure 2, lane 2); the upper band represents 60–70% of the total. In a rEPO variant that removes the O-linked glycosylation site only the lower band is present (Figure 2, lane 3). Treatment with *N*-glycanase, neuraminidase, and O-glycanase results in disappearance of the upper band and an increase in an amount of the lower band (Figure 3, lane 4). These results indicate that the upper band contains O-linked carbohydrate and that the lower one does not. The ratio of glycosylated to nonglycosylated rEPO did not vary significantly in different experiments or at different expression levels. This indicates that Ser126 is not always glycosylated by COS-1 cells and that the efficiency of glycosylation at

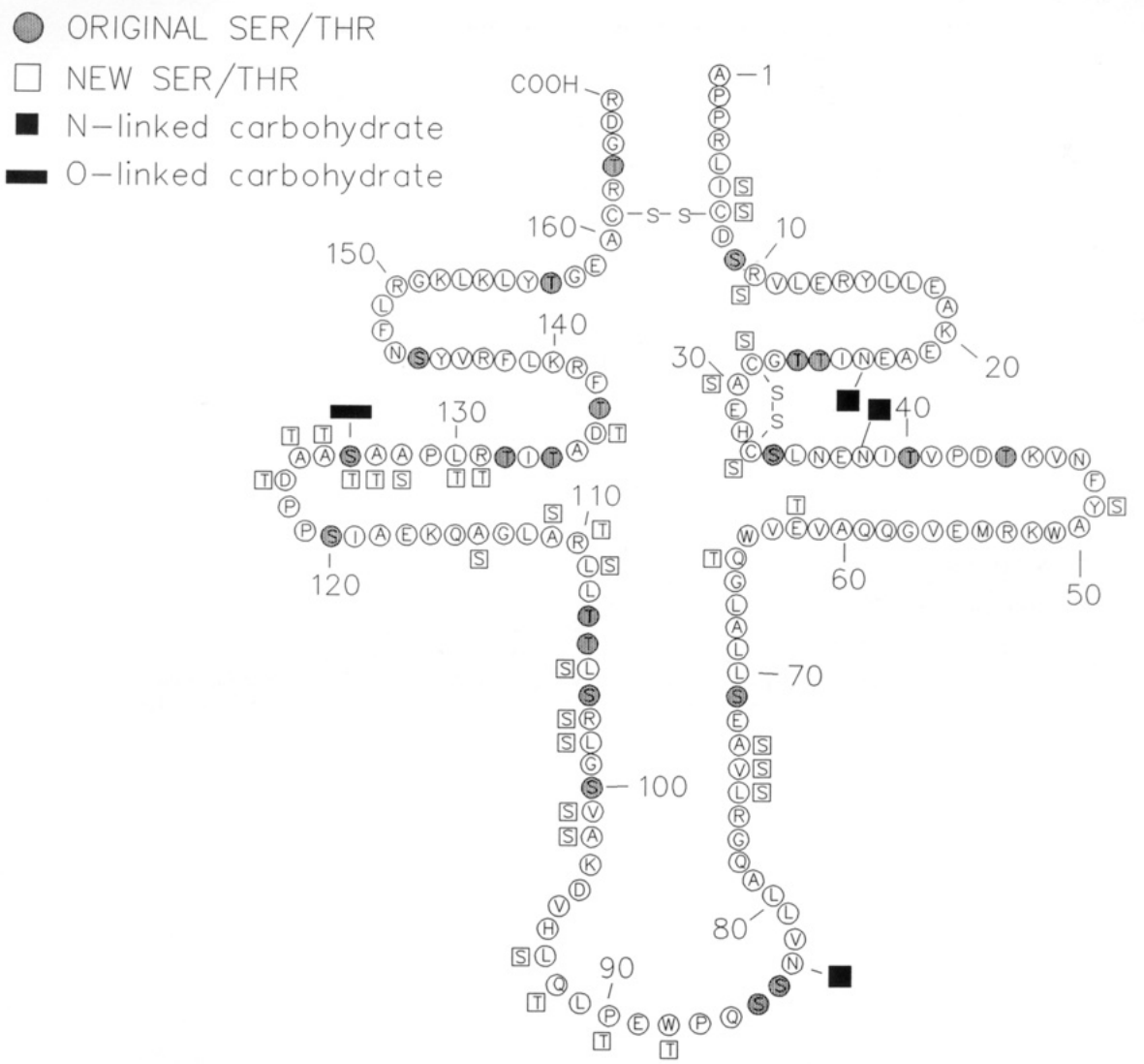

FIGURE 1: Locations of existing Ser and Thr residues and residues introduced by *in vitro* mutagenesis. The primary amino acid sequence of human EPO is shown. Existing Ser and Thr residues are filled circles. New Ser and Thr residues are boxed and adjacent to the position of the substitution.

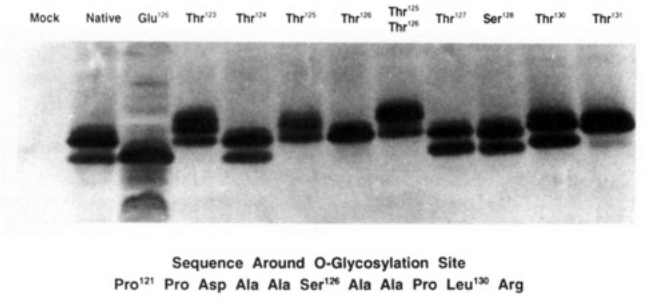

FIGURE 2: Effect of mutations around the glycosylated Ser126 on glycosylation efficiency. rEPO (native) or rEPO variants with the indicated mutations were expressed in COS cells. rEPO containing each variant in supernatants was immunoprecipitated, treated with *N*-glycanase, and subjected to Western analysis as described in Experimental Procedures.

Ser126 is determined primarily by elements within the rEPO molecule and not by other variables such as growth rate, culture conditions, or expression level.

We next examined all the Ser and Thr rEPO variants (Figure 1) to see whether any of them migrated with a size consistent with addition of a second O-linked oligosaccharide chain. Only two of the rEPO variants had a species with decreased electrophoretic mobility consistent with additional O-linked carbohydrate (Figure 2, lanes 4 and 6). Some of the remaining 30 variants had different proportions of glycosylated and nonglycosylated rEPO, but none of them had bands that had an electrophoretic mobility slower than rEPO with the human amino acid sequence (data not shown). This suggests that only the region around Ser126 contains the structural information required for O-glycosylation.

The rEPO variants that were larger in size contained Thr123 and Thr125 mutations. Both of these mutations are close to the existing O-glycosylation site at Ser126. This suggests that the region around residue 126, and not just residue 126, contains the necessary information for O-glycosylation. With both Thr123 and Thr125 variants there was no detectable nonglycosylated material, but there were species that migrated with the same mobility as O-glycosylated rEPO as well as several larger species. The fact that there are several different larger species is presumably due to heterogeneity in structure of the added oligosaccharide chains (see below), for example, variations in the amount of attached sialic acid. The Ser/Thr replacements at positions 124, 127, 128, 130, and 131 affected the proportion of glycosylated and nonglycosylated material but not the mobilities (Figure 2, lanes 5, 8, 9, 10, and 11). This suggests that some, but not all, positions in this region can be glycosylated by COS-1 cells.

(b) Characterization of the O-Linked Carbohydrate on the Thr125 EPO Variant. To understand the nature of the

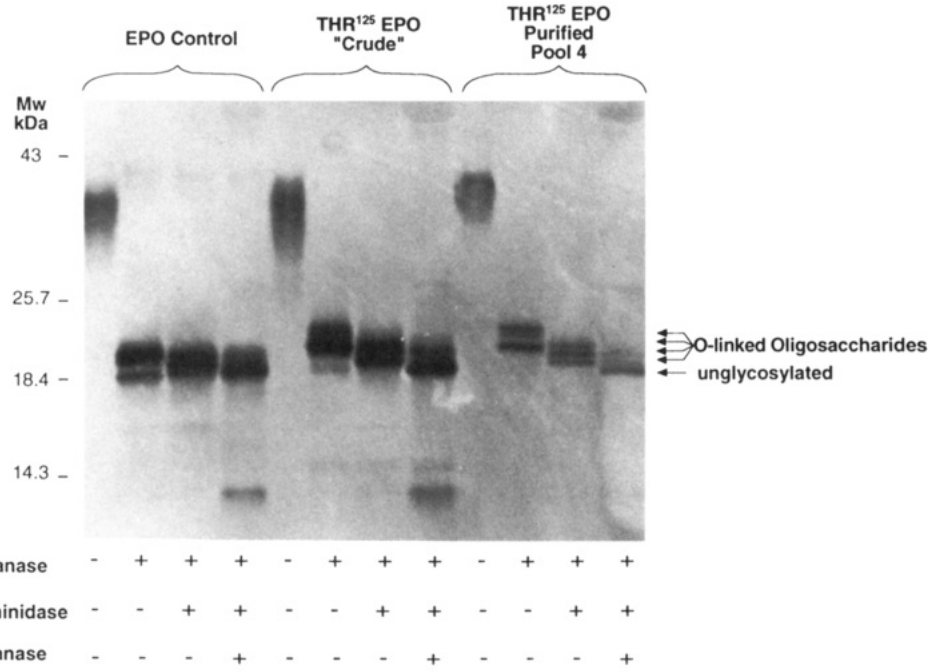

FIGURE 3: Glycanase treatment of rEPO and Thr125 EPO. Conditioned media containing rEPO (control), crude Thr125 EPO, or purified Thr125 EPO (pool 4) were treated with enzymes as shown. The 12.5% SDS–PAGE immunoblot of the samples was treated with EPO antibodies and developed as described in Experimental Procedures.

carbohydrate on the Thr125 variant, the variant was stably expressed in CHO cells. A supernatant containing this variant was treated with *N*-glycanase and subjected to Western analysis. A species with a slower mobility than any found in rEPO is apparent (compare Figure 3, lane 2, to lanes 6 and 10). Therefore, CHO cells also add additional carbohydrate to this variant. Treatment with neuraminidase (Figure 3, lanes 7 and 11) and *O*-glycanase (Figure 3, lanes 8 and 12) reduced the size to that of nonglycosylated material. These results indicate that the larger sized bands from the Thr125 variant were due to O-linked carbohydrate containing sialic acid.

The purified Thr125 EPO was separated into different pools depending on the number of negative charges (see Experimental Procedures). Human EPO has a theoretical maximum of 14 sialic acids, four on each of the three N-linked oligosaccharides and two on the O-linked oligosaccharide. Thus an extra O-linked chain could add two additional sialic acids. To determine the number of negative charges present, the different pools were compared to purified rEPO by isoelectric focusing. The rEPO standard contains EPO species with 10–13 sialic acids (Figure 4, lanes 1 and 5). Pools 2, 3, and 4 had species with increasing negative charge (lanes 3, 4, and 5). The Thr125 EPO pool with the most negatively charged material (pool 4) had species that migrated at positions with predicted sialic acid contents of 11–16 (lane 4). Isoelectric focusing of *N*-glycanase-treated material also allowed detection of two new bands that were more negatively charged than those found on a similarly treated rEPO standard (data not shown). These results suggest that the O-linked oligosaccharides contain up to two additional negative charges, presumably due to sialic acids.

To determine which positions on Thr125 EPO are glycosylated, Thr125 (EPO pool 2, 3, or 4) and rEPO were treated with trypsin and the peptides were separated by reverse-phase HPLC. Thr125 EPO from all three pools gave similar patterns. However, a comparison of rEPO and Thr125 EPO indicated peak height differences (peaks 1 and 3) and the presence of a new peak (peak2) in Thr125 EPO (Figure 5).

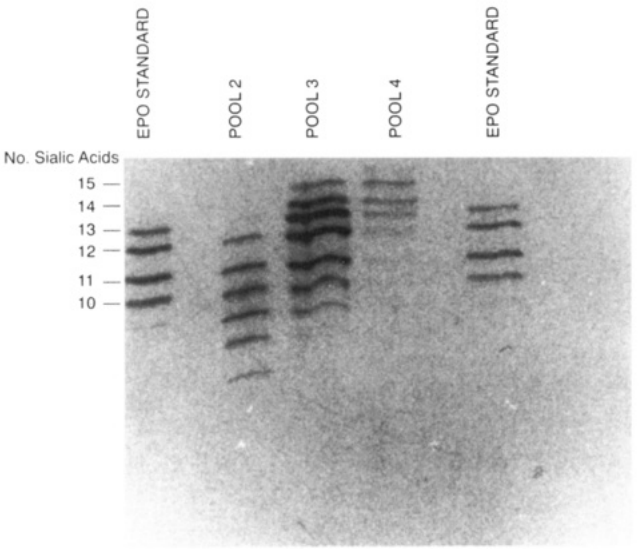

FIGURE 4: Isoelectric focusing of rEPO (standard) or purified Thr125 EPO. rEPO samples were analyzed by isoelectric focusing in polyacryalmide gels containing 5 M urea and pH 3–10 ampholytes. The proteins were visualized by staining with Coomassie Blue R-250.

The three peaks that differed between rEPO and Thr125 EPO (Figure 5, peaks 1, 2 and 3) were purified, and the sequences in each were determined. Peak 1 from Thr125 EPO contained 2 sequences: TITADTR, which represents residues 132–139, and LYTGEAxR, which represents residues 155–162. Both of these sequences were also found in rEPO. The residue at position 161 could not be assigned but is predicted to be a cysteine. Peak 3 from Thr125 EPO contained only the sequence VYSNFLR, which represents residues 144–150. This sequence was also found in peak 3 from rEPO. However, unlike Thr125 EPO, peak 3 from rEPO also contained a peptide with no signal at Ser126 and a normal signal at position 125. This result is consistent with carbohydrate attached to Ser126 in rEPO. Peak 2 from Thr125 EPO was not found in rEPO. It contained the sequence EIASPPDAxSAPPL, which represents residues 117–130 in Thr125 EPO. There was no signal

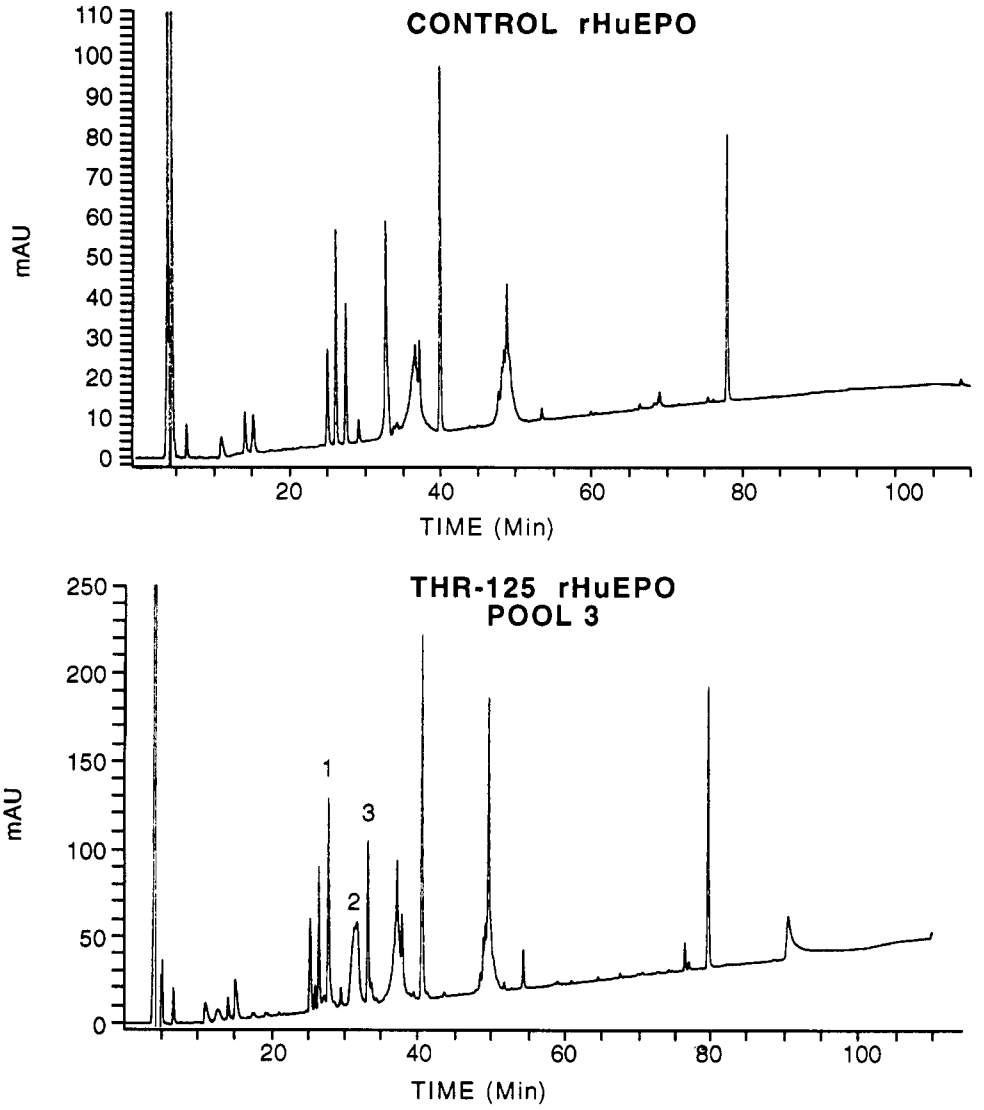

FIGURE 5: Separation of rEPO peptides produced by trypsin. Purified rEPO standard or purified Thr125 EPO (pool 3) was treated with trypsin and separated by HPLC as described in Experimental Procedures. Peaks that differed in size or location are indicated.

at Thr125, while a signal for Ser126 was observed. This result is consistent with carbohydrate attached to Thr125 and the presence of nonglycosylated Ser126 in this peptide. In addition to the differences in peaks 1, 2, and 3, there were other smaller peaks found in Thr125 EPO and not in rEPO. The peptides at 18.2, 29.3, 30, and 54.2 min were isolated and sequenced. Only the peptide at 29.3 min had a sequence related to the O-glycosylation site. The sequence EIAxPPDAxxAAPL was found, which also represents residues 117–130. The three positions at which no sequence was obtained are Ser120, Thr125, and Ser126. The amount of peptide used for sequence analysis was very low, and thus it is not clear whether these residues were modified or whether these residues were present but the amount of material released was below the limit of detection. However, the fact that this sequence was found suggests that it must be modified differently from the peptide with the same sequence that was found at 31.4 min (peak 2). Thus there are at least two different forms of Thr125 EPO; the major one is glycosylated at position 125 and not at 126.

There are several explanations for the observed increase in the amount of carbohydrate seen and the sequence results with the Thr125 variant. There could be two oligosaccharide chains, one attached to Thr125 and a second attached sometimes to Ser126; alternatively, there could be one chain at Thr125, but the chain is sometimes increased in mass. To determine what type of O-linked carbohydrate was attached to Thr125 EPO, pool 3 from Thr125 EPO was treated with trypsin, separated by rpHPLC, and then analyzed by mass spectrometry. The LC/MS scan was examined using the Sciex software; signals were displayed from specific ions, as well as from the entire mass range. This made it possible to locate the regions of the LC/MS runs where the mass of a tryptic peptide for amino acids 117–131 included one (Figure 6A) or two (Figure 6B) oligosaccharide chains on the peptide. Figure 6C presents the entire mass range for comparison. In Figure 6, we see that there are two different ions that eluted close to each other, but the ion representing the more highly glycosylated two-chain peptide eluted slightly earlier than the one-chain peptide. This would be consistent with the suggestion that the higher mass was due to addition of a second O-linked oligosaccharide chain which resulted in a lower hydrophobicity and not the suggestion that there was a single chain larger in size. Additional analysis of the complete spectra (Figure 6C) using Sciex software indicated that peptides could be identified that had one oligosaccharide chain with zero, one, or two sialic acids, as well as two chains with two, three, or four sialic acids (data not shown). Thus it appears that there are two O-linked classes of Thr125 EPO. One class corresponds to the peptide in Figure 5 detected at 29.3 min and is glycosylated at both Thr125 and Ser126, while the

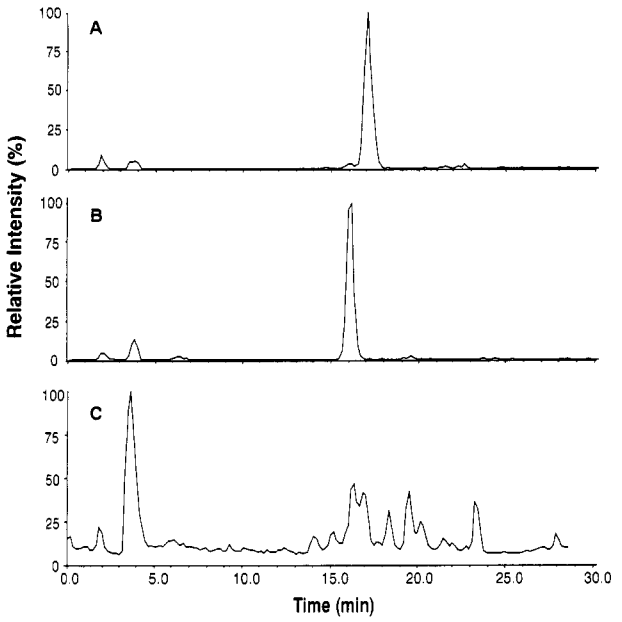

FIGURE 6: Ion-extraction display of LC/MS analysis of a tryptic digest of rEPO Thr 125 analog. Thr125 EPO was digested with trypsin and analyzed by LC/MS as described in Experimental Procedures. To locate the regions of the resulting separation containing the O-glycosylated peptides, an ion-extraction search was performed, with criteria set to find the region of the chromatogram where one would find all ions with mass to charge ratios in the range 300–2400 (panel C, essentially all ions), ions in the range 1221.5–1223.5 (panel A, the MH^{2+} ions for single chain, two sialic acids), and ions in the range 1033–1035 (panel B, the MH^{3+} ions for two chains, three sialic acids). In a separate experiment, the full spectrum for the time regions in panels A and B was examined to determine the presence of multiple levels of sialic acids for each.

other class corresponds to peak 2 and has a single oligosaccharide chain at Thr125, the same size as that found on rEPO. Furthermore, each of the oligosaccharide chains has some structural microheterogeneity.

(c) Mutations That Alter Secondary and Tertiary Structure of rEPO Also Affect O-Glycosylation. To see whether the tertiary or the secondary structure of rEPO is important for O-linked glycosylation, we examined the effect of removal of disulfide bonds on glycosylation. Human EPO has two disulfide bonds that join Cys7 to Cys161 and Cys29 to Cys33. The disulfide bonds are essential for activity (Wang et al., 1985). We have found that mutations in any of the cysteines reduce *in vitro* activity and that the EPO variants have an altered immunoreactivity with monoclonal antibodies that recognize folded and unfolded forms of EPO (data not shown). Thus these mutations affect the conformation of EPO. If the O-linked glycosylation machinery recognizes a primary amino acid sequence or a particular secondary structure not affected by tertiary structure, then the amount of carbohydrate added to mutants lacking disulfide bonds should be unaffected. However, if folding is required, then O-glycosylation should be inhibited in the mutants. As shown in Figure 7, disruption of either disulfide bond by a Cys to Ser substitution resulted in a decreased proportion of O-glycosylated to nonglycosylated rEPO. Sixty percent to 70% of rEPO produced in COS-1 cells contains O-linked glycosylation (Figure 7, lane 2). A Ser7 mutation resulted in only 5–10% O-linked glycosylation (lane 5), while either a Ser29 (lane 7) or a Ser33 (lane 9) mutation resulted in 30–40% glycosylation. This indicates that disruption of disulfide bonds inhibits O-linked carbohydrate addition and that the Cys7–Cys161 disulfide bond is more important for maintaining an O-glycosylation-competent structure than the Cys29–Cys33 disulfide bond. Antibody binding experiments indicate that the Ser7 mutation has a greater effect on protein folding than do Ser29 mutations (data not shown). Thus, the extent of unfolding affects the efficiency of O-glycosylation. Variants with mutations at other residues around the disulfide bonds such as at residue 6, 11, or 30 are O-glycosylated normally (lanes 4, 6, and 8). Therefore, the effect seems to be specific to the disulfide bonds and not to the sequence around the disulfide bonds. Other EPO variants were identified that disrupt rEPO structure as evidenced by decreased immunoreactivity in RIA, increased immunoreactivity with an antibody that reacts strongly with unfolded rEPO, and decreased biological activity. These mutants, Thr137 to Ile and Leu105 to Ser, both had a reduced O-glycosylation efficiency; approximately 10% of the material contained O-linked carbohydrate (data not shown). Thus, the effects of mutations on the efficiency of O-glycosylation were not limited to those that affect disulfide bonds. These results indicate that the tertiary fold and not just the local sequence context around rEPO is important for efficient O-glycosylation.

(d) Adjacent Prolines Enhance O-Glycosylation of EPO in Vivo. It has been suggested that prolines around O-linked glycosylation sites may be important for recognition by the glycosylation machinery (Lehle & Bause, 1984; Wilson et al., 1991). Prolines are found at some positions around Ser126 including residues 121, 122, and 129. We substituted an Asn for the Pro at position 121 (−5 relative to Ser126) and saw no effect on glycosylation (Figure 8, lane 4). This indicates that this proline is not essential for glycosylation at Ser126. We also introduced prolines at position 124, 125, or 127. The variant with a Pro124 substitution inhibited O-glycosylation; only 10–20% of the material contained O-linked carbohydrate (Figure 8, lane 5). However, a proline at either position 125 or 127 enhanced O-glycosylation. We could not detect any rEPO from these variants that lacked O-linked carbohydrate. This indicates that a proline at the −1 or +1 position can enhance O-glycosylation and that a Pro at −2 can inhibit O-glycosylation *in vivo*. A variant with a Pro124 Thr125 substitution was also constructed. With this variant we saw an increased amount of rEPO that contained larger amounts of carbohydrate. This is presumably due to an increase in the amount of material with oligosaccharide chains at both Thr125 and Ser126 (compare Figure 8, lanes 8 and 9). The Thr125 variant is glycosylated primarily at Thr125 but contains some carbohydrate at Ser126 (see above). Thus in this variant the proline at position 124 must enhance glycosylation at Ser126. It is noteworthy that the Thr123 variant also contains additional O-linked carbohydrate. There are prolines at both −2 and −1 relative to Thr123. These results suggest that a Pro at −2 can enhance glycosylation in some sequence contexts.

To examine the role of prolines on glycosylation in more detail, we have introduced Thr residues at positions 88 and 130 (Figure 1). There are existing proline residues at positions 87 and 129. Western blots on *N*-glycanase-treated material indicate that neither Thr88 nor Thr130 contains O-linked carbohydrate. For Thr88 EPO the mass of O-linked oligosaccharides added, as well as the efficiency of glycosylation, was unchanged (data not shown). With Thr130 EPO the percentage of O-glycosylation was reduced slightly but the mass of the oligosaccharides added was the same (Figure 2). These results suggest that a proline adjacent to a Thr is not by itself sufficient for addition of O-linked carbohydrate.

(e) Effect of Other Mutations Near Ser126 on O-Glycosylation of rEPO in Vivo. A Pro at position 124 inhibits O-glycosylation at position 126 (see above). To see whether

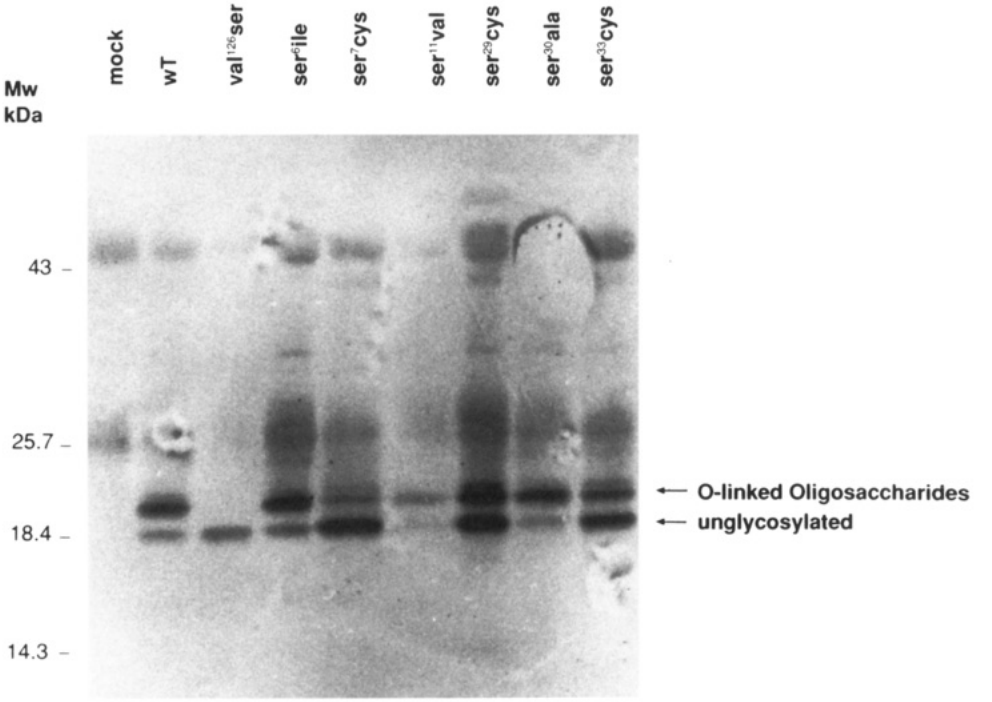

FIGURE 7: Effect of disulfide bonds on glycosylation efficiency at ser126. COS conditioned medium containing the indicated rEPO (wT) or rEPO variants was immunoprecipitated with anti-EPO polyclonal antibody, and the immunoprecipitate was treated with *N*-glycanase. Samples were then run on SDS–PAGE, blotted to nitrocellulose, and developed with anti-EPO antibodies as described in Experimental Procedures.

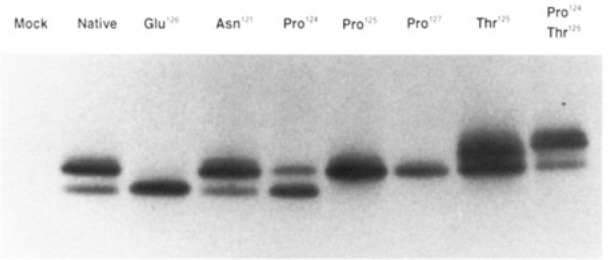

FIGURE 8: Effect of prolines on efficiency of glycosylation at Ser126. COS conditioned medium containing rEPO (Native) or the indicated EPO variants was immunoprecipitated with anti-EPO polyclonal antibody, and the immunoprecipitate was treated with *N*-glycanase. Samples were then run on SDS–PAGE, blotted to nitrocellulose, and developed with anti-EPO antibodies as described in Experimental Procedures.

this effect is specific to proline mutations, Val, Gly, and Thr were substituted at position 124. In all cases the amount of O-glycosylation at Ser126 was reduced to approximately 40–50% of the total (data not shown). Thus these substitutions also inhibited O-glycosylation, albeit not as severely as with Pro124. This indicates that there are only certain allowed substitutions at the –2 position in this sequence context that are conducive to efficient O-glycosylation. Other mutations around this region also inhibit glycosylation at Ser126, e.g., Thr127, Ser128, and Thr130 (Figure 2, lanes 9, 10, and 11). Thus there are amino acid substitutions at other positions that are also not allowed. We have identified some mutations that enhance glycosylation. Ser126 to Thr and Arg131 to Thr both result in greater that 90% O-glycosylation at Ser126 (Figure 2, lanes 7 and 12). The Arg131 to Thr mutation may enhance glycosylation by removing a charge from the +5 position. In this regard it is noteworthy that an Asp123 to Ala mutation had no apparent effect (data not shown). Thus if the charged amino acid residues around the glycosylation site inhibit O-glycosylation, the effect is position dependent. A Thr125 Thr126 mutant also resulted in an increased amount of O-glycosylated material and an increased mass of oligosaccharide. In this case there is an increase in the amount of material that contains two oligosaccharide chains. These results indicate that the percent occupancy of a glycosylation site can be modulated by the local sequence around the site.

DISCUSSION

Several studies have examined the relationship between primary sequence and O-glycosylation using synthetic substrates and purified enzymes (Lehle & Bause, 1984; O'Connell et al., 1992; Wang et al., 1992, 1993). These types of experiments are useful for examining primary sequence requirements for O-glycosylation but are less useful for examining conformational effects. The fact that there is no consensus sequence for O-glycosylation suggests that there are subtle conformation-dependent structures recognized by oligotransferases. For these reasons we undertook an *in vivo* approach in an intact molecule to see whether we could identify important conformational epitopes for O-glycosylation.

Our initial analysis was to see whether particular regions or secondary structures are required for O-glycosylation. We introduced Ser/Thr residues at selected regions of the rEPO molecule. The regions included predicted α-helices and turns. In two cases we introduced residues adjacent to prolines. In total, Ser or Thr substitutions were made at 33 other positions in rEPO. Only the variants with Thr123 or Thr125 substitutions resulted in addition of extra O-linked carbohydrate. There are 21 Ser and Thr residues in HuEPO. Only Ser126 contains O-linked carbohydrate. Thus of 54 different positions in rEPO only three could be O-glycosylated, and these were localized to the region around the existing O-glycosylation site. This suggests that only this region contains attributes that allow specific O-linked carbohydrate addition. It also indicates that glycosylation is driven not by the primary amino acid sequence around the site but by secondary or tertiary structure because at least three nearby positions can be glycosylated, and each is in a different sequence context.

The variant with a Thr125 mutation was glycosylated primarily at position 125, but it also could be glycosylated at positions 125 and 126 simultaneously. rEPO with Thr at residue 123 also has a greater mass of O-linked carbohydrate. We have not characterized this material in detail, so we do not know whether Thr123 or Ser126 is preferred. The results with the Thr125 variant indicate that while both Thr125 and Ser126 can be glycosylated simultaneously, Thr125 is preferred. We have observed an overall similar phenomenon in the characterization of O-glycosylation of Ser142 and Thr143 in another protein, stem cell factor (Derby et al., 1993). The observations noted here are mimicked in the stem cell factor results in all respects including HPLC separation of forms with differing masses corresponding to sialic acid differences. Recently O'Connell et al. (1992) reported that synthetic peptides containing an EPO sequence, residues 121–131, could be glycosylated at Thr126 using an enzyme purified from bovine colostrum. They also found that insertion of an additional Thr at position 127 resulted in glycosylation at only the second Thr. Wang et al. (1993) reported that two adjacent threonines could be simultaneously glycosylated *in vitro* but that a sequence containing a Ser adjacent to a Thr was glycosylated only at the Thr. Our results disagree with those of both groups. We find that a Thr-Thr sequence is glycosylated at both positions in agreement with Wang et al. (1993). However, we find that a Thr-Ser sequence can also be glycosylated at both positions but that a Thr-Thr sequence is glycosylated more efficiently. Interestingly, we were unable to see any additional glycosylation when we substituted a Thr for an Ala at position 127, resulting in a Ser-Thr sequence. In this case we do not know which of the two residues is glycosylated.

The fact that Thr125 is glycosylated more efficiently than Ser126 is of interest. One explanation is that position 125 is in a more favorable conformation than position 126. Another possibility is that there is a preference for Thr over Ser. In support of this possibility are the observations that a Thr126 variant is glycosylated more efficiently than rEPO which contains Ser126 [see Results and Delorme et al. (1992)] and that a Thr125/Thr126 variant is glycosylated more efficiently than a Thr125/Ser126 variant. It has been proposed that different enzymes may be responsible for Thr vs Ser glycosylation (Wang et al., 1992; O'Connell et al., 1992). However, more recent results indicate that a single purified enzyme can perform both glycosylations but that the sequence of the peptide is important for Ser glycosylation (Wang et al 1993). On balance the data support the proposal that Thr residues are better substrates than Ser.

Some rEPO lacks O-linked carbohydrate when expressed in either COS or CHO cells, but when O-linked oligosaccharides are present, they are only found at position 126 (Delorme et al., 1992; Tsuda et al., 1990). This suggests that glycosylation at Ser126 is specific but not efficient. This observation allowed us to design experiments to identify structural requirements for glycosylation at Ser126. We have found that the percent occupancy of the O-glycosylation site is decreased by mutations that inhibit protein folding. This suggests that addition to this O-glycosylation site is conformation dependent. Subcellular fractionation studies suggest that O-linked oligosaccharide addition occurs in the Golgi where proteins are already folded (Elhammer & Kornfeld, 1984; Hanover et al., 1982), which is consistent with the folding requirement that we have identified.

Prolines at the −1 and +1 positions enhance O-glycosylation. A proline, when present at the −2 position (Pro124), inhibits O-glycosylation. These results are in agreement with previous studies which suggest an importance of prolines (Lehle & Bause, 1984; Wilson et al., 1991). Interestingly, a Pro124 Thr125 variant is glycosylated more efficiently than a Thr125 mutation by itself. The fact that the Thr125 variant is glycosylated primarily at Thr125 suggests that the Pro124 mutation enhances glycosylation at Ser126 in this variant. These results suggest that the location and sequence context of prolines are important. We have also found that introduction of a Thr adjacent to an existing proline does not always result in addition of carbohydrate (Thr88 and Thr130) and that removal of a proline five residues away (position 121) has no effect. Thus a proline can enhance O-glycosylation at some sites but is not sufficient at other sites.

These results are all consistent with the effect on glycosylation of the protein conformation around the glycosylation site and not with recognition of a particular sequence context. It has been proposed that Ser/Thr residues in β-turns are the structures that are recognized (Aubert et al., 1976; Fiat et al., 1980). It is well known that structurally similar turns can have different primary amino acid sequences (Chou & Fasman, 1977). This could account for the heterogeneity in amino acid sequences found around O-glycosylation sites. Furthermore, the Ser126 in rEPO is not predicted to be part of a β-turn, and it is not efficiently glycosylated. However, the mutations that enhanced glycosylation (Pro at −1 or +1 and Arg 131 to Thr) are all substitutions that increase the β-turn-forming potential of the region. However, the enhanced glycosylation of the Arg131 to Thr mutation may be due to removal of the charged residue from the +5 position. On the other hand, there are only certain allowed substitutions. Ser, Gly, and Thr are amino acids found at high frequency in turns (Chou & Fasman, 1977), but their presence at certain positions around Ser126 in rEPO inhibits O-glycosylation (gly at −2, Thr at +1, Ser at +2). It may be that there are important position effects for turn-forming amino acids necessary to actually allow turns, or that introduction of these amino acids may place the Ser or Thr in the wrong part of the turn. Another possibility is that the type of β-turn may affect glycosylation efficiency and that the different substitutions cause different types of turns. For example, it has been shown that a Pro-Ser glycopeptide is likely to form a type I(III) turn, while a similar peptide lacking the Pro forms a type II turn (Hollosi et al., 1990). Thus a type I(III) turn may be glycosylated more readily than a type II turn. Further experiments will resolve these possibilities.

ACKNOWLEDGMENT

We wish to thank J. Grant, F. Wang, D. Green, and B. Aguero for expert technical assistance, T. Jones and E. Fisher for oligonucleotide synthesis, and K. Chen, S. Suggs, A. Harcourte, and J. Katzowitz for sequencing analysis.

REFERENCES

Aubert, J.-P., Biserte, G., & Loucheux-Lefebvre, M.-H. (1976) *Arch. Biochem. Biophys. 175*, 410–418.

Chou, P. Y., & Fasman, G. D. (1977) *J. Mol. Biol. 115*, 135–175.

Delorme, E., Lorenzini, T., Giffin, J., Martin, F., Jacobsen, F., Boone, T., & Elliott, S. (1992) *Biochemistry 31*, 9871–9876.

Derby, P. L., Wypych, J., Rush, R. S., Clogston, C. L., & Rohde, M. F. (1993) in *Techniques in Protein Chemistry V* (Crabbe, J. W., Ed.) pp 89–96, Adacemic Press, New York.

Elhammer, A., & Kornfeld, S. (1984) *J. Cell Biol. 99*, 327–331.

Fiat, A.-M., Jolles, J., Aubert, J.-P., Loucheux-Lefebvre, M.-H., & Jolles, P. (1980) *Eur. J. Biochem. 111*, 333–339.
Hanover, J. A., Elting, J., Mintz, G. R., & Lennarz, W. J. (1982) *J. Biol. Chem. 257*, 10172–10177.
Hollosi, M., Perczel, A., & Fasman, G. D. (1990) *Biopolymers 29*, 1549–1564.
Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367–382.
Lai, P.-H., Everett, R., Wang, F.-F., Arakawa, T., & Goldwasser, E. (1986) J. *Biol. Chem. 261*, 3116–3121.
Law, M. L, Cai, G.-Y., Lin, F.-K., Wei, Q., Huang, S.-Z., Hartz, J. H., Morse, H., Lin, C.-W., Jones, C., & Kao, F.-T. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6920–6924.
Lehle, L., & Bause, E. (1984) *Biochim. Biophys. Acta 799*, 246–251.
Lin, F.-K., Suggs, S., Lin, C.-H., Browne, J. K., Smalling, R., Egrie, J. C., Chen, K. K., Fox, G. M., Martin, F., Stabinski, Z., Badrawi, S. M., Lai, P.-H., & Goldwasser, E. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7580–7584.
O'Connell, B., Tabak, L. A., & Ramasubbu, N. (1991) *Biochem. Biophys. Res. Commun. 180*, 1024–1030.
O'Connell, B. C., Hagen, F. K., & Tabak, L. A. (1992) *J. Biol. Chem. 267*, 25010–25018.
Pless, D. D., & Lennarz, W. J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 134–138.
Strahl-Bolsinger, S., & Tanner, W. (1991) *Eur. J. Biochem. 196*, 185–190.
Tsuda, E., Kawanishi, G., Ueda, M., Masuda, S., & Sasaki, R. (1990) *Eur. J. Biochem. 188*, 405–411.
Wang, F. F., Kung, C. K.-H., & Goldwasser, E. (1985) *Endocrinology 116*, 2286–2292.
Wang, Y., Abernethy, J. L., Eckhardt, A. E., & Hill, R. L. (1992) *J. Biol. Chem. 267*, 12709–12716.
Wang, Y., Agrwal, N., Eckhardt, A. E., Stevens, R. D., & Hill, R. L. (1993) *J. Biol. Chem. 268*, 22979–22983.
Wigler, M., Pellicer, A., Silverstein, S., & Axel, R. (1978) *Cell 14*, 725–731.
Wilson, I. B. H., Gavel, Y., & Heijne, G. (1991) *Biochem. J. 275*, 529–534.
Yanisch-Perron, C., Viera, J., & Messing, J. (1985) *Gene 33*, 103–119.