Phylogenetic and mutational analyses reveal key residues for UDP-glucuronic acid binding and activity of β1,3-glucuronosyltransferase I (GlcAT-I)

MAGALI FONDEUR-GELINOTTE,[1,5] VIRGINIE LATTARD,[1,5] RAFAEL ORIOL,[2] ROSELLA MOLLICONE,[2] JEAN-CLAUDE JACQUINET,[3] GUILLERMO MULLIERT,[4] SANDRINE GULBERTI,[1] PATRICK NETTER,[1] JACQUES MAGDALOU,[1] MOHAMED OUZZINE,[1] AND SYLVIE FOURNEL-GIGLEUX[1]

[1]UMR 7561 CNRS-Université Henri Poincaré Nancy I, Faculté de Médecine, 54505 Vandœuvre-lès-Nancy cedex, France
[2]INSERM U602; Université de Paris XI, 94807 Villejuif, France
[3]UMR 6005 CNRS, UFR Sciences, Université d'Orléans, 45067 Orléans cedex 02, France
[4]UMR 7036 CNRS-Université Henri Poincaré Nancy I, Faculté des Sciences, 54505 Vandœuvre-lès-Nancy cedex, France

(RECEIVED January 10, 2006; FINAL REVISION March 9, 2006; ACCEPTED March 9, 2006)

Abstract

The β1,3-glucuronosyltransferases are responsible for the completion of the protein–glycosaminoglycan linkage region of proteoglycans and of the HNK1 epitope of glycoproteins and glycolipids by transferring glucuronic acid from UDP-α-D-glucuronic acid (UDP-GlcA) onto a terminal galactose residue. Here, we develop phylogenetic and mutational approaches to identify critical residues involved in UDP-GlcA binding and enzyme activity of the human β1,3-glucuronosyltransferase I (GlcAT-I), which plays a key role in glycosaminoglycan biosynthesis. Phylogeny analysis identified 119 related β1,3-glucuronosyltransferase sequences in vertebrates, invertebrates, and plants that contain eight conserved peptide motifs with 15 highly conserved amino acids. Sequence homology and structural information suggest that Y84, D113, R156, R161, and R310 residues belong to the UDP-GlcA binding site. The importance of these residues is assessed by site-directed mutagenesis, UDP affinity and kinetic analyses. Our data show that uridine binding is primarily governed by stacking interactions with the phenyl group of Y84 and also involves interactions with aspartate 113. Furthermore, we found that R156 is critical for enzyme activity but not for UDP binding, whereas R310 appears less important with regard to both activity and UDP interactions. These results clearly discriminate the function of these two active site residues that were predicted to interact with the pyrophosphate group of UDP-GlcA. Finally, mutation of R161 severely compromises GlcAT-I activity, emphasizing the major contribution of this invariant residue. Altogether, this phylogenetic approach sustained by biochemical analyses affords new insight into the organization of the β1,3-glucuronosyltransferase family and distinguishes the respective importance of conserved residues in UDP-GlcA binding and activity of GlcAT-I.

Keywords: glucuronosyltransferase family; phylogeny; site-directed mutagenesis; enzyme kinetics; UDP-glucuronic acid binding site; glycosaminoglycan synthesis

Supplemental material: see www.proteinscience.org

[5]These authors contributed equally to this work.

Reprint requests to: Sylvie Fournel-Gigleux, UMR CNRS 7561-Université Henri Poincaré Nancy 1, Faculté de Médecine, BP 184, 54505 Vandœuvre-lès-Nancy cedex, France; e-mail: sfg@medecine.uhp-nancy.fr; fax: +33-383-68-39-59.

Abbreviations: GAG, glycosaminoglycan; Gal, galactose; GlcA, glucuronic acid; GlcNAc, *N*-acetylglucosamine; GlcAT-I, galactose-β1,3-glucuronosyltransferase-I; HPLC, high performance liquid chromatography; *P. pastoris, Pichia pastoris*; UDP-GlcA, UDP-α-D-glucuronic acid; Xyl, xylose.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.062089106.

The human galactose-β1,3-glucuronosyltransferase I (GlcAT-I) belongs to a family of glycosyltransferases (family 43 in CAZy database [http://afmb.cnrs-mrs.fr/CAZY/]) implicated in the formation of GlcAβ1,3Gal linkages in different glycoconjugates. This enzyme completes the synthesis of the common linker region of glycosaminoglycans (GAGs) by transferring glucuronic acid (GlcA) onto the terminal galactose of the glycopeptide primer of proteoglycans. The GAG chains of proteoglycans regulate major biological processes such as cell proliferation and recognition, extracellular matrix deposition, and morphogenesis (Yanagishita and Hascall 1992; Lin and Perrimon 2002). Furthermore, there is growing evidence for the implication of proteoglycans in several disease processes including arthropathies (Schwartz and Domowicz 2002), Alzheimer's disease (van Horssen et al. 2003), and cancer (Wegrowski and Maquart 2004). Thus, elucidation of the structure and function of glycosyltransferases involved in GAG biosynthesis has recently attracted much interest because of their crucial role in several fundamental physiological and pathological processes.

GAG assembly initiates by the formation of the tetrasaccharide, GlcAβ1,3Galβ1,3Galβ1,4Xylβ, attached to the side-chain hydroxyl group of specific serine residues of core proteins of proteoglycans. This oligosaccharide serves as a primer for chain elongation by the alternate addition of *N*-acetylglucosamine (GlcNAc) or *N*-acetylgalactosamine (GalNAc) and GlcA residues to form either heparin/heparan-sulfates or chondroitin/dermatan-sulfates GAG chains, respectively. GlcAT-I catalyzes the final step of the formation of the linkage tetrasaccharide primer, which is essential for the conversion of a core protein into a functional proteoglycan, thus regulating the overall process of proteoglycan synthesis (Kitagawa et al. 1998). Two related β1,3-glucuronosyltransferase enzymes, referred to as GlcAT-P and GlcAT-S, have been identified, which share a high degree of sequence similarity with GlcAT-I and belong to the same gene family (Terayama et al. 1997; Marcos et al. 2002). GlcAT-P and GlcAT-S are required for the addition of GlcA onto the terminal Galβ1,4GlcNAc sequence of glycoproteins and glycolipids, respectively, producing the HNK1 (for Human Natural Killer) epitope precursor sequence GlcAβ1,3Galβ1,4GlcNAc. This trisaccharide sequence then serves as the acceptor site for the 3-*O*-sulfotransferase that subsequently generates the HNK1 epitope primarily expressed in brain. This carbohydrate epitope that is recognized by HNK1 monoclonal antibodies is found on many neural cell adhesion molecules such as NCAM, myelin-associated glycoprotein (McGarry et al. 1983), and glycolipids (Chou et al. 1986) and is presumed to be involved in cell–cell interactions such as adhesion, migration, and neurite extension.

The presence of GAGs has been demonstrated in the invertebrate organisms *Drosophila melanogaster* and *Caenorhabditis elegans* (Toyoda et al. 2000). Furthermore, the existence of a conventional linkage region tetrasaccharide sequence was recently established for these invertebrate GAG chains, suggesting that their fundamental structures and biosynthetic mechanisms are similar to the mammalian GAG chains. Recently, three related β1,3-glucuronosyltransferases have been cloned in *D. melanogaster* and designated DmGlcAT-I, DmGlcAT-BSI, and DmGlcAT-BSII (where BS stands for broad specificity; Kim et al. 2003). An ortholog gene of GlcAT-I (*sqv8*) has also been cloned from *C. elegans* and its defects caused morphological abnormality such as squashed vulva (Bulik et al. 2000).

Among the β1,3-glucuronosyltransferases, human GlcAT-I was the first cloned and has since been extensively studied in our laboratory and others (Kitagawa et al. 1998; Ouzzine et al. 2000a) due to its critical location in the biosynthetic pathway of GAGs and its potential as a pharmacological target (Venkatesan et al. 2004). Biochemical and structural analyses indicated that GlcAT-I is organized as a dimer, each subunit with a Rossman-like fold divided into two regions connected by the so-called DXD motif (D195–D196–D197 in GlcAT-I) (Ouzzine et al. 2000b; Pedersen et al. 2000). The N-terminal region (residues 26–74) comprises the UDP-sugar binding region and is terminated by the DDD sequence involved in the coordination of Mn^{2+} divalent cations essential for GlcAT-I activity (Gulberti et al. 2003). The C-terminal region (75–335) includes the acceptor substrate binding site and is terminated by a C-terminal domain extending to the other molecule in the dimer, that is thought to be important for substrate recognition (Gulberti et al. 2005).

The aim of this study was to identify crucial residues involved in UDP-α-D-glucuronic acid (UDP-GlcA) recognition and β1,3-glucuronosyltransferase activity. A previous study emphasized the key role of H308 in governing the specificity of GlcAT-I toward the nucleotide–sugar (Ouzzine et al. 2002). In order to better understand the recognition process of the donor substrate, we develop here a phylogenetic approach, that allowed us to identify 119 related β1,3-glucuronosyltransferase sequences in vertebrates, invertebrates, and plants. Multiple sequence alignments revealed conserved peptide motifs and amino acids, stimulating the evaluation of the function of these potentially important residues. Systematic site-directed mutagenesis of these residues in the human GlcAT-I led us to delineate their respective importance in UDP-GlcA binding and in β1,3-glucuronosyltransferase activity.

Results

Phylogenetic analysis

Phylogeny analysis identified a total of 119 β1,3-glucuronosyltransferase-like enzymes. Thirty-two were already

present in EMBL/GenBank and 87 were reconstructed in silico from expression sequence tags (EST) and whole genome shotgun (WGS) banks (see online supplemental data). The phylogenetic analysis was first carried on the 119 sequences (not shown) and gave a clear separation in three main groups: vertebrates, invertebrates, and plants, with two or more subfamilies in each group. Then a second analysis was carried on 40 selected sequences representing the main subfamilies of each of the three subgroups (Fig. 1). All the vertebrate β1,3-glucuronosyltransferase sequences could be clearly ascribed to one of the three subfamilies GlcAT-I, GlcAT-P, and GlcAT-S. These three subfamilies were found in all vertebrates including fish, amphibians, birds, and mammals and result from two duplication events of a single ancestral gene. The first duplication of the ancestral vertebrate gene split apart the GlcAT-I subfamily from the other two and the second duplication separated GlcAT-P from GlcAT-S. The mean inter-subfamily genetic distance between the GlcAT-P and GlcAT-S is the shortest one (0.27 ± 0.01), suggesting that this duplication event occurred recently (Fig. 1). Three urochordate β1,3-glucuronosyltransferase-like enzymes (*Ciona intestinalis*, *Ciona savignyi*, and *Oikopleura dioica*) could not be ascribed to any of the GlcAT-I, GlcAT-P, or GlcAT-S, because they appear to branch out from the common evolutionary path before the duplication events at the origin of the three vertebrate subfamilies. Three *D. melanogaster* β1,3-glucuronosyltransferases responsible for the synthesis of the GAG–protein linkage region

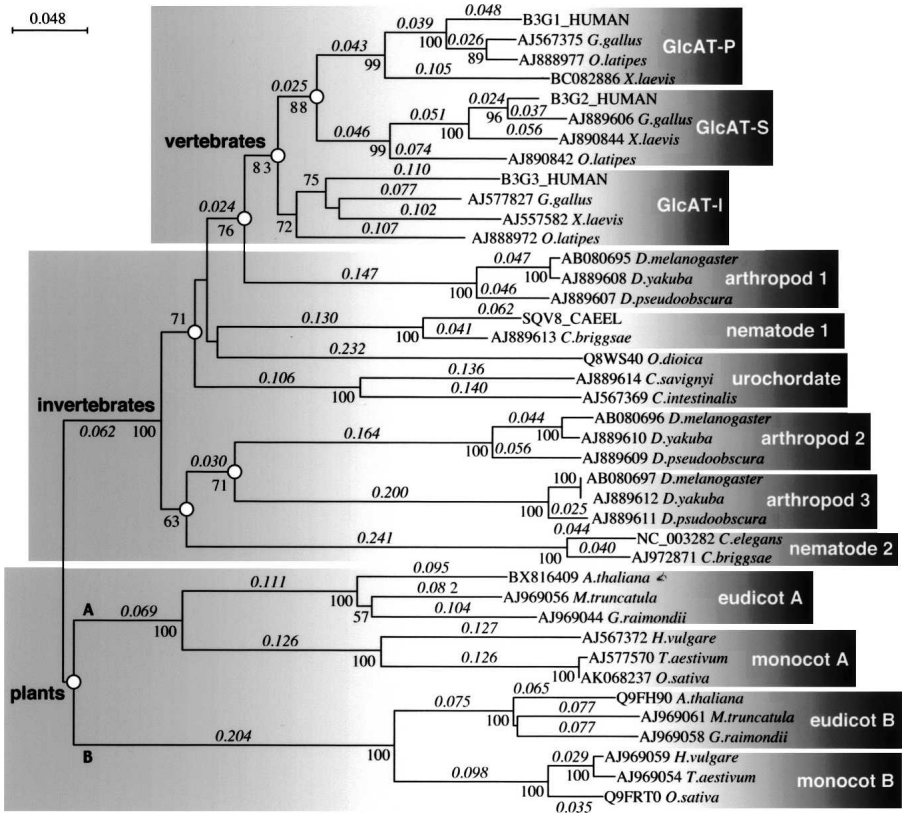

Figure 1. Neighbor-joining phylogenetic tree of β1,3-glucuronosyltransferases. Forty sequences were selected out from the total 119 different β1,3-glucuronosyltransferase-like sequences based on the following criteria: (1) They were the more representative vertebrate sequences comprising one mammal (*Homo sapiens*), one bird (*Gallus gallus*), one amphibian (*Xenopus laevis*), and one fish (*Oryzias latipes*); (2) they represented the urochordate (*Oikopleura dioica*, *Ciona intestinalis*, and *Ciona savignyi*); (3) they were present in the three arthropod branches (*Drosophila melanogaster*, *Drosophila yakuba*, and *Drosophila pseudoobscura*); (4) they were present in two nematode species (*Caenorhabditis elegans* and *Caenorhabditis briggsae*); and (5) they were present in the main two groups of flowering plants (eudicots *Arabidopsis thaliana*, *Medicago truncatula*, and *Gossypium raimondii*; monocots *Hordeum vulgare*, *Triticum aestivum*, and *Oryza sativa*). The analysis was carried out on the segment of 295 positions comprised between the first position of the conserved motif 1 and the last position of the conserved motif 8 of the complete multialignment (Fig. 2). One hundred and fifty-six positions (53%) were selected in 13 G-blocks (Castresana 2000). The open circles show the duplication events that suggest the existence of a general trend from the more recent duplication node (GlcAT-P and GlcAT-S) to the more ancient duplication node (plant A and plant B), by a progressive increase of genetic distances (figures in italics, *above* branches) of vertebrate, invertebrate, and plant subfamilies. The robustness of the tree branches was tested by bootstrap calculations from 500 sets of data. Bootstrap values higher than 50% are considered significative and are reported at the *left* of each node. The scale bar represents the number of substitutions per site for a unit branch length.

of proteoglycans were recently cloned and characterized (Kim et al. 2003). Our analysis indicated that these *D. melanogaster* sequences belong to three separate branches and could not be ascribed to either of the GlcAT-I, GlcAT-P, or GlcAT-S vertebrate subfamilies. These three enzymes clustered together with the other arthropod β1,3-glucuronosyltransferase-like enzymes. A similar phenomenon was observed with the nematode sequences (*C. elegans* and *C. briggsae*) that clustered together in at least two separate branches (Fig. 1). The presence of a third nematode branch was suggested by other *C. elegans* sequences (O17708, O62113, O02304, not shown). Finally, β1,3-glucuronosyltransferase-like enzymes with similar conserved peptide motifs and similar inter-motif distances, as described below, were also found in two plant subfamilies (A and B) containing monocot and eudicot flowering plants (Figs. 1, 2). Overall, the genetic distances between subfamilies suggest the existence of a progressive trend for the time of appearance of the different subfamilies: GlcAT-P ≅ GlcAT-S < GlcAT-I < arthropod 1 < nematode 1 ≅ urochordate ≅ arthropod 2 ≅ arthropod 3 < nematode 2 < plant A ≅ plant B. The mean genetic distance between the A and B plant subfamilies (0.66 ± 0.03) is much longer than the mean genetic distance between GlcAT-P and GlcAT-S subfamilies (0.27 ± 0.01), suggesting that the duplication at the origin of the A and B plant subfamilies is more ancient (Fig. 1).

Conserved peptide motifs of β1,3-glucuronosyltransferases

Eight conserved peptide motifs were defined in the ClustalW alignment of all the vertebrate β1,3-glucuronosyltransferase-like sequences (Fig. 2). Motifs 1–4 and 8

```
Conserved motifs      ------1-------     -----2-----     -----3-----     ------4------5/6/7------8------
donor binding sites          Y84               D113      R156  R161            DDDN194-7          H308  R310
B3G3_HUMAN ; GlcAT-I IYVVTPTYARLVQK<12>VPRLHWLLVED<41>PRGVEQRNKAL<22>GVVYFADDDNTYS<100>CTRVLVWHTRTEK 313
AJ577827 G.gallus    IFVVTPTYARPVQK<12>VPSLHWLVVED<41>ARGVEQRNRAL<12>GVVYFADDDNTYS<100>CTQVLVWHTRTEK 239
AJ557582 X.laevis    IYVVTPTYARPHQL<12>VPSLHWILVED<41>PRGVEQRNEAL<12>GVVYFADDDNTYS<100>CTKVWVWHTRTEK 294
AJ888972 O.latipes   IFVITPTYARLVQK<12>VPQLHWIVVED<41>PRGVQQRNEGL<17>GVVYFADDDNTYS<101>CTKVLVWHTRTEK 316
B3G1_HUMAN ; GlcAT-P IHVVTPTYSRPVQK<12>VPNLHWLVVED<41>PRGTMQRNLAL<14>GVVYFADDDNTYS<102>CTKILVWHTRTEK 316
AJ567375 G.gallus    IHVITPTYSRPVQK<12>VPNLHWILVED<41>PRGTMQRNLAL<14>GVVYFADDDNTYS<102>CTKILVWHTRTEK 316
BC082886 X.laevis    IYVITPTYSRPVQK<12>VVNLHWIVVED<41>PRGTTQRNLGL<14>GVVYFADDDNTYS<102>CTKVLVWHTRAER 324
AJ888977 O.latipes   IHIITPTYSRPVQK<12>VPNLHWILVED<41>PRGTMQRNLAL<14>GIVYFADDDNTYS<103>CTKILVWHTRTEK 317
B3G2_HUMAN ; GlcAT-S IYAITPTYSRPVQK<12>VAQLHWILVED<35>PRATEQRNAGL<14>GVLFFADDDNTYS<101>CTKVLVWHTRTEK 305
AJ889606 G.gallus    IYAITPTYSRPVFT<12>VARLHWILVED<35>PRATEQRNAGL<15>GVLFFADDDNTYS<101>CTKVLVWHTRTEK 286
AJ890844 X.laevis    IFAITPTYSRPVQK<12>VPRLHWILVED<35>PRATEQRNAGL<22>GVVFFADDDNTYS<101>CTKVLVWHTRTEK 312
AJ888977 O.latipes   IHIITPTYSRPVQK<12>VPNLHWILVED<41>PRGTMQRNLAL<14>GIVYFADDDNTYS<101>CTKILVWHTRTEK 317
AB080695 D.melano. 1 IYAVTPTYPRPAQK<12>LPHLHWIIVED<43>PRGVEQRNLAL<12>SIVFFMDDDNSYS<104>CTDVLVWHTRTEK 284
AJ889608 D.yacuba  1 IYAVTPTYPRPAQK<12>LPHLHWIIVED<43>PRGVEQRNLAL<12>SIVFFMDDDNSYS<104>CTDVLVWHTRTEK 284
AJ889607 D.pseudo. 1 IYAITPTYARPAQK<12>LPNLHWIIVED<43>PRGVEQRNMAL<12>AIVFFMDDDNSYS<104>CRDVLVWHTRTEK 286
SQV8_CAEEL C.elega.1 IYFITPTHFRAAQR<12>VPNLHWIVVED<41>PRGVEQRNLAL<12>GVVYFGDDDNTYD<102>CTKVYVWHTRTEK 327
AJ889613 C.briggsae1 IYFVTPTHFRAAQK<12>VPNLHWIVVED<41>PRGVEQRNAAL<12>GVVYFGDDDNTYD<102>CTKVYVWHTRTEK 327
Q8WS40   O.dioica   IYLITPTYARLTQK<12>VENLHWIIIED<43>PRGVQQRNAGL<12>GYTYFLDDDNTYD<103>CRDILVWHTRTEK 301
AJ881614 C.savignyi  VRQSLPTYSRLTQK<12>VPRFHWILIED<31>TKDLQTRNTAL<12>GVLYFMDDDNTYG<100>CTEVNVWHTQTVK 290
AJ567369 C.intestin. IHAITSTYARLTQK<12>LSNFHWILIED<37>VKDLLTRNAAL<11>GIVYFMDDDNTYD<-99>CTEVYVWHTQTMK 276
AB080696 D.melano. 2 IYFVTPTYPRREQI<12>IPRLHWLVADD<36>PRGVANRRAAL<10>GILYFGDDDNTYD<100>CTEILVWHTQTKS 308
AJ889610 D.yacuba  2 IYFVTPTYPRREQI<12>VPRLHWLVADD<36>PRGVANRRAAL<10>GILYFGDDDNTYD<100>CTEILVWHTQTKS 314
AJ889609 D.pseudo. 2 IYFVTPTYPRREQI<12>VPRLHWLVAND<36>PRGVANRRAAL<10>GVLYFGDDDNTYD<100>CSQILVWHTQTKS 324
AB080697 D.melano. 3 LYIITPTYRRPEQL<12>VVNLLWLVIED<37>PRGVSNRNRGL<09>GVLYFADDDNTYD<100>CRDILTWHTQTKK 289
AJ889612 D.yacuba  3 LYIITPTYRRPEQL<12>VVNLLWLVIED<37>PRGVSNRNRGL<09>GVLYFADDDNTYD<100>CRDILTWHTQTKK 289
AJ889611 D.pseudo. 3 LYIITPTYRRPEQL<12>VANLLWLVIED<37>PRGVSNRNRGL<09>GVLYFADDDNTYD<100>CRDILTWHTQTKK 294
NC_003282 C.elegans2 IIVVTPTYKRMTRI<12>IKDLHWIIVED<31>RRGWYQRTMAL<17>GVVYFGDDDNSYD<106>DREILVWHTKTST 229
AJ972871 C.briggsae2 IIVVTPTYKRMTRI<12>VKDLHWIVIED<31>RRGWYQRTMAL<17>GVVYFGDDDNSYD<106>DREIFVWHTKTST 280
BX816409 A.thalia.A1 LIVVTPTYNRAMQA<13>ESPVLWIVVEG<31>DRGVHQRNTAL<09>GIVYFADDDNIYS<115>CSSILNWHLHLDA 372
AJ969056 M.trunca.A1 LIIVTPTYNRSFQS<13>PPPLLWVVVEM<31>DRGVHQRNKAL<09>GIVYFADDDNVYS<114>CKKIMNWHLHLNV 419
AJ969044 G.raimon.A1 LIVVTPTYNRGFQA<13>KPPLVWIVVEE<31>DPRVHQRNAAL<09>GIVFFADDDNVYT<119>CSKVMNWHLHLDT 372
AJ567372 H.vulgareA6 LIVVTATTVRPHQA<13>PPPLLWIVAEW<31>KIIVCQKNNAI<09>GIVHFADEERAYS<115>CTRIMVWNFDLEP 339
AJ577570 T.aestiv.A6 LIVVTITSARPQQA<13>PPPLLWLVVEW<31>KIAVCQRNNAI<09>GIVHFADEERSYM<115>CNRVMVWNFALEP 329
AK068237 O.sativa A6 LIIVTITTVRPQQA<13>QSPLLWLVVEW<31>KIAVCQRNTAI<09>GIMHFADEERSYM<115>CNRVMVWNFNLEP 429
Q9FH90 A.thaliana B8 VIVVTPTYVRTFQA<13>PYDLVWIVVEA<35>KLETKMRLHAL<09>GIVMFADDSNMHS<145>GRRVLLWWLRVEA 418
AJ969061 M.trunca.B8 VIVVTPTYVRTFQA<13>PYDLIWIVVEA<35>KVESLMRLHAL<09>GIVMFADDSNMHN<143>GRHVLLWWLRVEA 414
AJ969058 G.Raimon.B8 VIVVTPTYVRTFHA<13>PYDLVWIVVEA<34>KLESKMRLRAL<09>GIVMFADDSNMHS<141>GRQVLLWWIRVEA 416
AJ969059 H.vulgareB8 VLVVTPTYSRAFQA<13>PYPLTWIVVEA<35>ATENRMRLHAL<09>GVVVFADDSNVHS<140>GKKVLLWWLRVEA 453
AJ969054 T.aestiv.B8 VLVVTPTYSRAFQA<13>PYPLTWIVVEA<35>ATENRMRLHAL<09>GVVVFADDSNVHS<140>GKKVLLWWLRVEA 453
Q9FRT0   O.sativa B8 VLVVTPTYSRAFQA<13>PYPLTWIVVEA<35>ATENRMRLHAL<09>GVIVFADDSNVHS<151>GKKILLWWLRVEA 456
```

Figure 2. ClustalW sequence alignment of the donor substrate-related conserved peptide motifs 1–4 and 8 of the 40 sequences analyzed in Figure 1. The relative positions of the eight conserved peptide motifs are illustrated in the *first* line. The positions of the donor-related conserved amino acids in the human GlcAT-I are illustrated in the *second* line. For each motif, positions conserved at >50% are shown with a gray background, and the highly conserved donor-related amino acid positions within each motif are shown as white letters on a solid background. The shading of alignments is based on a chemical alphabet comprising five groups: acidic or amide (E,D,Q,N), hydrophobic (I,L,V,M), aromatic (F,Y,W), basic (R,H,K), and hydroxyl (S,T), and the remaining four amino acids (A,G,P,C) were analyzed separately. This alphabet is based on frequencies of evolutionary replacement among amino acids, chemical characterizations, and minimal base differences between codons. Amino acids of the same group were considered equivalent for the definition of conserved positions. The number of intermotif amino acids is given between the "<" and ">" symbols. The gap between motifs 4 and 8 contains the acceptor substrate-related motifs 5, 6, and 7 and the corresponding inter-motif distances. The sequences of the motifs and the size of the intermotif distances are relatively well conserved in all the β1,3-glucuronosyltransferase-like proteins.

contain the expected donor substrate binding amino acids as predicted from the crystal structure of the human GlcAT-I (Pedersen et al. 2000; Negishi et al. 2003): Y84 in motif 1; D113 in motif 2; R156 and R161 in motif 3; D194, D195, D196, and N197 in motif 4, and H308 and R310 in motif 8. Motifs 5, 6, and 7 contain the expected acceptor substrate-related amino acid positions: E227 in motif 5; W243, R247, and D252 in motif 6, and E281 in motif 7 (not shown). These eight conserved motifs are constituted by 89 amino acids (27%) and contain 15 highly conserved residues (4%) among the 335 amino acids of the human GlcAT-I sequence. They were also present in all the invertebrate and plant β1,3-glucuronosyltransferase-related sequences analyzed, although they are less conserved in invertebrates and plants (Fig. 2). In addition to the 119 sequences described in the present paper, many other invertebrate and plant partial sequences contained β1,3-glucuronosyltransferase-like conserved motifs, suggesting that these sequences are largely distributed among eukaryota. They were indeed found in most invertebrates comprising other urochordates (*Halocyntia roretzi* AV382946), other arthropods (*Locusta migratoria* CO821541, *Homarus americanus* CN852790, *Callinectes sapidus* CV162017, *Boophylus microplus* CK190607), other nematodes (*Meloidogine incognita* BM882152, *Meloidogyne chitwoodi* CB830291, *Ancylostoma caninum* BF250404, *Brugia malayi* AA585037, *Strongyloides stercoralis* BE224696, *Trichinella spiralis* BG353857), trematodes (*Schistosoma mansoni* AAC46895, *Schistosoma japonicum* AAP05952), and a medusae (*Hydra magnipapillata* DN136848). They were also present among other eudicots (*Mesembryanthenum crystallinum* AI823190, *Prunus armeniaca* CB823217, *Glycine max* BQ297554, *Gossypium arboreum* BG440534, *Vitis vinifera* CF208104, *Lactuca sativa* BQ853287), other monocots (*Sorghum propinquum* BG487898), coniferopsida (*Pinus taeda* BQ702040, AW784032), and briophyta (*Piscomitrella patens* BJ594624, BJ595061). However, these enzymes are probably not present in all eukaryota since we could not find β1,3-glucuronosyltransferase-like motifs among fungi, protozoa, or algae.

Site-directed mutagenesis of conserved amino acid residues

In order to further assess the functional role of the conserved amino acids identified by phylogenetic comparison, we conducted systematic mutational analyses of these residues in human GlcAT-I, a representative member of the β1,3-glucuronosyltransferase family. In this study, we targeted the amino acids predicted to interact with the donor substrate. For this purpose, we expressed the recombinant GlcAT-I in *Pichia pastoris* and compared activities, kinetic parameters, and UDP binding properties of wild-type and mutant proteins.

Analysis of the three-dimensional structure of GlcAT-I–UDP-GlcA complex (Fig. 3) suggests that the position of uracil group is fixed through a ring-stacking interaction with Y84 and a hydrogen bond with D113 (Pedersen et al. 2000, 2002). Both conservative and nonconservative mutations of these residues were carried out and the consequences were evaluated. Upon expression in *P. pastoris*, immunoblot analysis indicated that the Y84F mutant was produced at higher level than the wild-type protein, while the Y84A mutant was slightly less expressed (Fig. 4A). The activity of the recombinant enzymes was normalized to the amount of expressed GlcAT-I protein and values are presented in Figure 4B. Interestingly, substitution of Y84 to alanine completely abrogated the enzyme activity, whereas the conservative mutation of Y84 to a phenylalanine residue restored enzyme activity up to 60% of the wild-type enzyme (Fig. 4B). Determination of the kinetic parameters of this mutant indicated a moderate increase in K_m value toward UDP-GlcA (threefold; Table 1). The observation that the conserved Y84 residue has diverged to histidine in

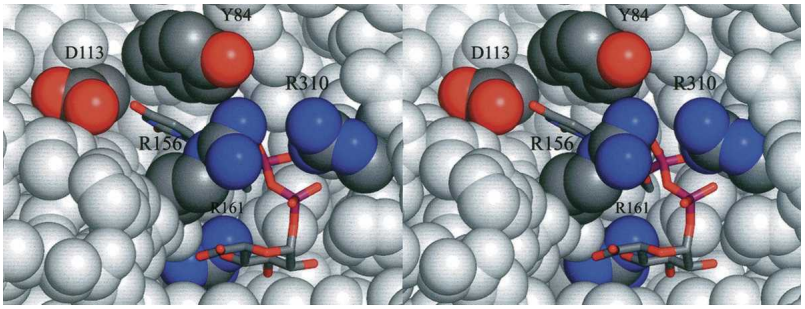

Figure 3. Stereo view of the human GlcAT-I donor substrate-binding site. The structure of GlcAT-I is shown in complex with UDP-GlcA (PDB code 1FGG). Conserved amino acid residues predicted to interact with the uridine base (D113 and Y84), with the pyrophosphate (R156 and R310), and with the GlcA group (R161) are indicated. The structure of UDP-GlcA is represented in colored sticks.

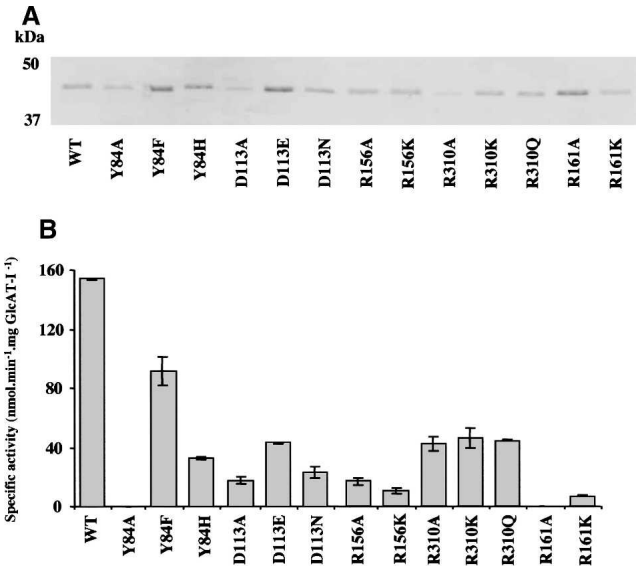

Figure 4. Heterologous expression and activity of wild-type human GlcAT-I and mutants expressed in yeast *P. pastoris*. (*A*) SDS-PAGE and immunoblot analysis of membrane fraction of recombinant yeast cells expressing wild-type (WT) and GlcAT-I mutants; 2 μg membrane proteins were loaded per lane. (*B*) Glucuronosyltransferase activities were determined toward Galβ1,3Gal as acceptor substrate. Values are expressed per mg GlcAT-I protein and are the mean ± SD of three determinations.

C. elegans β-1,3-glucuronosyltransferase (Fig. 2) led us to determine the functional consequences of a similar substitution in GlcAT-I. Interestingly, the Y84H mutant has retained significant activity (21%, Fig. 4B) with a K_m value double that of the wild type (Table 1). UDP binding experiments showed that, as expected, the wild-type protein exhibited efficient binding to UDP-beads, which was prevented by addition of UDP in the incubation mixture (Fig. 5A), whereas Y84A exhibited impaired binding to UDP-beads, resulting in the presence of ~50% of the protein in the supernatant. In contrast, Y84F mutant was able to bind efficiently to UDP-beads as the wild-type protein (Fig. 5B). Altogether, these results indicated that the tyrosine residue at position 84 plays a crucial role in GlcAT-I activity and UDP-GlcA binding. This function appeared to be mainly attributable to the phenyl ring, likely via stacking interactions with the nucleotide base, since a phenylalanine or a histidine residue were able to replace, to some extent, the wild-type tyrosine.

Next, we analyzed the role of D113 predicted to interact with N-3 of uracil. Upon expression in *P. pastoris*, the D113A mutant was produced in a slightly lower amount compared with the wild-type enzyme whereas D113E and D113N mutants were produced in higher or similar amounts, respectively (Fig. 4A). We found that substitution of D113 by alanine and by asparagine strongly decreased the enzyme activity (to 11% and 15% of the wild-type GlcAT-I, respectively; Fig. 4B). The K_m value toward UDP-GlcA for the D113N mutant was increased 13-fold, suggesting that the amide group of asparagine is much less favorable to hydrogen-bond formation with N-3 of uracil compared with the carboxyl group of aspartate. On the other hand, mutation of D113 by glutamic acid affected the enzyme activity to a lesser extent when compared with the D113A or D113N mutants (reduced to 27% of the wild-type enzyme). Determination of the kinetic parameters of the D113E mutant showed a threefold increase in K_m value toward UDP-GlcA. These results suggest that the nucleotide-binding site of GlcAT-I could accomodate a longer carboxyl side chain at this position, although this substitution results in a lower catalytic efficiency of the enzyme. UDP binding experiments showed that the three aspartate-substituted mutants exhibited reduced capacity to bind to the beads, consistent with the increased K_m values of these mutants (Fig. 5B). These results indicate that mutations of D113 impaired, but did not abolish, the UDP binding capacity of GlcAT-I to UDP-beads. These observations are consistent with the important contribution of Y84 residue together with D113 in the positioning of the base of the nucleotide–sugar.

We next examined the role of the basic residues R156 and R310, which are predicted to interact with the pyrophosphate group of UDP (Fig. 3). Atom NH1 of R156 is in hydrogen-bond distance from O2B of the α-phosphate, whereas atom NH1 of R310 is predicted to

Table 1. *Kinetic analysis of wild-type GlcAT-I and mutants toward UDP-GlcA*

Protein	K_m UDP-GlcA (mM)
Wild type	0.06 ± 0.004
Y84A	N.D.
Y84F	0.18 ± 0.01
Y84H	0.14 ± 0.02
D113A	N.D.
D113E	0.19 ± 0.03
D113N	0.76 ± 0.08
R156A	N.D.
R156K	N.D.
R310A	0.21 ± 0.03
R310K	0.06 ± 0.01
R310Q	0.04 ± 0.006
R161A	N.D.
R161K	1.27 ± 0.2

Membranes of recombinant yeast cells (50 μg per assay) were incubated with increasing concentrations of UDP-GlcA (0–20 mM) at a constant concentration of Galβ1,3Gal (20 mM). Kinetic parameters were determined by nonlinear regression analysis of the data fitted to Michaelis-Menten rate equation. The results are the mean values of two independent determinations on assays performed in duplicate.
N.D., not determined (under limit of detection).

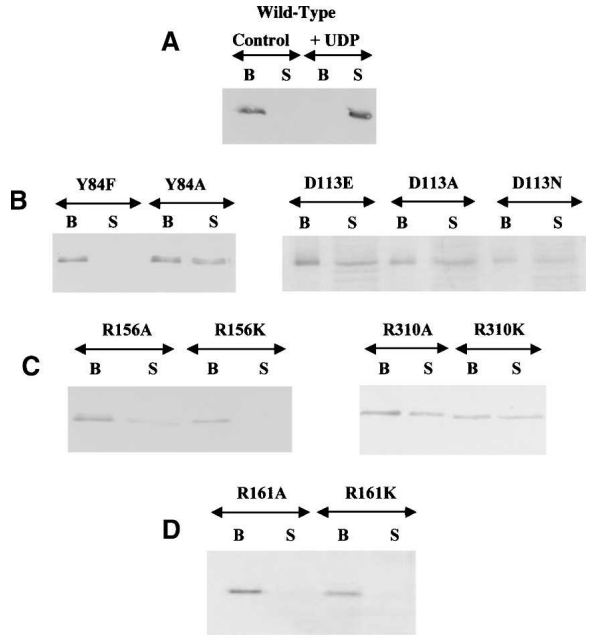

Figure 5. UDP binding of wild-type human GlcAT-I and mutants expressed in yeast *P. pastoris*. UDP-beads were incubated in the presence of solubilized membrane proteins from yeast cells expressing wild-type GlcAT-I and mutants. Proteins bound to UDP-beads (B) and present in the supernatant (S) were separated by centrifugation and subjected to SDS-PAGE followed by immunoblot analysis using anti-GlcAT-I antibodies. Figures are representative of at least three separate experiments.

interact with O1B of the β-phosphate. The role of these residues was investigated by expressing mutants with nonconservative (R to A) and conservative (R to K) substitutions. R156A and R156K mutants were expressed at similar amounts as the wild-type protein (Fig. 4A). These mutations strongly reduced GlcAT-I activity by 90% and 95%, respectively (Fig. 4B), precluding determination of kinetic parameters. However, binding studies revealed that although the mutations were detrimental to the enzyme activity, the mutants R156A and R156K have retained most of their capacity to bind to UDP-beads (Fig. 5C).

On the other hand, mutations of R310 were less deleterious than that of R156 in terms of enzyme activity (Fig. 4B). Substitution of R310 to alanine or lysine reduced the enzyme activity by 70%. K_m value of R310A toward UDP-GlcA was increased about threefold compared with wild-type, whereas that of the R310K mutant was not significantly affected (Table 1). We also substituted R310 by a glutamine residue present at this position in urochordate and arthropod β1,3-glucuronosyltransferase sequences (Fig. 2). Interestingly, this mutation produced about the same effect as the arginine to alanine or lysine substitution (Fig. 4B). Furthermore, the K_m value for the R310Q mutant was unchanged, similarly to the R310K mutant. Binding studies indicated that mutations at position 310 impaired, but did not abrogate, binding of GlcAT-I to UDP-beads (Fig. 5B). Altogether, these data provide evidence for distinguishable roles of the two active site arginine R156 and R310 residues. We show that mutation of R156 produced deleterious effects on GlcAT-I activity, which may not primarily result from impaired interactions of this residue with the UDP part of the nucleotide–sugar, whereas substitutions of R310 affected, to a lesser extent, either binding of the donor substrate or enzyme activity.

Finally, we determined the impact of mutation of R161, which was invariant in all species considered in this study. The R161A and R161K mutants were expressed in higher and slightly lower levels than the wild-type protein, respectively (Fig. 4A). Substitution of the positively charged arginine residue by the neutral alanine amino acid completely inactivated GlcAT-I, precluding determination of kinetic parameters for this mutant. Replacement of R161 by the positively charged lysine residue severely compromised GlcAT-I activity (reduced to 5% of wild type) together with a 20-fold increase in K_m, consistent with the strong conservation of this amino acid (Fig. 4B; Table 1). On the other hand, none of the mutations modified binding to UDP-beads (Fig. 5D). This was in agreement with the predicted role of the guanidino group of R161 in hydrogen-bonding interactions with the 3-OH group of glucuronic acid and not with the UDP part of the donor substrate. Altogether, these data emphasized the critical role of R161 in the function of GlcAT-I, consistent with the strong conservation of this amino acid residue (Fig. 2).

Discussion

The implication of GAGs in a wide range of biological processes has led to a considerable interest in the understanding of the glycosyltransferases involved in their biosynthetic pathway. GlcAT-I, which catalyzes a key step of GAG synthesis, belongs to the β1,3-glucuronosyltransferase family. The aim of this study was to acquire a better knowledge of the organization of these glycosyltransferases by phylogenetic analyses. The sequence similarities observed between the different members of this family suggest that they are dictated by the need for UDP-GlcA recognition, an assumption supported by structural analyses. In the present study, we were able to define the importance of conserved amino acids and to delineate their respective contribution in the recognition of the different parts of the donor substrate by conducting systematic mutagenesis analyses, affinity binding experiments, and kinetic studies.

Phylogeny suggests that the GTA–β1,3-glucuronosyltransferases (for review, see Unligil and Rini 2000) are

rather old enzymes that have been well conserved through evolution, since β1,3-glucuronosyltransferase-like sequences were found in all the eukaryotes tested from vertebrates to invertebrates and plants, with the exception of the most primitive eukaryotes (fungi, protozoa, and algae). However, the two gene duplication events at the origin of the three vertebrate subfamilies (GlcAT-I, GlcAT-P, and GlcAT-S) are relatively recent and have probably occurred early in the vertebrate lineage, since only vertebrates express these three subfamilies. Recent studies have demonstrated that GAGs of the chondroitin- and heparan-sulfate types are present in *D. melanogaster* and in *C. elegans* and bear structural features common to vertebrate polysaccharides (Toyoda et al. 2000). A β1,3-glucuronosyltransferase from *D. melanogaster* referred to as DmGlcAT-I has been shown to present a restricted specificity toward Galβ1,3Galβ-*O*- derivatives, similarly to its vertebrate GlcAT-I counterpart (Kim et al. 2003). Two other drosophila β1,3-glucuronosyltransferases exhibited a broader specificity, and their role in the synthesis of glycosaminoglycans, glycolipids, and glycoproteins has been proposed, although the function of these enzymes was less clearly defined than that of DmGlcAT-I (Kim et al. 2003). Our phylogenetic analysis and functional studies suggest that the implication of the β1,3-glucuronosyltransferases in the completion of glycosaminoglycan linkage region and in the synthesis of the sulfoglucuronyl HNK1 epitope represents an example of convergent evolution, since similar acceptor substrate specificities have evolved in a parallel manner in vertebrates and invertebrates. However, the separation in distinct subfamilies of invertebrates is less apparent than in the vertebrates. In the nematode *C. elegans*, functional studies of the glycosyltransferase encoded by *sqv8* have provided convincing evidence for the involvement of this enzyme in the synthesis of the GAG–protein linkage region (Bulik et al. 2000). In addition, our phylogenetic analysis predicts the presence of other related β1,3-glucuronosyltransferases in this invertebrate organism. It may be hypothesized that these enzymes are involved in the synthesis of HNK1 epitopes as these glycoepitopes have also been described in *C. elegans* (Haltiwanger and Lowe 2004). However, this hypothesis awaits further investigation.

Furthermore, the present phylogenetic analysis led us to include several putative plant β1,3-glucuronosyltransferases in the same gene family. Although a large number of genes encoding UDP-glycosyltransferases have been recently identified in plant genomes, to our knowledge none has been ascribed to this β1,3-glucuronosyltransferase family (Ross et al. 2001). Interestingly, a glucuronosyltransferase has been shown recently to be a major actor of pectin biosynthesis, and thus is being implicated in plant adhesion mechanisms (Iwai et al. 2002). This glucuronosyltransferase is related to the vertebrate EXT2 involved in heparan-sulfate polymerization. As yet, no physiological function has been assigned to the β1,3-glucuronosyltransferases identified in our study. This would be of special interest with regard to plant cell adhesion events.

Multiple alignments of the 119 β1,3-glucuronosyltransferase sequences from distant species highlighted conserved peptide motifs that contain highly conserved amino acids. Among the most conserved from plants to vertebrates, is a tyrosine residue in motif 1 (at position 84 in GlcAT-I). Comparison of the three-dimensional structure of the crystallized glycosyltransferases has provided evidence that the human GlcAT-I and GlcAT-P display a GTA fold with a structurally conserved N-terminal nucleotide–sugar binding domain (Coutinho et al. 2003). This domain consists of a general central parallel β-strand arrangement surrounded by a variable number of helices on both sides, reminiscent of the Rossman fold (Breton et al. 2006). Crystal analysis of GlcAT-I indicated that the side chain of Y84 maintains the uridine base in a parallel ring-stacking arrangement (Pedersen et al. 2000). We document in this study a complete loss of activity resulting from tyrosine to alanine substitution at position 84, substantiating the critical function of Y84. Interestingly, the replacement of Y84 by a phenylalanine significantly restored GlcAT-I activity and binding properties of the enzyme. The presence of a histidine residue in place of tyrosine in the *sqv8*-encoded *C. elegans* glucuronosyltransferase led us to analyze the effect of this substitution in GlcAT-I. Interestingly, the Y84Q mutant retains significant activity with a K_m value toward UDP-GlcA slightly increased compared to that of the wild type. Altogether, these data provide evidence for the major contribution of the phenyl group of Y84 in π-stacking interactions with the uracil base, and that this structural element represents a key feature of the β1,3-glucuronosyltransferase family. Consistently, the side chain of Y93 in GlcAT-P (corresponding to Y84 in GlcAT-I) has been shown to be involved in stacking interaction with the uridine base ring of the donor substrate (Kakuda et al. 2004). The contribution of aromatic stacking interactions in binding the uracil base has also been highlighted in the case of several other glycosyltransferases, including the bovine β4-galactosyltransferase (Gastinel et al. 1999).

Beyond structural similarities, a conserved feature among GTA enzymes consists of the presence of an aspartate residue, which coordinates N-3 of the uracil base (Coutinho et al. 2003). Indeed, a conserved aspartate in motif 2 (corresponding to D113 in GlcAT-I) was identified among the β1,3-glucuronosyltransferases, although this residue was not strictly invariant, especially in plants. Analysis of the X-ray crystal structure of the

human GlcAT-P in complex with UDP indicated that D122 is in hydrogen-bond distance to N-3 of uridine (Kakuda et al. 2004). We show here that replacement of D113 of GlcAT-I by the longer carboxylate chain of glutamic acid or by the amide group of asparagine strongly reduced the enzyme activity together with an increase in K_m value toward UDP-GlcA. A mutational analysis of GlcAT-P fused to glutathione-*S*-transferase (GST) showed that substitution of D122 to alanine decreased the transferase activity to <5% of the wild-type fusion protein (Ohtsubo et al. 2000). However, we observed that none of the conservative or nonconservative mutations of D113 in GlcAT-I resulted in the suppression of the enzyme activity or UDP binding. Altogether, these results indicate that this position may tolerate some amino acid changes, without completely compromising the function of the enzyme.

The GTA glycosyltransferases share a common, though not sequence-invariant ribose/Mn^{2+} coordinating carboxylates, generally termed DXD motif that is found in the conserved motif 4 of the β1,3-glucuronosyltransferase family. We have previously highlighted that the organization of this motif is very similar between SpsA from *Bacillus subtilis* and GlcAT-I with the two last adjacent aspartates of a XDD sequence playing a major role in ribose/Mn^{2+} interactions (Tarbouriech et al. 2001; Gulberti et al. 2003). In plants, D194 and D195 are the best conserved. The biological significance of this finding remains to be investigated.

In the present study, we investigated the role of the two active site arginine residues R156 and R310 located in the vicinity of the pyrophosphate bridge of UDP-GlcA. Crystal structure analysis of GlcAT-I suggested that the linkage region between uridine and GlcA plays a key role in the glycosyltransferase catalytic mechanism (Pedersen et al. 2002; Negishi et al. 2003). In addition to coordinating the Mn^{2+} ions in a conformation typical of bidendate divalent metal-bound nucleotides, the phosphates directly interact with the protein. In UDP–*N*-acetylglucosaminyltransferase I, the α-phosphate makes a salt bridge with R117 and a hydrogen bond to the amide nitrogen of V321, and the β-phosphate hydrogen bonds to the hydroxyl group of S322 (Unligil et al. 2000). In GlcAT-I, atom NH1 of R156 is in hydrogen-bond distance from O2β of the α-phosphate, whereas atom NH1 of R310 may interact with O1β of the β-phosphate (Pedersen et al. 2002). Similarly, in GlcAT-P, phosphoric acid moieties of UDP are recognized by R165 and R313 through hydrogen bonds (Kakuda et al. 2004). The functional importance of R156 was supported by the strong conservation of this residue among the β1,3-glucuronosyltransferase family members and by mutational analysis of this residue in GlcAT-I. Our results indicated that replacement of R156 by alanine or lysine strongly impairs the enzyme activity. Nevertheless, we show that all the mutants have retained most of their ability to bind to the UDP-affinity gel, leading to the suggestion that the interactions between R156 and the pyrophosphate group may not be essential in this respect. Recent crystal analysis of the GlcAT-I–UDP-GlcA complex has highlighted potential interactions between R156 and the GlcA part of the donor substrate (Pedersen et al. 2002). Our results support the idea that the importance of the interactions of R156 with GlcA may predominate over those with the α-phosphate in the functional role of this residue.

On the other hand, we found that R310 partially contributes to the donor substrate interaction and activity. This assertion is substantiated by the observation that the mutations reduced, but did not abrogate, enzyme activity, nor obviate binding of the protein to UDP-beads. These results are consistent with studies performed on GlcAT-P fused to GST (Ohtsubo et al. 2000). In the fusion protein, mutation of R313 (corresponding to R310 in GlcAT-I) did not abolish enzyme activity, leading the authors to the conclusion that the phosphate moiety partially contributes to the donor substrate interaction but is not essential. Supporting this idea, the arginine residue at position 310 in GlcAT-I or 313 in GlcAT-P is replaced by a glutamine in the *D. melanogaster* orthologs forms (DmGlcAT-BSI and BSII; Kim et al. 2003), whereas R156 is invariant in the β1,3-glucuronosyltransferase family. Furthermore, the R310Q GlcAT-I mutant constructed in the present work retains significant activity, indicating that amino acid substitution at this position does not dramatically impair the function of the enzyme.

The position of GlcA in GlcAT-I appears to be determined through a hydrogen-bond network with several conserved residues (Pedersen et al. 2002). We previously emphasized the critical role of the conserved H308 in UDP-GlcA recognition (Ouzzine et al. 2002). Here, we provide evidence for the strict conservation of R161 in all species considered, including plants, and we show that it is essential for enzyme function, likely through hydrogen-bond formation with the 4-OH group of GlcA (Pedersen et al. 2002). To our knowledge, no structural or functional information on the role of this residue is available in other related β1,3-glucuronosyltransferase members.

In conclusion, we describe the phylogenetic organization of the β1,3-glucuronosyltransferase family and suggest that the function of these enzymes in the synthesis of glycosaminoglycan and glucuronosyl epitopes is probably conserved throughout eukaryotic evolution, since conserved β3-glucuronosyltransferase peptide motifs and conserved inter-motif distances were found in most eukaryote species analyzed. The physiological function of the plant β1,3-glucuronosyltransferase-like proteins remains to be investigated. In addition, we have demonstrated that the amino acid residues belonging to the UDP-GlcA binding site perform distinguishable roles and

exhibit different functional importance. We found that D113 and R310 residues are promiscuous to amino acid substitutions. However, R156 and R161 residues, which are strictly conserved among β1,3-glucuronosyltransferases, are critical for activity. In addition, our results emphasize the major contribution of Y84 in UDP binding and activity. Altogether, this work suggests that the conservation of specific residues through evolution may ensure UDP-GlcA donor substrate specificity and the β1,3-glucuronosyltransferase activity of these enzymes.

Materials and methods

Materials

UDP-GlcA, UDP, and anti-rabbit alkaline phosphatase-conjugated immunoglobulins were from Sigma (L'Isle d'Abeau). UDP[^{14}C]-GlcA was from Perkin-Elmer. CYMAL-5 (cyclohexyl-pentyl-β-D-maltoside) was from Interchim. UDP-Fractogel beads were provided by Calbiochem (Merck). Bacterial and yeast culture media were from Difco (Becton Dickinson). Protein assay reagent was obtained from Bio-Rad. The restriction enzymes were provided by New England Biolabs. T4 DNA ligase was from GIBCOBRL-LifeTechnologies. The *P. pastoris* expression system and competent *Escherichia coli* cells were purchased from Invitrogen. The QuikChange site-directed mutagenesis kit was from Stratagene.

Phylogenetic analysis and sequence alignment

Thirty-two β1,3-glucuronosyltransferase ortholog sequences from different species were retrieved from data banks by Psi-Blast (Altschul et al. 1997) with the GlcAT-I (B3G3_HUMAN), GlcAT-P (B3G1_HUMAN), and GlcAT-S (B3G2_HUMAN) sequences. All these sequences belong to the CAZy GT-43 family (http://afmb.cnrs-mrs.fr/CAZY/index.html).

A systematic search for new β1,3-glucuronosyltransferase sequences was performed by TblastN using EST and WGS data banks as previously described for sialyltransferases (Harduin-Lepers et al. 2005). DNA contigs were made by Cap3 (Huang and Madan 1999) for each new β1,3-glucuronosyltransferase-like gene, within each species. New complete protein sequences were analyzed for the presence of the N-terminal transmembrane domain by PHD-htm (Combet et al. 2000) and for the presence of the eight β1,3-glucuronosyltransferase conserved peptide motifs. Eighty-seven new β1,3-glucurunosyltransferase-like sequences were found by these approaches and were submitted to EMBL: AJ557581-2; AJ567371-5; AJ577238-9; AJ577360-2; AJ577569-70; AJ577582-3; AJ577827-8; AJ888971-82; AJ889606-17; AJ890836-45; AJ969044-65; AJ971040-3; AJ971295, and AJ971871-5 (see online supplementary table).

Protein and DNA alignments were performed by ClustalW (Thompson et al. 1994) and saved in Pir format. The Pir alignment was used for the selection of informative positions by G-Blocks (Castresana 2000). Phylogeny analysis was carried out with Phylowin (Galtier et al. 1996) using neighbor joining, observed distances, and 500 bootstrap replicates (Felsenstein 1985). The length of the horizontal branches are a measure of the genetic distance between nodes or between the peripheral side of each branch and a node, based on the number of substitutions per site for a unit branch length (see scale bar in Fig. 1). The genetic distance between two subfamilies is the sum of the branch distances from the periphery of the first subfamily to the first common node plus the distance from this node to the periphery of the second subfamily.

Plasmid construction and site-directed mutagenesis

Cloning of GlcAT-I cDNA and subcloning into pPICZB (Invitrogen) for the heterologous expression of the full-length protein in the methyltrophic yeast *P. pastoris* have been previously described (Ouzzine et al. 2000b). Construction of amino acid-substituted mutants of GlcAT-I was performed using the QuikChange site-directed mutagenesis kit (Stratagene) according to the recommendations of the manufacturer. Mutants were systematically checked by sequencing. The sequence of the sense and antisense mutation primers is indicated in Table 2. The various mutants were individually expressed in the yeast *P. pastoris* as described below.

Heterologous expression in the yeast P. pastoris

Each pPICZ-derived vector containing the native or mutant sequence for GlcAT-I was transformed into the *P. pastoris* SMD1168 yeast strain using the *P. pastoris* Easy Comp Transformation kit (Invitrogen). Transformants were selected on YPD plates containing yeast extract, peptone, dextrose, and 100 μg/mL of Zeocin (Invitrogen). The cells were grown in buffered glycerol/complex medium (BMGY medium) prior to induction by methanol

Table 2. *Sequence of the primers used for site-directed mutagenesis*

Mutant	Primer	Sequence
Y84A	Sense	5′-GTTACCCCCACC**GCT**GCCAGGCTGGTA-3′
	Antisense	5′-TACCAGCCTGGC**AGC**GGTGGGGGTAAC-3′
Y84F	Sense	5′-GTTACCCCCACC**TTT**GCCAGGCTGGTA-3′
	Antisense	5′-TACCAGCCTGGC**AAA**GGTGGGGGTAAC-3′
Y84H	Sense	5′-GTTACCCCCACC**CAT**GCCAGGCTGGTA-3′
	Antisense	5′-TACCAGCCTGGC**ATG**GGTGGGGGTAAC-3′
D113A	Sense	5′-GTGCTGGTGGAG**GCT**GCTGAGGGTCCC-3′
	Antisense	5′-GGGACCCTCAGC**AGC**CTCCACCAGCAC-3′
D113E	Sense	5′-GTGCTGGTGGAG**GAG**GCTGAGGGTCCC-3′
	Antisense	5′-GGGACCCTCAGC**CTC**CTCCACCAGCAC-3′
D113N	Sense	5′-GTGCTGGTGGAG**AAT**GCTGAGGGTCCC-3′
	Antisense	5′-GGGACCCTCAGC**ATT**CTCCACCAGCAC-3′
R156A	Sense	5′-TGGGTTCATCCC**GCT**GGTGTCGAGCAG-3′
	Antisense	5′-CTGCTCGACACC**AGC**GGGATGAACCCA-3′
R156K	Sense	5′-TGGGTTCATCCC**AAG**GGTGTCGAGCAGC-3′
	Antisense	5′-CTGCTCGACACC**CTT**GGGATGAACCCAG-3′
R310A	Sense	5′-GTGTGGCATACT**GCG**ACAGAGAAGCCC-3′
	Antisense	5′-GGGCTTCTCTGT**CGC**AGTATGCACACC-3′
R310K	Sense	5′-GTGTGGCATACT**AAG**ACAGAGAAGCCC-3′
	Antisense	5′-GGGCTTCTCTGT**CTT**AGTATGCACACC-3′
R310Q	Sense	5′-GTGTGGCATACT**CAG**ACAGAGAAGCCC-3′
	Antisense	5′-GGGCTTCTCTGT**CTG**AGTATGCACACC-3′
R161A	Sense	5′-GGTGTCGAGCAG**GCG**AACAAGGCCCTG-3′
	Antisense	5′-CAGGGCCTTGTT**CGC**CTGCTCGACACC-3′
R161K	Sense	5′-GGTGTCGAGCAG**AAG**AACAAGGCCCTG-3′
	Antisense	5′-CAGGGCCTTGTT**CTT**CTGCTCGACACC-3′

(2% [v/v]) for 48 h at 30°C in a rotary shaker (215 rpm). Yeast cells were harvested by centrifugation at 3000*g* for 10 min, broken with glass beads, and further submitted to differential centrifugations as previously described (Ouzzine et al. 1999). The 100,000*g* pellet corresponding to the membrane fraction was resuspended by Dounce homogenization in sucrose-Hepes buffer (0.25 M sucrose, 5 mM HEPES at pH 7.4). Protein concentration was evaluated by the method of Bradford (1976).

Glucuronosyltransferase activity

Galβ1,3Gal was used as acceptor substrate for the measurement of activity of the recombinant human GlcAT-I, as previously described (Lattard et al. 2006). Standard reactions were performed in 100 mM acetate buffer (pH 6.5) containing 10 mM $MnCl_2$, 50 μg membrane protein, 20 mM Galβ1,3Gal, 5 mM UDP-GlcA in a total volume of 50 μL at 37°C for 60 min. We verified that these conditions ensured linearity of the enzyme activity versus time. The reaction product was analyzed by high performance liquid chromatography (HPLC) after chromophore labeling by reductive amination using aniline (Wang et al. 1984). Briefly, the incubation medium (50 μL) was mixed with 5 μL of reagent mixture containing 1 mmol aniline, 350 μL methanol, 35 mg sodium cyanoborohydride, and 41 μL acetic acid, and incubated at 80°C for 30 min. After cooling, 200 μL of water and 1 mL of chloroform were added and mixed for 30 sec. The aniline derivatives were partitioned into the aqueous phase, and the labeled product was analyzed by HPLC on a reverse phase C-18 column (4.6 × 150 mm, 4 μm; Waters) at a detection wavelength of 240 nm. The mobile phase was composed of 5% (v/v) acetonitrile, 0.05% (v/v) 1,4-diethylaminobutane in water (pH 8.0) and used at a flow rate of 0.5 mL/min.

Kinetic parameters were evaluated from initial velocity values of the reaction performed as described above, using varying concentrations of UDP-GlcA (0–20 mM) at a constant concentration of Galβ1,3Gal (20 mM). K_m and V_{max} values were determined by nonlinear least square regression analysis of the data fitted to Michaelis-Menten rate equation ($v = V_{max} \times [S] / K_m + [S]$) using the curvefitter program of Sigmaplot 9.0 (Segel 1975).

UDP-beads binding

The membrane fraction from yeast cells expressing recombinant wild-type or mutant GlcAT-I (500 μg protein) was treated by CYMAL-5 (10 mg/mL 1% [v/v]) at 4°C and submitted to 100,000*g* ultracentrifugation (1 h). The supernatant containing the solubilized material was used for UDP binding experiments according to Gulberti et al. (2003). Briefly, a 20-μL aliquot of UDP-beads (10 μmol UDP/mL gel) was incubated at 4°C for 30 min with the detergent-solubilized material in 100 mM acetate buffer (pH 6.5) containing 10 mM Mn^{2+}. Incubations were also performed in the presence of UDP used as a potential competitor of GlcAT-I binding to UDP-beads. Following incubation, the beads were harvested by centrifugation, the supernatant was saved, and the beads were washed once in acetate buffer solution and then boiled in Laemmli sample buffer. Both the supernatant and the bound material associated with the beads were then analyzed by SDS-PAGE and immunoblotting (Laemmli 1970).

Electrophoresis and protein analysis

Protein samples were separated on SDS-12% (w/v) polyacrylamide gels and electrotransferred to ImmobilonP membranes. The blot was developed using a polyclonal anti-peptide antibody and alkaline phosphatase-conjugated anti-rabbit immunoglobulins as secondary antibodies, as previously described (Ouzzine et al. 2000b). The amount of recombinant wild-type and mutant GlcAT-I protein expressed in membrane fractions of yeast cells was evaluated from Western blot analysis using a calibration curve established with 0.01–0.10 μg purified GlcAT-I protein (Lattard et al. 2006) and run on the same gel, as described (Gulberti et al. 2005). Scanning densitometry was performed using ScionImage software.

Molecular modeling

The model structure of the GlcAT-I complex with UDP-GlcA was built by energy minimization and molecular dynamics calculations based on the initial structure of GlcAT-I (PDB code 1FGG; Pedersen et al. 2000).

Electronic supplemental material

Supplementary material for Electronic Edition: 1, EMBL acc number.doc.

Acknowledgments

This work was supported by grants from ''Fonds National pour la Science'' (ACI no. 0693 ''Biologie Cellulaire, Moléculaire et Structurale'' BMCS152 2004), from Association Nationale pour la Recherche (GT-GAG NT05-3_42251/2005), from IT2B ''Centre National de la Recherche Scientifique'' (CNRS)-''Institut National de la Santé et de la Recherche Médicale'' (INSERM), and from PRO-A INSERM and PHRC Régional programs. This work was also funded partially by CNRS (GDR 2590) and by ''Association pour la Recherche sur le Cancer'' (ARC, 3611).

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72:** 248–254.

Breton, C., Snajdrova, L., Jeanneau, C., Koea, J., and Imberty, A. 2006. Structures and mechanisms of glycosyltransferases superfamily. *Glycobiology* **16:** 29–37.

Bulik, D.A., Wei, G., Toyoda, H., Kinoshita-Toyoda, A., Waldrip, W.R., Esko, J.D., Robbins, P.W., and Selleck, S.B. 2000. *sqv-3*, *-7*, and *-8*, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis. *Proc. Natl. Acad. Sci.* **97:** 10838–10843.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17:** 540–552.

Chou, D.K., Ilyas, A.A., Evans, J.E., Costello, C., Quarles, R.H., and Jungalwala, F.B. 1986. Structure of sulfated glucuronyl glycolipids in the nervous system reacting with HNK-1 antibody and some IgM paraproteins in neuropathy. *J. Biol. Chem.* **261:** 11717–11725.

Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. 2000. NPS@: Network protein sequence analysis. *Trends Biochem. Sci.* **25:** 147–150.

Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. 2003. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328:** 307–317.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using bootstraps. *Evolution Int. J. Org. Evolution* **39:** 783–791.

Galtier, N., Gouy, M., and Gautier, C. 1996. SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12:** 543–548.

Gastinel, L.N., Cambillau, C., and Bourne, Y. 1999. Crystal structures of the bovine β4-galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J.* **18:** 3546–3557.

Gulberti, S., Fournel-Gigleux, S., Mulliert, G., Aubry, A., Netter, P., Magdalou, J., and Ouzzine, M. 2003. The functional glycosyltransferase signature sequence of the human β1,3-glucuronosyltransferase is a XDD motif. *J. Biol. Chem.* **278:** 32219–32226.

Gulberti, S., Lattard, V., Fondeur, M., Jacquinet, J.C., Mulliert, G., Netter, P., Magdalou, J., Ouzzine, M., and Fournel-Gigleux, S. 2005. Phosphorylation and sulfation of oligosaccharide substrates critically influence the activity of human β1,4-galactosyltransferase 7 (GalT-I) and β1,3-glucuronosyltransferase I (GlcAT-I) involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol. Chem.* **280:** 1417–1425.

Haltiwanger, R.S. and Lowe, J.B. 2004. Role of glycosylation in development. *Annu. Rev. Biochem.* **73:** 491–537.

Harduin-Lepers, A., Mollicone, R., Delannoy, P., and Oriol, R. 2005. The animal sialyltransferases and sialyltransferase-related genes: A phylogenetic approach. *Glycobiology* **15:** 805–817.

Huang, X. and Madan, A. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* **9:** 868–877.

Iwai, H., Masaoka, N., Ishii, T., and Satoh, S. 2002. A pectin glucuronyltransferase gene is essential for intercellular attachment in the plant meristem. *Proc. Natl. Acad. Sci.* **99:** 16319–16324.

Kakuda, S., Shiba, T., Ishiguro, M., Tagawa, H., Oka, S., Kajihara, Y., Kawasaki, T., Wakatsuki, S., and Kato, R. 2004. Structural basis for acceptor substrate recognition of a human glucuronyltransferase, GlcAT-P, an enzyme critical in the biosynthesis of the carbohydrate epitope HNK-1. *J. Biol. Chem.* **279:** 22693–22703.

Kim, B.T., Tsuchida, K., Lincecum, J., Kitagawa, H., Bernfield, M., and Sugahara, K. 2003. Identification and characterization of three *Drosophila melanogaster* glucuronyltransferases responsible for the synthesis of the conserved glycosaminoglycan-protein linkage region of proteoglycans. Two novel homologs exhibit broad specificity toward oligosaccharides from proteoglycans, glycoproteins, and glycosphingolipids. *J. Biol. Chem.* **278:** 9116–9124.

Kitagawa, H., Tone, Y., Tamura, J., Neumann, K.W., Ogawa, T., Oka, O., Kawasaki, T., and Sugahara, K. 1998. Molecular cloning of glucuronosyltransferase I involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol. Chem.* **273:** 6615–6618.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature* **227:** 680–685.

Lattard, V., Fondeur-Gelinotte, M., Gulberti, S., Jacquinet, J.C., Boudrant, J., Netter, P., Magdalou, J., Ouzzine, M., and Fournel-Gigleux, S. 2006. Purification and characterization of a soluble form of the recombinant human galactose-β1,3-glucuronosyltransferase I expressed in the yeast *Pichia pastoris*. *Protein Expr. Purif.* **47:** 131–143.

Lin, X. and Perrimon, N. 2002. Developmental roles of heparan sulfate proteoglycans in *Drosophila*. *Glycoconj. J.* **19:** 363–368.

Marcos, I., Galan, J.J., Borrego, S., and Antinolo, G. 2002. Cloning, characterization, and chromosome mapping of the human GlcAT-S gene. *J. Hum. Genet.* **47:** 677–680.

McGarry, R.C., Helfand, S.L., Quarles, R.H., and Roder, J.C. 1983. Recognition of myelin-associated glycoprotein by the monoclonal antibody HNK-1. *Nature* **306:** 376–378.

Negishi, M., Dong, J., Darden, T.A., Pedersen, L.G., and Pedersen, L.C. 2003. Glucosaminylglycan biosynthesis: What we can learn from the X-ray crystal structures of glycosyltransferases GlcAT1 and EXTL2. *Biochem. Biophys. Res. Commun.* **303:** 393–398.

Ohtsubo, K., Imajo, S., Ishiguro, M., Nakatani, T., Oka, S., and Kawasaki, T. 2000. Studies in the structure-function relationship of the HNK-1 associated glucuronosyltransferase, GlcAT-P, by computer modeling and site-directed mutagenesis. *J. Biochem.* **128:** 283–291.

Ouzzine, M., Magdalou, J., Burchell, B., and Fournel-Gigleux, S. 1999. Expression of functionally active human hepatic UDP-glucuronosyltransferase (UGT1A6) lacking the N-terminal signal sequence in the endoplasmic reticulum. *FEBS Lett.* **454:** 187–191.

Ouzzine, M., Antonio, L., Burchell, B., Netter, P., Fournel-Gigleux, S., and Magdalou, J. 2000a. Importance of histidine residues for the function of the human liver UDP-glucuronosyltransferase UGT1A6: Evidence for the catalytic role of histidine 370. *Mol. Pharmacol.* **58:** 1609–1615.

Ouzzine, M., Gulberti, S., Netter, P., Magdalou, J., and Fournel-Gigleux, S. 2000b. Structure/function of the human Ga1β1,3-glucuronosyltransferase. Dimerization and functional activity are mediated by two crucial cysteine residues. *J. Biol. Chem.* **275:** 28254–28260.

———. 2002. The donor substrate specificity of the human β1,3-glucuronosyltransferase I (GlcAT-I) towards UDP-glucuronic acid is determined by two crucial histidine and arginine residues. *J. Biol. Chem.* **277:** 25439–25445.

Pedersen, L.C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T.A., and Negishi, M. 2000. Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. *J. Biol. Chem.* **275:** 34580–34585.

Pedersen, L.C., Darden, T.A., and Negishi, M. 2002. Crystal structure of β1,3-glucuronyltransferase I in complex with active donor substrate UDP-GlcUA. *J. Biol. Chem.* **277:** 21869–21873.

Ross, J., Li, Y., Lim, E., and Bowles, D.J. 2001. Higher plant glycosyltransferases. *Genome Biol.* **2:** http://genomebiology.com.

Schwartz, N.B. and Domowicz, M. 2002. Chondrodysplasias due to proreoglycan defects. *Glycobiology* **12:** 57–68.

Segel, I.H. 1975. *Enzyme kinetics.*. John Wiley & Sons, New York.

Tarbouriech, N., Charnock, S.J., and Davies, G.J. 2001. Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: A comparison with related NDP-sugar glycosyltransferases. *J. Mol. Biol.* **314:** 655–661.

Terayama, K., Oka, S., Sejki, T., Miki, Y., Nakamura, A., Kozutsumi, Y., Takio, K., and Kawasaki, T. 1997. Cloning and functional expression of a novel glucuronosyltransferase involved in the biosynthesis of the carbohydrate epitope HNK1. *Proc. Natl. Acad. Sci.* **94:** 6093–6098.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Toyoda, H., Kinoshita-Toyoda, A., and Selleck, S.B. 2000. Structural analysis of glycosaminoglycans in *Drosophila* and *Caenorhabditis elegans* and demonstration that *tout-velu*, a *Drosophila* gene related to EXT tumor suppressors, affects heparan sulfate *in vivo*. *J. Biol. Chem.* **275:** 2269–2275.

Unligil, U.M. and Rini, J.M. 2000. Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **10:** 510–517..

Unligil, U.M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H., and Rini, J.M. 2000. X-ray crystal structure of rabbit *N*-acetylglucosaminyltransferase I: Catalytic mechanism and a new protein superfamily. *EMBO J.* **19:** 5269–5280.

van Horssen, J., Wesseling, P., van den Heuvel, L.P., de Waal, R.M., and Verbeek, M.M. 2003. Heparan sulphate proteoglycans in Alzheimer's disease and amyloid-related disorders. *Lancet Neurol.* **2:** 482–492.

Venkatesan, N., Barré, L., Benani, A., Netter, P., Magdalou, J., Fournel-Gigleux, S., and Ouzzine, M. 2004. Stimulation of proteoglycan synthesis by glucuronosyltransferase-I gene delivery: A strategy to promote cartilage repair. *Proc. Natl. Acad. Sci.* **101:** 18087–18092.

Wang, W.T., Le Donne, N.C., Ackerman, B., and Sweely, C.C. 1984. Structural characterisation of oligosaccharides by high-performance liquid chromatography, fast-atom bombardment-mass and exoglycosidase digestion. *Anal. Biochem.* **141:** 366–381.

Wegrowski, Y. and Maquart, F.X. 2004. Involvement of stromal proteoglycans in tumour progression. *Crit. Rev. Oncol. Hematol.* **49:** 259–268.

Yanagishita, M. and Hascall, V.C. 1992. Cell surface heparan sulfate proteoglycans. *J. Biol. Chem.* **267:** 9451–9454.