The EMBO Journal vol.9 no.10 pp.3171 – 3178, 1990

Analysis of the substrate binding sites of human galactosyltransferase by protein engineering

Daisuke Aoki, Hubert E.Appert[1], Dennis Johnson[2,3], Shan S.Wong[2] and Michiko N.Fukuda

La Jolla Cancer Research Foundation, La Jolla, CA 92037, USA, [1]Department of Surgery, Medical College of Ohio at Toledo, Toledo, OH 43699, USA and [2]Department of Chemistry, University of Lowell, Lowell, MA 01854, USA

[3]Present address: School of Medicine, Ohio State University, Columbus, OH 43210, USA

Communicated by C.G.Gahmberg

An expression vector, pIN-GT, encoding the soluble form of β1,4-galactosyltransferase (GT) has been constructed from human GT cDNAs and the pIN-III-*ompA*$_2$ expression vector. *Escherichia coli* strain SB221 harboring the pIN-GT plasmid produces and secretes a fusion protein consisting of the *omp*A signal and GT. The expression of GT was detected by assaying enzymatic activity as well as by Western blotting using anti-GT antibodies. The recombinant GT was purified to homogeneity by *N*-acetylglucosamine–Sepharose affinity chromatography. The NH_2-terminal peptide sequence of purified GT confirmed the cleavage site of the fusion protein by bacterial signal peptidase. This expression system was utilized to produce mutant forms of GT in order to identify specific amino acids involved in substrate binding sites. Photoaffinity labeling of GT with UDP-galactose analog, 4-azido-2-nitrophenyluridylylpyrophosphate (ANUP), followed by cyanogen bromide (CNBr) cleavage revealed that ANUP bound to a fragment of GT composed of amino acid residues from Asp276 to Met328. Within this peptide segment, Tyr284, Tyr287, Tyr309, Trp310 and Trp312 were separately substituted into Gly and Tyr287 into Phe by site-directed mutagenesis. Enzymatic activity assay showed drastic reduction of the activity in all of the mutants except that Tyr287→Phe remained as active as wild-type GT. Kinetic studies of the mutated GT showed that Tyr284, Tyr309 and Trp310 are critically involved in the *N*-acetyglucosamine binding and Tyr309 is involved in UDP-galactose binding as well. These results indicate that these tyrosines and tryptophans in GT are essential for the binding of the acceptor *N*-acetylglucosamine, and that UDP-galactose also binds to residue(s) nearby where *N*-acetylglucosamine binds.

Key words: affinity labeling/enzyme kinetics/galactosyltransferase/glycoprotein/mutations

Introduction

β1,4-galactosyltransferase (GT[1], EC 2.4.1.22) is an enzyme that transfers galactose from UDP-galactose to terminal *N*-acetylglucosamine in glycoproteins and glycolipids (Ram and Munjal, 1985). GT is primarily present in the *trans*-cisternae of the Golgi complex in a membane-bound form (Roth and Berger, 1982; Smith and Brew, 1977; Strous *et al.*, 1985). The enzyme also exists in a soluble form in body fluids such as milk, colostrum and serum. The soluble form of the enzyme is derived from the membrane form by proteolytic cleavage (Smith and Brew, 1977; Strous *et al.*, 1985) at Arg77 (Masri *et al.*, 1988). The catalytic domain, or soluble form, of GT is composed of 323 amino acid residues. Presumably within this domain, GT has binding sites for its substrates, donor (UDP-galactose) and acceptor (*N*-acetylglucosamine-terminated glycoconjugates). Previous chemical modification studies of GT showed that cysteine (Magee and Ebner, 1974; Wong and Wong, 1984), tryptophan (Clymer *et al.*, 1976) and tyrosine (Geren *et al.*, 1975; Silvia and Ebner, 1984; Takase and Ebner, 1981, 1984) are involved in the GT–substrate interactions. However, the location of the amino acids involved in substrate binding could not be determined by the method used in previous investigations.

The cDNAs encoding GT have been isolated from human (Appert *et al.*, 1986a; Masri *et al.*, 1988), bovine (Narimatsu *et al.*, 1986; D'Agostaro *et al.*, 1989) and murine sources (Nakazawa *et al.*, 1988; Shaper *et al.*, 1988). The availability of GT cDNA opens a way to investigate structure and function relationships of this enzyme by recombinant DNA techniques. To accomplish such studies, we chose a bacterial expression system that does not naturally produce GT.

This report describes the production of enzymatically active recombinant GT in *Escherichia coli* harboring the GT expression vector and the identification of UDP-galactose and *N*-acetylglucosamine binding sites by site-directed mutagenesis.

Results

Construction of pIN-GT vector and production of recombinant GT in E.coli

An expression vector for GT has been made from cDNA encoding the soluble form (catalytic domain) of GT. The *Rsr*II site of the cDNA coincided with the proteolytic cleavage site used to produce the soluble form of the enzyme. A 995 bp cDNA *Rsr*II–*Hin*dIII fragment encoding the soluble form of the enzyme was inserted into the *Eco*RI and *Hin*dIII sites of pIN-III-*ompA*$_2$ to construct the pIN-GT expression vector (see Figure 1). *E.coli* SB221 cells were transformed by the pIN-GT vector, and expression of the fusion protein was induced by isopropyl-β-D-thiogalactoside (IPTG). In order to examine production of GT by the SB221 cells, the total cell lysate, the periplasmic fraction, and the culture medium were subjected to Western blotting analysis. The cell lysate of the transformed SB221 gave several bands which are reactive with anti-human GT antibodies (Figure 2, lane 1). Among these bands, the molecular weight of the largest component was estimated to be 36 kd. This value

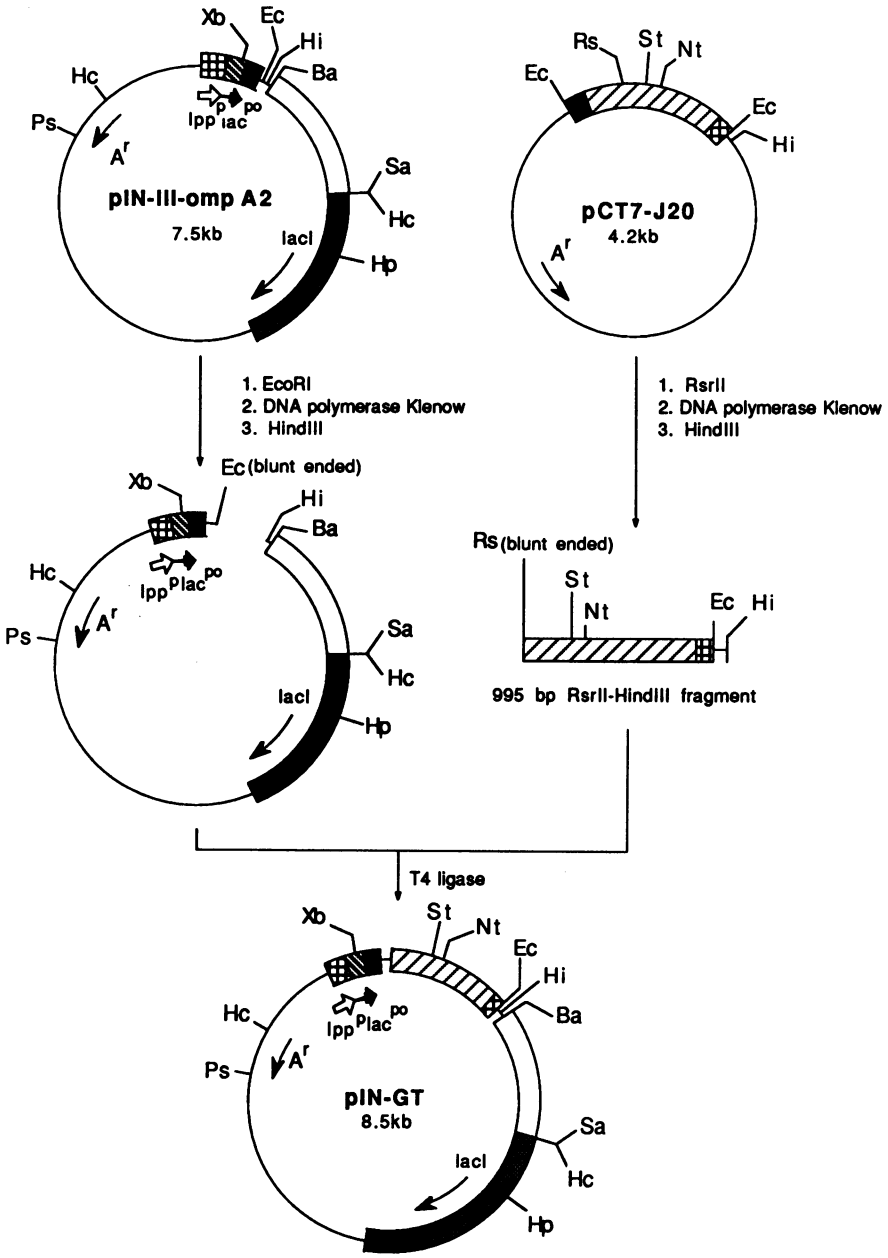


Fig. 1. Construction of pIN-GT vector. CT7−J20 (1286 bp) containing the full-length coding sequence (▨, 1200 bp) for human GT has been inserted at the *Eco*RI site of Bluescript. This plasmid, designated as pCT7−J20, was digested by *Rsr*II, treated with Klenow and digested with *Hin*dIII. A 995 bp fragment which consisted of the coding sequence of the soluble form of the enzyme, flanked by 13 bp of 3′ untranslated sequence and an extra 12 bp of polylinker sequence from the vector at the 3′ end was ligated at *Eco*RI (converted to blunt end) and *Hin*dIII sites of the pIN-III-*omp*A2 expression vector. To adjust the translation reading frame in order to produce an in-frame fusion molecule, pIN-III-*omp*A_2 was selected out of the pIN-III-*omp*A series (Ghrayeb *et al.*, 1984). Ec, *Eco*RI; Nt, *Not*I; Hi, *Hin*dIII; Rs, *Rsr*II; St, *Sst*I; Ba, *Bam*HI; Xb, *Xba*I; Hc, *Hinc*II; Sa, *Sal*I; Hp, *Hpa*I; Ps, *Pst*I site; lpp^p, *lpp* promoter; lac^{po}, *lac* promoter−operator; A^r, ampicillin resistance gene.

is consistent with 35.9 kd calculated from the 323 amino acids of the soluble form of the enzyme plus an extra three amino acids expected for the protein produced by the pIN-GT vector. Low molecular weight bands in lane 1 may be degradation products due to bacterial proteases. Recombinant GT was released from SB221 cells, as both the periplasmic fraction (lane 2) and the culture medium (lane 3) contained the major 36 kd immunoreactive bands. Cells transformed with the pIN-III-*omp*A_2 vector alone, lacking the GT insert, did not produce any immunoreactive proteins (Figure 2, lanes 4, 5 and 6).

Enzymatic activity of GT was detected in SB221 cells transformed with the vector containing the GT insert. Each 1 ml of bacterial culture yielded 0.29 and 0.50 munits GT activity in the culture medium upon induction with IPTG for 4 and 16 h, respectively. GT activity in the periplasmic fraction obtained from the same culture was 0.029 and 0.44 munits, respectively. Neither the periplasmic fraction nor culture medium obtained from SB221 cells transformed with the pIN-III-*omp*A_2 vector alone showed any GT activity.

Purification of the GT secreted from E.coli harboring pIN-GT

Since the transformed SB221 cells secreted a functionally active enzyme into the culture medium, this fact allowed us to use *N*-acetylglucosamine−Sepharose affinity chromato-

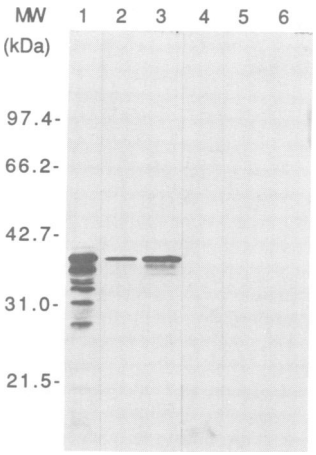


Fig. 2. Western blotting of the proteins produced by *E.coli* SB221 cells using anti-human GT antibodies. **Lane 1**, total cell lysate; **lane 2**, periplasmic fraction; **lane 3**, culture medium. These fractions were prepared from SB221 transformed with pIN-GT. **Lanes 4, 5** and **6** show total cell lysates, periplasmic fraction and medium which were obtained from cells harboring the pIN-III-*ompA*$_2$ plasmid with no insert.

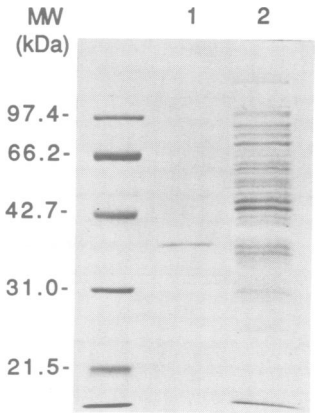


Fig. 3. Purification of the GT from the culture medium of *E.coli* SB221 harboring the pIN-GT plasmid. SDS−PAGE of the culture medium and purified enzyme is shown. The gel was stained with Coomassie blue. **Lane 1**: the enzyme purified from the culture medium of SB221 harboring pIN-GT by affinity chromatography of *N*-acetylglucosamine−Sepharose column (see Materials and methods for purification). **Lane 2**: the culture medium. Mol. wt markers are phosphorylase b (97.4 kd), bovine serum albumin (66.2 kd), ovalbumin (42.7 kd), carbonic anhydrase (31.0 kd) and soy bean trypsin inhibitor (21.5 kd).

graphy for purification of the recombinant enzyme. The affinity purified enzyme exhibited a single band having molecular weight of 36 kd on SDS−PAGE (Figure 3, lane 1). Furthermore, this band reacted with anti-human GT antibodies in Western blotting analysis (data not shown; see also Figure 2). The purified enzyme was subjected to amino acid sequencing by Edman degradation. Phenyl-thiohydantoin analysis showed that the NH_2-terminus of the purified protein consisted of Ala-Glu-Leu, which agrees with the amino acid sequence expected from cleavage of the GT fusion protein construct by bacterial signal peptidase (Figure 4).

Analysis of the oligosaccharides synthesized by the recombinant GT

In order to characterize the enzymatic reaction catalyzed by recombinant GT, the oligosaccharide products were analyzed by HPLC. The HPLC profile of the [^{3}H]oligosaccharides synthesized by the recombinant enzyme suggests the addition of [^{3}H]galactose to each acceptor saccharide (see Figure 5, panels A and C). The pattern obtained by using recombinant enzyme (panel C) resembled qualitatively as well as quantitatively the profile obtained by using human milk GT (panel B). Digestion of the oligosaccharide produced by the recombinant GT (hatched area in Figure 5B) with β-galactosidase from jack beans, which preferentially cleaves the Galβ1→4GlcNAc linkage (Kobata and Ginsburg, 1972), shows the release of monosaccharide representing [^{3}H]-galactose (panel D). These results indicate that galactose was transferred from UDP-galactose to the non-reducing terminal *N*-acetylglucosamine through a β1→4 linkage, and that the recombinant enzyme has specificity identical to the naturally produced GT.

Photoaffinity labeling of GT by ANUP

In order to identify the UDP-galactose binding site in GT, photoaffinity labeling of GT with a UDP-galactose analog, 4-azido-2-nitrophenyluridylylpyrophosphate (ANUP) (Lee *et al.*, 1983), was performed. A mixture of GT and [^{14}C]ANUP was exposed under UV light, so that ANUP

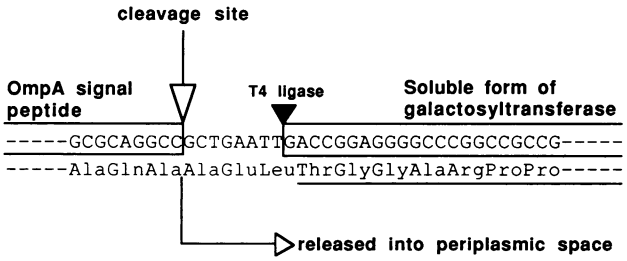


Fig. 4. Nucleotide sequence and corresponding amino acid sequence at the junction between the *ompA* signal peptide and the soluble form of GT in pIN-GT. The arrow indicates the cleavage site of the *ompA* signal peptide. The authentic soluble form of GT is indicated by underlining. The amino acids Ala−Glu represent extra amino acids derived from the polylinker region of pIN-III-*ompA*$_2$; the following Leu is encoded by a codon which consists of nucleotides TT from the pIN-III-*ompA*$_2$ and nucleotide G from the GT cDNA.

covalently linked to GT at the UDP-galactose binding site through photoexcitation. ^{14}C-labeled GT was then cleaved by CNBr. As shown in Figure 6, the peptide fragments were separated by HPLC into 14 fractions, and most of the radioactivity was found in fractions 10 and 11. Amino acid composition of fraction 11 (Table I) agrees with the composition of the peptide derived from Asp276−Met328, which was predicted from the nucleotide sequence of GT cDNA (Masri *et al.*, 1988, see Figure 7). Amino acid composition of fraction 10 was not assignable to any predicted GT fragment. It was, however, reasonable to infer that this fragment resulted from the partial CNBr cleavage of fraction 11, because studies done on the unlabeled enzyme showed that fractions 10 and 11 had the same NH_2-terminal amino acid sequence.

Expression of mutated GT in E.coli

Previous chemical modification studies of GT have suggested the involvement of cysteine, tyrosine and tryptophan in substrate binding (Clymer *et al.*, 1976; Geren *et al.*, 1975; Magee and Ebner, 1974; Silvia and Ebner, 1984; Takase and Ebner, 1981; Wong and Wong, 1984). The peptide

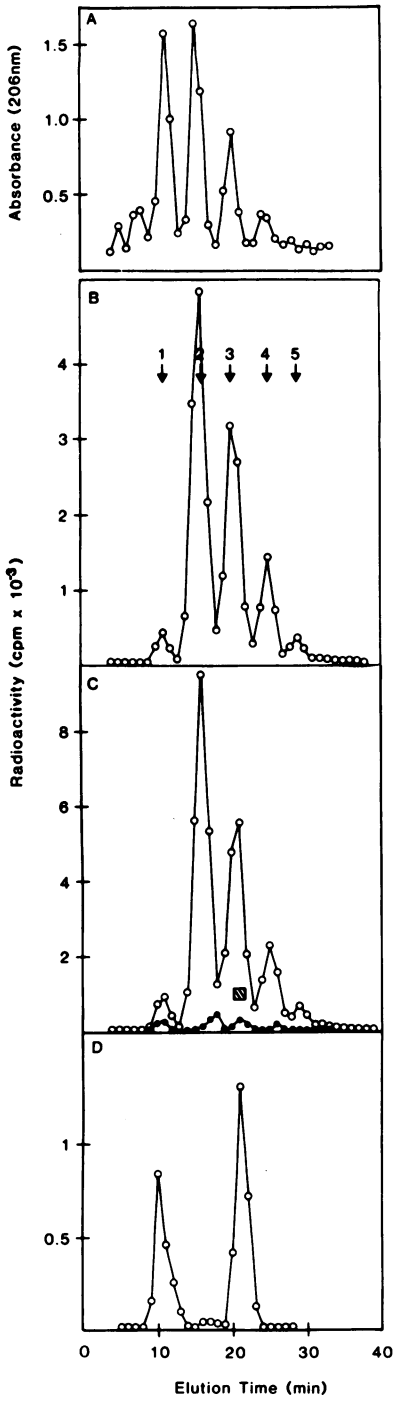
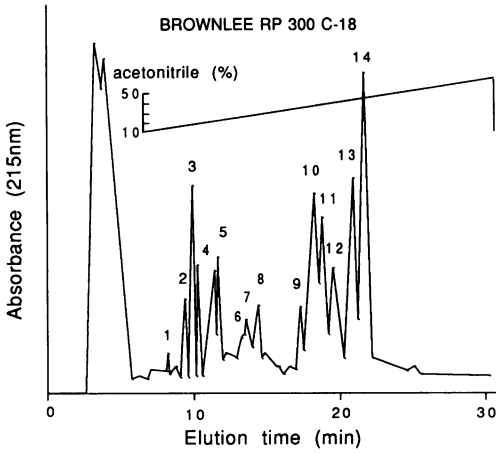


Fig. 5. HPLC profiles of oligosaccharides synthesized from UDP-[^{3}H]galactose and acceptor oligosaccharides by the action of recombinant GT. **Panel A**: HPLC profile of *N*-acetylglucosamine and chitooligosaccharides which were used as acceptors for the enzyme reaction. Absorbance at 206 nm was monitored. **Panel B**: HPLC profile of the oligosaccharide product synthesized by the GT purified from human milk. **Panel C**: HPLC profile of the oligosaccharide products synthesized by the recombinant enzyme. The recombinant enzyme and UDP-[^{3}H]galactose were incubated with (−○−) and without (−●−) acceptor. **Panel D**: HPLC profile of the digest of [^{3}H]oligosaccharide product with jack bean β-galactosidase. The material shown in the panel C hatched bar (2×10^4 c.p.m.) was treated with 30 munits of β-galactosidase (Sigma) at 37°C for 2 h in 0.1 M Na-citrate buffer pH 4.0. Arrow 1 indicates the elution position of [^{3}H]galactose used as a standard. Arrow 2 indicates the position of [^{3}H]Galβ1→4GlcNAc; 3, 4 and 5 indicate the elution positions of Galβ1→4(GlcNAc)$_2$, Galβ1→4(GlcNAc)$_3$ and Galβ1→4(GlcNAc)$_4$, respectively.



peak	cpm
1	100
2	55
3	100
4	85
5	145
6	170
7	130
8	105
9	375
10	1580
11	930
12	295
13	320
14	325

Fig. 6. HPLC purification of CNBr fragments of [^{14}C]ANUP-labeled GT. Human milk GT was labeled with [^{14}C]ANUP and then cleaved by CNBr. Peptides were separated by a reverse-phase HPLC column Brownlee RP-300 C-18. Radioactivity in each fraction is shown in the Table.

Table I. Amino acid composition of GT peptide (Fraction 11) obtained by CNBr cleavage

Amino acids	% Composition	Number of residues per peptide (estimated from % composition)	Number of residues from Asp276 to Met328 (deduced from cDNA sequencing)
Asx	15.18	7	8
Thr	3.4	1	1
Ser	7.56	3	3
Glx	9.66	4	4
Pro	5.96	2	2
Gly	15.05	7	8
Ala	3.94	1	1
1/2-Cys	–	–	0
Val	6.54	3	3
Met	–	–	1
Ileu	3.61	1	2
Leu	6.17	3	4
Tyr	6.09	3	3
Phe	7.22	3	7
His	1.05	0	0
Lys	3.7	1	2
Arg	4.85	2	2
Trp	–	–	2

fragment which was labeled by [^{14}C]ANUP contains three tyrosines and two tryptophans but no cysteine (Figure 7). To determine which tyrosine and tryptophan residues in this sequence are involved in the enzymatic activity, six mutants of GT were prepared (Table II). In each mutant, either tyrosine or tryptophan was replaced in turn by phenylalanine or glycine, and each mutant placed in pIN expression vector. *E.coli* SB221 cells were transformed by the pIN-GT plasmid containing the mutations. Western blotting demonstrated that *E.coli* SB221 harboring the mutated GT plasmids produced 36 kd protein reactive with anti-GT antibodies in cell lysates and in the culture media (data not shown).

Enzyme activity assays of each culture medium showed that all of the mutated GT, except the Y287F mutant, demonstrated dramatically decreased GT activity. Y287F was active at wild-type levels, suggesting that the inactivation of the other GT mutants must be caused by the replace-

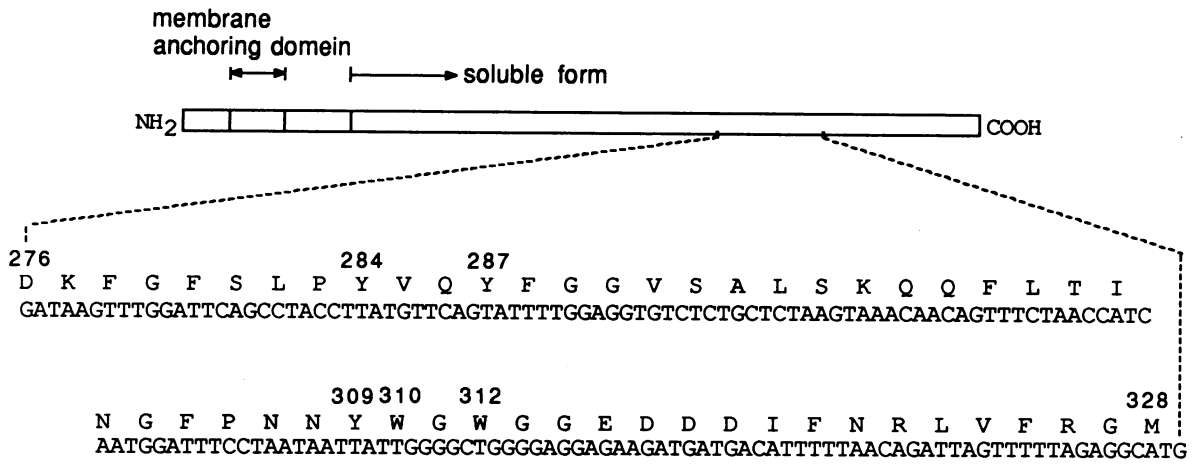


Fig. 7. Amino acid sequence of the CNBr peptide (Asp276–Met328) labeled with [^{14}C]ANUP. Upper bar represents the full-length of human β1,4-galactosyltransferase composed of 400 amino acid residues. Amino acid sequence (symbolized) of the CNBr peptide (Asp276–Met328) labeled with [^{14}C]ANUP and its corresponding nucleotide sequence (Masri *et al.*, 1988) are shown below. The numbered amino acid, Tyr(Y)287, Tyr(Y)309, Trp(W)310 and Trp(W)312 show mutation sites in the present study.

ment of each tyrosine or tryptophan residue by glycine, and not by an unplanned artifact.

In the cases of Y284G, Y309G and W310G, measurable activity could not be obtained, unless high concentrations of acceptor chitobiose were used. The concentrations required to detect GT activity in these mutants caused substrate inhibition in the expressed wild-type GT. This fact does not allow us to compare directly the apparent activity between wild-type and mutated GTs under the same conditions. In order to assay the GT mutants, the concentration of chitobiose and UDP-galactose was increased, but two other mutants, Y287G and W312G, still showed activity too low to perform the kinetic study.

Kinetic studies of mutated GT

To examine whether lowered GT activity of the mutants is due to low affinity for the substrates, kinetic analysis was carried out according to the previously reported kinetic studies on naturally produced GT (Khatra *et al.*, 1974; Bell *et al.*, 1976). As shown in Figure 8, the reciprocal of the velocity versus the reciprocal of varied chitobiose concentrations were plotted at several fixed concentrations of UDP-galactose.

From each double-reciprocal plot, the intercepts and slopes were replotted against the reciprocal of UDP-galactose concentrations. The kinetic constants for UDP-galactose and chitobiose were then calculated and are given in Table II. The dissociation constants and Michaelis constants for the wild-type are consistent with those reported previously for naturally produced GT (Khatra *et al.*, 1974; Bell *et al.*, 1976). In comparison, the K_a of Y284G is only ~3-fold and its K_{ia} is also within the same range, indicating that the Y284G mutant has almost the same affinity for UDP-galactose. However, K_b and K_{ib} of Y284G are increased 1000-fold, indicating that this mutant has a drastic reduction in its affinity for chitobiose. On the other hand, the K_a and K_{ia} values of Y309G are 30-fold larger than those of the wild-type, suggesting that Y309G has reduced affinity for UDP-galactose. The mutant Y309G also showed increased values of K_b and K_{ib}, indicating that Y309G has reduced affinity for chitobiose. W310G mutant also has reduced affinity for chitobiose as its K_b and K_{ib} are 300 times less than those of the wild-type. These results suggest that Tyr284, Tyr309 and Trp310 are critically involved in the binding of *N*-acetylglucosamine. Furthermore, Tyr309 is likely involved in the UDP-galactose binding.

Table II. Kinetic constants of GT and mutated GTs

	Kinetic constants (mM)			
	UDP-galactose		Chitobiose	
	K_a[a]	K_{ia}	K_b	K_{ib}
Wild-type	0.15	0.11	0.35	0.26
Y284G[b]	0.43	0.17	316	122
Y287G	ND[c]	ND	ND	ND
Y309G	5.3	2.7	376	196
Y310G	0.51	0.38	97.6	70.9
W312G	ND	ND	ND	ND
Y287F	0.15	0.075	0.34	0.17

[a] K_a and K_b are the Michaelis constants for UDP-galactose and chitobiose, respectively; and K_{ia} and K_{ib} are dissociation constants for UDP-galactose and chitobiose from E–Mn^{2+}–UDP-galactose and E–Mn^{2+}–chitobiose complexes, respectively.
[b] A GT mutant which has replacement of tyrosine (Y)284 by glycine (G) is designated as Y284G. Following mutants are designated in the same way.
[c] ND: GT activity is too low to determine the kinetic constants.

Discussion

The present paper describes construction of an expression vector for human GT and experiments which demonstrate the production of enzymatically active GT in *E.coli* and identify the substrate binding sites in GT.

Expression of GT was detected in the periplasmic fraction and in the culture medium of *E.coli* harboring the pIN-GT vector by Western blotting (Figure 2) as well as GT activity assay. The recombinant GT was purified to homogeneity from the culture medium by one-step affinity chromatography using an *N*-acetylglucosamine–Sepharose column (Figure 3).

Successful expression of biologically active human proteins using the pIN-III-*omp*A vectors has been reported for growth hormone (Hsiung *et al.*, 1986), superoxide dismutase (Takahara *et al.*, 1988) and macrophage colony-stimulating

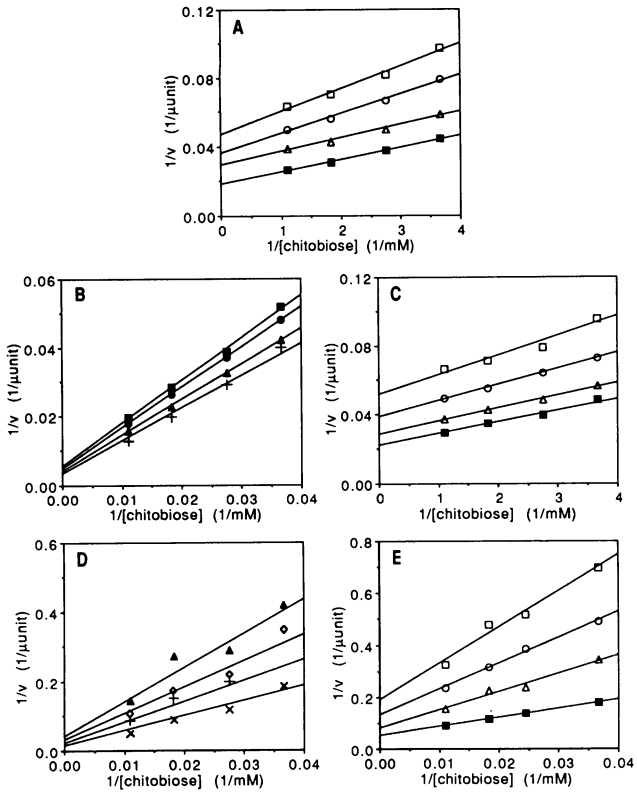


Fig. 8. Kinetic analysis of GT and its mutated forms. GT activity measured at the varied concentrations of chitobiose is expressed in double-reciprocal plots at the several fixed concentrations of UDP-galactose of: open squares, 0.07315 mM; open circles, 0.11 mM; open triangles, 0.22 mM; closed squares, 0.44 mM; closed circles, 0.66 mM; closed triangles, 1.32 mM; open diamonds, 2.0625 mM; +, 3.3 mM; and X, 8.25 mM. **Panel A**, wild-type (unmutated form); **panel B**, Y284G; **panel C**, Y287F; **panel D**, Y309G; and **panel E**, W310G.

factor (Libby *et al.*, 1987). In the pIN-III-*omp*A vectors (Ghrayeb *et al.*, 1984), the cloned foreign gene is inserted immediately after the *omp*A signal peptide sequence, which is under the control of both the *lpp* promoter and the *lac* promoter. Since the *lac* promoter fragment also contains the operator region, the cloned gene expression can be induced by a *lac* inducer such as IPTG. Upon gene expression, the fusion protein of the *omp*A signal peptide and foreign gene product should be targeted to reach the outer membrane, and the NH_2-terminal *omp*A signal peptide should be cleaved off by bacterial signal peptidase, which allows the product protein to be released into the periplasmic space, and further, on cell lysis, into the culture medium (Takagi *et al.*, 1988). The correct disulfide bonds and the proper secondary structure of secreted human growth hormone indicates that the *E.coli* periplasm can provide an environment which facilitates efficient disulfide bond formation and proper folding of the protein (Hsiung *et al.*, 1986). Production of enzymatically active GT utilizing the pIN-III-*omp*A vector adds an another example for successful folding using this system.

In order to produce enzymatically active GT, we removed a cDNA segment encoding the membrane-anchoring domain from a full-length of GT cDNA for the following reasons. First, it was assumed that the NH_2-terminal peptide region up to the proteolytic cleavage site is not necessary for the catalytic function of GT, because the soluble form of the enzyme has complete catalytic activity. Second, if a fusion protein contains a membrane-anchoring domain in addition to the *omp*A signal peptide, the COOH-terminal region of GT most likely will remain in the host cell's cytoplasm and will be degraded by bacterial proteases. The soluble form of GT is expected to be found in the periplasm, since the *omp*A signal peptide is cleaved by bacterial signal peptidase (Ghrayeb *et al.*, 1984).

Very little is known about the structure and function relationships of glycosyltransferases, despite the fact that they are one of the largest groups of functionally related enzymes. So far, GT, α2,6-sialyltransferase, α1,3-galactosyltransferase and α1,3-*N*-acetylgalactosaminyltransferase have been cloned and sequenced (Appert *et al.*, 1986b; Joziasse *et al.*, 1989; Masri *et al.*, 1988; Nakazawa *et al.*, 1988; Narimatsu *et al.*, 1986; Shaper *et al.*, 1988; Weinstein *et al.*, 1987; Yamamoto *et al.*, 1990). There are common structural features among these glycosyltransferases (Paulson and Colley, 1989): they are type II integral membrane proteins, with a short NH_2-terminal cytoplasmic tail followed by an uncleaved combination signal–membrane anchoring sequence, a short stem region and a large catalytic domain including the COOH-terminus. There is, however, no similarity in nucleotide sequence or amino acid sequence between these glycosyltransferases and it is therefore difficult to predict which peptide segment is performing certain functions such as binding to nucleotide sugars.

Previous enzyme kinetics and chemical modification studies on GT had limitations with respect to their ability to identify the substrate binding sites (Silvia and Ebner, 1984; Takase and Ebner, 1984; Wong and Wong, 1984). In this context, it should be noted that photoaffinity labeling by UDP-sugar analog is indispensable for localizing the peptide segment that includes substrate binding sites (Figures 6 and 7 and Table I). The expression vector system described in this report has a potential for defining more precisely the critical parts of the enzyme where substrate interactions occur by employing site-directed mutagenesis.

In the present study, we demonstrated that the mutation of Tyr284, Tyr309 and Trp310 into Gly greatly reduces the affinity for chitobiose and the Tyr309 mutation also reduced the affinity for UDP-galactose (Figure 8, Table II). This evidence together with photoaffinity labeling data (Figure 6 and Table I), strongly suggest that substrate binding sites are within the identified segment (Figure 7), and the mutated tyrosines and tryptophan are essential for the binding of *N*-acetylglucosamine and/or UDP-galactose. Topologically close binding of donor and acceptor as shown here seems reasonable as GT catalyzes the transfer of galactose from donor to acceptor. These data suggest the possibility that Trp249–Trp250–Tyr251 and nearby peptides in α1→3-galactosyltransferase (Joziasse *et al.*, 1989) are involved in substrate binding.

The expression system utilized allowed us to detect enzymatic activity which has been reduced more than 1000 times by mutation, because of the absolute absence of background activity in host cells. Further study on GT expressed in *E.coli* will be expected to identify the catalytic site, the Mn^{2+} binding site, and residues interacting with α-lactalbumin.

Materials and methods

Cells and plasmids

E.coli SB221 (*hsdR, leuB6, lacY, thi, trpE5, lpp/F', lacI, proAG, lacZYA*) (Coleman *et al.*, 1985) was the host strain for the recombinant plasmids.

The plasmid pIN-III-*ompA*$_2$ (Ghrayeb *et al.*, 1984) was kindly provided by Dr M.Inouye, Robert Wood Johnson Medical School at Rutgers, Piscataway, NJ, USA.

Construction of pIN-GT vector

The restriction enzymes, *Not*I and *Rsr*II were purchased from Stratagene (La Jolla, CA). *Eco*RI, *Hin*dIII, DNA polymerase I Klenow fragment and T4 ligase were from Bethesda Research Laboratories (Bethesda, MD). The human GT cDNAs, CT7 and J20, have been isolated as described previously (Masri *et al.*, 1988). Since neither CT7 nor J20 encompass the full coding region of GT, a full-length GT cDNA, CT7−J20, was constructed by joining a 5′ portion of the CT7 insert (599 bp) and the 3′ portion of the J20 insert (*Not*I−*Eco*RI, 824 bp) at the *Not*I site. pCT7-J20, CT7−J20 subcloned in Bluescript (Stratagene, La Jolla, CA), was digested with *Rsr*II, and the *Rsr*II site was filled in to form blunt ends, using DNA polymerase I Klenow fragment and dGTP, dCTP and dTTP. The blunt ended, open plasmid was then digested with *Hin*dIII so that a 995 bp *Rsr*II(blunt)−*Hin*dIII fragment was released. pIN-III-*ompA*$_2$ was digested with *Eco*RI, the resulting overhanging nucleotides were filled in to form blunt ends with Klenow fragment, and then the plasmid was digested with *Hin*dIII. The *Rsr*II(blunt)−*Hin*dIII insert was ligated at *Eco*RI(blunt) and *Hin*dIII sites of the pIN-III-*ompA*$_2$ vector. This construction placed the insert in the correct orientation for expression, fused in-frame with the *omp*A signal sequence. The resulting vector, pIN-GT, was used for transformation of *E.coli* SB221 cells to express recombinant protein.

Gel electrophoresis and Western blotting

SB221 cells were transformed with the pIN-GT plasmid and cultured in M9-CA medium containing ampicillin (50 μg/ml) until $OD_{500\ nm}$ reached 0.5. IPTG was added to 0.05 mM and cells were cultured at 37°C for 4 or 16 h. The pelleted SB221 cells were lysed with 2 mg/ml of lysozyme (Sigma), and this preparation was designated as the total cell lysate. The periplasmic fraction was prepared as described previously (Hsiung *et al.*, 1986). The culture supernatant was concentrated to ~5% using Centricon 10 device (Amicon, MA). All fractions were solubilized in Laemmli sample buffer and subjected to 10% SDS−PAGE (Laemmli, 1970). Electrotransfer to nitrocellulose filters was performed according to Towbin *et al.* (1979). The filters were soaked with 3% bovine serum albumin in 10 mM Tris−HCl pH 8.0, containing 150 mM NaCl (TBS) for 4 h, then with TBS containing rabbit anti-human GT antiserum (diluted 1:500) (Berger *et al.*, 1981), which was kindly provided by Dr E.G.Berger of the Medizinisch-Chemisches Institut, University of Berne, Switzerland, for 2 h at room temperature. The filters were washed three times with TBS containing 0.05% Triton X-100 and were incubated with TBS containing alkaline phosphatase conjugated goat anti-rabbit IgG (Sigma, 1:1000) for 1 h. After washing the filters with TBS, alkaline phosphatase on the filters was reacted with 5-bromo-4-chloro-3-indoylphosphate in the presence of nitroblue tetrazolium in 100 mM Tris−HCl buffer pH 9.5 containing 100 mM NaCl and 5 mM $MgCl_2$, and immunoreactive components on the filters were visualized.

Assay of GT and kinetic analysis

The activity of GT was measured in a reaction mxiture (20 μl) of 0.4125 mM UDP-galactose (7×10^6 c.p.m./μmol), 10 mM *N*-acetylglucosamine, 10 mM $MnCl_2$, 0.05% bovine serum albumin and 5 μl enzyme solution in 25 mM Na-cacodylate buffer pH 7.4. After incubating at 37°C for 20−30 min, 0.4 ml of QAE-Sephadex (1 vol wet gel and 5 vol of water) suspension was added. The gel and reaction were mixed thoroughly by rotating the tube at room temperature for 10 min, and the supernatant (200 μl) obtained after centrifugation was counted for radioactivity. In each enzyme reaction, a control assay lacking the acceptor was carried out to obtain the precise quantity of UDP-galactose incorporation.

Kinetic analysis was performed as described above, except that di-*N*-acetylchitobiose (0.27−90.9 mM) was used as acceptor and concentration of UDP-galactose varied between 0.07315 mM and 8.25 mM. Unless otherwise stated, data were represented in μunits per 5 μl of enzyme solution. One unit is defined as 1 μmol of galactose transferred per minute.

N-Acetylglucosamine−Sepharose affinity column chromatography

Sepharose 4B was activated by CNBr and was conjugated to 6-amino-1-hexyl-2-acetoamino-2-deoxy-β-D-glucopyranoside as described (Barker *et al.*, 1972). The culture filtrate of SB221 harboring pIN-GT was dialyzed against water, adjusted to 25 mM Na-cacodylate buffer, pH 7.4 containing 8 mM UMP, 25 mM $MnCl_2$, 1 mM β-mercaptoethanol and 200 μM PMSF and was applied to the *N*-acetylglucosamine−Sepharose column equilibrated with the same buffer. The GT which bound to the column was eluted with the same buffer containing 8 mM *N*-acetylglucosamine and 25 mM EDTA, but without UMP and $MnCl_2$, as described by Barker *et al.* (1972). The fractions containing GT activity were pooled and concentrated using Micro-ProDiCon (Bio-Molecular Dynamics, OR). The affinity-purified enzyme was examined for its purity by SDS−PAGE and was also subjected to NH_2-terminal sequence analysis.

Peptide NH_2 terminal sequencing

The NH_2-terminal sequence of fusion protein was determined by automatic protein sequencing (gas phase sequencer model 470 with model 120 on line HPLC, Applied Biosystems) at the Department of Biology, University California San Diego.

Oligosaccharide analysis by HPLC

Chitooligosaccharides and product oligosaccharides were analyzed by HPLC (Model 5000, Varian) using a column (4 mm × 25 cm) of Lichrosorb-NH_2 (EM Science). The column was equilibrated with 75% acetonitrile and 25% 10 mM KH_2PO_4 pH 5.0, and oligosaccharides were eluted by an isocratic gradient to 50% acetonitrile over 50 min at 1 ml/min. Eluate was collected every 1 min and monitored by UV absorbance at 206 mM for chitooligosaccharides or by radioactivity for ^{3}H-labeled oligosaccharides. ^{3}H-labeled oligosaccharide products were prepared as follows. The reaction mixture (80 μl) consisted of 63.5 pmol UDP-[^{3}H]galactose (specific activity 31.5 Ci/mmol), 2 mg acceptor oligosaccharides, 10 mM $MnCl_2$ in 20 mM Na-cacodylate buffer, pH 7.4, 20 μl of 50 mM ATP and 20 μl enzyme solution. After incubation at 37°C for 16 h, the reaction mixture was applied to a column of Amberlite MB-3 (Sigma) to remove salt. Radioactive components eluted with water were pooled and were analyzed by HPLC using the condition described above.

Photoaffinity labeling of GT

GT was purified to homogeneity from human milk as described previously (Appert, 1986a). ^{14}C-labeled ANUP was prepared as described previously (Lee *et al.*, 1983). A reaction mixture (200 μl) was composed of 0.4 mM [^{14}C]ANUP (1.0 μCi/μmol), 180 mM $MnCl_2$, 5 mM β-mercaptoethanol and 0.4 mg purified GT in 75 mM glycine-glycine buffer pH 8.5. The solution was placed in a 300 μl quartz cuvette and was exposed to a long wave UV light (115 V, 60 Hz, 2.5 A) for 5 min. The procedure was repeated several times to prepare a large quantity of labeled enzyme. [^{14}C]ANUP-labeled GT (3 mg) was dissolved in 300 μl of 0.2 M *N*-ethylmorpholine acetate buffer pH 8.3, containing 6 M guanidine. Disulfide groups were reduced by dithiothreitol and alkylated by 4-vinylpyridine. Reduced and alkylated protein was purified by reverse-phase HPLC on a C-18 column, using a linear gradient of 0.1% TFA in water to 0.1% TFA in acetonitrile propanol (3:1). Three milligrams of GT, labeled with ANUP and subsequently reduced and alkylated, was cleaved with CNBr as follows. The dry protein was dissolved in 70 μl of formic acid (88%) and then diluted to 88 μl with water such that the final concentration of formic acid was 70%. Approximately a 100-fold excess of CNBr was added (one crystal). The mixture was then incubated in the dark under nitrogen for 20 h, after which time the solution was dried, wet filmed with 10−30 μl of amino ethanol and dried again. The sample was taken up in 50 μl of formic acid and immediately diluted 1:3 with the running buffer for HPLC. Elution was carried out by a linear gradient of 0.1% TFA in water:acetonitrile (9:1) versus 0.1 TFA in water:acetonitrile (1:1) by increasing acetonitrile concentration 2% per min (Figure 6).

Site-directed mutagenesis

In order to introduce desired mutation, a 775 bp GT cDNA fragment was excised from pIN-GT by *Sst*I and *Hin*dIII and subcloned in M13mp18 phage vector (Vieira and Messing, 1982). This M13 phage DNA was amplified in *E.coli* CJ236 (*dut*1 *ung*1 *thi*-1 *relA*1/pCJ105 (*cam*^{r}F′)) which is deficient in dUTPase and uracil *N*-glycosylase (Kunkel, 1985), so that M13 DNA incorporates uracil in the place of thymine. The uracil-containing single-stranded phage DNA was purified from the culture supernatant and used as template (Kunkel, 1985). Synthetic oligonucleotides designed for each mutation were phosphorylated with T4 polynucleotide kinase (Bethesda Research Laboratories, Bethesda, MD). The phosphorylated oligonucleotide (7.5 pmol) was annealed with 0.5 pmol of template DNA in 10 μl of 40 mM Tris−HCl pH 7.5 containing 10 mM $MgCl_2$ and 50 mM NaCl. Double-stranded DNA was then synthesized by incubating the annealed template and primer together with 3.5 U of T4 DNA polymerase (Stratagene, La Jolla, CA), 7 U of T4 DNA ligase and 500 μM of each of dATP, dGTP, dCTP and dTTP in 50 μl of 68 mM Tris−HCl pH 7.5 containing 10 mM $MgCl_2$, 1.0 mM ATP and 5 mM dithiothreitol for 5 min at 0°C, 5 min at 22°C and 3 h at 37°C. Wild-type *E.coli* JM105 was transformed by the synthesized double-stranded phage DNA to subclone mutated GT. Mutation was confirmed by nucleotide sequencing by dideoxy chain termination

method (Sanger *et al.*, 1977) using Sequenase (United States Biochem Co., Cleveland, OH). Mutated GT insert was ligated back at the *Sst*I and *Hin*dIII site of pIN-GT vector for expression of mutated GT in *E.coli* SB221 cells. The following oligonucleotides were used for each mutation: Y284G AATACTGAACACCAGGTAGGCTGA; Y287G CACCTCCAAAACC-CTGAACATAAG; Y287F GACACCTCCAAAGAACTGAACATAA; Y309G CCCCAGCCCCAACCATTATTAGGAAA; W310G TCCCCA-GCCTCCATAATT; W312G TTCTCCTCCCCCGCCCCAATAATT.

Acknowledgements

The authors thank Dr E.G.Berger for providing anti-human galactosyl-transferase antibodies; Drs R.Maki, C.Der (La Jolla Cancer Research Foundation) for helpful discussions and for the plasmids. We thank K.A.Masri and C.Dubois for their assistance through this work. We thank M.Williamson (University of California, San Diego) for peptide sequencing, and Dr M.Williams (La Jolla Cancer Research Foundation) for his critical reading of this manuscript. This work was supported by National Institutes of Health grants DK 37016, CA 34014 and S07RR05700.

References

Appert,H.E., Rutherford,T.J., Tarr,G.E., Wiest,J.G., Thomford,N.R. and McCorquodale,D.J. (1986a) *Biochem. Biophys. Res. Commun.*, **139**, 163–168.

Appert,H.E., Rutherford,T.J., Tarr,G.E., Thomford,N.R. and McCorquodale,D.J. (1986b) *Biochem. Biophys. Res. Commun.*, **138**, 244–229.

Barker,R., Olsen,K.W., Chaper,J.H. and Hill,R.L. (1972) *J. Biol. Chem.*, **247**, 7135–7147.

Bell,J.E., Beyer,T.A. and Hill,R.L. (1976) *J. Biol. Chem.*, **251**, 3003–3013.

Berger,E.G., Mandel,T. and Schilt,U. (1981) *J. Histochem. Cytochem.*, **29**, 364–370.

Clymer,D.S., Geren,C.R. and Ebner,K.E. (1976) *Biochemistry*, **15**, 1093–1097.

Coleman,J.M., Inukai,M. and Inouye,M. (1985) *Cell*, **43**, 351–360.

D'Agostaro,G., Bandiak,B. and Tropak,M. (1989) *Eur. J. Biochem.*, **183**, 211–217.

Geren,C.R., Magee,S.C. and Ebner,K.E. (1975) *Biochemistry*, **14**, 1461–1463.

Ghrayeb,J., Kimura,H., Takahara,M., Hsiung,H., Masui,Y. and Inouye,M. (1984) *EMBO J.*, **3**, 2437–2442.

Hsiung,H.M., Mayne,N.G. and Becker,G.W. (1986) *Bio/Technology*, **4**, 991–995.

Joziasse,D.H., Shaper,J.H., Van der Eijnden,D.H., Van Tunen,A.J. and Shaper,N.L. (1989) *J. Biol. Chem.*, **264**, 14290–14297.

Khatra,B.S., Herries,D.G. and Brew,K. (1974) *Eur. J. Biochem.*, **44**, 537–560.

Kobata,A. and Ginsburg,V. (1972) *J. Biol. Chem.*, **247**, 1525–1529.

Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.

Laemmli,U.K. (1970) *Nature*, **277**, 680–685.

Lee,T.K., Wong,L.-J.C. and Wong,S.S. (1983) *J. Biol. Chem.*, **258**, 13166–13171.

Libby,R.T., Braedt,G., Kronheim,S.R., March,C.J., Urdal,D.L., Chiaverotti,T.A., Tushinski,R.J., Michizuki,D.Y., Hopp,T.P. and Cosman,D. (1987) *DNA*, **6**, 221–229.

Magee,S.C. and Ebner,K.E. (1974) *J. Biol. Chem.*, **249**, 6992–6998.

Masri,K.A., Appert,H.E. and Fukuda,M.N. (1988) *Biochem. Biophys. Res. Commun.*, **157**, 657–663.

Nakazawa,K., Ando,T., Kimura,T. and Narimatsu,H. (1988) *J. Biochem.*, **104**, 165–168.

Narimatsu,H., Sinha,S., Brew,K., Okayama,H. and Qasba,P.K. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4720–4724.

Paulson,J.C. and Colley,K.J. (1989) *J. Biol. Chem.*, **264**, 17615–17618.

Ram,B.P. and Munjal,D.D. (1985) *CRC Crit. Rev. Biochem.*, **17**, 257–311.

Roth,J. and Berger,E.G. (1982) *J. Cell Biol.*, **92**, 223–229.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Shaper,N.L., Hollis,G.F., Douglas,J.G., Kirsch,I.R. and Shaper,J.H. (1988) *J. Biol. Chem.*, **263**, 10420–10428.

Silvia,J.S. and Ebner,K.E (1984) *J. Biol. Chem.*, **255**, 1262–11267.

Smith,C.A. and Brew,K. (1977) *J. Biol. Chem.*, **252**, 7294–7299.

Strous,G.J., van Berkhof,P., Willamsen,R., Gueze,H.J. and Berger,E.G. (1985) *J. Cell Biol.*, **97**, 723–727.

Takagi,H., Morinaga,Y., Tsuchiya,M., Ikemura,H. and Inoue,M. (1988) *Bio/Technology*, **6**, 948–950.

Takahara,M., Sagai,H., Inouye,S. and Inouye,M. (1988) *Bio/Technology*, **6**, 195–198.

Takase,K. and Ebner,K.E. (1981) *J. Biol. Chem.*, **256**, 7269–2776.

Takase,K. and Ebner,K.E. (1984) *Curr. Topics Cell. Regul.*, **24**, 51–62.

Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350–4354.

Trayer,I.P. and Hill,R.L. (1971) *J. Biol. Chem.*, **246**, 6666–6675.

Wong,L.J.C. and Wong,S.S. (1984) *Int. J. Biochem.*, **16**, 913–917.

Weinstein,J., Lee,E.U., McEntee,K., Lai,P.-H. and Paulson,J.C. (1987) *J. Biol. Chem.*, **262**, 17735–17743.

Yamamoto,F., Marken,J., Tsuji,T., White,T., Clausen,H. and Hakamori,S. (1990) *J. Biol. Chem.*, **265**, 1146–1151.

Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259–268.

Zsebo,K.M., Lu,H.S., Fieschko,J.C., Goldstein,L., Davis,J., Duker,K., Suggs,S.V., Lai,P.S. and Bitter,G.A. (1986) *J. Biol. Chem.*, **261**, 5858–5865.

Received on May 4, 1990