

# Characterization of a UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase with an unusual lectin domain from the platyhelminth parasite *Echinococcus granulosus*

Teresa FREIRE\*, Cecilia FERNÁNDEZ†<sup>1</sup>, Cora CHALAR‡, Rick M. MAIZELS†, Pedro ALZARI§, Eduardo OSINAGA\*<sup>2</sup> and Carlos ROBELLO\*

\*Departamento de Bioquímica, Facultad de Medicina, Universidad de la República, Av. Gral. Flores 2125, Montevideo CP 11800, Uruguay, †Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh, U.K., ‡Sección Bioquímica, Instituto de Biología, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, and §Unité de Biochimie Structurale, URA 2185 CNRS, Institut Pasteur, Paris, France

As part of a general project aimed at elucidating the initiation of mucin-type O-glycosylation in helminth parasites, we have characterized a novel ppGalNAc-T (UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase) from the cestode *Echinococcus granulosus* (Eg-ppGalNAc-T1). A full-length cDNA was isolated from a library of the tissue-dwelling larval stage of the parasite, and found to code for a 654-amino-acid protein containing all the structural features of ppGalNAc-Ts. Functional characterization of a recombinant protein lacking the transmembrane domain showed maximal activity at 28 °C, in the range 6.5–7.5 pH units and in the presence of Cu<sup>2+</sup>. In addition, it transferred GalNAc to a broad range of substrate peptides, derived from human mucins and O-glycosylated parasite proteins, including acceptors containing only serine or only threonine residues. Interestingly, the C-terminal region of Eg-ppGalNAc-T1 bears a highly unusual lectin domain, considerably longer than the one from other members of the family, and including only one of the

three ricin B repeats generally present in ppGalNAc-Ts. Furthermore, a search for conserved domains within the protein C-terminus identified a fragment showing similarity to a recently defined domain, specialized in the binding of organic phosphates (CYTH). The role of the lectin domain in the determination of the substrate specificity of these enzymes suggests that Eg-ppGalNAc-T1 would be involved in the glycosylation of a special type of substrate. Analysis of the tissue distribution by *in situ* hybridization and immunohistochemistry revealed that this transferase is expressed in the hydatid cyst wall and the subtegumental region of larval worms. Therefore it could participate in the biosynthesis of O-glycosylated parasite proteins exposed at the interface between *E. granulosus* and its hosts.

**Key words:** *Echinococcus granulosus*, helminth parasite, O-glycosylation, UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase.

## INTRODUCTION

Helminth parasites are long-lived metazoa, causing chronic infections of medical and veterinary importance. These organisms, which are not a natural group, belong to mainly two phyla: the Nematoda (roundworms) and the Platyhelminthes (flatworms), with the latter including the classes Cestoda (tapeworms) and Trematoda (flukes). The dog tapeworm *Echinococcus granulosus* is an agent of hydatid disease, a major zoonosis on a worldwide scale. Cystic echinococcosis (hydatidosis), caused by the larval stage of the parasite, is acquired from the ingestion of eggs excreted with dog faeces and produces clinical disease in human and economical losses to the livestock industry. The larva dwells in the viscera of intermediate hosts; it has the form of a fluid-filled cyst, bounded by a cyst wall. The hydatid fluid contains host proteins as well as parasite excretion/secretion products. The cyst wall comprises an innermost 'germinal layer' of live parasite tissue, which synthesizes an outer, carbohydrate-rich 'laminated layer'. The latter structure is unique to the genus *Echinococcus* and its biosynthesis represents a major metabolic activity of the germinal

layer; it plays a key role in the establishment and persistence of infection by preventing the access of host cells to the live parasite. The germinal layer also gives origin, through budding towards the interior of the cyst, to the larval worms or protoscolecemes. These stages are capable of infecting dogs and maturing to adult worms; for this reason, the cysts containing protoscolecemes are said to be fertile [1].

Parasite glycoconjugates, mainly those present on the surface and in secretion products, appear to play critical roles in the interaction of helminths with their hosts. In particular, O-glycans and mucin-like molecules have been implicated in host recognition and avoidance of immune responses [2]. This is the case, for example, for O-linked glycans present in the glycocalyx of cercariae from the trematode *Schistosoma mansoni* that would be involved in the penetration of the mammalian host, and of an extensively characterized family of mucin-like proteins participating in immune evasion, which are constituents of both the surface coat and secretion products of infective larvae from the nematode *Toxocara canis* [3]. For cestodes, a detailed study has recently

Abbreviations used: dOSM, deglycosylated ovine submaxillary mucin; EST, expressed sequence tag; ppGalNAc-T, UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase; RT, reverse transcriptase.

<sup>1</sup> Present address: Cátedra de Inmunología, Facultad de Química, Universidad de la República, Montevideo, Uruguay.

<sup>2</sup> To whom correspondence should be addressed (email eosinaga@fmed.edu.uy).

The expressed sequence tag and full-length cDNA sequences corresponding to Eg-ppGalNAc-T1 have been submitted to the DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under the accession numbers BI244133 and AY353720 respectively. We have also submitted sequence data to the above-mentioned databases for the following sequences: gly-3 (AF031833), gly-4 (AF031834), gly-5a (AF031835), gly-5b (AF031836), gly-5c (AF031837), gly-6a (AF031838), gly-7 (AF031841), gly-8 (AF031842) and gly-9 (AF031843).

demonstrated that a major antigen from the laminated layer of *E. multilocularis* is a mucin-type glycosylated protein [4].

Over the past years, we have been involved in the study of the initiation pathway of mucin-type O-glycosylation in helminth parasites. In this context, we described the presence of the simple mucin type Tn antigen (Thr/Ser-O-GalNAc), one of the most specific human tumour-associated structures [5], in larval and adult tissues of *E. granulosus* [6] and, subsequently, in other species belonging to the two main helminth phyla [7,8], thus making the interesting observation that truncated O-glycosylation appears to be widespread among these organisms. We also started to analyse the biosynthesis of Tn structures by evaluating ppGalNAc-T (UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase) activity (EC 2.4.1.41) in *T. canis*, *Fasciola hepatica* and *E. granulosus* [7,8]. Furthermore, during an ongoing characterization of the transcriptome of *E. granulosus* larval stages [9], we isolated a cDNA clone coding for a novel ppGalNAc-T. The enzymes from this family, which catalyse the first step in the biosynthesis of O-glycans, i.e. the transfer of GalNAc to serine or threonine residues in polypeptides, represent key regulatory factors to define the repertoire of such structures expressed by a cell [10]. They belong to the family 27 of retained nucleotide-diphospho-sugar transferases based on amino acid sequence similarities [11–13]. To date, 14 distinct members have been cloned in mammals [14–28] and it is predicted that most of these isoforms will have different functions, in view of the kinetic properties and unique substrate specificities described for several of them [29]. It has been estimated that ppGalNAc-Ts underwent gene duplication before the divergence of deuterostomes and protostomes [10]. The family has indeed been identified, *in silico* and biochemically, in the free-living nematode *Caenorhabditis elegans* [30] and in the fruit fly *Drosophila melanogaster* [31], both members of the Ecdysozoa, one of the major superphyla of protostomes [32,33]. Orthologous isoforms have also been described between mammalian ppGalNAc-Ts and the enzymes from these two model organisms [10,31]. Moreover, Wojczyk et al. [34] have recently characterized a ppGalNAc-T from the protozoan *Toxoplasma gondii* and have mentioned the existence of additional isoforms in this parasite.

In the present study, we describe a novel *E. granulosus* ppGalNAc-T (designated Eg-ppGalNAc-T1). We analyse the structural features of the protein predicted from the cDNA, functionally characterize the corresponding recombinant enzyme and study its localization in the parasite larval stages. From an evolutionary perspective, it is worth noting that Eg-ppGalNAc-T1 is the first ppGalNAc-T to be described from a representative of the Lophotrochozoa, the sister group of the Ecdysozoa, i.e. the other major superphylum of protostomes.

## EXPERIMENTAL

### Parasite materials

Hydatid cysts were obtained from infected bovines at local abattoirs in Uruguay. Protoscoleces were recovered by aseptic aspiration of the cyst contents, whereas cyst walls (germinal and laminated layers) were carefully dissected from open cysts and separated from host tissue using forceps. Protoscoleces and cyst walls were thoroughly washed with sterile PBS and fixed for histological studies in freshly prepared 4% (w/v) paraformaldehyde in PBS for 1 h. The material was then washed several times in PBS, dehydrated through increasing concentrations of methanol in PBS and embedded in paraffin wax. Tissue sections of 5–7  $\mu\text{m}$  were cut, mounted on to polylysine-coated slides, dried and stored at 4 °C.

### Cloning and sequence analysis of Eg-ppGalNAc-T1

A cDNA clone containing the full-length sequence of Eg-ppGalNAc-T1 was isolated from a spliced-leader cyst wall library in the context of an *E. granulosus* EST (expressed sequence tag) project. cDNA was prepared by RT (reverse transcriptase)-PCR and directionally cloned into pSPORT1 (Gibco BRL, Paisley, Renfrewshire, Scotland, U.K.) as described previously [9]. The GenBank® accession numbers of the EST and full-length cDNA corresponding to Eg-ppGalNAc-T1 are BI244133 and AY353720 respectively. Amino acid sequences of ppGalNAc-Ts were aligned with ClustalW (1.82).

### Expression in COS-7 cells and purification of Eg-ppGalNAc-T1

A partial sequence of Eg-ppGalNAc-T1 lacking the predicted transmembrane domain was PCR-amplified from the recombinant pSPORT plasmid and cloned in an expression vector using Gateway Technology according to the manufacturer's instructions (Invitrogen, Leek, The Netherlands) as outlined below. The PCR product was designed to encode a recombinant enzyme including amino acids 59–654 of the protein sequence translated from the cDNA; it was amplified with *Taq* DNA polymerase and the primers N1F, 5'-gggacaagttgtacaaaaagcaggcttcgaaggatagaa-cctatGAAACTCTGAAGGCACCTTCAG-3' and N2R, 5'-gggg-accacttgtacaagaaagctgggtcctaAGATGTCTAGCCGTGGTGAC-TG-3' (lower-case letters indicate the sequences required for the Gateway TM BP clonase reaction; Invitrogen, Carlsbad, CA, U.S.A.). The purified PCR product was inserted into a donor vector by BP recombination (through the *attB* and *attP* recombination sites) to create entry clones. The 10  $\mu\text{l}$  BP cloning reaction [2  $\mu\text{l}$  of BP reaction buffer, 1  $\mu\text{l}$  of pDONR201 vector at 150 ng/ $\mu\text{l}$ , 1  $\mu\text{l}$  (approx. 40 ng) PCR product, 2  $\mu\text{l}$  of BP Clonase Enzyme mix, TE (10 mM Tris/HCl and 1 mM EDTA, pH 8) added to a final volume of 10  $\mu\text{l}$ ] was incubated overnight at room temperature (25 °C). The BP reaction was then treated with 1  $\mu\text{l}$  of proteinase K (2  $\mu\text{g}/\mu\text{l}$ ) for 10 min at 37 °C, and 1  $\mu\text{l}$  was transformed into 50  $\mu\text{l}$  of library efficiency *E. coli* DH5 $\alpha$  chemically competent cells. Transformants were selected on Luria-Bertani plates containing 50  $\mu\text{g}/\text{ml}$  kanamycin and the positive clones were confirmed by PCR and sequencing. Finally, Eg-ppGalNAc-T1 in entry clones was transferred into a destination vector by LR recombination (through the *attL* and *attR* recombination sites) to create expression clones: the 10  $\mu\text{l}$  of LR cloning reaction [2  $\mu\text{l}$  of LR reaction buffer, 1  $\mu\text{l}$  of linearized pDEST 12.2 vector at 150 ng/ $\mu\text{l}$ , 1  $\mu\text{l}$  (approx. 200 ng) entry clone, 2  $\mu\text{l}$  of LR Clonase Enzyme mix, TE added to a final volume of 10  $\mu\text{l}$ ] was incubated overnight at 25 °C, then treated with proteinase K and transformed into DH5 $\alpha$  cells as described above. Transformants were recovered on Luria-Bertani-ampicillin (100  $\mu\text{g}/\text{ml}$ ) plates. A positive clone (pDEST12.2-EgT1) was selected, sequenced and used for protein expression.

COS-7 cells were grown to 90% confluency in Dulbecco's modified Eagle's medium (Life Technologies), containing 10% (v/v) fetal bovine serum, 5% 200 mM pyruvate and 5% 200 mM glutamine at 37 °C in an atmosphere of 5% CO<sub>2</sub> and electroporated with pDEST12.2-EgT1. After selection of transformants with 1 mg/ml neomycin, the presence and expression of the gene were confirmed by RT-PCR and Western-blot analysis. The recombinant protein was affinity purified from the cell extracts on a column prepared with the purified IgG fraction of a polyclonal antiserum raised against a fragment of Eg-ppGalNAc-T1 catalytic domain (see below) immobilized on to CNBr-activated Sepharose. Cell extracts were re-circulated for 4 h at 4 °C; the column was washed with PBS containing 0.7 M NaCl and

**Table 1** Eg-ppGalNAc-T1 activity towards synthetic peptide acceptors

Polypeptidic acceptor	Amino acid sequence	Reference	ppGalNAc-T activity (pmol GalNAc/mg)	Relative activity (%)
Human MUC1	DTRPAPGSTA	[48]	57.4 ± 3.0	18
Human MUC2	PTTTPITTTTV	[49]	56.8 ± 4.9	18
Human MUC5AC	KGGGTTSTTSAP	[50]	53.6 ± 0.2	17
Human MUC5B	VLTTTATPTA	[51]	51.4 ± 1.5	16
<i>T. brucei</i> mucin	SSLSSFFASSAVG	[52]	58.4 ± 2.0	19
<i>T. cruzi</i> mucin	KPPTTTTTTKPP	[53]	54.2 ± 11.1	17
<i>S. mansoni</i> mucin	ISTSPSPNITTTT	[54]	312.4 ± 36.9	100
dOSM	–	–	46.8 ± 11.3	15
Without acceptor	–	–	12.0 ± 3.0	4

eluted using 0.1 M triethylamine buffer (pH 10.5). The purified fractions were immediately neutralized with phosphate buffer and found to yield a single band of the mass predicted for the recombinant protein, when analysed by SDS/PAGE (10% gel) followed by silver staining or Western blotting with the anti-Eg-ppGalNAc-T1 serum, using conventional methods. A pool of these fractions (0.25 mg/ml) was used for the characterization of enzymic activity.

#### Functional characterization of Eg-ppGalNAc-T1

The ppGalNAc-T activity was initially assessed using a poly-acceptor substrate as described previously [8]. Briefly, the standard reaction mixture (50 µl) contained 23 µl of enzyme preparation, 5 µg of dOSM (deglycosylated ovine submaxillary mucin) and 150 µM UDP-[<sup>3</sup>H]GalNAc (approx. 80 000 d.p.m. and 17 000 d.p.m./pmol respectively) in 50 mM imidazole/HCl (pH 7.2), 10 mM MnCl<sub>2</sub> and 0.5% Triton X-100. The enzymic activity towards dOSM was evaluated at different temperatures, pH conditions (in the range 5.0–9.0) and in the presence of various bivalent cations (MnCl<sub>2</sub> was substituted with 10 mM of the chloride salt from each cation). In addition, the transfer of GalNAc was analysed in synthetic peptides corresponding to the tandem repeats of highly glycosylated proteins from *Trypanosoma cruzi*, *Trypanosoma brucei*, *S. mansoni* and human mucins (MUC1, MUC2, MUC5AC and MUC5B) (Table 1). For these assays, each mixture containing 2 mM peptide was incubated at 28 °C for 2 h and the reaction was terminated by the addition of 50 µl of 250 mM EDTA. The glycosylated peptides were separated from unreacted UDP-[<sup>3</sup>H]GalNAc on a 1 ml AG1X-8 column (Cl<sup>-</sup> form; Sigma, St. Louis, MO, U.S.A.), eluted with 2.6 ml of water, and the radioactivity was measured with a Beckman LS Analyzer scintillation counter. Activity was expressed as pmol of transferred [<sup>3</sup>H]GalNAc · (mg of protein)<sup>-1</sup> · h<sup>-1</sup>. All experiments were performed in triplicate and corrected for background values, measured from reaction mixtures lacking the corresponding acceptor substrate.

#### Preparation of polyclonal antibodies

Fragments from Eg-ppGalNAc-T1 corresponding to portions of the catalytic domain (amino acids 357–461) or the C-terminal region (amino acids 544–654) were PCR-amplified from the pSPORT recombinant plasmid using, respectively, the following sense and antisense primer pairs: EgCD-F, 5'-cgggatccTGGGG-TGGTGAGAACTTGGAG-3' and EgCD-R, 5'-cgaagcttctaAGG-AGCAATGTTGTCAAGAAACC-3' or EgSP-F, 5'-cgggatccTG-GGTCTTCAGTGAACCCACC-3' and EgSP-R, 5'-cgaagcttAC-GCAGATGTCTAGCCGTGG-3' (lower-case letters indicate restriction sites for *Bam*HI and *Hind*III in the forward and re-

verse primers respectively). After digestion with *Bam*HI and *Hind*III, each product was cloned into the pQE30 vector (Qiagen, Chatsworth, CA, U.S.A.), so as to encode for a protein carrying a His<sub>6</sub> tail at the N-terminus. The recombinants were expressed in *E. coli* M15 by induction with 0.5 mM isopropyl β-D-thiogalactoside, purified over Ni<sup>2+</sup>-nitrilotriacetate columns under denaturing conditions according to the manufacturer's instructions (Qiagen), and used to prepare rabbit antisera. A subcutaneous priming of 200 µg of each protein in complete Freund's adjuvant (Sigma) was followed by two injections of 200 µg in incomplete Freund's adjuvant (Sigma), after 2 and 4 weeks. The antisera were collected 2 weeks after the last antigen dose and checked by ELISA. The IgG fraction from the serum, raised against the catalytic domain, was column-purified over Protein A and used to isolate the recombinant enzyme as described above. The sera and purified antibodies were stored at –80 °C until use.

#### In situ hybridization studies

Sense and antisense digoxigenin-labelled riboprobes were generated from the subclones constructed to prepare the protein fragments used to raise antisera. These span nt 1071–1383 (312 bp, coding for amino acids 357–461) and 1632–1965 (334 bp, coding for amino acids 544–654) of Eg-ppGalNAc-T1 cDNA. Parasite sections were deparaffinated and rehydrated through 100, 75, 50 and 25% methanol in PBS containing 0.1% Tween 20 (PBST). After two washes of 10 min in PBST, they were incubated for 15 min with proteinase K (1 µg/ml in PBST), washed twice for 5 min in PBST, refixed for 20 min with 4% paraformaldehyde in PBST containing 0.1% glutaraldehyde and washed twice with PBST. Then, the sections were heated at 65 °C for 10 min to inactivate endogenous alkaline phosphatases and prehybridized for 2 h at 50 °C with hybridization solution [50% (v/v) deionized formamide, 1.3 × SSC (1 × SSC is 0.3 M NaCl and 0.03 M sodium citrate), pH 5.3, 5 mM EDTA, 50 µg/ml yeast tRNA, 0.2% Tween 20, 0.5% CHAPS, 100 µg/ml heparin]. Hybridization was performed overnight in a humid chamber at 50 °C with 2 ng/µl of denatured (95 °C, 3 min) digoxigenin-labelled RNA probe in 50 µl of hybridization solution. After hybridization, the sections were washed twice at 50 °C for 20 min with 2 × SSC and then with 1 × SSC, pH 7.2, then twice at room temperature for 15 min with 1 × SSC. They were subsequently rinsed twice for 10 min with MAB-T [0.1% Tween 20 in MAB (100 mM maleic acid, pH 7.5, and 150 mM NaCl)], and blocked with MAB containing 1% BSA (MAB-B) for 30 min, followed by MAB-B including 5% (v/v) heat-inactivated sheep serum (blocking solution) for 2 h. The sections were incubated overnight at 4 °C with alkaline phosphatase-conjugated anti-digoxigenin antibody (Boehringer Mannheim, Mannheim, Germany), diluted to 1/500 in blocking solution. After washing with MAB-T, the sections were rinsed twice with developing buffer (100 mM NaCl, 100 mM Tris/HCl, pH 9.5, 50 mM MgCl<sub>2</sub> and 0.1% Tween 20) and incubated with Nitro Blue Tetrazolium and 5-bromo-4-chloroindol-3-yl phosphate in the same buffer. After colour development, the sections were washed again twice with PBST, fixed in 4% paraformaldehyde in PBS for 30 min, mounted under a coverslip using 70% (v/v) glycerol. Digital images of parasite tissues were captured under identical conditions using a CoolSNAP-Pro Monochrome Digital kit (Media Cybernetics, San Diego, CA, U.S.A.); contrast and brightness were corrected with the Image Pro Plus software (Media Cybernetics).

#### Immunohistochemical analysis

The sections were deparaffinated and rehydrated as described for *in situ* hybridization studies. After the inactivation of endogenous

peroxidase activity with 3% H<sub>2</sub>O<sub>2</sub> in 70% ethanol for 15 min, they were treated with blocking solution [0.1% Tween 20, 5% (w/v) BSA, 1% heat-inactivated horse serum, 2% (w/v) glycine in PBS] for 45 min, and incubated overnight at 4 °C with antisera raised against recombinant fragments corresponding to Eg-ppGalNAc-T1 catalytic domain or C-terminus, diluted 1:100 in PBS. At the end of the incubation period, the sections were washed in PBS and incubated with goat anti-rabbit antibodies conjugated with Alexa-Fluor 488 (final dilution 1:400; Molecular Probes, Eugene, OR, U.S.A.) for 1.5 h at room temperature, washed with PBS and mounted in antifade mounting medium (Citifluor, London, U.K.). They were examined with a Nikon E800 microscope equipped with epifluorescence and fitted with the appropriate filters. Digital images were captured within 24 h after immunostaining using a CoolSNAP-Pro Monochrome Digital camera without any contrast or brightness corrections.

## RESULTS

### Eg-ppGalNAc-T1 is a typical member of the ppGalNAc-T family

A new member of the ppGalNAc-T family (family 27 of glycosyltransferases, GT27, [12]) has been cloned from an *E. granulosus* cyst wall cDNA library, and named Eg-ppGalNAc-T1. The presence of the *Echinococcus* spliced leader [35] at the 5'-end of the clone and of a putative polyadenylated tail at its 3'-end indicate that it corresponds to a full-length cDNA. Conceptual translation reveals an open reading frame of 1962 nt encoding for a 654 amino acid protein with a predicted molecular mass of 74.8 kDa (Figure 1A). Similarly to all members of the ppGalNAc-T family, the translated polypeptide is predicted to be a typical type II membrane protein comprising an N-terminal cytoplasmic region (20 amino acids), a transmembrane anchor (23 amino acids), a stem region (72 amino acids) and a putative catalytic domain (344 amino acids, between residues 116 and 459), which includes a glycosyltransferase 2 (GT2) domain [36], spanning positions 151–345. Figure 1(A) shows the distribution of domains in the sequence of the *E. granulosus* transferase together with an alignment with the entire *C. elegans* ppGalNAc-T family. To avoid nomenclature confusions, it may be pertinent to mention that the two structural motifs identified by Hagen et al. [37] within the catalytic domain of ppGalNAc-Ts and named GT-1 and Gal/GalNAc-T are also present in Eg-ppGalNAc-T1 (residues 147–260 and 327–390 respectively).

The catalytic domain of Eg-ppGalNAc-T1 has 62% similarity with the closest member of the family, namely, the isoform gly-9 from *C. elegans* [30]; it includes the seven critical acidic residues (at positions 159, 188, 243, 247, 352, 361 and 364), two histidine residues (245 and 386) that appear to be involved in the enzymic activity of other ppGalNAc-Ts, as determined by site-directed mutagenesis studies [37] and four conserved cysteine residues (138, 372, 381 and 449). It also contains the sequon DXH (at positions 243–245) that would be a binding site for the bivalent cation and the sugar donor; and the motif DXXXXXWGGENXE (352–364), with three of the seven carboxylates and other critical residues in the subsequence WGG (reviewed in [10]).

### Eg-ppGalNAc-T1 contains an unusual lectin domain

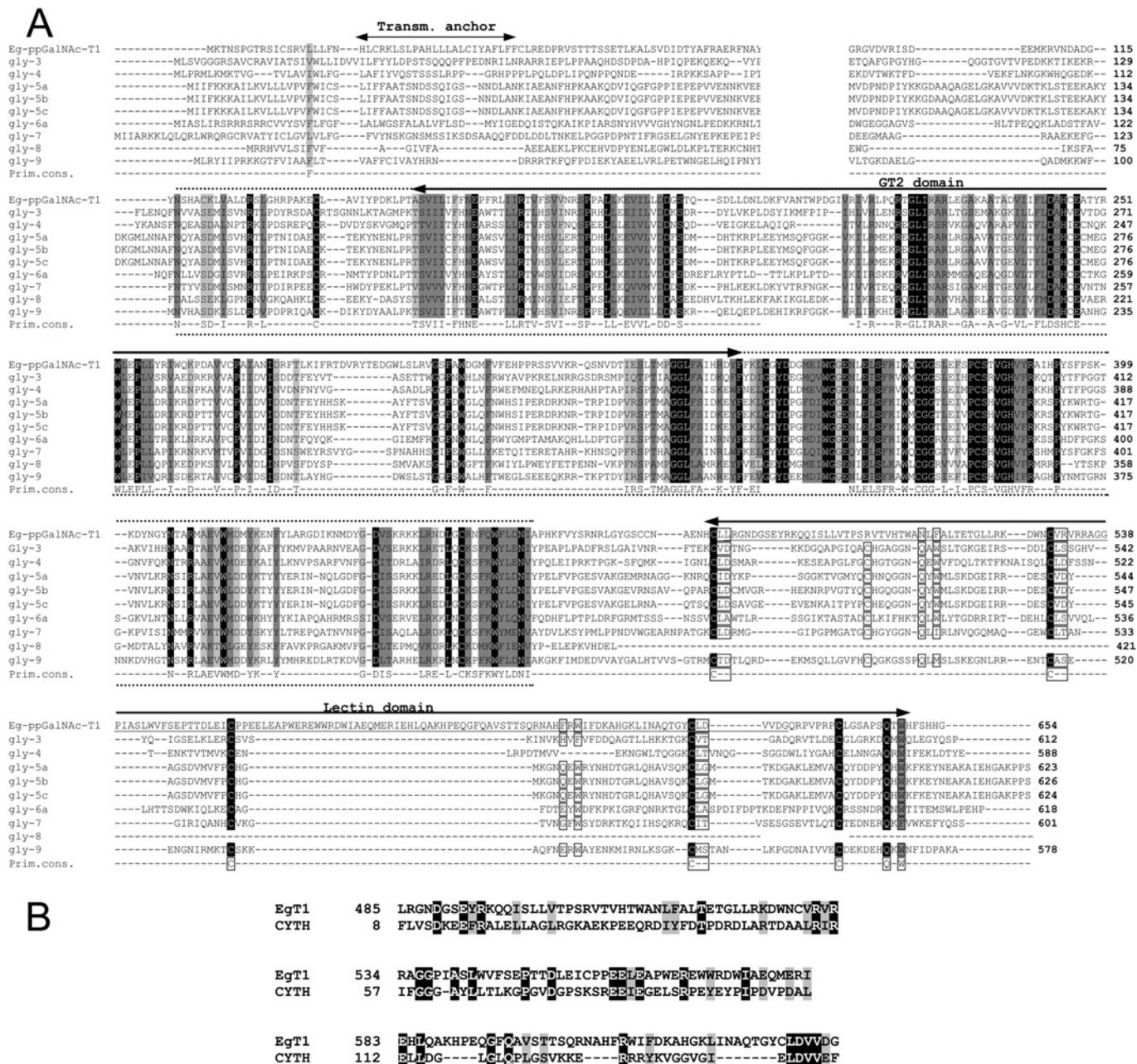
Similarly to virtually all known members of this protein family, the C-terminus of the *Echinococcus* transferase corresponds to a putative lectin domain. However, this has some highly unusual features in the novel enzyme. The lectin domains of ppGalNAc-Ts, although very diverse in overall sequence, are generally of the

'ricin B'-type (classified as carbohydrate-binding module 13; see Coutinho et al. [13]); they are composed of three homologous subdomains ( $\alpha$ ,  $\beta$  and  $\gamma$ ) of approx. 40 amino acids, each of them containing a characteristic 'QXW' pattern and a disulphide bond, whose first cysteine residue is followed by hydrophobic and polar amino acids (usually, a 'CLD/E' motif). These structural elements appear to be essential for proper folding and lectin function [38]. Eg-ppGalNAc-T1 has only one typical ricin subdomain, the one corresponding to the  $\gamma$  repeat (between amino acids 614 and 650) that includes the CLD (625–627) and QTW (646–648) motifs, as well as a conserved cysteine residue (639). The  $\alpha$  repeat, in turn, contains the CLL (483–485) and NLF (514–516) sequons, which align with CLD and QXF from most homologous enzymes; however, no second cysteine residue capable of forming a disulphide bond is present. In addition, although some traits of a putative  $\beta$  subdomain can be identified [CVR (530–532), cysteine (556) and a conserved tryptophan residue (608)], this repeat appears unique in that it is approx. 50 amino acids longer than the corresponding region of other ppGalNAc-Ts. A search for conserved protein domains in the C-terminus of Eg-ppGalNAc-T1 (amino acids 460–654) yielded no homology with ricin B, despite the presence of the  $\gamma$  subdomain and other elements of this lectin motif. Interestingly, the fragment comprising residues 532–607 (76 amino acids), which corresponds to the portion 'inserted' within the  $\beta$  repeat of other ppGalNAc-Ts, shows 33% similarity with a recently defined protein domain specialized in the binding of organic phosphates (CYTH) [39] and includes three of the four basic residues predicted to participate in phosphate binding (arginine residues at positions 532, 534 and 607) (Figure 1B).

### Eg-ppGalNAc-T1 has a distinct metal dependence and broad substrate specificity

A recombinant protein lacking the putative transmembrane domain of the enzyme was expressed in COS-7 cells. Transfection of the gene was confirmed by RT-PCR and expression of the predicted protein by Western blotting using a polyclonal antiserum raised against a fragment of Eg-ppGalNAc-T1 (Figure 2A). The absence of the mRNA and expressed protein in untransfected controls was similarly checked in parallel. The recombinant protein was then immunoaffinity-purified to near homogeneity (Figure 2B) and used to characterize the transferase activity with dOSM, a poly-acceptor of GalNAc residues. Maximal activity was observed at 28 °C when compared with 20 and 37 °C (Table 2) and over the range 6.5–7.5 pH units (Figure 3). Analysis of the metal requirement of Eg-ppGalNAc-T1 showed behaviour distinct from that previously reported for other members of the family. In fact, Mn<sup>2+</sup> has always been identified as the best activator of ppGalNAc-Ts, and the transferase capacity was found to be affected by the cation species added to the reaction medium. For Eg-ppGalNAc-T1, a similar level of activity was observed with several cations, including Mn<sup>2+</sup>, whereas Cu<sup>2+</sup> appeared to be the best cofactor (Table 3).

Eg-ppGalNAc-T1 activity was subsequently characterized using a panel of synthetic peptides (Table 1). No sequences designed from potential natural substrates of the enzyme were included because no GalNAc-O-linked glycopeptides have been identified yet in *E. granulosus*. We assessed the transferase activity of peptides derived from O-glycosylated proteins of *S. mansoni*, the protozoan parasites *T. cruzi* and *T. brucei* and from human mucins (MUC1, MUC2, MUC5AC and MUC5B). The enzyme was capable of transferring GalNAc to threonine or serine residues in acceptor substrates, as shown by its activity in peptides



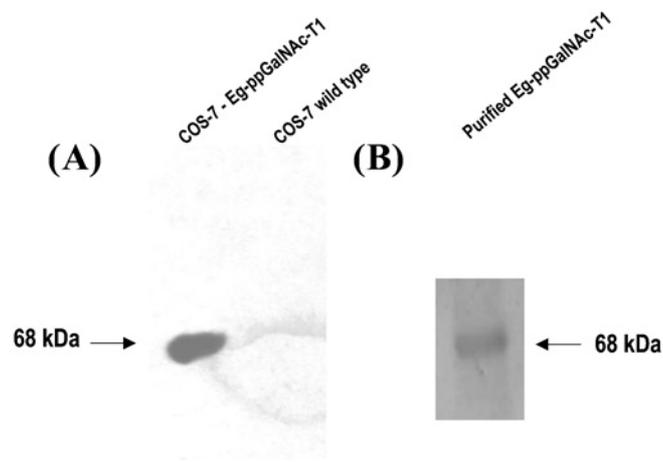
**Figure 1** Sequence analysis of Eg-ppGalNac-T1

(A) Multiple sequence alignment of the predicted amino acid sequence from Eg-ppGalNac-T1 cDNA with the *C. elegans* ppGalNac-T family. Protein sequences were aligned for maximum identity by introducing gaps using ClustalW (1.82) as a progressive tool. Amino acids are shaded black if identical, dark grey if strongly similar and light grey if weakly similar. Conserved structural domains are highlighted with arrows and the localization of the catalytic domain is indicated by dotted lines. Boxed amino acids represent conserved residues from the lectin domain. The putative CYTH domain of Eg-ppGalNac-T1 is underlined. The GenBank® accession numbers of the aligned sequences are AY353720 (Eg-ppGalNac-T1), AF031833 (gly-3), AF031834 (gly-4), AF031835 (gly-5a), AF031836 (gly-5b), AF031837 (gly-5c), AF031838 (gly-6a), AF031841 (gly-7), AF031842 (gly-8) and AF031843 (gly-9). (B) Alignment of the C-terminal region from Eg-ppGalNac-T1 and the CYTH consensus. Protein sequences were aligned for maximum identity by introducing gaps using ClustalW (1.82) as a progressive tool. Amino acids are shaded in black if identical and in grey if similar.

containing only threonine (MUC2, MUC5B, *T. cruzi* mucin) or serine (*T. brucei* mucin) residues. In addition, although all tested peptides were good substrates, Eg-ppGalNac-T1 preferentially glycosylated that derived from a *S. mansoni* glycoprotein. The background activity measured in the absence of acceptor could derive from hydrolysis of UDP-GalNac without transfer to peptide, as has been seen with other ppGalNac-Ts (Table 1).

**Eg-ppGalNac-T is expressed in the hydatid cyst wall and protoscolex**

The expression domains of the *Eg-ppGalNac-T1* gene in parasite larval stages were analysed using digoxigenin-labelled riboprobes which hybridize with the catalytic domain or the C-terminus of the enzyme (Figure 4). A clear difference was noted when comparing the reactivity of the corresponding antisense reagents: the signal

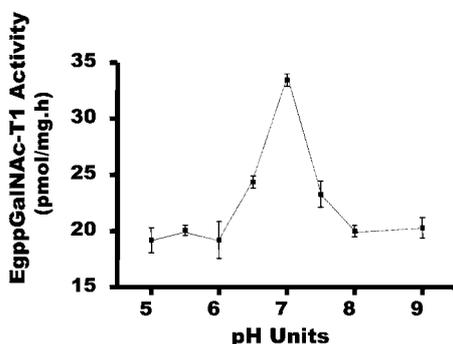


**Figure 2** Expression of Eg-ppGalNAc-T1 in COS-7-transfected cells and purification of the recombinant enzyme

(A) A partial sequence of Eg-ppGalNAc-T1 lacking the predicted transmembrane domain was expressed in COS-7 cells. The expression of Eg-ppGalNAc-T1 was confirmed by Western blot using a rabbit antiserum raised against a fragment of Eg-ppGalNAc-T1 catalytic domain. No reactivity was observed when the antiserum was used to probe an extract from non-transfected cells. (B) The recombinant protein was affinity-purified from the cell extracts over a column prepared with the purified IgG fraction from the antiserum used in (A), immobilized on to CNBr-activated Sepharose, and analysed by SDS/PAGE (10% gel) followed by silver staining.

**Table 2** Temperature dependence of Eg-ppGalNAc-T1 activity

Temperature (°C)	ppGalNAc-T activity (pmol GalNAc/mg)	Relative activity (%)
20	16.9 ± 1.5	49
28	34.8 ± 0.3	100
37	16.7 ± 1.1	48



**Figure 3** pH dependence of Eg-ppGalNAc-T1 activity

The transferase activity was evaluated with dOSM and UDP-(<sup>3</sup>H)-GalNAc at various pH values. Results are expressed in pmoles of transferred <sup>3</sup>H-GalNAc · (mg of protein)<sup>-1</sup> · h<sup>-1</sup>. Bars correspond to the S.D. of three determinations.

from the probe directed against the conserved catalytic domain appeared faster, was considerably more intense and showed a broader distribution than the one recognizing the C-terminus, indicating that the former would detect transcripts from more than one ppGalNAc-T isoform, whereas the latter would only react with Eg-ppGalNAc-T1. Both probes hybridized with the germinal layer (Figures 4A, 4B, 4D and 4E), and the signal from the

**Table 3** Cation dependence of Eg-ppGalNAc-T1 activity

Cation (10 mM)	ppGalNAc-T activity (pmol GalNAc/mg)	Relative activity (%)
Mn <sup>2+</sup>	40.0 ± 3.2	100
Cu <sup>2+</sup>	63.9 ± 1.5	160
Mg <sup>2+</sup>	35.4 ± 0.6	89
Ni <sup>2+</sup>	34.1 ± 0.6	85
Co <sup>2+</sup>	33.3 ± 1.4	83
Zn <sup>2+</sup>	31.9 ± 0.5	80
Ca <sup>2+</sup>	30.2 ± 0.2	76
-	5.4 ± 1.2	14

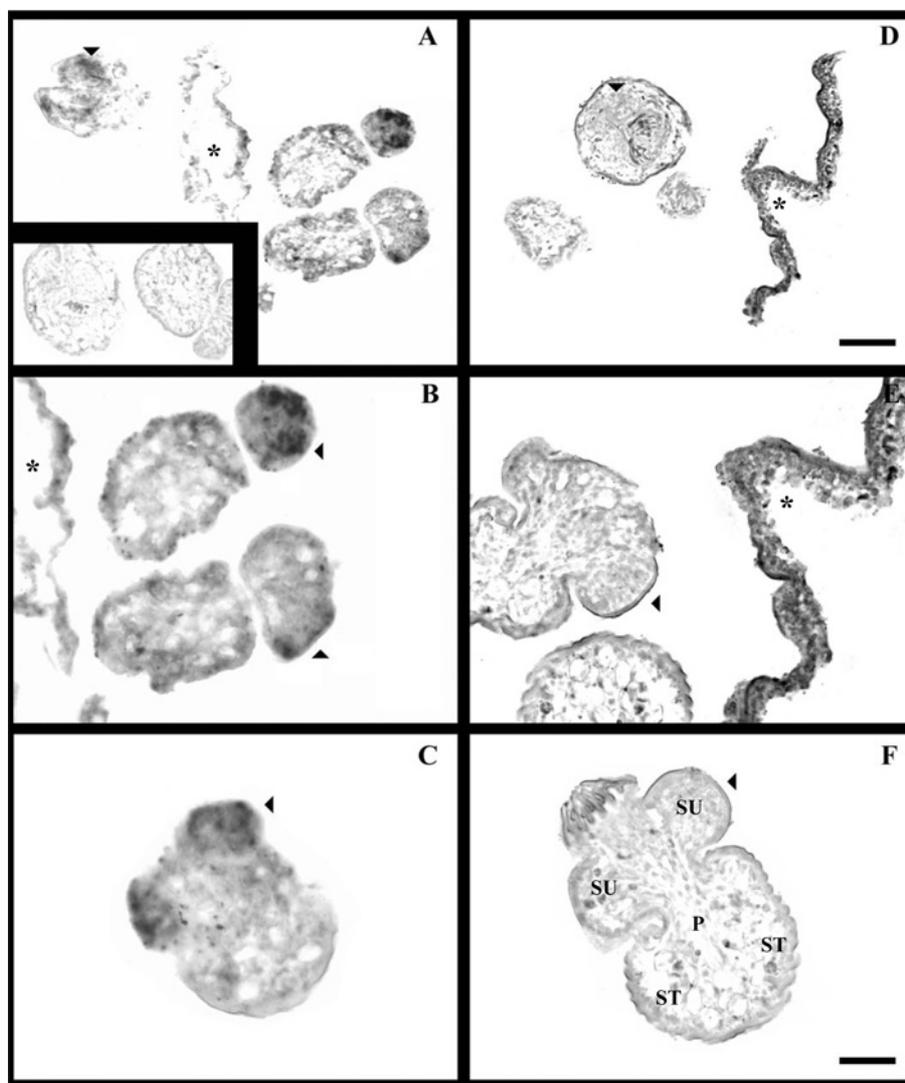
one specifically recognizing Eg-ppGalNAc-T1 was particularly intense in this parasite structure. In some sections (Figures 4D and 4E), a faint reaction was also observed between the specific probe and the subtegumental region of protoscolexes, thus suggesting that Eg-ppGalNAc-T1 is expressed in the larval worms as well as in the cyst wall. In addition, the probe against the conserved domain hybridized with the protoscolex parenchyma and, very strongly, with their suckers (Figures 4A–4C), indicating that at least one different transferase would also be expressed in *E. granulosus*.

The distribution of Eg-ppGalNAc-T1 in the same parasite materials was studied by immunohistochemistry with polyclonal antibodies raised against recombinant fragments corresponding to the same regions of the enzyme catalytic domain or C-terminus, used to prepare the riboprobes (Figure 5). The pattern of the protein was similar to the one observed for the transferase transcripts. In fact, the reactivity of the germinal layer with both sera (Figures 5A, 5B and 5D) showed a homogeneous distribution of Eg-ppGalNAc-T1 in this structure, whereas the signal obtained with the serum against the catalytic domain indicated that one or more additional ppGalNAc-Ts would be expressed in the protoscolex parenchyma and the suckers (Figure 5B). In addition, the two sera clearly reacted with the subtegumental region of the larval worms (Figures 5A–5C), confirming that Eg-ppGalNAc-T1 is also present in protoscolexes.

## DISCUSSION

The present study describes a novel ppGalNAc-T from the platyhelminth parasite *E. granulosus*, which, despite being a typical member of this protein family, has some unusual structural and functional features, as discussed below. The protein sequence predicted from the isolated cDNA includes all the elements characteristic of these transferases. In particular, its catalytic domain appears closely related to that of *C. elegans* gly-9, an isoform whose catalytic activity has not yet been demonstrated [30]. gly-9 was recently found to group in a statistically significant clade with *D. melanogaster* PGANT1 (FlyBase CG8182), when putative ppGalNAc-Ts from *C. elegans*, *D. melanogaster* and mammals were analysed looking for subfamilies of orthologous genes [31]. It may be worth noting that no mammalian enzyme was identified that clustered with these transferases [10,31].

In addition, expression of Eg-ppGalNAc-T1 in COS-7 cells produced a catalytically active protein, with maximal activity towards a multiacceptor of GalNAc at pH 6.5–7.5, 28 °C and in the presence of Cu<sup>2+</sup>, although all assayed cations were good co-factors. The optimum pH was within the range observed for other members of the family (e.g. 7.2–8.6 and 6.8–8.2 were determined for transferases isolated from bovine colostrum [40]



**Figure 4** Expression of Eg-ppGalNAc-T1 in *E. granulosus* larval stages: *in situ* hybridization analysis

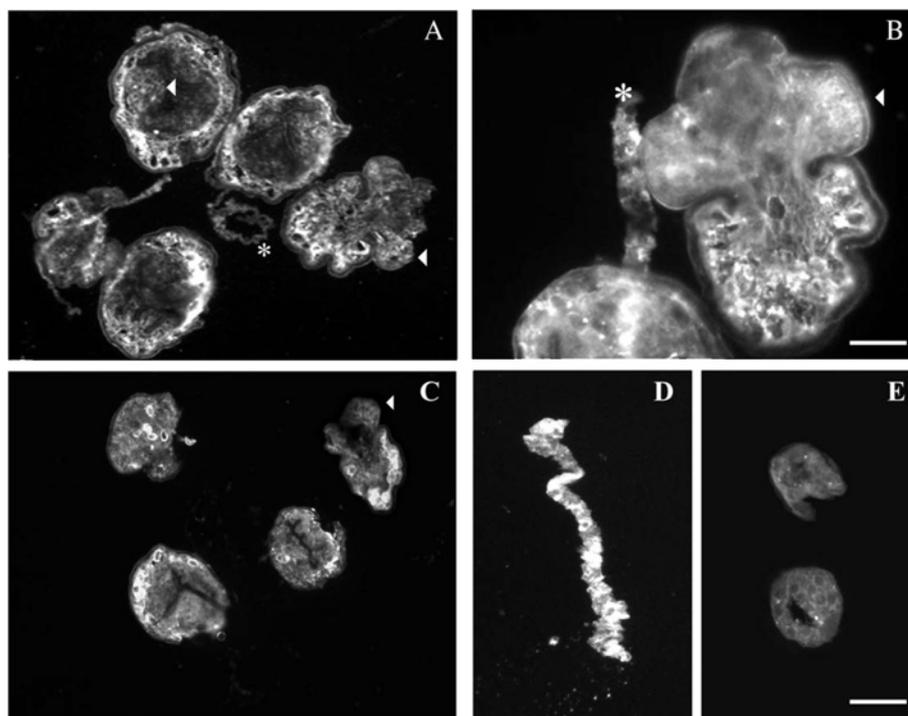
Protoscolex and cyst wall sections were hybridized with digoxigenin-labelled anti-sense RNA probes directed against the catalytic domain (A–C) or the C-terminus (D–F) of the enzyme; the reaction was developed using an alkaline phosphatase-conjugated anti-digoxigenin antibody. A protoscolex section assayed in parallel with a sense probe (inset to A) showed no reactivity. Arrowheads indicate the protoscolex suckers and asterisks the germinal layer. Localization of other structures mentioned in the text is marked in the evaginated larval worm shown in F (SU, suckers; P, parenchyma; ST, subtegument). Scale bars: (A, D), 100  $\mu\text{m}$ ; (B, C, E, F), 50  $\mu\text{m}$ .

and porcine submaxillary gland [41] respectively). The temperature dependence of Eg-ppGalNAc-T1 was similar to that previously reported for *C. elegans*: at 20 and 37 °C, the activity decreased to 50% of the value measured at 28 °C. In *C. elegans*, the observed maximal activity of worm homogenates and of purified recombinant GLY-3 between 20 and 25 °C, and the sharp drop over 30 °C or below 15 °C were attributed to the organism being a soil nematode that thrives at 16–25 °C [30]. This is not the case with *E. granulosus*, a tissue-dwelling parasite with mammalian definitive and intermediate hosts. Therefore the interpretation of the biological significance of this behaviour is not straightforward, and it is possible that it does not reflect a property of the native enzyme.

Eg-ppGalNAc-T1 showed an unusual requirement of a metal cofactor: (i) it was activated by a whole range of bivalent cations, (ii) GalNAc transfer was highest in the presence of  $\text{Cu}^{2+}$ , and (iii) it was of approximately the same order for all other assayed cations (75–100% relative to the activity with  $\text{Mn}^{2+}$ ).

Previous studies performed with mammalian ppGalNAc-Ts and metals other than  $\text{Mn}^{2+}$  identified  $\text{Co}^{2+}$  and  $\text{Cu}^{2+}$  as good activators, although not as efficient as  $\text{Mn}^{2+}$ ; in addition, significant variations were usually observed in the activity of a given enzyme with different cations (see e.g. [40–42]). Therefore the metal dependence of Eg-ppGalNAc-T1 would be a distinctive functional trait. Such behaviour was not expected since the conserved histidine and acidic residues that are proposed to be involved in the co-ordination of  $\text{Mn}^{2+}$  in ppGalNAc-Ts [10] are also present in this *Echinococcus* protein. It is difficult to speculate about the co-ordination chemistry of other putative cofactors since no crystal structure has been solved yet for a ppGalNAc-T. Analysis of the cation dependence of the native *Echinococcus* protein could be informative as to whether our result reflects what happens *in vivo*.

The most distinctive trait of the newly described enzyme is probably its unique C-terminus, which includes only one typical ricin repeat and bears no overall homology with this lectin motif. It has



**Figure 5** Expression of Eg-ppGalNAc-T1 in *E. granulosus* larval stages: immunohistochemical analysis

Protoscolex and cyst wall sections were incubated with polyclonal antibodies raised against recombinant fragments corresponding to the catalytic domain (**A, B**) or the C-terminus (**C, D**) of the enzyme; the reaction was developed using a goat anti-rabbit conjugated to Alexa-Fluor 488. A protoscolex section, assayed in parallel with PBS (negative control), showed no reactivity (**E**). Arrowheads indicate the protoscolex suckers and asterisks the germinal layer. See Figure 4(F) for the localization of other structures mentioned in the text. Scale bars: (**A, C, E**), 100  $\mu\text{m}$ ; (**B, D**), 50  $\mu\text{m}$ .

recently been noted that, in spite of the diversity of ppGalNAc-T lectin domains, the patterns of CLD and QXW motifs are conserved in orthologous transferases displaying unique and identical catalytic properties *in vitro* [31]. Together with detailed structure–function analysis of murine and rat isoforms [37,43], this observation indicates that the lectin domain is critically involved in determining the glycosylated substrate specificity of a given enzyme. Interestingly, the C-terminus of Eg-ppGalNAc-T1 showed some similarity with the recently described CYTH domain [39], defined from the catalytic domains of adenylate cyclases and thiamine triphosphatases (and named after them, as underlined). CYTH occurs either by itself or in combination with other motifs and is believed to define a novel superfamily of domains specialized in the binding of organic phosphates. This would be achieved through the interaction of four conserved basic residues with acidic phosphate moieties, three of which are present in the *Echinococcus* transferase. Therefore if all the structural elements identified are put together, the C-terminus of Eg-ppGalNAc-T1 can be described as a ‘hybrid’ domain consisting of two ricin B lectin repeats (one putative  $\alpha$  and one typical  $\gamma$  subdomain), separated by a fragment including features of a phosphate-binding motif, which, in turn, appears to be inserted within a third ricin repeat. Considering that the lectin domain modulates the substrate specificity of a transferase, it is tempting to speculate that this unusual domain would have evolved to catalyse the glycosylation of phosphopeptides. In the present work, we initiated the study of the specificity of Eg-ppGalNAc-T1, and found that it glycosylated a broad range of peptide substrates, including acceptors containing only serine or only threonine residues. The activity was remarkably high with a peptide designed from a *S. mansoni* Thr/Ser-rich protein. Studies on whether this could be related to a natural substrate of the

enzyme, as well as analysis of its capacity to transfer GalNAc to phosphorylated and/or glycosylated peptides are in progress; these will allow elucidation of the involvement of the ricin and CYTH domains in the determination of Eg-ppGalNAc-T1 specificity.

Analysis of the tissue distribution of Eg-ppGalNAc-T1 by *in situ* hybridization and immunohistochemistry revealed that this transferase is expressed in the subtegumental region of protoscolexes and the germinal layer of the cyst wall. This latter location is not surprising, taking into account the source of the cDNA clone, as well as pioneer biochemical studies supporting the existence of mucin-type components in the laminated layer [44] recently confirmed by detailed chemical analysis for *E. multilocularis* [4]. The presence of Eg-ppGalNAc-T1 in the germinal layer indicates that it would be involved in the biosynthesis of the laminated layer, whose constituents are exocytosed towards the interface with the intermediate host. The tegument from the basal region of protoscolexes is also coated with O-glycans, described as a mucopolysaccharide glycocalyx [1], and, according to our previous study, Tn and sialyl-Tn are expressed in the protoscolex tegument [6]. The fact that no ppGalNAc-T was detected with our probes in the tegument of the larval worms raises the possibility that the biosynthesis of the glycocalyx is carried out in the subtegumental region and the O-glycosylated products subsequently exported. A more detailed analysis, including ultrastructural studies, could clarify this point.

Eg-ppGalNAc-T1 is the only member of the family identified so far in the transcriptome of *E. granulosus* larval stages. Interestingly, our results suggest that at least one additional ppGalNAc-T is expressed in the protoscolex parenchyma and in the suckers. The presence of a transferase in the parenchyma appears consistent with our report describing Tn and sialyl-Tn

antigens in that location [6]. As to the suckers, these surface-associated structures allow the developing worm to grasp the intestinal epithelium and attach at the base of the villi in the dog gut, through an interaction that probably involves parasite glycoproteins. The existence of a different isoform in protoscolexes is further supported when comparing the substrate specificity of Eg-ppGalNAc-T1 with our initial results of ppGalNAc-T activity in this organism: the MUC2-derived peptide behaved as the best substrate of a protoscolex lysate [7], whereas it was not the preferred acceptor of the isolated transferase. The expression of several ppGalNAc-Ts in other metazoa indicates that this would also be the case for *Echinococcus* and related organisms. In this respect, it is worth noting that at least four ppGalNAc-Ts can be identified when searching the database from the *S. mansoni* EST genome project (available at <http://bioinfo.iq.usp.br/schisto>); this survey of the parasite transcriptome is estimated to account for 92% of a 14000 gene complement [45].

To summarize, the present study describes the structural and functional characterization of Eg-ppGalNAc-T1, a novel member of the ppGalNAc-T family with some unique features, which could be involved in the biosynthesis of O-glycosylated proteins exposed at the interface between *E. granulosus* and its hosts. If Eg-ppGalNAc-T1 is indeed involved in the synthesis of the laminated layer, its distinct properties could make it an interesting target for the design of drugs; in this way, it might be possible to interfere with the assembly of a parasite structure, which is known to be essential for the establishment of a successful infection in intermediate hosts. The fact that a particular ppGalNAc-T isoform has been found to be required for development and viability in *D. melanogaster* [31,46,47] would support the feasibility of this hypothesis.

This work was supported in part by Comisión Sectorial de Investigación Científica (Uruguay) and the ECOS France–Uruguay Programme. C. F. was awarded a Travel Research Fellowship from the Wellcome Trust.

## REFERENCES

- Thompson, R. C. A. (1995) Biology and systematics of *Echinococcus*. In Biology and Systematics of *Echinococcus* (Lymbery, A. J. and Thompson, R. C. A., eds.), pp. 1–50, CAB International, Wallingford
- Theodoropoulos, G., Hicks, S., Corfield, A., Miller, B. and Carrington, S. (2001) The role of mucins in host–parasite interactions: part II – helminth parasites. *Trends Parasitol.* **17**, 130–135
- Loukas, A., Hintz, M., Linder, D., Mullin, N., Parkinson, J., Tetteh, K. and Maizels, R. (2000) A family of secreted mucins from the parasitic nematode *Toxocara canis* bears diverse mucin domains but shares similar flanking six-cysteine repeat motifs. *J. Biol. Chem.* **275**, 39600–39607
- Hülsmeier, A., Gehrig, P., Geyer, R., Sack, R., Gottstein, B., Deplazes, P. and Köhler, P. (2002) A major *Echinococcus multilocularis* antigen is a mucin-type glycoprotein. *J. Biol. Chem.* **277**, 5742–5748
- Hakomori, S. (1989) Aberrant glycosylation in tumors and tumor-associated carbohydrate antigens. *Adv. Cancer Res.* **52**, 257–331
- Alvarez-Errico, D., Medeiros, A., Miguez, M., Casaravilla, C., Malgor, R., Carmona, C., Nieto, A. and Osinaga, E. (2001) O-glycosylation in *Echinococcus granulosus*: identification and characterization of the carcinoma associated Tn antigen. *Exp. Parasitol.* **98**, 100–109
- Casaravilla, C., Freire, T., Malgor, R., Medeiros, A., Osinaga, E. and Carmona, C. (2003) Mucin-type O-glycosylation in helminth parasites from major taxonomic groups: evidence for widespread distribution of the Tn antigen (GalNAc-Ser/Thr) and identification of UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase activity. *J. Parasitol.* **89**, 709–715
- Freire, T., Casaravilla, C., Carmona, C. and Osinaga, E. (2003) Mucin type O-glycosylation in *Fasciola hepatica*: characterization of carcinoma associated Tn and sialyl-Tn antigens and evaluation of UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase activity. *Int. J. Parasitol.* **33**, 47–56
- Fernandez, C., Gregory, W. F., Loke, P. and Maizels, R. M. (2002) Full-length-enriched cDNA libraries from *Echinococcus granulosus* contain separate populations of oligo-capped and trans-spliced transcripts and a high level of predicted signal peptide sequences. *Mol. Biochem. Parasitol.* **122**, 171–180
- Ten Hagen, K., Fritz, T. and Tabak, L. (2003) All in the family: the UDP-GalNAc:polypeptide N-acetylglucosaminyltransferases. *Glycobiology* **13**, 1R–16R
- Campbell, J. A., Davies, G. J., Bulone, V. and Henriksat, B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **326**, 929–939
- Campbell, J. A., Davies, G. J., Bulone, V. and Henriksat, B. (1998) Erratum. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **329**, 719
- Coutinho, P. M., Deleury, E., Davies, G. J. and Henriksat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**, 307–317
- White, T., Bennett, E. P., Takio, K., Sorensen, T., Bonding, N. and Clausen, H. (1995) Purification and cDNA cloning of a human UDP-N-acetyl- $\alpha$ -D-galactosamine:polypeptide N-acetylglucosaminyltransferase. *J. Biol. Chem.* **270**, 24156–24165
- Bennett, E. P., Hassan, H. and Clausen, H. (1996) cDNA cloning and expression of a novel human UDP-N-acetyl- $\alpha$ -D-galactosamine. Polypeptide N-acetylglucosaminyltransferase, GalNAc-T3. *J. Biol. Chem.* **271**, 17006–17012
- Bennett, E. P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitriou, J., Hollingsworth, M. A., Merx, G., van Kessel, A. G. et al. (1998) Cloning of a human UDP-N-acetyl- $\alpha$ -D-galactosamine:polypeptide N-acetylglucosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. *J. Biol. Chem.* **273**, 30472–30481
- Bennett, E. P., Weghuis, D. O., Merx, G., van Kessel, A. G., Eiberg, H. and Clausen, H. (1998) Genomic organization and chromosomal localization of three members of the UDP-N-acetylglucosamine:polypeptide N-acetylglucosaminyltransferase family. *Glycobiology* **8**, 547–555
- Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merx, G., van Kessel, A. G., Olofsson, S. and Clausen, H. (1999) Cloning and characterization of a close homologue of human UDP-N-acetyl- $\alpha$ -D-galactosamine:polypeptide N-acetylglucosaminyltransferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy. *J. Biol. Chem.* **274**, 25362–25370
- White, K. E., Lorenz, B., Evans, W. E., Meitinger, T., Strom, T. M. and Econs, M. J. (2000) Molecular cloning of a novel human UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase, GalNAc-T8, and analysis as a candidate autosomal dominant hypophosphatemic rickets (ADHR) gene. *Gene* **246**, 347–356
- Toba, S., Tenno, M., Konishi, M., Mikami, T., Itoh, N. and Kurosaka, A. (2000) Brain-specific expression of a novel human UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase (GalNAc-T9). *Biochim. Biophys. Acta* **1493**, 264–268
- Zara, J., Hagen, F. K., Ten Hagen, K. G., Van Wuyckhuysse, B. C. and Tabak, L. A. (1996) Cloning and expression of mouse UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase-T3. *Biochem. Biophys. Res. Commun.* **228**, 38–44
- Hagen, F. K., Ten Hagen, K. G., Beres, T. M., Balys, M. M., Van Wuyckhuysse, B. C. and Tabak, L. A. (1997) cDNA cloning and expression of a novel UDP-N-acetyl-D-galactosamine:polypeptide N-acetylglucosaminyltransferase. *J. Biol. Chem.* **272**, 13843–13848
- Ten Hagen, K. G., Hagen, F. K., Balys, M. M., Beres, T. M., Van Wuyckhuysse, B. and Tabak, L. A. (1998) Cloning and expression of a novel, tissue specifically expressed member of the UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase family. *J. Biol. Chem.* **273**, 27749–27754
- Ten Hagen, K. G., Tetaert, D., Hagen, F. K., Richet, C., Beres, T. M., Gagnon, J., Balys, M. M., Van Wuyckhuysse, B., Bedi, G. S., Degand, P. et al. (1999) Characterization of a UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase that displays glycopeptide N-acetylglucosaminyltransferase activity. *J. Biol. Chem.* **274**, 27867–27874
- Ten Hagen, K. G., Bedi, G. S., Tetaert, D., Kingsley, P. D., Hagen, F. K., Balys, M. M., Beres, T. M., Degand, P. and Tabak, L. A. (2001) Cloning and characterization of a ninth member of the UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase family, ppGalTase-T9. *J. Biol. Chem.* **276**, 17395–17404
- Guo, J. M., Zhang, Y., Cheng, L., Iwasaki, H., Wang, H., Kubota, T., Tachibana, K. and Narimatsu, H. (2002) Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase family, ppGalNAc-T12. *FEBS Lett.* **524**, 211–218
- Zhang, Y., Iwasaki, H., Wang, H., Kudo, T., Kalka, T., Hennes, T., Kubota, T., Cheng, L., Inaba, N., Gotoh, M. et al. (2003) Cloning and characterization of a new human UDP-N-acetyl- $\alpha$ -D-galactosamine:polypeptide N-acetylglucosaminyltransferase, designated ppGalNAc-T13, that is specifically expressed in neurons and synthesizes GalNAc  $\alpha$ -serine/threonine antigen. *J. Biol. Chem.* **278**, 573–584

- 28 Wang, H., Tachibana, K., Zhang, Y., Iwasaki, H., Kameyama, A., Cheng, L., Guo, J., Hiruma, T., Togayachi, A., Kudo, T. et al. (2003) Cloning and characterization of a novel UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase, pp-GalNAc-T14. *Biochem. Biophys. Res. Commun.* **300**, 738–744
- 29 Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J. and Clausen, H. (2000) The lectin domain of UDP-*N*-acetyl- $\beta$ -galactosamine:polypeptide *N*-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. *J. Biol. Chem.* **275**, 38197–38205
- 30 Hagen, F. K. and Nehrke, K. (1998) cDNA cloning and expression of a family of UDP-*N*-acetyl- $\beta$ -galactosamine:polypeptide *N*-acetylgalactosaminyltransferase sequence homologs from *Caenorhabditis elegans*. *J. Biol. Chem.* **274**, 8268–8277
- 31 Schwientek, T., Bennett, E. P., Flores, C., Thacker, J., Hollmann, M., Reis, C. A., Behrens, J., Mandel, U., Keck, B., Schafer, M. A. et al. (2002) Functional conservation of subfamilies of putative UDP-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyltransferases in *Drosophila*, *Caenorhabditis elegans*, and mammals. One subfamily composed of (I)235Aa is essential in *Drosophila*. *J. Biol. Chem.* **277**, 22623–22638
- 32 Aguinaldo, A., Turbeville, J., Linford, L., Rivera, M., Garey, J. and Lake, J. (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature (London)* **387**, 489–493
- 33 Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'homme, B. and de Rosa, R. (2000) The new animal phylogeny: reliability and implications. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4453–4456
- 34 Wojczyk, B., Stwora-Wojczyk, M., Hagen, F., Striepen, B., Hang, H., Bertozzi, C., Roos, D. and Spitalnik, S. (2003) cDNA cloning and expression of UDP-*N*-acetyl- $\beta$ -galactosamine:polypeptide *N*-acetylgalactosaminyltransferase T1 from *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **131**, 93–107
- 35 Brehm, K., Jensen, K. and Frosch, M. (2000) mRNA trans-splicing in the human parasitic cestode *Echinococcus multilocularis*. *J. Biol. Chem.* **275**, 38311–38318
- 36 Charnock, S. and Davies, G. (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38**, 6380–6385
- 37 Hagen, F. K., Hazes, B., Raffo, R., deSa, D. and Tabak, L. A. (1999) Structure–function analysis of the UDP-*N*-acetyl- $\beta$ -galactosamine:polypeptide *N*-acetylgalactosaminyltransferase. Essential residues lie in a predicted active site cleft resembling a lactose repressor fold. *J. Biol. Chem.* **274**, 6797–6803
- 38 Hazes, B. (1996) The (QxW)3 domain: a flexible lectin scaffold. *Protein Sci.* **5**, 1490–1501
- 39 Iyer, L. and Aravind, L. (2002) The catalytic domains of thiamine triphosphatase and CyaB-like adenylyl cyclase define a novel superfamily of domains that bind organic phosphates. *BMC Genomics* **3**, 33
- 40 Elhammer, A. and Kornfeld, S. (1986) Purification and characterization of UDP-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyltransferase from bovine colostrum and murine lymphoma BW5147 cells. *J. Biol. Chem.* **261**, 5249–5255
- 41 Wang, Y., Abernethy, J., Eckhardt, A. and Hill, R. (1992) Purification and characterization of a UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase specific for glycosylation of threonine residues. *J. Biol. Chem.* **267**, 12709–12716
- 42 Sugiura, M., Kawasaki, T. and Yamashina, I. (1982) Purification and characterization of UDP-GalNAc:polypeptide *N*-acetylgalactosamine transferase from an ascites hepatoma, AH 66. *J. Biol. Chem.* **257**, 9501–9507
- 43 Tenno, M., Kézdy, F., Elhammer, A. and Kurosaka, A. (2002) Function of the lectin domain of polypeptide *N*-acetylgalactosaminyltransferase 1. *Biochem. Biophys. Res. Commun.* **298**, 755–759
- 44 Kilejian, A. and Schwabe, C. (1971) Studies on the polysaccharides of the *Echinococcus granulosus* cyst, with observations on a possible mechanism for laminated membrane formation. *Comp. Biochem. Physiol. B* **40**, 25–36
- 45 Verjovski-Almeida, S., DeMarco, R., Martins, E. A., Guimaraes, P. E., Ojopi, E. P., Paquola, A. C., Piazza, J. P., Nishiyama, Jr, M. Y., Kitajima, J. P., Adamson, R. E. et al. (2003) Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nat. Genet.* **35**, 148–157
- 46 Flores, C. and Engels, W. (1999) Microsatellite instability in *Drosophila spellchecker1* (MutS homolog) mutants. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2964–2969
- 47 Ten Hagen, K. and Tran, D. (2002) A UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase is essential for viability in *Drosophila melanogaster*. *J. Biol. Chem.* **277**, 22616–22622
- 48 Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E. N. and Wilson, D. (1990) Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J. Biol. Chem.* **265**, 15286–15293
- 49 Gum, Jr, J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. *J. Biol. Chem.* **269**, 2440–2446
- 50 Porchet, N., Pigny, P., Buisine, M., Debailleul, V., Degand, P., Line, A. and Aubert, J. P. (1995) Human mucin genes: genomic organization and expression of MUC4, MUC5AC and MUC5B. *Biochem. Soc. Trans.* **23**, 800–805
- 51 Desseyn, J. L., Guyonnet-Dupérat, V., Porchet, N., Aubert, J. P. and Laine, A. (1997) Human mucin gene MUC5B, the 10.7-kb large central exon encodes various alternate subdomains resulting in a super-repeat. Structural evidence for a 11p15.5 gene family. *J. Biol. Chem.* **272**, 3168–3178
- 52 Nolan, D. P., Jackson, D. G., Windle, H. J., Pays, A., Geuskens, M., Michel, A., Voorheis, H. P. and Pays, E. (1997) Characterization of a novel, stage-specific, invariant surface protein in *Trypanosoma brucei* containing an internal, serine-rich, repetitive motif. *J. Biol. Chem.* **272**, 29212–29221
- 53 Acosta-Serrano, A., Almeida, I. C., Freitas-Junior, L. H., Yoshida, N. and Schenkman, S. (2001) The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. *Mol. Biochem. Parasitol.* **114**, 143–150
- 54 Menrath, M., Michel, A. and Kunz, W. (1995) A female-specific cDNA sequence of *Schistosoma mansoni* encoding a mucin-like protein that is expressed in the epithelial cells of the reproductive duct. *Parasitology* **111**, 477–483

Received 4 December 2003/6 May 2004; accepted 14 May 2004

Published as BJ Immediate Publication 14 May 2004, DOI 10.1042/BJ20031877