Structure-Function Analysis of Human α1,3-Fucosyltransferase

AMINO ACIDS INVOLVED IN ACCEPTOR SUBSTRATE SPECIFICITY*

(Received for publication, June 28, 1995, and in revised form, January 12, 1996)

Zhenghai Xu¶, Loc Vo§, and Bruce A. Macher¶†

From the Department of Chemistry and Biochemistry, San Francisco State University, San Francisco, California 94132

A series of molecular biology experiments were carried out to identify the catalytic domain of two human α1,3/4-fucosyltransferases (fucosyltransferases (FucTs) III and V), and to identify amino acids that function in acceptor substrate binding. Sixty-one and 75 amino acids could be eliminated from the N terminus of FucTs III and V, respectively, without a significant loss of enzyme activity. In contrast, the truncation of one or more amino acids from the C terminus of FucT V resulted in a dramatic or total loss of enzyme activity. Results from the truncation experiments demonstrate that FucT III62–361 (containing amino acids 62–361) and FucT V76–374 (containing amino acids 76–374) are active, whereas shorter forms of the enzymes were inactive.

The shortest, active forms of the enzymes are more than 93% identical at the predicted amino acid level, but have distinct acceptor substrate specificities. Thus, FucT III is an α1,4-fucosyltransferase, whereas FucT V is an α1,3-fucosyltransferase with disaccharide substrates. All but one of the amino acid sequence differences between the two proteins occur near their N terminus. Results obtained from domain swapping experiments demonstrated that the single amino acid sequence difference near the C terminus of these enzymes did not alter the enzyme’s substrate specificity. However, swapping a region near the N terminus of the truncated form of FucT III into an homologous region in FucT V produced a protein with both α1,3- and α1,4-fucosyltransferase activity. This region contains 8 of the amino acid sequence differences that occur between the two proteins.

α1,3/4-Fucosyltransferases (FucTs) catalyze the final step in the synthesis of a wide spectrum of fucosylated glycoconjugates including those that function in leukocyte adhesion (1, 2). Cell type-specific and developmental antigens are also products of these enzymes, as are tumor-associated antigens. These proteins are members of a large group of proteins (i.e. glycosyltransferases) that are responsible for the synthesis of glycolipids, glycoproteins, and proteoglycans (for a recent review, see Ref. 3). Glycosyltransferases are type II membrane proteins with four recognized protein domains: cytoplasmic, transmembrane, stem, and catalytic (4, 5). The presence of a stem region in these proteins is based on the observation that soluble, catalytically active forms occur naturally and are formed by proteolytic cleavage of the full-length enzyme in what is thought to be an extended protein domain (i.e. the stem region). The catalytic region is proposed to be a globular domain that contains the active site of the enzyme. Recently, we (6) have utilized a PCR based approach to determine which portions of the N- and C-terminal domains of murine and marmoset α1,3-galactosyltransferases can be eliminated without loss of activity (i.e. catalytic domain). The location of the catalytic domain of other glycosyltransferases is unknown. In this report, the location of the catalytic domain of two members of the human FucTs, FucT III and V, is presented. These two proteins were chosen for analysis because they share a high level of amino acid sequence similarity and yet have distinct acceptor substrate specificity (7, 8). As illustrated in Fig. 1, there are only 30 amino acid differences between the full-length sequences of these two proteins. In addition to these sequence variations, FucT V contains two regions not found in FucT III (shown as dashes in the FucT III sequence). A majority of the amino acid sequence differences between FucT III and V occur at the N terminus of these enzymes, with only one amino acid difference occurring over the last 200 amino acids. Therefore, precisely locating the catalytic domain should also identify the number of amino acid residues that account for the distinct acceptor substrate specificities of FucT III and V. The results presented in this study provide a precise location of the catalytic domains of FucT III and V and demonstrate there are less than 25 differences in the amino acid sequences of the catalytic domain of the two enzymes. Portions of the N-terminal domain from the truncated enzymes have been swapped producing chimeric proteins that were analyzed for acceptor specificity. These analyses demonstrated that the distinct acceptor specificity of the FucTs results from as few as 8 amino acid sequence differences.

EXPERIMENTAL PROCEDURES

Materials—Plasmids containing FucT sequences were previously described (9). Primers were synthesized by the Microchemical Core Facility (San Diego State University, San Diego, CA). The cloning vector pGIR-201 (10) was kindly provided by Dr. Kurt Drickamer, Columbia University, New York. The expression vector pSVL was purchased from Pharmacia (Piscataway, N.J.). Dulbecco’s modified Eagle’s medium with 4500 mg/liter glucose, phosphate-buffered saline, penicillin, streptomycin, and fetal calf serum were obtained from HyClone Laboratories Inc. (Logan, Utah). GDP-α1,3/4-Fucose (6.45 Ci/mmol) was obtained from DuPont NEN (Boston, MA). Dowex anion exchange resin 1 × 2, 200–400 mesh, Ca form, was obtained from Bio-Rad. C14 Sep-Pak cartridges were purchased from Fisher Scientific. The type 1 and type 2 disaccharides were purchased from Sigma. The 8-methoxycarbonyloxyethyl glycoside acceptors were kindly provided by Dr. Ole Hindsgaul (Department of Chemistry, University of Alberta).

Construction of Truncated Forms of Protein A-FucTs Fusion Proteins—The vector pROTA was designed to direct the synthesis of secretable, protein A fusion proteins in eukaryotic cells (e.g. COS-7...
cells) (11). Primers (1-10, 15, and 17) were used for the synthesis of truncated protein A fusion forms of FucT III and V (Table I). The PCR product from each set of primers was digested with EcoRI (Promega, Madison, WI), separated on a 1% agarose gel (Fisher Scientific, Fair Lawn, NJ), and transferred to a sheet of nitrocellulose. The protein A-IgG binding domain of the chimeric proteins was detected by incubating the blot with the following reagents: (i) biotinylated rabbit IgG and (ii) goat anti-rabbit-alkaline phosphatase substrate reagent. 

**Western Blot Analysis**—In some experiments, an alkaline phosphatase and an alkaline phosphatase substrate reagent.

**Expression Studies Using the Cloning Vector pGIR-201 and the Eukaryotic Expression Vector pSVL**—To obtain a soluble, secreted form of FucT without a protein A domain, Fuc-T DNA constructs were prepared using primers 11–13 (FucT V containing aa 44–361 and aa 59–361). The upstream primers contain an Xba I restriction site and the downstream primer 14 contains a Sac I restriction site. The PCR products were subcloned into the Xba I and Sac I restriction sites downstream from the dog insulin signal peptide gene of pGIR-201. The FucT construct, along with the dog insulin signal peptide gene, was isolated by digestion with Nhe I and Sac I and cloned into the Xba I and Sac I sites of the mammalian expression vector pSVL. The resulting plasmid (designated Fuc-T-pGIR-201-pSVL) was used to transfect COS-7 cells, and the FucT protein product was secreted into the cell culture medium. The medium was used as an enzyme source after being concentrated 50-fold by ultrafiltration (Amicon, Danvers, MA).

**Truncation of FucTs III and V**—The current model of the protein domain structure of glycosyltransferases indicates that a significant portion of the N terminus of these proteins is not involved in the binding of substrates or the catalytic process (4, 5). Therefore, our first goal was to determine what portion of the N terminus of FucT III and V could be eliminated without loss of activity. In a previous study we demonstrated that soluble forms of FucT III and V missing their first 51 and 62 amino acids, respectively, retained activity. To more precisely define the number of N-terminal amino acids that are required for enzyme activity a PCR based approach was used to create truncated forms of FucT V that lacked larger portions of its N-terminal amino acid sequence. These constructs were pre-

**RESULTS**

Truncation of FucTs III and V—The current model of the protein domain structure of glycosyltransferases indicates that a significant portion of the N terminus of these proteins is not involved in the binding of substrates or the catalytic process (4, 5). Therefore, our first goal was to determine what portion of the N terminus of FucT III and V could be eliminated without loss of activity. In a previous study we demonstrated that soluble forms of FucT III and V missing their first 51 and 62 amino acids, respectively, retained activity. To more precisely define the number of N-terminal amino acids that are required for enzyme activity a PCR based approach was used to create truncated forms of FucT V that lacked larger portions of its N-terminal amino acid sequence. These constructs were pre-

**Fig. 1. A comparison of the predicted amino acid sequences of FucTs III and V.** Sequences are those previously reported (7, 8).
pared as protein A chimeras and were assayed after their isolation by affinity chromatography on IgG-agarose beads as described under “Experimental Procedures.” As shown in Table II, several truncated forms of FucT V were prepared by PCR and each assayed for enzyme activity.

The results presented in Table II demonstrate that the truncated forms of FucT V with ≥299 amino acids (aa 76–374) retained activity; shorter forms were inactive. The studies shown in Table II were done with a disaccharide (type 2) acceptor which is known to be a poorer acceptor than the corresponding H-type structure. To rule out the possibility that the inactive proteins had simply lost the ability to utilize a single acceptor, each chimeric protein identified as inactive with the disaccharide acceptor was tested with several acceptors (results not shown). Only one protein (i.e., FucT V with a 76-amino acid deletion at the N terminus aa 77–374) had activity with any of the acceptors. This enzyme was minimally active (−5% as active as the longer enzyme forms) with an H-type 2 acceptor. Finally, the other active forms of FucT V were analyzed with a range of acceptor substrates and found to have an acceptor substrate specificity similar to the full-length enzyme as previously reported (9) (not shown).

Based on the FucT V results, a more limited analysis of FucT III constructs was done. Thus, a FucT III containing amino acids 62–361 had an activity equivalent to that obtained with forms of the enzyme previously characterized (9), whereas a shorter form (aa 67–361) was inactive. Therefore, a form of FucT III with ≥300 amino acids (aa 62–361) retained activity. Assays with various acceptors demonstrated that these forms of FucT III had a substrate specificity identical to that previously reported (7, 9) (not shown).

The current model of the protein domain structure of glycosyltransferases indicates that the C-terminal portion of a glycosyltransferase constitutes the catalytic domain, but little is known about the importance of amino acids at the C terminus for catalytic activity. To investigate this issue, FucT V proteins were prepared which lack one or two of the C-terminal amino acids of the full-length enzyme. The results presented in Table II demonstrate that removal of one amino acid from the C terminus of FucT V drastically alters catalytic activity. A protein missing two of the C-terminal amino acids was inactive, even when tested with an H-type 2 acceptor.

It is possible that the lack of detectable enzyme activity for the shorter FucT III and V constructs is due to the fact that the COS cells do not secrete these forms of the proteins, or that these forms are rapidly degraded. To rule out these possibilities the medium from COS cells, transfected with plasmids containing inserts encoding various FucT constructs, was mixed with IgG-agarose beads and the bound proteins were analyzed on Western blots. Fig. 2 shows that the inactive constructs (Table II) were produced and secreted into the medium. Furthermore, the relative amounts of inactive chimeric proteins were similar to that of the active proteins. Finally, these proteins appear to have the expected molecular weight and thus, inactivity does not appear to be due to proteolytic degradation.

Protein A–FucTs versus non-Protein A–FucTs—The truncation studies just described were done with chimeric forms of the FucTs which contained a protein A, IgG binding domain at their N terminus. To investigate whether this N-terminal mod-
The following concentrations were used for each acceptor: 1 and 4, 5 mM; 2, 3, 5, and 6, 1 mM. R1, O-(CH2)8-COOCH3. NT, not tested. A dash indicates an activity ≤ 10%.

Table III

| Substrate | A-FucT III | FucT III | A-FucT V | FucT V |
|-----------|------------|----------|----------|-------|
| Gal1β1,4GlcNAc 1 | — | NT | 40 | 70 |
| Fucα1,2Galβ1,4GlcNAc1-R1, 2 | — | — | 100 | 100 |
| NeuAcα2,3Galβ1,4GlcNAc1-R1, 3 | — | — | 25 | 15 |
| Galβ1,3GlcNAc 4 | 35 | 40 | — | — |
| Fucα1,2Galβ1,3GlcNAc1-R1, 5 | 100 | 100 | 35 | 60 |
| NeuAcα2,3Galβ1,3GlcNAc1-R1, 6 | 90 | 70 | — | — |

The following concentrations were used for each acceptor: 1 and 4, 5 mM; 2, 3, 5, and 6, 1 mM. R1, O-(CH2)8-COOCH3. NT, not tested. A dash indicates an activity ≤ 10%.

Table III

| Substrate | A-FucT III | FucT III | A-FucT V | FucT V |
|-----------|------------|----------|----------|-------|
| Gal1β1,4GlcNAc 1 | — | NT | 40 | 70 |
| Fucα1,2Galβ1,4GlcNAc1-R1, 2 | — | — | 100 | 100 |
| NeuAcα2,3Galβ1,4GlcNAc1-R1, 3 | — | — | 25 | 15 |
| Galβ1,3GlcNAc 4 | 35 | 40 | — | — |
| Fucα1,2Galβ1,3GlcNAc1-R1, 5 | 100 | 100 | 35 | 60 |
| NeuAcα2,3Galβ1,3GlcNAc1-R1, 6 | 90 | 70 | — | — |

Fig. 2. Western blot analysis of protein A-FucT chimeric protein sequences separated by SDS-PAGE. Proteins were purified by affinity chromatography on IgG-agarose, separated by SDS-PAGE, transferred to nitrocellulose, and detected with an alkaline phosphatase detection system as described under "Experimental Procedures." Lane 1, FT III (62–361); Lane 2, FT III (67–361); Lane 3, FT V (77–374); Lane 4, FT V (77–374); Lane 5, FT V (78–374); Lane 6, FT V (76–372); Lane 7, FT V (76–373).

Fig. 3. A comparison of the predicted amino acid sequences of the N-truncated forms of FucTs III (62–361) and V (76–374). Sequences are based on those previously reported (7, 8). Underlined amino acid residues in FucT III are those found to differ (i.e. L to F and I to A) in the constructs described in this report compared to those previously reported. The position of Cys143 and Cys150 of FucT III and V, respectively, is shown in outline type.
evidence that regions near Lys341 and Lys351 of bovine galactosyltransferase (14) have used a chemical modification approach to obtain nucleotide sugar binding site. For example, Yadav and Brew formation on amino acid residues that may form part of the substrate binding and catalytic sites of glycosyltransferases. The binding. Aoki et al. (16) used a site-directed mutagenesis approach to obtain evidence that Cys340 is involved in UDP-Gal binding by FucT V. This led us to carry out a series of domain swapping experiments. The results of the domain swapping experiments allow us to conclude that: (i) the single amino acid differences at the C terminus of FucT III and V do not affect acceptor substrate specificity, (ii) a protein with either region 1 or 2 of FucT III has a type 1 acceptor specificity, (iii) the combination of region 1 of FucT III with region 2 of FucT V produces an enzyme with both type 1 and type 2 acceptor specificity, and (iv) a protein containing amino acids in region 1 of FucT V does not have a type 2 acceptor specificity, whereas one containing the 12 amino acids of region 2 of FucT V does. Taken together these results demonstrate that the amino acids in regions 1 and 2 of FucT III and V are critically involved in defining the acceptor substrate specificity of these two enzymes.

The recognition of complex oligosaccharides by proteins, including lectins, antibodies, and enzymes is accomplished primarily through interactions with particular hydroxyl groups on the carbohydrate, but van der Waals interactions also occur in most cases. The interactions occur through hydrogen bonds between the sugar hydroxyls and side chains of amino acids. In a recent study we evaluated the ability of FucTs III and V to utilize several modified forms (deoxygenterated and containing modified amino acids) of the GlcNAc residue of type 1 and type 2 disaccharides as acceptors. An important result from this analysis was that both enzymes had an absolute requirement for a hydroxyl group at carbon C-6 of galactose. The minimum energy conformations of type 1 and 2 disaccharides have been recognized to influence the affinity of the α2,6-sialyltransferase for CMP-sialic acid (18). Finally, we (19) have recently completed a study that demonstrates that Cys340 and Cys354 of FucT III and V, respectively, are in or near the binding site for GDP-Fuc. Our current efforts are directed at more precisely defining the amino acids that affect acceptor specificity.

The protein domain structure proposed several years ago for glycosyltransferases indicates that the C-terminal portion of these enzymes contains their catalytic domain. The results we have obtained in this study demonstrate that about 20% of the N-terminal amino acids of the FucT III and V sequences are not required for enzyme activity and therefore, constitute the other three recognized protein domains (i.e., cytoplasmic, transmembrane, and stem). In contrast, truncation of the C terminus of these enzymes results in their inactivation. We have reported similar results for two forms of α1,3-galactosyltransferase (6). In all of these cases approximately 300 amino acids have been found to be required for enzyme activity. However, this does not appear to be the minimum length required by glycosyltransferases since forms of β1-4 galactosyltransferase missing 127 out of 400 amino acids are active (15, 16).

Our truncation studies also demonstrate that the two amino acid segments that occur in FucT V which make it 13 amino acids longer than FucT III are located between the transmembrane domain and the catalytic domain (i.e., in the stem region). Thus, FucT V has a stem region that is 42 amino acids long, whereas FucT III's stem region is approximately 13 amino acids shorter. This result is reminiscent of the observation made by Joziasse et al. (20) that mice produce three forms of α1,3-galactosyltransferase which differ only in the length of their stem regions.

The most important result of the truncation study was that active forms of FucT III and V only differ at 23 out of about 300 amino acid residues. Since these enzymes have distinct acceptor specificities, it allowed us to conclude that some or all of the amino acid differences occurring between the two enzymes must account for their distinct acceptor specificities. This led us to carry out a series of domain swapping experiments. The results of the domain swapping experiments allow us to conclude that: (i) the single amino acid differences at the C terminus of FucT III and V do not affect acceptor substrate specificity, (ii) a protein with either region 1 or 2 of FucT III has a type 1 acceptor specificity, (iii) the combination of region 1 of FucT III with region 2 of FucT V produces an enzyme with both type 1 and type 2 acceptor specificity, and (iv) a protein containing amino acids in region 1 of FucT V does not have a type 2 acceptor specificity, whereas one containing the 12 amino acids of region 2 of FucT V does. Taken together these results demonstrate that the amino acids in regions 1 and 2 of FucT III and V are critically involved in defining the acceptor substrate specificity of these two enzymes.

The recognition of complex oligosaccharides by proteins, including lectins, antibodies, and enzymes is accomplished primarily through interactions with particular hydroxyl groups on the carbohydrate, but van der Waals interactions also occur in most cases primarily through stacking of the underface of pyranose residues with aromatic amino acids (21). Many of the interactions occur through hydrogen bonds between the sugar hydroxyls and side chains of amino acids. In a recent study we evaluated the ability of FucTs III and V to utilize several modified forms (deoxygenterated and containing modified amino acids) of the GlcNAc residue of type 1 and type 2 disaccharides as acceptors. An important result from this analysis was that both enzymes had an absolute requirement for a hydroxyl group at carbon C-6 of galactose. The minimum energy conformations of type 1 and 2 disaccharides have been recognized to...
have different molecular topographies. Since a correctly oriented Gal residue appears to be essential for enzyme activity (based on the absolute requirement of its C6-hydroxyl group), the minimum energy conformations of type 1 and 2 structures, in effect, invert the relative orientation of the Gal and GlcNAc residues of the disaccharides by approximately 180°. Therefore, the positions of the NHAc and CH2OH groups of the GlcNAc residues are effectively interchanged. The present study suggests that the active site of FuCT III and V forms a pocket capable of discriminating between OH-6 of the Gal residue and either the NHAc or CH2OH group of the GlcNAc residue, and this differentiation is realized by interaction with amino acids in the N-terminal domain of these enzymes.

Another interesting result from the domain swapping experiments was that the amino acids (Asp\textsuperscript{336} and Ala\textsuperscript{349} of FuCT III and V, respectively) that differ near the C terminus of FuCTs III and V did not alter their activity or acceptor specificity compared to the parent enzymes. Based on a report by Nishihara et al. (22) we had anticipated that the first domain swap shown in Fig. 4 would produce an inactive protein. These workers (22) had reported that the coding region of FuCT III of some Lewis negative individuals contained a single nucleotide base change that resulted in an Asp\textsuperscript{336} → Ala mutation. This would produce a FuCT III that contained a catalytic region similar to the first domain swap protein shown in Fig. 4. Thus, we had predicted that this domain swapped protein would be inactive. Recently, the same research group reported that their original sequencing results were incorrect and that the actual mutation in these Lewis negative individuals is His\textsuperscript{256} → Lys (23).

DNA sequencing of the FuCT III and V revealed that all of the FuCT V constructs had a sequence that corresponded to that previously reported for the full-length enzyme, whereas all of the FuCT III constructs contained a nucleotide base difference compared to the sequence previously reported for the full-length enzyme. These differences represent natural variations in the DNA sequence obtained from different sources of DNA. Our original template for cloning FuCT III was human placental DNA, whereas the source of template for the original report on FuCT III was the human tumor cell line A431 (7). Regardless of the origin, the resulting changes in the amino acid sequence for the FuCT III proteins prepared for our studies did not have a major effect on enzyme activity or acceptor substrate specificity. This is in contrast to other single amino acid substitutions detected in Lewis negative individuals (22–27). Furthermore, the domain swap construct that had an altered acceptor substrate specificity did not contain these amino acids and thus, these amino acids do not seem to be involved in acceptor substrate recognition.

During the review of our manuscript, Legault et al. (28) published a study which also demonstrated that a discrete peptide fragment within α1,3,4-fucosyltransferases is responsible for discriminating among different oligosaccharide acceptor substrates. They identified a so-called “hypervariable region” in the fucosyltransferases that contains as few as 11 amino acids and participates in the binding of type I acceptor substrates. This area is very near the region we have found to affect acceptor substrate specificity. In contrast to the work presented here, Legault et al. (28) utilized full-length constructs of the enzymes and relied largely on cell surface staining with antibodies to various type I and II carbohydrate epitopes to analyze the effect of swapping different domains between fucosyltransferases. In spite of the differences between our approach and those of Legault et al., (28), the conclusions drawn are similar. This adds strength to the concept that a small peptide region at the N terminus of the enzymes’ sequence-constant C terminus is critical for determining acceptor substrate specificity. Future studies will determine which amino acids in this region are critical for substrate recognition.

The results presented here offer some useful insights into the active site of glycosyltransferases. Identification of the amino acids that control acceptor substrate recognition will refine the domain structure that has defined glycosyltransferases for several years. Since several glycosyltransferases recognize either a type 1 or, more often, a type 2 acceptor it will be interesting to determine if a common set of amino acid residues can be defined among a group of glycosyltransferases that have similar acceptor substrate specificities.

Acknowledgments—We appreciate the helpful discussions and input of Drs. Cheryl Stults, Ole Hindsgaul, and Theodora de Vries and Eric Holmgren. We also thank Dr. Kurt Drickamer for providing the plasmid pgIR-201 and Marie Mardinko for help with the Western blot analyses.

REFERENCES

1. Macher, B. A., Holmes, E. H., Swiedler, S. J., Stults, C. L. M., and Srnka, C. A. (1991) Glycobiology 1, 577–584.
2. de Vries, T., and van den Eijnden, D. H. (1992) Histochem. J. 24, 761–770.
3. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.
4. Holmes, E. H., Xu, Z., Sherwood, A., and Macher, B. A. (1994) J. Biol. Chem. 269, 5334–5341.
5. Holmes, E. H., Xu, Z., Sherwood, A., and Macher, B. A. (1995) J. Biol. Chem. 270, 8145–8151.