Glycoconjugate Journal 15, 537–548 (1998)

REVIEW

Cell surface galactosyltransferase: current issues*

Barry D. Shur‡, Susan Evans and Qingxian Lu

Department of Cell Biology, Emory University School of Medicine, Atlanta, GA 30322, USA and Department of Biochemistry and Molecular Biology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

Introduction§

The diversity of carbohydrate structures both amazes and overwhelms us, as does the diversity of the enzymes that are responsible for their synthesis. Cells express only a limited subset of the glycosyltransferases and glycosidases that are available to them, and do so in tissue-specific and spatially-specific patterns. In this regard, how particular enzymatic components of the glycosylation apparatus become and remain segregated in specific subcellular organelles is a question that impacts on issues ranging from cell biology to pharmaceutical development. Clearly, amino acid sequences contained within and adjacent to the transmembrane domain are responsible, at least in part, for the selective retention of specific glycosyltransferases to the Golgi compartments. The specific amino acid residues that are required, and how they cooperate with cytosolic proteins to localize unique glycosyltransferases to defined membrane organelles continues to be an area of intense study [1–3].

Compounding this complexity of glycosyltransferase expression, are studies that show that some glycosyltransferases are localized on the plasma membrane in addition to their conventional Golgi location [4, 5]. This prompted the hypothesis that glycosyltransferases may function as cell adhesion molecules by binding their complementary oligosaccharide substrates on adjacent cell surfaces or in the extracellular matrix [6, 7]. Since sugar nucleotides are normally not present in the extracellular fluids, it was presumed that the surface-localized glycosyltransferase-oligosaccharide complex would form a stable adhesive bond. It remains possible, however, that cells could make a suitable sugar donor available to catalyze the reaction, thus dissociating the adhesive complex if desired [8, 9].

Whether, in fact, glycosyltransferases are constituents of the cell surface remained unclear for some time, as it was difficult to eliminate completely all possible contamination from intracellular sources (for review, see [4, 5]). As has been discussed previously, a wide variety of observations confirm the presence of a few specific glycosyltransferases on the cell surface, most notably β1,4-galactosyltransferase (GalTase). The evidence supporting this includes the observations that GalTase can be detected on the surface of specific cell types at the light, confocal, and electron microscopic levels through the use of monospecific polyclonal and monoclonal antibodies raised against affinity-purified and/or bacterially-expressed, recombinant proteins [10–19]; that preparations of purified plasma membranes possess GalTase enzyme activity that cannot be attributed to Golgi contamination [12, 20]; as well as by a number of other observations that have been discussed extensively [4, 5].

Cell surface GalTase functions as a cell adhesion molecule during fertilization and development

The realization that GalTase is indeed present on the cell surface led to an analysis of surface GalTase function. A wide range of GalTase-specific perturbants, including competitive GalTase substrates, antibodies raised against affinity-purified recombinant GalTase, GalTase modifier proteins, exoglycosidase pretreatments, and others, suggest that GalTase can function as a cell surface receptor for extracellular glycoside ligands, thereby facilitating selected

* Presented at the Second Electronic Glycoscience Conference (EGC2) held on the Internet, September 9–20, 1996.

‡ To whom correspondence should be addressed. Department of Cell Biology, Emory University School of Medicine, Anatomy/Physiology, Room 100, 1648 Pierce Drive, Atlanta, GA 30322, USA. E-mail: barry@cellbio.emory.edu

§ The Editors recognize that the hypotheses presented in this review are based primarily on the work of the author and are, in part, controversial. However, the importance of these concepts to glycobiology warrant a full and open discussion. The Editors therefore invite scientists interested in this topic to submit either letters to the editor or review articles which present opposing views to those expressed in this article.

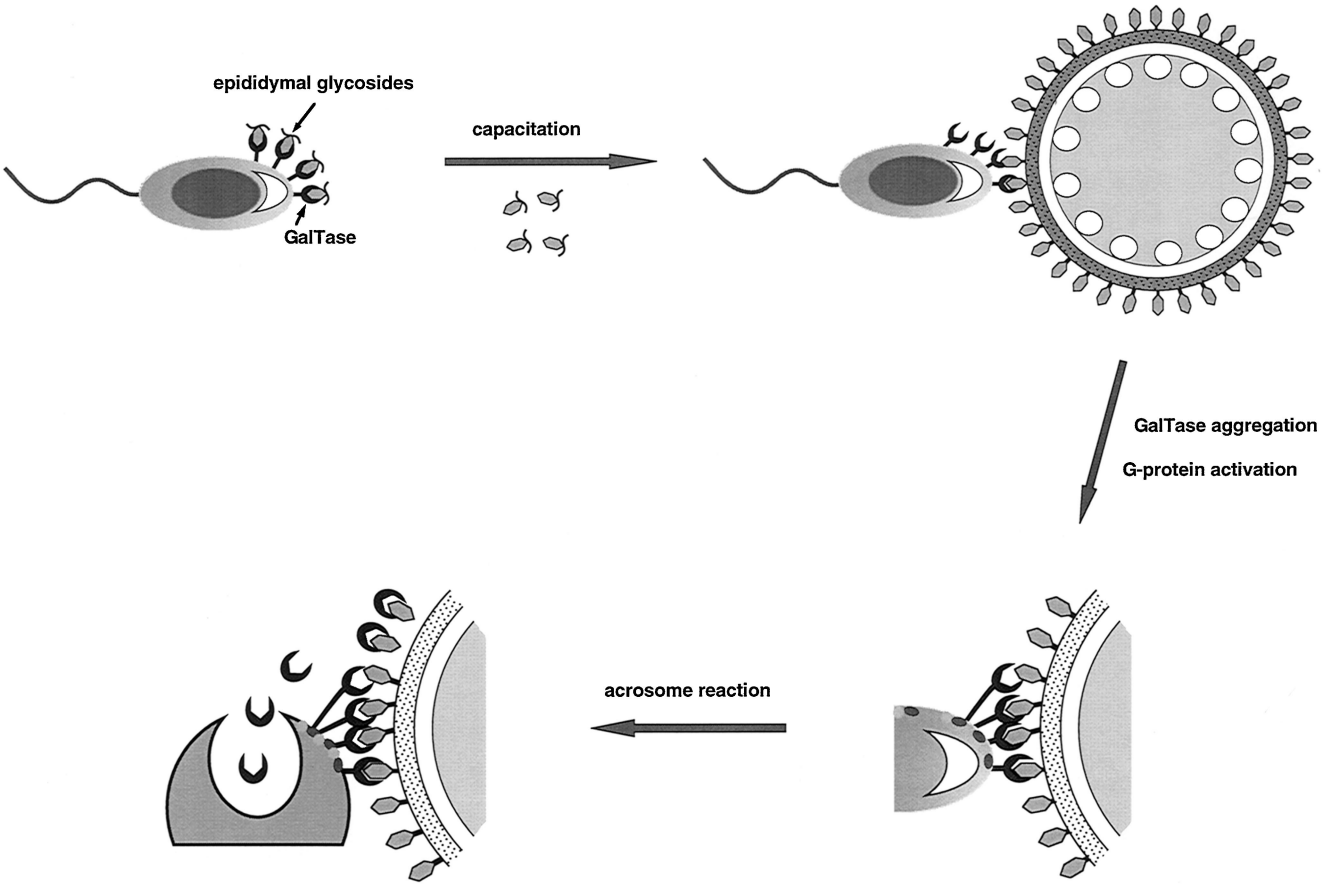

Figure 1. GalTase is restricted to the dorsal, anterior aspect of the sperm plasma membrane, where it is masked by epididymally-derived lactosaminoglycans. After capacitation, these glycosides are shed from the sperm surface, exposing GalTase, making it available to bind its *N*-acetylglucosamine-terminating ligand on the egg coat. Multivalent oligosaccharide chains cross-link GalTase, which triggers activation of a heterotrimeric G-protein complex that associates directly with the GalTase cytoplasmic domain. The resulting acrosome reaction releases hydrolytic enzymes, including *N*-acetylglucosaminidase, which aids sperm penetration through the egg coat by removing potential binding sites for sperm-localized GalTase.

cellular interactions. The ability of a variety of reagents, all of which have different modes of action, but all of which are GalTase-specific, to inhibit cell–cell and cell–matrix interactions strongly supports the hypothesis that surface GalTase functions as a cell adhesion molecule (for review, see [4, 5]). This hypothesis has been confirmed by the ability to manipulate, both positively and negatively, GalTase expression on the cell surface in transfected cells as well as in transgenic animals [18, 21–26].

The initial studies focused on surface GalTase function during homotypic intercellular adhesions [4, 5]. However, the best studied systems suggest that surface GalTase functions as a receptor for oligosaccharide ligands in the extracellular matrix, most notably during sperm binding to the egg coat and during cell interactions with the basal lamina.

Sperm GalTase function during murine fertilization

A number of laboratories have reported that sperm GalTase functions as a gamete receptor during mouse fertilization by binding to oligosaccharide ligands in the egg coat, or zona pellucida [27–31]. GalTase is found on the surface of mouse sperm in the expected location for a zona receptor – a discrete domain on the dorsal, anterior aspect of the sperm head, where it behaves as an integral membrane protein [11]. In the cauda epididymis, polylactosaminoglycans are secreted that bind sperm, thus masking the GalTase active site. During capacitation, these competitive glycoconjugates are shed from the sperm surface, making GalTase available to bind its oligosaccharide ligand in the zona pellucida [27, 28] (Figure 1).

The importance of GalTase in sperm–zona binding has been demonstrated using a series of reagents that block GalTase or the GalTase recognition site on the zona pellucida, all of which inhibit sperm-zona binding [11, 27–29]. The sperm GalTase ligand, or substrate, in the zona has been identified and shown to be the same class of O-linked oligosaccharides on the ZP3 glycoprotein known to possess sperm-binding activity [32]. Sperm GalTase selectively binds O-linked oligosaccharides on ZP3; it does not bind

other ZP3 oligosaccharides or those on other zona glycoproteins, even though they are recognized by nonsperm GalTase [32]. This illustrates a critical point that will be discussed below in more detail, *ie*, the presence of a terminal *N*-acetylglucosamine is necessary but not sufficient to confer binding by sperm GalTase. The membrane-bound surface enzyme has a much more stringent substrate specificity than does the Golgi-localized GalTase.

Binding of ZP3 to sperm induces the acrosome reaction by cross-linking GalTase residues in the sperm plasma membrane via multivalent oligosaccharides [33]. Cross-linking of GalTase activates a pertusis toxin-sensitive, heterotrimeric G-protein complex that is directly associated with the GalTase cytoplasmic domain, thus triggering the acrosome reaction [34].

As a consequence of the acrosome reaction, GalTase is redistributed to the lateral sperm head where its function remains unknown [35]. One possibility is that GalTase may stabilize sperm adhesion to the zona during the initial stages of penetration. In any event, the presence of GalTase on acrosome-reacted sperm could conceivably impede sperm penetration through the zona by binding to exposed *N*-acetylglucosamine residues on ZP3 oligosaccharides. This potential problem is eliminated by the release of an acrosomal *N*-acetylglucosaminidase that removes the GalTase-binding site from ZP3 in the vicinity of the penetrating sperm [36]. After the fertilizing sperm activates the egg, *N*-acetylglucosaminidase is released from egg cortical granules, globally destroying the GalTase-binding sites on ZP3 and accounting for the loss of ZP3's sperm-binding activity following fertilization [37].

These studies predict that elevating the expression of GalTase on the sperm surface in appropriate transgenic animals should lead to increased binding of ZP3, increased G-protein activation, and increased sensitivity to zona-induced acrosome reactions. Similarly, one would predict that eliminating GalTase from the sperm surface through homologous recombination should lead to reduced ZP3 binding and to sperm that are refractory to zona-induced acrosome reactions. As presented below, both of these predictions have been tested and confirmed [23, 25]. However, even though GalTase null sperm are unable to bind ZP3 or undergo an acrosome reaction, GalTase-null males are still fertile. This raises interesting issues, discussed later, regarding the physiological/evolutionary significance of the ZP3-GalTase acrosome reaction cascade.

GalTase is found on all mammalian sperm tested, including human, where it is confined to the plasma membrane consistent with a role in gamete recognition [11, 38–40]. Anti-GalTase antibodies and other GalTase-specific perturbations inhibit sperm–egg binding in other species [41–43], most notably bovine (Tengowski, Shur and Schatten, unpublished observations), suggesting that sperm GalTase may facilitate sperm–egg binding in other mammals. Similarly, the ZP3 ligand is expressed in the ovaries of many mammals [44, 45]. This then raises the issue of what determines species specificity, if molecules similar to GalTase and ZP3 function in fertilization in other species. One possibility, is that the membrane environment influences the substrate specificity of surface GalTase [22, 32, 46–48], which may impact recognition of ZP3 oligosaccharides in a species-specific manner. Alternatively, other sperm surface components may complex with GalTase to impart species specificity to the process. In this regard, it is noteworthy that GalTase-null sperm are still able to bind to the egg coat, even though they are unable to undergo an acrosome reaction [25]. This then suggests that GalTase cooperates with other zona-binding proteins to form a multimeric gamete receptor [49, 50].

In addition to GalTase, a number of other sperm proteins have been suggested to function as a ZP3 binding component [51]. The two others most frequently cited are sp56 [52, 53] and p95 [54, 55].

sp56 is a sperm peripheral membrane protein that was first identified by ZP3-photoaffinity labeling and claimed to confer sperm ZP3-binding activity and species specificity of sperm–egg recognition in mice [52, 53]. Protein immunoblot analysis of egg or embryo extracts show that purified sp56 binds to unfertilized eggs but not to fertilized embryos, and preincubation of eggs with purified sp56 blocks sperm–zona binding [53]. These data are consistent with sp56 having zona binding activity. However, the physiological role of sp56 in zona binding remains unclear, since it can only be removed from ZP3 affinity columns by strong denaturants (*ie*, urea) and it is a peripheral membrane protein. Recently, sp56 has been shown to be a soluble constituent of the acrosomal matrix and not a cell surface component, which argues that sp56 does not function during initial gamete recognition [56]. However, this raises the interesting possibility that sp56 may facilitate the binding of acrosome-reacted sperm to the zona prior to penetration.

Another mouse sperm protein claimed to physically associate with ZP3 is p95, a phosphotyrosine-containing protein of M_r 95 000 [54, 55]. This protein has recently been demonstrated to be a testis-specific tyrosine phosphorylated form of hexokinase [57]. A putative human analog, hu9, has been isolated and shown to have protein kinase activity, and peptides deduced from its sequence inhibit sperm–egg binding [58]. How these peptides interfere with recognition of sperm-binding oligosaccharides on ZP3 and whether the human clone is, in fact, the mouse homolog rather than the previously identified c-mer proto-oncogene [59, 60], is a matter of active debate.

Surface GalTase function during cell interactions with the basal lamina

GalTase-specific perturbants produce striking teratological abnormalities in developing chick embryos [61], implying that surface GalTase participates in cell interactions during

embryogenesis. To understand the function of surface GalTase during embryonic development, investigators have turned to more accessible *in vitro* systems.

The cumulative evidence suggests that GalTase functions on embryonic cells as a receptor for glycoside ligands in the basal lamina [4, 5]. Cells initially adhere to basal lamina components in a GalTase-independent manner [13, 62, 63], which is presumably mediated by the integrins [64]. Signals induced by the basal lamina, perhaps via the integrins, result in increased GalTase expression on the cell surface, which becomes associated with the cytoskeleton and localizes to the newly formed leading edge of the cell [13, 65]. Here, GalTase mediates lamellipodia formation and subsequent migration by binding to basal lamina oligosaccharides. This is shown by the ability of competitive GalTase substrates, blocking antibodies, and other GalTase-specific perturbants to inhibit the spreading and migration of a variety of cell types on basal lamina matrices *in vitro* [13, 17, 18, 22, 46, 62, 63]. That these *in vitro* studies reflect, at least to some degree, the requirement of surface GalTase *in vivo*, is shown by the ability of anti-GalTase antibodies to inhibit the migration of chick embryo cranial neural crest cells after microinjection into their migratory pathway [66].

In an analogous manner, GalTase is expressed on the growth cones of developing neurites, where it facilitates neurite outgrowth on laminin matrices, in conjunction with the integrins, from PC12 cells [14]. Similar results for dorsal root ganglion neurons have been documented by others [67, 68].

The binding site for GalTase in the basal lamina has been identified as N-linked oligosaccharides within the E8 domain of laminin; the same domain that is responsible in large part for the cell migration and neurite outgrowth properties of laminin [46]. As was the case for the sperm surface enzyme, PC12 surface GalTase shows a much more stringent substrate specificity than does soluble non-PC12 GalTase, which galactosylates all laminin fragments [46].

These studies predict that the rate of cell migration and neurite outgrowth should be affected by altering the expression of surface GalTase. This was borne out by the observation that the migration rate of stably-transfected mouse fibroblasts is regulated by the amount of GalTase expressed on the cell surface; increasing GalTase expression leads to enlarged lamellipodia [24], whereas decreasing GalTase expression by producing a dominant negative protein reduces lamellipodia size and stability [22, 24]. In an analogous manner, increasing surface GalTase expression on PC12 cells increases neurite formation [26].

Molecular biology of GalTase

Studies of fertilization and of cell-matrix interactions clearly support a role for GalTase as a signal transducing receptor for extracellular glycosides. What is less clear, though, are the mechanisms by which GalTase is targeted to the cell surface and elicits intracellular signals upon ligand binding. This has been an area of active debate, but the cloning of the GalTase gene products has opened up new approaches to this problem [69, 70].

The gene for GalTase encodes two similar, but not identical, proteins [69, 70]. Both GalTase proteins have a type II membrane configuration, analogous to all other glycosyltransferases cloned to date [71], with a relatively short N-terminal cytoplasmic domain, a signal sequence/transmembrane domain, and a large C-terminal lumenal or extracellular catalytic domain. The two GalTase proteins have identical catalytic and transmembrane domains, but differ in their cytoplasmic domains due to translation from two different in-frame AUGs that result from differential transcription initiation of two size classes of RNAs [69, 70]. The shorter protein has a cytoplasmic tail of only 11 amino acids, whereas the longer species has a 24 amino acid cytoplasmic domain.

Analysis of the 5′ sequences upstream of the two in-frame ATGs has identified positive and negative regulatory elements that likely impact upon the tissue-specific expression of the two different GalTase mRNAs [72, 73]. Binding sites for the ubiquitous SP1 transcription factor are scattered throughout the 5′ flanking sequences, and probably contribute to the expression of both size classes of GalTase mRNAs. More complex, though, are the presence of putative positive and negative regulatory sites thought to bind tissue-specific transcription factors, the cooperative binding of which results in the cell type-specific expression of the long and short GalTase transcripts.

Whether the two different GalTase isoforms resulting from the two different GalTase transcripts have functionally distinct roles within the cell or, rather, reflect tissue-specific proteins with similar, if not identical, biosynthetic functions remains an issue of active debate [72, 74]. A number of observations suggest that the short GalTase isoform is normally confined to the Golgi complex, where it serves a purely biosynthetic function; it is this isoform that is specifically upregulated during lactation to participate in lactose biosynthesis. On the other hand, there is reason to believe that the long GalTase isoform can function both biosynthetically in the Golgi complex, and due to its alternate cytoplasmic domain, can also function as a signal transducing receptor on the cell surface [74] (Figure 2). It is important to emphasize that this hypothesis does not suggest that all glycoprotein galactosylation requires the expression of the short GalTase isoform, since it is clear that the long isoform can also function biosynthetically. The salient distinction is that it is only the long isoform that can function as a signal transducing receptor for extracellular glycoside ligands. An alternative viewpoint is that the two GalTase isoforms show identical subcellular distributions [75, 76], and that differences in the cytoplasmic domains of these proteins are solely a byproduct of differential processing of the two GalTase mRNAs [72, 73]. To assess whether, in fact, the two different GalTase isoforms are related in any

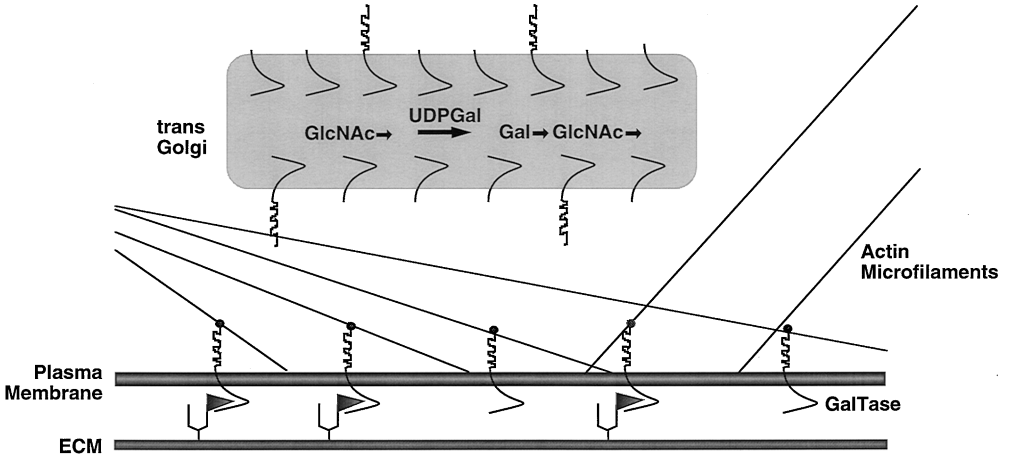

Figure 2. Cells express two similar GalTase isoforms with different cytoplasmic domains. The longer isoform contains an additional 13 amino acid extension (represented by the wavy line) that distinguishes it from the shorter isoform. Although both isoforms can function biosynthetically in the Golgi complex, it is hypothesized that only the longer isoform is able to function as a signal transducing receptor for extracellular glycoside ligands.

way to GalTase's two distinct subcellular locations and/or functions, a variety of experimental predictions have been tested.

Prediction I: The expression of the two different GalTase mRNAs should correlate, in some way, with the specific activities of the two different subcellular pools of GalTase.

The relative distribution of GalTase in the Golgi and cell surface pools shows enormous cell type variation. At one extreme are sperm, where all of the GalTase is localized on the plasma membrane since sperm no longer synthesize glycoproteins and no longer have any need for a Golgi complex [11]. Lactating mammary gland represents the other extreme, in that it expresses predominantly the Golgi-derived biosynthetic form required for lactose synthesis [12]. Interestingly, spermatogenic cells make only the long GalTase transcript [77], whereas lactating mammary gland is greatly enriched for the short transcript [12, 72]. Similarly, cell lines with an abundant Golgi complex actively involved in the elaboration of an extracellular matrix show proportionally more short transcript than long transcript [12, 78]. These observations are consistent with the hypothesis that the long GalTase protein functions as the biologically active isoform on the cell surface, in addition to its traditional biosynthetic function in the Golgi. In contrast, the short isoform functions solely in intracellular glycoconjugate biosynthesis [74].

One test of this hypothesis is to examine the expression of the two GalTase transcripts in cells in which the pools of Golgi and surface GalTase can be differently regulated. The differentiation of F9 embryonal carcinoma cells into a secretory epithelium is accompanied by a dramatic increase in GalTase specific activity due to elaboration of the Golgi complex [12]. In contrast, the plasma membrane GalTase specific activity remains constant during differentiation. Nuclease protection analysis shows that the expression of short GalTase mRNA parallels the specific activity of the Golgi pool of GalTase, whereas the expression of long GalTase mRNA remains constant as does the plasma membrane specific activity.

It is noteworthy that the long GalTase transcript is predominately (or exclusively) expressed in L cells, brain, and spermatocytes, cells which are actively engaged in glycoprotein biosynthesis [72]. In other cell types, the expression of the short transcript correlates with glycoprotein galactosylation, including differentiating F9 embryonal carcinoma cells, PYS cells, virgin mammary gland, and 3T3 fibroblasts [12, 21, 22, 78]. Thus, although it is clear that both isoforms can function biosynthetically, the critical distinction between the two isoforms, according to that proposed here, is that it is only the long isoform that is able to function as a cell surface receptor for extracellular glycosides.

Prediction II: Antibodies unique to the long isoform should immunostain the cell surface.

Antibodies specific for the 13 amino acid sequence unique to the cytoplasmic domain of the long isoform have been used to define the subcellular distribution of the long GalTase isoform by immunofluorescence confocal microscopy [19]. As predicted, a portion ($\sim$60%) of the long isoform is localized to the Golgi complex, whereas the remainder appears on the cell surface. Interestingly, the surface distribution of the long isoform shows cell-type specificity consistent with GalTase's purported role in cellular interactions, being expressed at cell-cell junctions in adherent cells and at the leading and trailing edges of migrating cells.

Prediction III: Overexpressing the long GalTase isoform should lead to increased expression of GalTase on the cell surface.

A more direct test of the relationship between the two GalTase isoforms and their subcellular distribution and function is to determine whether, in fact, overexpression of the long GalTase isoform results in increased expression of GalTase on the cell surface, relative to cells overexpressing the short GalTase isoform.

Reports using cDNAs encoding the murine, human, and bovine long GalTase isoforms suggest that the long isoform is the biologically relevant form on the cell surface. For example, multiple, independent, stably-transfected cell lines that overexpress the murine long GalTase isoform have increased GalTase specific activity on the cell surface and on purified plasma membranes, as compared to the GalTase specific activity of the total cell lysate [12]. In contrast, cell lines overexpressing the short GalTase isoform to similar degrees do not show increased GalTase specific activity on the cell surface, relative to that in the cell lysate. Similarly, others have reported that expressing either the human [79] or bovine [80] long GalTase isoform in transfected cells results in GalTase expression on the cell surface, in addition to the intracellular compartment. As was reported with the murine isoform [12], when the human short isoform is expressed in transfected cells, it is confined to the Golgi complex [79].

In one report, both bovine GalTase isoforms were suggested to be expressed on the cell surface [75], although in the one clone analyzed there appeared to be quantitatively more long GalTase on the cell surface than short GalTase. However, clonal variation was not addressed, nor was it determined which GalTase isoform was being produced [75]. In any event, it is possible, if not likely, that the relative expression of the different isoforms on the cell surface is dependent upon the relative degree of expression as well as on the cell type being used for expression.

Perhaps more compelling is the observation that certain murine cells can be experimentally manipulated to selectively increase the expression of the endogenous long GalTase mRNA without affecting the expression of the short mRNA. Under these conditions, cell surface-associated GalTase activity is selectively increased, whereas intracellular GalTase activity remains unaffected [22].

Prediction IV: The cytoplasmic and transmembrane domains of the long, but not the short, GalTase isoform should target a reporter protein to the cell surface.

An early report was unable to detect a reporter protein on the cell surface when fused to either the long or short GalTase cytoplasmic and transmembrane domains [76]. However, since only a small portion of GalTase is expressed on the surface, it is not surprising that these studies failed to detect surface expression given the low level of transgene expression in this system. This issue has been recently revisited using higher resolution assays. The cytoplasmic and transmembrane domains of the long isoform are able to target bacterial chloramphenicol acetyltransferase (CAT) to the cell surface as assayed by fluorescence activated cell sorting using anti-CAT antibodies. In contrast, CAT is confined to the Golgi complex when fused to the cytoplasmic and transmembrane domains of the short GalTase isoform (Figure 3).

Prediction V: Overexpressing the long, but not the short, GalTase isoform should interfere with GalTase-dependent cellular interactions.

Although the previous observations show that the long GalTase isoform is expressed on the surface of normal cells and can be overexpressed on the plasma membrane in appropriately transfected cells, they do not show that the long isoform is the biologically relevant form that functions as a surface receptor. The most appropriate, and rigorous way to address this issue, is to assess the consequences of gain-of-function and loss-of-function mutations. This has been approached using both *in vitro* and *in vivo* systems.

As discussed earlier, overexpressing surface GalTase on PC12 cells or 3T3 fibroblasts leads to increased neurite outgrowth [26] and increased lamellipodia formation [24], respectively. The enlarged lamellipodia have the counterproductive effect of retarding the rate of cell migration, presumably due to increased adhesion to the underlying matrix.

In vivo analyses have centered on transgenic mice that overexpress the long GalTase isoform. The great majority of these homozygous mice die within the first day of birth. Interestingly, every organ examined (*ie*, small intestine, large intestine, kidney, liver, tongue, lung) is characterized by severe morphological abnormalities. For unknown reasons, $\sim$10% of the homozygotes survive the neonatal period, but these females are unable to lactate due to defective mammary gland development during the neonatal stage [21]. Glands from transgenic females overexpressing surface GalTase are characterized by abnormal and reduced ductal development with a concomitant reduction in alveolar expansion during pregnancy, and a dramatic reduction in the expression of milk-specific proteins. Mammary epithelial cultures derived from transgenic mammary glands are unable to form anastomosing networks of epithelial cells and fail to express milk-specific proteins, as do wild-type mammary cultures. These results suggest that surface GalTase is an important mediator of mammary cell interaction with the extracellular matrix. Furthermore, since perturbing surface GalTase levels inhibits the expression of mammary-specific gene products, GalTase may function as, or impact upon, signal transducing receptors that are required for mammary gland differentiation. Importantly for the hypothesis discussed here, although overexpression of the long

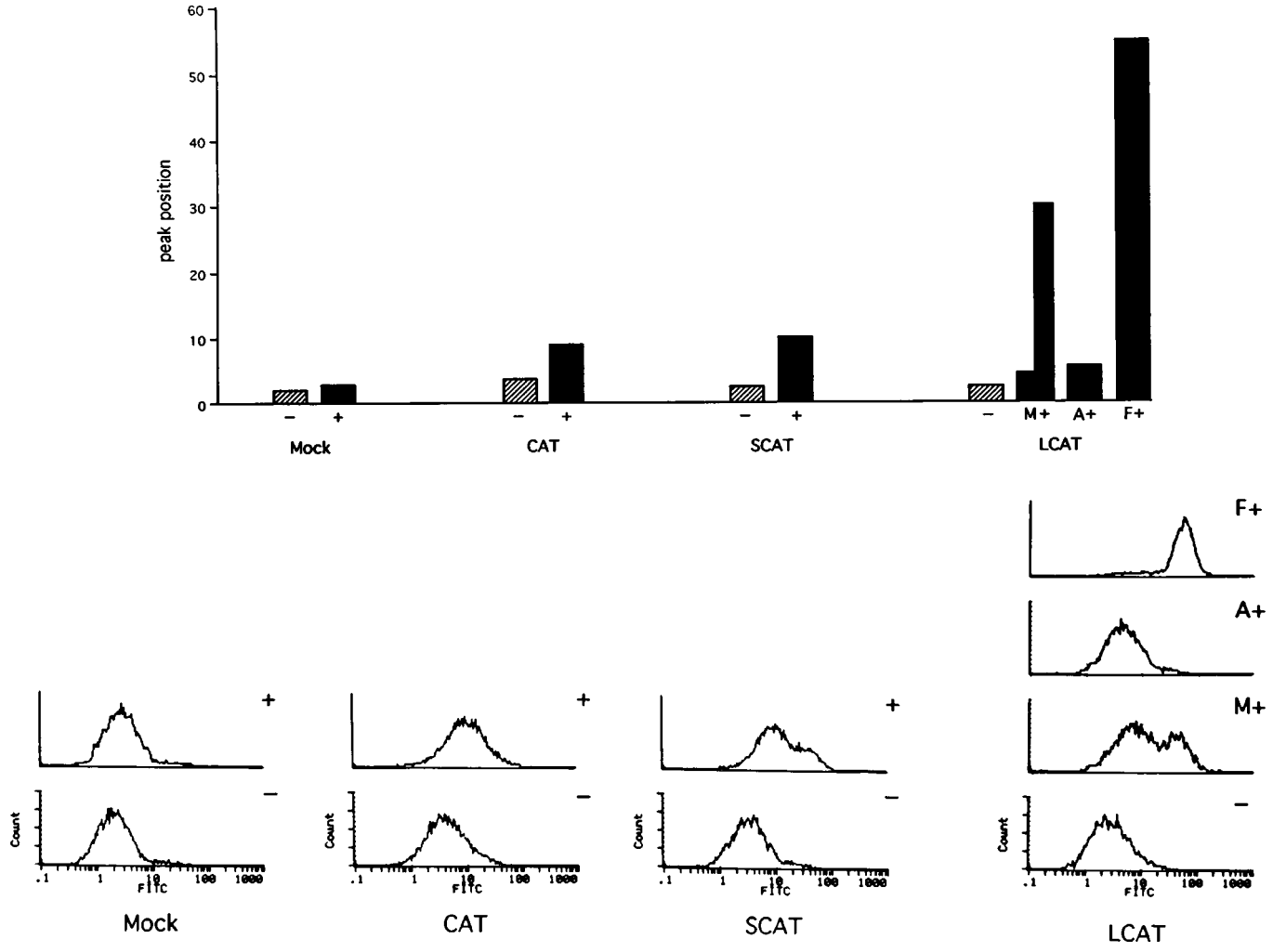

Figure 3. The cytoplasmic and transmembrane domains of the long isoform, but not the short isoform, are able to target a reporter protein to the cell surface. Cells transfected with empty vector (mock), or vector containing either bacterial chloramphenicol acetyltransferase (CAT), or CAT fused in-frame with the cytoplasmic and transmembrane domains of the short GalTase isoform (SCAT), or long GalTase isoform (LCAT), under the control of an inducible promoter, were cultured under induced (+) or non-induced (–) conditions. CAT expression on the cell surface was determined by fluorescence activated cell sorting using anti-CAT antibodies and fluorescently labeled secondary antibodies. The fluorescence profiles are shown in the lower panels, whereas the peak fluorescence intensity is graphed above. Inducing CAT and SCAT expression leads to some fluorescence detection on the cell surface, possibly due to cell leakage or over expression. However, expressing the LCAT protein leads to high levels of CAT expression on the cell surface. Interestingly, a biphasic expression pattern is seen (mixed, M), reflecting the lowly expressing cells still adherent to the matrix (A) and the highly expressing cells released from the matrix (F) due to the LCAT protein behaving as a dominant negative mutation in surface GalTase function.

GalTase isoform leads to embryonic and neonatal lethality, apparently due to defective organogenesis, mice overexpressing the short GalTase isoform show normal viability, embryonic development, and lactation [21].

Despite the fact that these animals are homozygous lethals, the heterozygotes survive apparently normally. This has allowed an assessment of elevated GalTase expression on the sperm surface [23]. Sperm exist as two genotypes in the heterozygote, but sharing of mRNA and protein between spermatids produces a phenotypically homogeneous population of sperm [81]. Sperm from these males have elevated GalTase expression on the surface, bind more radiolabeled ZP3 ligand, and are hypersensitive to zona-induced acrosome reactions [23]. In contrast, sperm from males forced to overexpress the short isoform in the testis behave normally.

Prediction VI: Expressing the long GalTase isoform on GalTase-null cells should create a surface GalTase-dependent phenotype.

One prediction would be that expressing the long GalTase isoform in cells that do not normally express GalTase on the surface would lead to a GalTase-dependent phenotype, *ie*,

a gain-of-function. HeLa cells express very low levels of surface GalTase [82] and were, therefore, transfected with murine cDNAs encoding the long GalTase isoform. HeLa cell transfectants express murine GalTase both within the Golgi complex and on the cell surface, as assayed by enzyme activity and with antibodies raised against recombinant murine GalTase.

HeLa cells expressing murine GalTase on their surface adhere more strongly to laminin substrates than do control transfectants [18] – a phenotype consistent with GalTase's reported role in lamellipodia formation and cell spreading [13, 17, 62, 63]. In contrast, when cells are plated on fibronectin substrates, which do not contain binding sites for GalTase, cell spreading is low and similar among all cell types. Cell spreading on laminin is inhibited by anti-GalTase antibodies, Fab fragments, and a variety of other GalTase-specific perturbants. Thus, expression of heterologous long GalTase produces a surface GalTase-dependent phenotype, confirming its function as a cell adhesion molecule.

Prediction VII: Inhibiting the function of the long, but not short, GalTase isoform should interfere with GalTase-dependent cellular interactions.

Loss-of-function mutations can be of two types: dominant negative constructs that inhibit the ability of surface GalTase to function, or elimination of surface GalTase by homologous recombination. A dominant negative mutation in surface GalTase function has been produced by expressing a truncated GalTase protein corresponding to the cytoplasmic and transmembrane domains of the long GalTase isoform, but devoid of the catalytic domain [22]. The truncated long GalTase protein displaces the endogenous full-length surface GalTase from its cytoskeletal association [65], forcing cells to round up and lose their adhesion to laminin matrices [13]. Expressing the truncated long GalTase protein has no effect on fibronectin adhesion, to which adhesion is GalTase-independent. Cells expressing the analogous truncated short GalTase protein show no phenotype on laminin or fibronectin [22]. The requirement for long GalTase to associate with the cytoskeleton in order to function as a surface receptor is virtually identical to that reported for other cell surface receptors, such as the cadherins [83].

Prediction VIII: Eliminating the expression of the long, but not the short, GalTase isoform should interfere with GalTase-dependent cellular interactions.

The most compelling data regarding the differential function of the two GalTase isoforms comes from GalTase-null animals created by homologous recombination [25]. Surprisingly, mice made homozygous null for both GalTase isoforms progress through embryogenesis apparently normally, but die during neonatal development due to galactosylation-related defects in the function of pituitary hormones and digestive enzymes [84, 85]. Nevertheless, a few GalTase-null mice survive the neonatal period and have been used to assess the effects of GalTase deficiency on sperm function.

Sperm from males devoid of both GalTase isoforms have negligible levels of GalTase, and as expected, are unable to bind the ZP3 ligand or undergo a ZP3-induced acrosome reaction, as are wild-type sperm (Figure 4). In contrast, GalTase-null sperm undergo the acrosome reaction normally in response to calcium ionophore, which bypasses the requirement for ZP3 binding. The inability of GalTase-null sperm to undergo a ZP3-induced acrosome reaction renders them physiologically inferior to wild-type sperm, since they are relatively unable to penetrate the egg coat and fertilize the oocyte *in vitro* [25].

Since eliminating both GalTase isoforms leads to a loss of surface GalTase expression as well as to a loss of Golgi-based galactosylation, it is possible that the inability of GalTase-null sperm to bind ZP3, undergo an acrosomal reaction, and penetrate the zona pellucida is the secondary result of defective galactosylation during spermatogenesis, rather than being the direct result of surface GalTase deficiency. This possibility was tested by creating mice deficient specifically in the long GalTase isoform – the form hypothesized here to be responsible for GalTase function as a signal transducing receptor. A point mutation was introduced into the first in-frame ATG from which the long isoform initiates translation. Cells bearing this point mutation would still transcribe mRNAs encoding the long and short isoforms normally, but due to inactivation of the first in-frame translation initiation codon, only the short isoform would be translated from both long and short transcripts.

Sperm from long isoform-null males have near normal levels of GalTase enzyme activity and are indistinguishable from wild-type sperm when analysed by indirect immunofluorescence using antibodies against the extracellular domain of recombinant GalTase. Furthermore, glycoprotein galactosylation in testis homogenates is similar to that in wild-type as assessed by RCA-1 lectin blotting of testicular proteins [25]. Although galactosylation appears normal, the sperm are unresponsive to either anti-GalTase IgG or zona pellucida glycoproteins (Figure 4). However, they respond normally to calcium ionophore. Thus, in every respect assayed, except for galactosylation, sperm null for either both isoforms or for the long isoform are indistinguishable from one another. These data show that the inability of GalTase-null sperm to undergo a ZP3-induced acrosome reaction is a direct result of the lack of the long GalTase isoform on the sperm surface, and is not the secondary result of defective galactosylation during spermatogenesis. Furthermore, the cytoplasmic domain unique to the long GalTase isoform that is responsible for association of the G-protein complex [34] appears to be absolutely required for ZP3-dependent signal transduction.

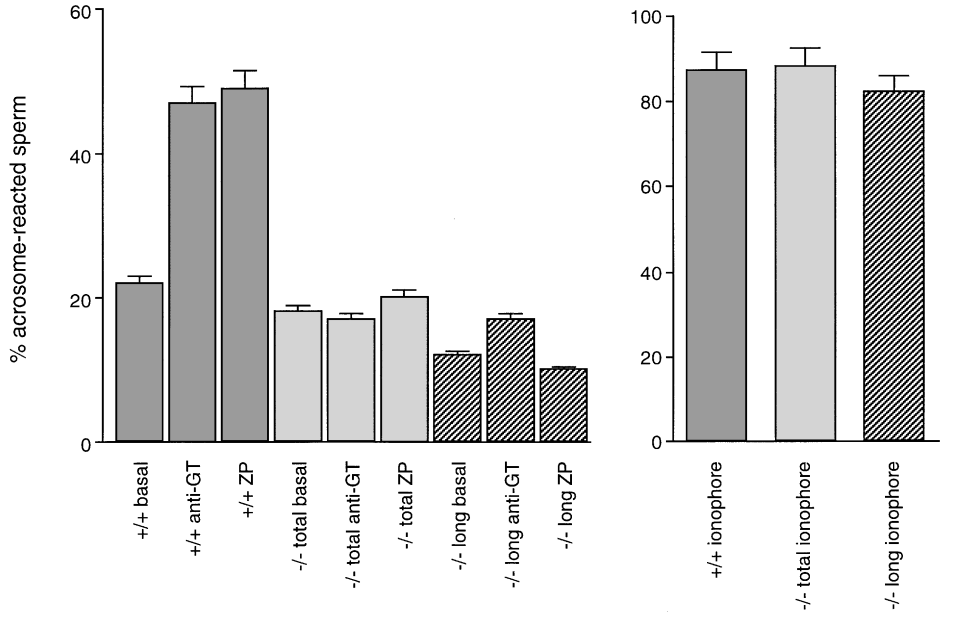

Figure 4. Sperm from GalTase-null (–/– total) males are unable to undergo an acrosome reaction in response to either anti-GalTase IgG (anti-GT) or zona pellucida glycoproteins (ZP), whereas wild-type ($+/+$) sperm do. Importantly, sperm from males that are forced to express the short GalTase isoform (–/– long), and which show normal galactosylation during spermatogenesis, are still unable to undergo an acrosome reaction in response to anti-GalTase IgG or zona pellucida glycoproteins. This illustrates that the cytoplasmic domain of the long isoform, with its characteristic 13 amino acid extension, is required for GalTase to function as a sperm surface receptor. Sperm of all three genotypes respond normally to calcium ionophore, which bypasses the requirement for ZP binding and GalTase cross-linking (from [25]).

The acrosome reaction has traditionally been viewed as a prerequisite for sperm penetration through the zona pellucida. However, GalTase-null sperm are fertile, even though they cannot undergo ZP3-induced acrosome reactions. Similarly, sperm made null for the acrosomal protease, acrosin, are also fertile [86, 87]. Both of these observations suggest that the entire ZP3-GalTase-G-protein activated release of acrosomal exocytosis is dispensable for fertilization. This then begs the question as to whether the presence of this receptor-ligand complex offers some physiological or competitive advantage to sperm. Consistent with this possibility, GalTase-null sperm are only $\sim 7\%$ efficient as wild-type in penetrating the zona pellucida and fertilizing the oocyte. Thus, in a noncompetitive environment where all sperm are phenotypically similar, as in the GalTase-null male, a few sperm eventually find their way through the zona pellucida, by either spontaneous acrosome reactions, breaks in the zona pellucida, or other scenarios, and fertilize the egg. However, when compared to wild-type sperm, this process is extremely inefficient, suggesting that in a competitive environment where not all sperm may be functionally equivalent, the integrity of the GalTase receptor complex is advantageous for successful fertilization. This has recently been shown to be the case for acrosin-null sperm, since when acrosin-null sperm are mixed with wild-type sperm, either by mixed inseminations or in chimeric testes, the wild-type sperm completely out competed the ability of acrosin-null sperm to fertilize eggs [87].

It is particularly interesting that GalTase-null sperm still bind to the zona pellucida, even though they show low levels of binding ZP3 ligand in solution. This implies that other sperm components cooperate with GalTase to mediate gamete recognition. The most pleasing model at present is that GalTase sees ZP3 oligosaccharides in the context of another sperm surface component that is responsible for the initial docking of sperm to the egg coat, similar to the concerted action of selectins and integrins in lymphocyte adhesion to the endothelium [88]. With this in mind, it would be of interest to determine if other sperm surface components thought to facilitate adhesion to the zona are responsible for the binding of GalTase-null sperm, such as PH-20 [49], zonadhesin [50], and/or those surface components that require the chaperone, calmegin, for proper folding and expression [89].

Other aspects that may contribute to GalTase expression on the cell surface

The studies discussed above indicate that the long GalTase isoform appears to be responsible for GalTase's ability to function as a signal transducing receptor on the cell surface. However, it is not entirely clear whether properties intrinsic to the long GalTase isoform are also responsible for its expression on the cell surface. Some data [12, 79], but certainly not all [76], are consistent with this possibility, such as the ability of the long GalTase cytoplasmic and

transmembrane domains to target a reporter protein to the cell surface (Figure 3). This suggests that the cytoplasmic domain of the long GalTase isoform overrides, in some way, the Golgi retention signal in the transmembrane domain and transports a portion of GalTase from the Golgi complex to the cell surface. Since it is well known that the length of the cytoplasmic domain impacts whether a protein resides in the Golgi complex or is transported to the cell surface [90], others have suggested that changes in the hydrophobic nature of the GalTase transmembrane domain may be a rate-limiting step in surface expression [75]. It is unclear, however, how fatty acylation may occur on the transmembrane domain of a type II membrane protein.

Interestingly, it has been demonstrated that the surface isoform is posttranslationally modified by the addition of carbohydrate side-chains in a post-Golgi compartment [80]. Whether this is a byproduct of transport out of the Golgi and to the cell surface, or rather, is a recognition motif involved in targeting a portion of the protein to the cell surface has yet to be determined.

Specificity of surface GalTase function

In addition to resolving the mechanisms underlying GalTase transport to the cell surface, it is equally important to clarify the specificity of surface GalTase binding. A finding common to all systems examined thus far, and which is important for our understanding of surface GalTase function, is that the surface enzyme has a much more restricted substrate specificity than does the Golgi-localized biosynthetic enzyme. In sperm [32], fibroblasts [22], PC12 cells [46], melanoma cells [47], and adrenal carcinoma cells [48], the surface enzyme recognizes only one predominant glycoprotein ligand. On the other hand, the Golgi enzyme is able to recognize a broad spectrum of *N*-acetylglucosamine-terminating oligosaccharides, either within the Golgi or on the cell surface. The presence of a terminal *N*-acetylglucosamine, the acceptor monosaccharide recognized by GalTase, is therefore necessary but not sufficient to confer binding by surface GalTase. This may reflect membrane-dependent recognition of other determinants in the oligosaccharide chain or within the protein backbone, as suggested for other carbohydrate-binding proteins and enzymes [91]. Equally likely is that specificity is dictated by adjacent plasma membrane components that cooperate with surface GalTase to form a multimeric receptor complex, as suggested for sperm GalTase. In any event, the enzymatic name 'β1,4-galactosyltransferase' is misleading in that it only refers to the linkage catalyzed by this enzyme (*ie*, galactose-β1,4-*N*-acetylglucosamine) when its UDP-galactose substrate is made available. Rather, we now know that the surface enzyme is distinct from the Golgi enzyme in that it recognizes only a selected subgroup of the possible glycoprotein substrates.

Surface GalTase function as a signal transducing receptor

Whatever the basis underlying surface expression, it is evident that the long GalTase isoform functions as a surface receptor for extracellular glycoside ligands. What is perhaps more intriguing, is that the surface isoform also functions as a signal transducing receptor, rather than as a passive carbohydrate-binding lectin-like protein [21, 34]. Other extracellular matrix receptors transmit intracellular signals following ligand binding [92], and GalTase may be another example of this. For example, aggregation of sperm GalTase by multivalent ligands activates a pertusis toxin-sensitive G-protein cascade [34]. The cytoplasmic domain of the long isoform, but not of the short isoform, is able to bind directly to a heterotrimeric G-protein complex containing the relevant pertusis toxin-sensitive G_i subunit. Specificity is shown by the fact that the same amino acid residues are unable to bind heterotrimeric G-proteins when their sequence is randomized [34]. Traditionally, G-protein-coupled receptors have been thought to be serpentine glycoproteins spanning the lipid bilayer multiple times. However, there is a growing family of single membrane spanning proteins that bind and activate heterotrimeric G-proteins [93, 94], of which GalTase may be the newest member.

Besides its ability to activate a G-protein cascade, the cytoplasmic domain of GalTase is known to associate with the cytoskeleton [65], an association that is required for GalTase [22], as well as for other cell surface proteins [83], to function as a surface receptor. It is entirely unclear, at present, how GalTase is linked to the cytoskeleton, be it direct to actin microfilaments or through intermediate coupling proteins as is the case for the integrins [64] and the cadherins [83].

The cytoplasmic domain of GalTase is also known to be serine/threonine phosphorylated [95], and a putative GalTase-associated kinase has been identified [96]. Whether, in fact, phosphorylation of the cytoplasmic domain is regulated, or impacts upon the interaction with heterotrimeric G-proteins and/or the cytoskeleton awaits testing. In any event, it is intriguing to realize that an otherwise biosynthetic enzyme may have been recruited to the cell surface in order to function as a signal transducing receptor for extracellular glycoside ligands. Since the long GalTase isoform is also present within the Golgi complex, it is appealing to think that it may participate in signaling across the Golgi membrane as well.

References

1 Yamaguchi N, Fukuda MN (1995) *J Biol Chem* **270**: 12170–6.

2 Nilsson T, Hoe ME, Slusarewicz P, Rabouille C, Watson R, Hunte F, Watzele G, Berger EG, Warren G (1994) *EMBO J* **13**: 562–74.

3 Tang BL, Low SH, Wong SH, Hong W (1995) *Eur J Cell Biol* **66**: 365–74.
4 Shur BD (1993) *Current Opinion Cell Biology* **5**: 854–63.
5 Shur BD (1991) *Glycobiology* **1**: 563–75.
6 Roseman S (1971) *Chem Phys Lipids* **5**: 270–97.
7 Roth S, McGuire EJ, Roseman S (1971) *J Cell Biol* **51**: 536–47.
8 Turley EA, Roth S (1979) *Cell* **17**: 109–15.
9 Begovac PC, Shi Y-X, Mansfield D, Shur BD (1994) *J Biol Chem* **269**: 31793–9.
10 Bayna EM, Shaper JH, Shur BD (1988) *Cell* **53**: 145–57.
11 Shur BD, Neely CA (1988) *J Biol Chem* **263**: 17706–14.
12 Lopez LC, Youakim A, Evans S, Shur BD (1991) *J Biol Chem* **266**: 15984–91.
13 Eckstein DJ, Shur BD (1989) *J Cell Biol* **108**: 2507–17.
14 Begovac PC, Shur BD (1990) *J Cell Biol* **110**: 461–70.
15 Taatjes DJ, Roth J, Shaper NL, Shaper JH (1992) *Glycobiology* **2**: 570–89.
16 Suganuma T, Muramatsu H, Muramatsu T, Ihida K, Kawano JI, Murata F (1991) *J Histochem Cytochem* **39**: 299–309.
17 Romagnano L, Babiarz B (1990) *Devel Biol* **141**: 254–61.
18 Nguyen T, Hinton D, Shur BD (1994) *J Biol Chem* **269**: 28000–9.
19 Youakim A, Dubois D, Shur BD (1994) *Proc Natl Acad Sci USA* **91**: 10913–17.
20 Lopez LC, Maillet C, Oleszkowicz K, Shur BD (1989) *Molec Cell Biol* **9**: 2370–7.
21 Hathaway HJ, Shur BD (1996) *Development* **122**: 2859–72.
22 Evans S, Lopez LC, Shur, BD (1993) *J Cell Biol* **120**: 1045–57.
23 Youakim A, Hathaway HJ, Miller DJ, Gong X, Shur BD (1994) *J Cell Biol* **126**: 1573–84.
24 Appeddu PA, Shur BD (1994) *Proc Natl Acad Sci USA* **91**: 2095–9.
25 Lu Q-X, Shur BD (1997) *Development* **124**: 4121–31.
26 Huang Q, Shur BD, Begovac PC (1995) *J Cell Sci* **108**: 839–47.
27 Shur BD, Hall NG (1982) *J Cell Biol* **95**: 567–73.
28 Shur BD, Hall NG (1982) *J Cell Biol* **95**: 574–79.
29 Lopez LC, Bayna EM, Litoff D, Shaper NL, Shaper JH, Shur BD (1985) *J Cell Biol* **101**: 1501–10.
30 Benau DA, Storey BT (1988) *Biol Reprod* **39**: 235–44.
31 Benau DA, McGuire EJ, Storey BT (1990) *Mol Reprod Dev* **25**: 393–9.
32 Miller DJ, Macek MB, Shur BD (1992) *Nature* **357**: 589–93.
33 Macek MB, Lopez LC, Shur BD (1991) *Devel Biol* **147**: 440–4.
34 Gong X, Dubois DH, Miller DJ, Shur BD (1995) *Science* **269**: 1718–21.
35 Lopez LC, Shur BD (1987) *J Cell Biol* **105**: 1663–70.
36 Miller DJ, Gong X, Shur BD (1993) *Development* **118**: 1279–89.
37 Miller DJ, Gong X, Decker G, Shur BD (1993) *J Cell Biol* **123**: 1431–40.
38 Fayrer-Hosken RA, Caudle AB, Shur, BD (1991) *Molec Reprod Devel* **28**: 74–8.
39 Miller DJ, Cross NL, Vazquez-Levin M, Shur BD (1992) In *Comparative Spermatology–20 Years After* (Baccetti B, ed) pp 569–74. Raven Press.
40 Larsen JL, Miller, DJ (1997) *Biol Reprod* **57**: 442–53.
41 Berger T, Davis A, Wardrip NJ, Hedrick JL (1989) *J Reprod Fertil* **86**: 559–65.
42 Takasaki S, Mori E, Hirano T, Furukawa K, Amano J, Mori T, Hedrick JL, Wardrip NJ, Kobata A (1993) *J Reprod Devel* **39**: 39–40.
43 Fayrer-Hosken RA, McBride CE, Brackett BG (1987) *Biol Reprod* **36**: 53.
44 Ringuette MJ, Sobieski DA, Chamow SM, Dean J (1986) *Proc Natl Acad Sci USA* **83**: 4341–5.
45 Chamberlin, ME, Dean J (1990) *Proc Natl Acad Sci USA* **87**: 6014–18.
46 Begovac PC, Hall D, Shur BD (1991) *J Cell Biol* **113**: 637–44.
47 Passaniti A, Hart GW (1990) *Cancer Res* **50**: 7261–71.
48 Penno MB, Passaniti A, Fridman R, Hart GW, Jordan C, Kumar S, Scott AF (1989) *Proc Natl Acad Sci USA* **86**: 6057–61.
49 Primakoff P, Hyatt H, Myles DG (1985) *J Cell Biol* **101**: 2239–44.
50 Hardy DM, Garbers DL (1995) *J Biol Chem* **270**: 26025–8.
51 Snell WJ, White JM (1996) *Cell* **85**: 629–37.
52 Cheng A, Le T, Palacios M, Bookbinder LH, Wassarman PM, Suzuki F, Bleil JD (1994) *J Cell Biol* **125**: 867–78.
53 Bookbinder LH, Cheng A, Bleil JD (1995) *Science* **269**: 86–9.
54 Leyton L, Saling P (1989) *Cell* **57**: 1123–30.
55 Leyton L, LeGuen P, Bunch D, Saling PM (1992) *Proc Natl Acad Sci USA* **89**: 11692–5.
56 Foster JA, Friday BB, Maulit MT, Blobel C, Winfrey VP, Olson GE, Kim KS, Gerton GL (1997) *J Biol Chem* **272**: 12714–22.
57 Kalab P, Visconti P, Leclerc P, Kopf GS (1994) *J Biol Chem* **269**: 3810–17.
58 Burks DJ, Carballada R, Moore HDM, Saling PM (1995) *Science* **269**: 83–6.
59 Bork P (1996) *Science* **271**: 1431–2.
60 Tsai JY, Silver LM (1996) *Science* **271**: 1432–4.
61 Shur BD (1977) *Devel Biol* **58**: 23–39.
62 Runyan RB, Maxwell GD, Shur BD (1986) *J Cell Biol* **102**: 432–41.
63 Runyan RB, Versalovic J, Shur BD (1988) *J Cell Biol* **107**: 1863–71.
64 Hynes RO (1992) *Cell* **69**: 11–25.
65 Eckstein DJ, Shur BD (1992) *Exp Cell Res* **201**: 83–90.
66 Hathaway HJ, Shur BD (1992) *J Cell Biol* **117**: 369–82.
67 Thomas WA, Schaefer AW, Treadway RM Jr (1990) *Development* **110**: 1101–14.
68 Riopelle RJ, Dow KE (1991) *Brain Res* **541**: 265–72.
69 Shaper NL, Hollis GF, Douglas JG, Kirsch IR, Shaper JH (1988) *J Biol Chem* **263**: 10420–8.
70 Russo RN, Shaper NL, Shaper JH (1990) *J Biol Chem* **265**: 3324–31.
71 Joziasse DH (1992) *Glycobiology* **2**: 271–7.
72 Harduin-Leper A, Shaper JH, Shaper, NL (1993) *J Biol Chem* **268**: 14348–59.
73 Rajput B, Shaper NL, Shaper JH (1996) *J Biol Chem* **271**: 5131–42.
74 Evans SC, Youakim A, Shur BD (1995) *BioEssays* **17**: 261–8.
75 Masibay AS, Balaji PV, Boeggeman EE, Qasba PK (1993) *J Biol Chem* **268**: 9908–16.
76 Russo RN, Shaper NL, Taatjes DJ, Shaper JH (1992) *J Biol Chem* **267**: 9241–47.
77 Pratt SA, Shur BD (1993) *Devel Biol* **156**: 80–93.
78 Kudo T, Narimatsu H (1995) *Glycobiology* **5**: 397–405.
79 Dinter A, Berger EG (1995) *Glycoconj J* **12**: 448.
80 Teasdale RD, Matheson F, Gleeson PA (1994) *Glycobiology* **4**: 917–28.

81 Pratt SA, Scully NF, Shur BD (1993) *Biol Reprod* **49**: 470–82.
82 Watzele G, Bachofner R, Berger EG (1991) *Eur J Cell Biol* **56**: 451–8.
83 Kintner C (1992) *Cell* **69**: 225–36.
84 Lu Q, Hasty P, Shur BD (1997) *Devel Biol* **181**: 257–67.
85 Asano M, Furukawa K, Kido M, Matsumoto S, Umesaki Y, Kochibe N, Iwakura Y (1997) *EMBO J* **16**: 1850–57.
86 Baba T, Azuma S, Kashiwabara S, Toyoda Y (1994) *J Biol Chem* **269**: 31845–9.
87 Adham IM, Nayernia K, Engel W (1997) *Molec Reprod Dev* **46**: 370–6.
88 Lasky LA (1992) *Science* **258**: 964–9.
89 Ikawa M, Wada I, Kominami K, Watanabe D, Toshimori K, Nishimune Y, Okabe M (1997) *Nature* **387**: 607–11.
90 Munro S (1995) *EMBO J* **14**: 4695–704.
91 Baranski TJ, Koelsch G, Hartsuck JA, Kornfeld S (1991) *J Biol Chem* **266**: 23365–72.
92 Juliano RL, Haskill S (1993) *J Cell Biol* **120**: 577–85.
93 Okamoto T, Katada T, Murayama Y, Ui M, Ogata E, Nishimoto I (1990) *Cell* **62**: 709–17.
94 Sun H, Seyer JM, Patel TB (1995) *Proc Natl Acad Sci USA* **92**: 2229–33.
95 Strous GJ, van Kerkhof P, Fallon RJ, Schwartz AL (1987) *Eur J Biochem* **169**: 307–11.
96 Bunnell BA, Heath LS, Adams DE, Lahti JM, Kidd VJ (1990) *Proc Natl Acad Sci USA* **87**: 7467–71.

Received 16 September 1997, revised 20 October 1997, accepted 27 October 1997