US 20120064541A1

(19) **United States**

(12) **Patent Application Publication**

Künzler et al.

(10) **Pub. No.: US 2012/0064541 A1**

(43) **Pub. Date: Mar. 15, 2012**

(54) **N-GLYCAN CORE BETA-GALACTOSYLTRANSFERASE AND USES THEREOF**

(75) Inventors: **Markus Künzler**, Neuenhof (CH); **Markus Aebi**, Wettingen (CH); **Lain Wilson**, Vienna (AT); **Alexander Walter Titz**, Kreuzlingen (CH); **Michael Hengartner**, Zurich (CH); **Alex Butschi**, Hinwil (CH)

(73) Assignees: **ETH ZÜRICH**, Zürich (CH); **UNIVERSITÄT FÜR BODENKULTUR WIEN**, Vienna (AT); **UNIVERSITÄT ZÜRICH**, Zürich (CH)

(21) Appl. No.: **13/322,505**

(22) PCT Filed: **May 28, 2010**

(86) PCT No.: **PCT/EP2010/003249**

§ 371 (c)(1), (2), (4) Date: **Nov. 25, 2011**

(30) **Foreign Application Priority Data**

May 28, 2009 (EP) 09007139.0

Publication Classification

(51) **Int. Cl.**

G01N 33/53	(2006.01)
C07K 14/00	(2006.01)
C07H 21/02	(2006.01)
C12N 9/96	(2006.01)
C12P 19/00	(2006.01)
C12N 1/19	(2006.01)
C12N 1/21	(2006.01)
C12N 5/10	(2006.01)
C07K 16/40	(2006.01)
C12N 5/12	(2006.01)
C07H 21/04	(2006.01)
C12N 15/63	(2006.01)

(52) **U.S. Cl.** **435/7.1**; 536/24.5; 530/395; 536/23.2; 435/188; 435/320.1; 435/254.2; 435/254.21; 435/254.23; 435/252.33; 435/419; 435/325; 435/348; 530/387.9; 435/331; 435/72

(57) **ABSTRACT**

The present invention relates to new galactosyltransferases, nucleic acids encoding them, as well as recombinant vectors, host cells, antibodies, uses and methods relating thereto.

Fig. 1

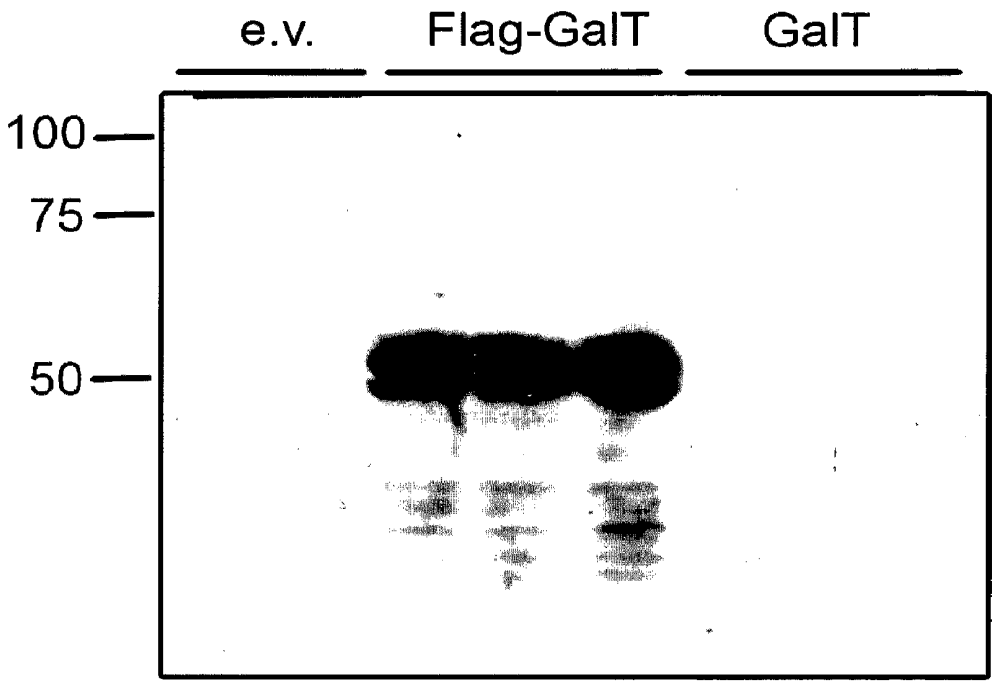

e.v.
Flag-GalT
GalT
100
75
50

Fig. 2

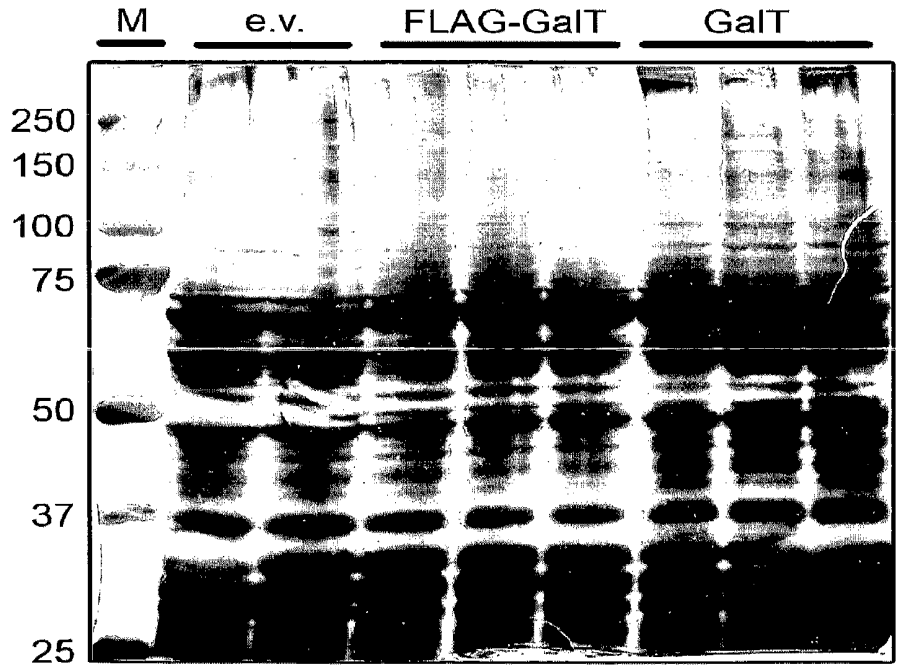

M
e.v.
FLAG-GalT
GalT
250
150
100
75
50
37
25

Fig. 3

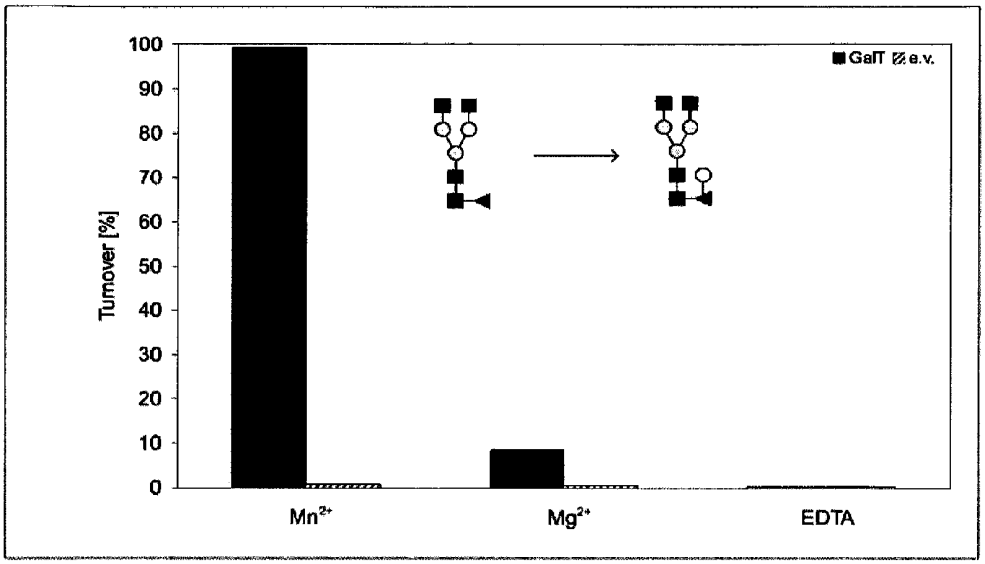

GalT e.v.
Turnover [%]
100
90
80
70
60
50
40
30
20
10
0
Mn2+
Mg2+
EDTA

Fig. 4

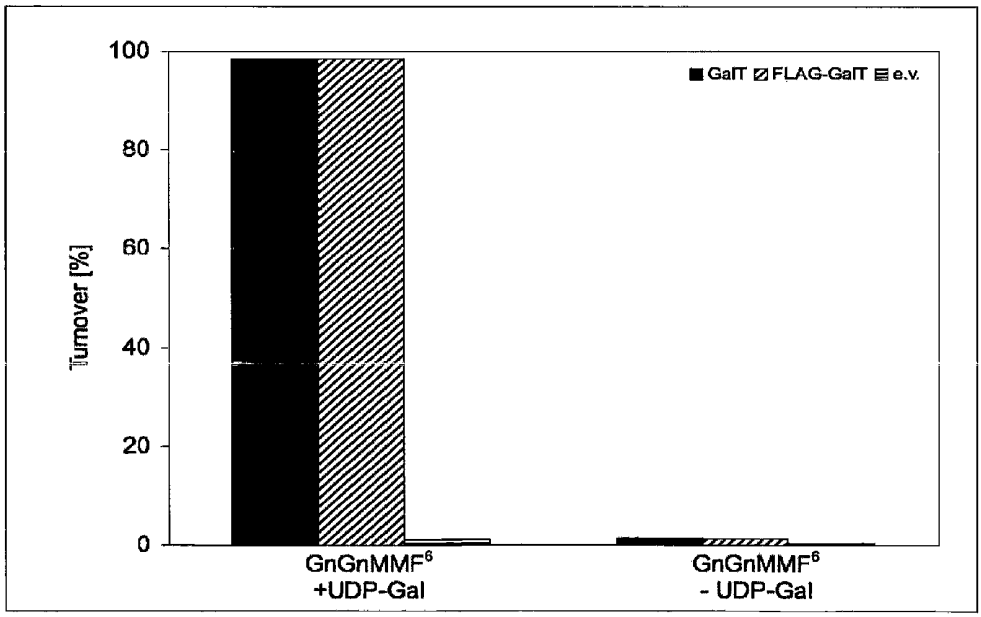

GalT FLAG-GalT e.v.
Turnover [%]
100
80
60
40
20
0
GnGnMMF6
+UDP-Gal
GnGnMMF6
- UDP-Gal

Fig. 5

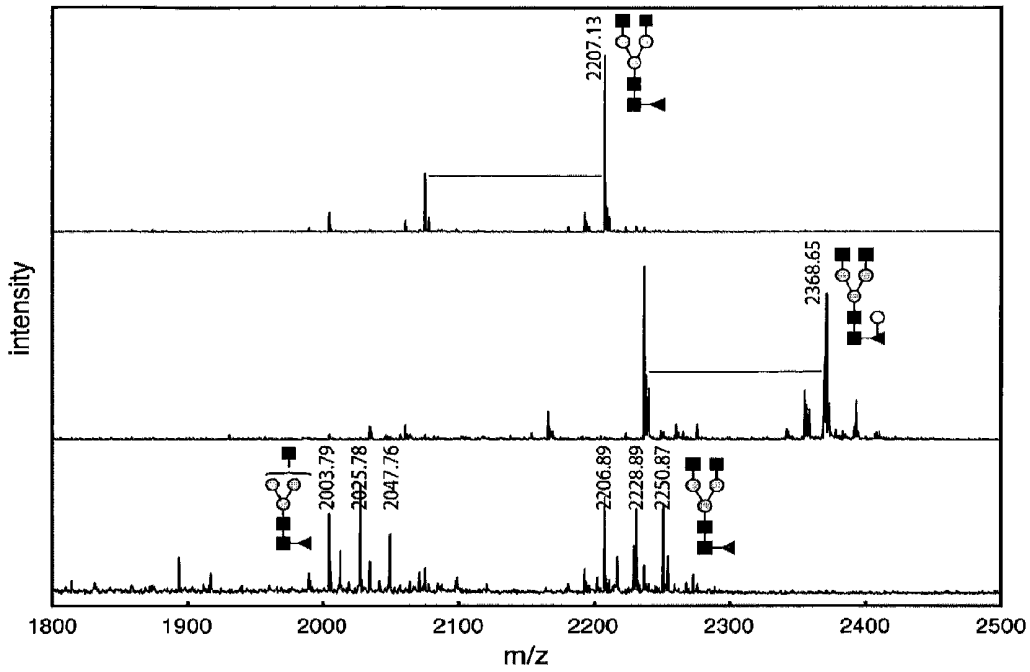

2207.13
2368.65
2003.79
2025.78
2047.76
2206.89
2228.89
2250.87
intensity
1800
1900
2000
2100
2200
2300
2400
2500
m/z

Fig. 6

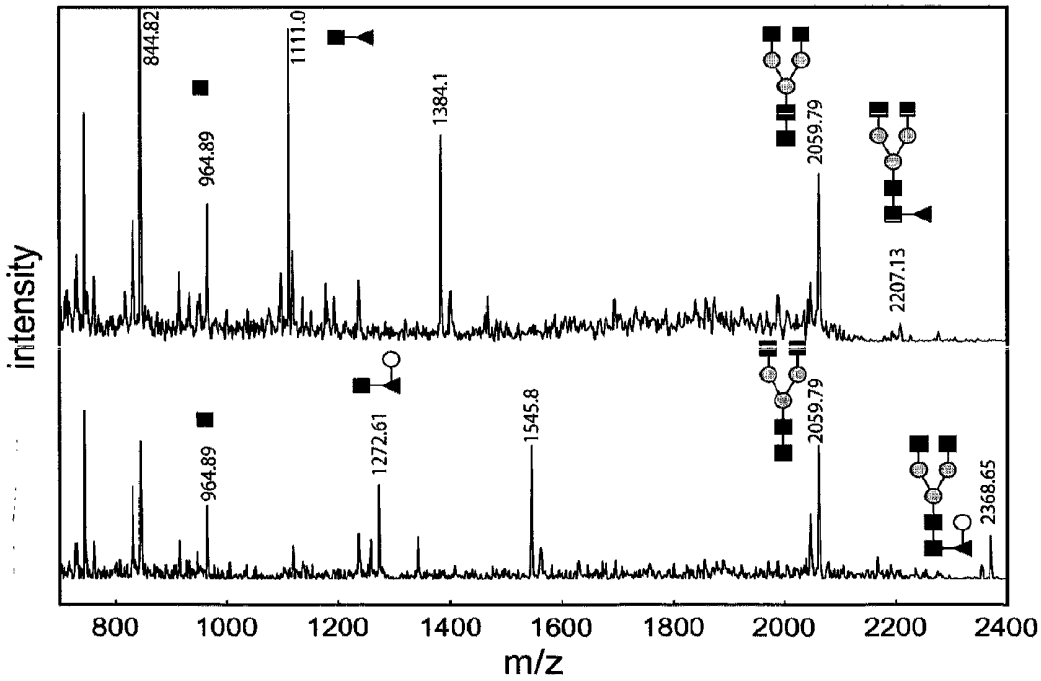

844.82
964.89
1111.0
1384.1
2059.79
2207.13
964.89
1272.61
1545.8
2059.79
2368.65
intensity
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z

Fig. 7

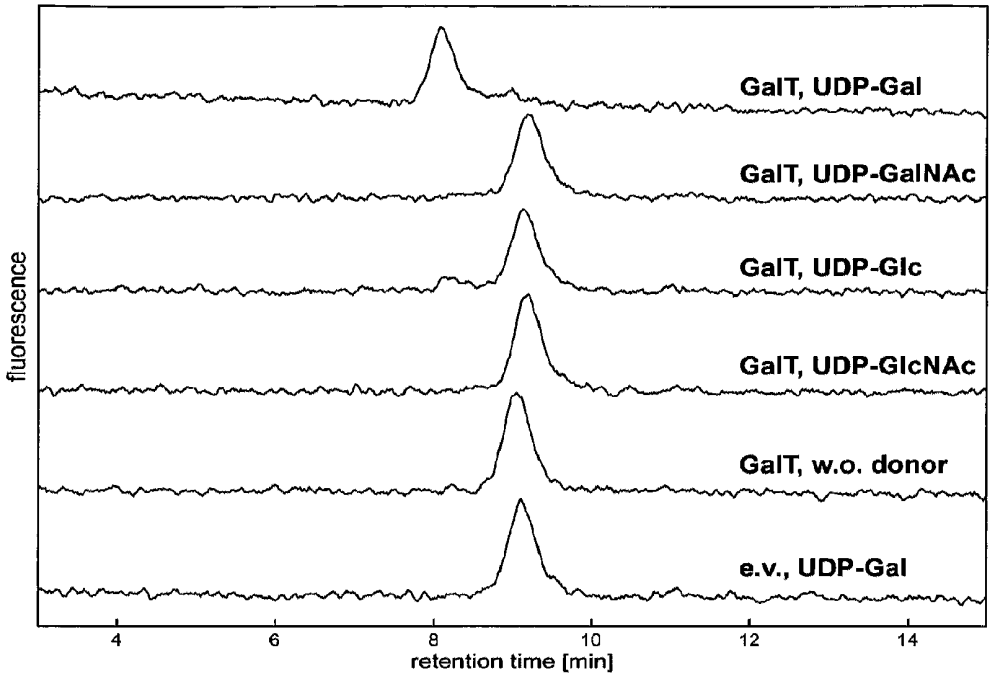

GalT, UDP-Gal
GalT, UDP-GalNAc
GalT, UDP-Glc
GalT, UDP-GlcNAc
GalT, w.o. donor
e.v., UDP-Gal
fluorescence
4
6
8
10
12
14
retention time [min]

Fig. 8

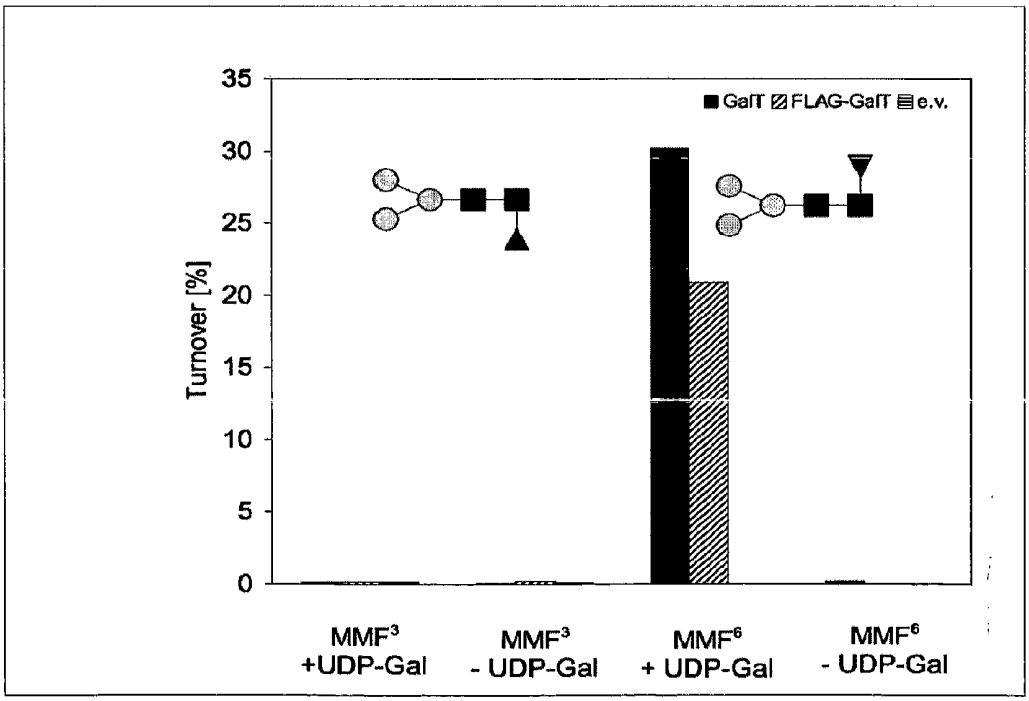

GalT FLAG-GalT e.v.
Turnover [%]
35
30
25
20
15
10
5
0
MMF3 +UDP-Gal
MMF3 - UDP-Gal
MMF6 + UDP-Gal
MMF6 - UDP-Gal

Fig. 9a

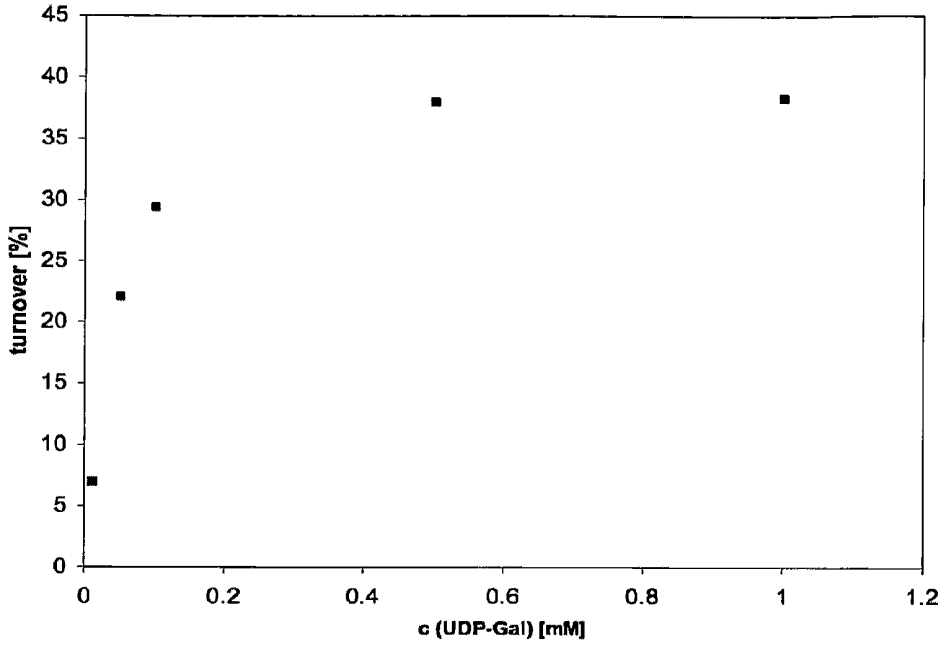

45
40
35
30
25
20
15
10
5
0
turnover [%]
0
0.2
0.4
0.6
0.8
1
1.2
c (UDP-Gal) [mM]

Fig. 9b

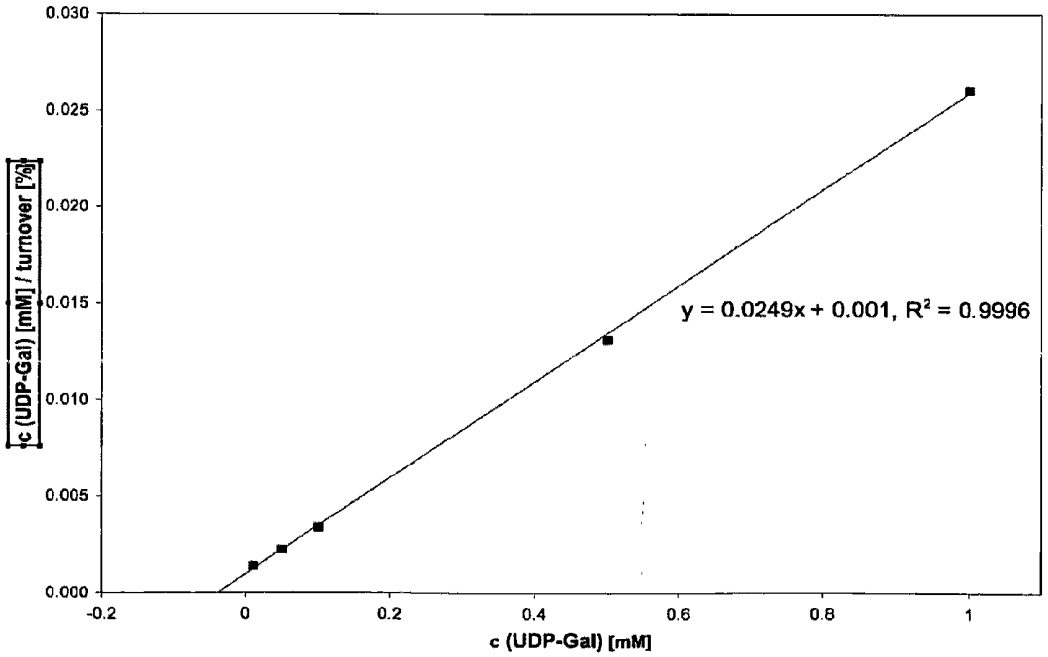

0.030
0.025
0.020
0.015
0.010
0.005
0.000
c (UDP-Gal) [mM] / turnover [%]
y = 0.0249x + 0.001, R² = 0.9996
-0.2
0
0.2
0.4
0.6
0.8
1
c (UDP-Gal) [mM]

Fig. 10

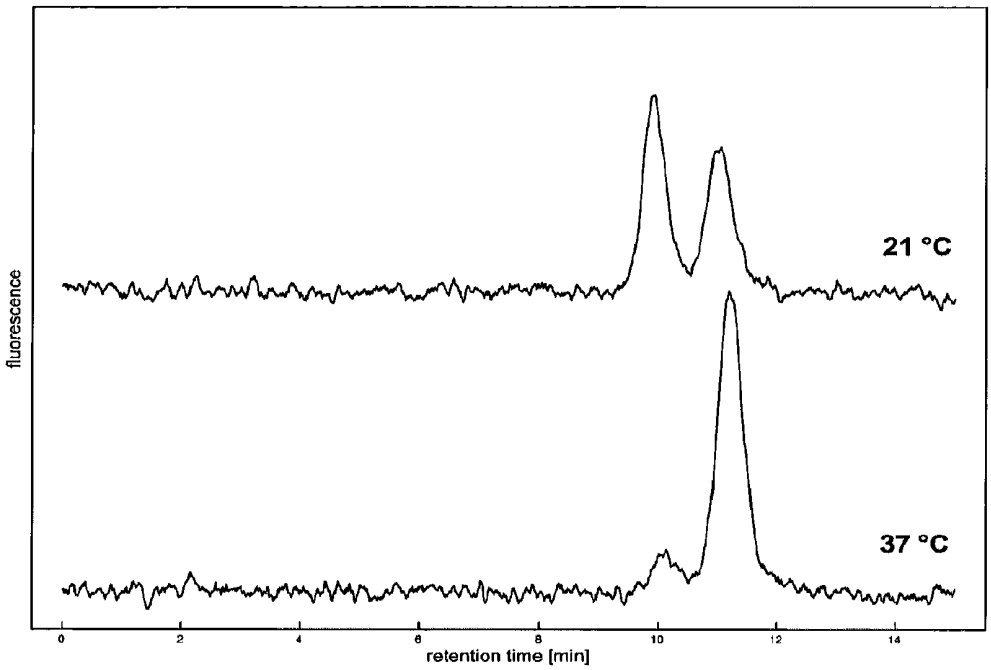

21 °C
37 °C
fluorescence
0
2
4
6
8
10
12
14
retention time [min]

Fig. 11

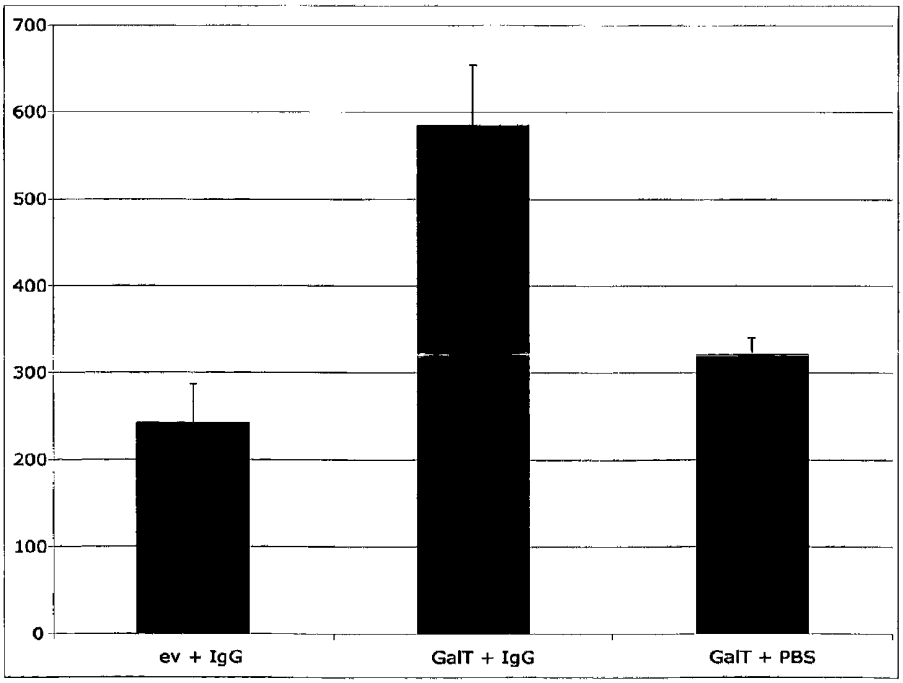
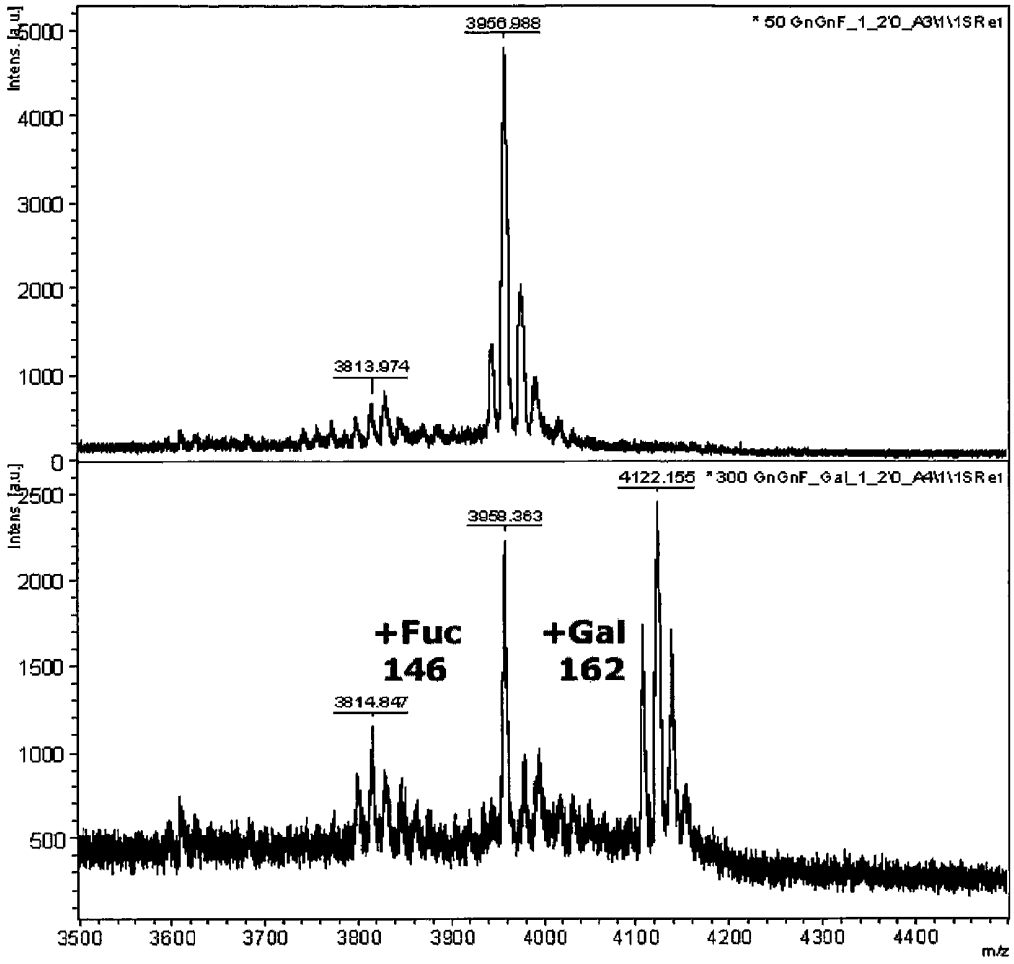

700
600
500
400
300
200
100
0
ev + IgG
GalT + IgG
GalT + PBS

Fig. 12


Intens. [a.u.]
3956.988
* 50 GnGnF_1_2'0_A3\1\1SRe1
3813.974
4122.155
* 300 GnGnF_Gal_1_2'0_A4\1\1SRe1
3958.363
+Fuc
146
+Gal
162
3814.847
m/z

N-GLYCAN CORE BETA-GALACTOSYLTRANSFERASE AND USES THEREOF

[0001] The present invention relates to new galactosyltransferases, nucleic acids encoding them, as well as recombinant vectors, host cells, antibodies, uses and methods relating thereto.

[0002] The "roundworms" or "nematodes" are the most diverse phylum of pseudocoelomates and one of the most diverse of all animals. Nematode species are difficult to distinguish; over 80,000 have been described, of which over 15,000 are parasitic. It has been estimated that the total number of roundworm species might be more than 500,000. Nematodes are ubiquitous in freshwater, marine and terrestrial environments. The many parasitic forms include pathogens in most plants, animals and also in humans.

[0003] *Caenorhabditis elegans* is a model nematode and is unsegmented, vermiform, bilaterally symmetrical, with a cuticle integument, four main epidermal cords and a fluid-filled pseudocoelomate cavity. In the wild, it feeds on bacteria that develop on decaying vegetable matter. Hannemann et al. (Glycobiology, 16, 874, 2006) isolated and structurally characterized D-galactopyranosyl-β-1,4-L-fucopyranosyl-α-1,6-D-GlcNAc (Gal-Fuc) epitopes at the core of N-glycans from *Caenorhabditis elegans*. The N-glycosylation pattern of *Caenorhabditis elegans* was recently reviewed in Paschinger et al. (Carbohydrate Res., 343, 2041, 2008).

[0004] It is the object of the present invention to provide new means for the recombinant production of Gal-Fuc-containing (poly/oligo)saccharides and Gal-Fuc-containing glycoconjugates. An additional object is to provide new uses for Gal-Fuc-containing poly/oligosaccharides and Gal-Fuc-containing glycoconjugates.

[0005] In a first aspect, the object is solved by an isolated and purified nucleic acid selected from the group consisting of:

[0006] (i) a nucleic acid comprising at least a nucleic acid sequence selected from the group consisting of nucleic acid sequences listed in SEQ ID NOs: 1, 3, 5, 7 and 9, preferably SEQ ID NO 1;

[0007] (ii) a nucleic acid having a sequence of at least 60, 65, 70 or 75% identity, preferably at least 80, 85 or 90% identity, more preferred at least 95% identity, most preferred at least 98% identity with a nucleic acid sequence selected from the group consisting of nucleic acid sequences listed in SEQ ID NOs 1, 3, 5 and 7, preferably SEQ ID NO: 1;

[0008] (iii) a nucleic acid that hybridizes to a nucleic acid of (i) or (ii);

[0009] (iv) a nucleic acid, wherein said nucleic acid is derivable by substitution, addition and/or deletion of one of the nucleic acids of (i), (ii) or (iii);

[0010] (v) a fragment of any of the nucleic acids of (i) to (iv), that hybridizes to a nucleic acid of (i).

[0011] In a preferred aspect the isolated and purified nucleic acid selected from the group consisting of:

[0012] (i) a nucleic acid comprising at least a nucleic acid sequence selected from the group consisting of nucleic acid sequences listed in SEQ ID NOs: 1, 3, 7 and 9 as well as the first 1428 nucleic acids of SEQ ID NO: 5, preferably SEQ ID NO 1;

[0013] (ii) a nucleic acid having a sequence of at least 60, 65, 70 or 75% identity, preferably at least 80, 85 or 90% identity, more preferred at least 95% identity, most preferred at least 98% identity with a nucleic acid sequence selected from the group consisting of nucleic acid sequences listed in SEQ ID NOs 1, 3 and 7 as well as the first 1428 nucleic acids of SEQ ID NO: 5, preferably SEQ ID NO: 1;

[0014] (iii) a nucleic acid that hybridizes to a nucleic acid of (i) or (ii);

[0015] (iv) a nucleic acid, wherein said nucleic acid is derivable by substitution, addition and/or deletion of one of the nucleic acids of (i), (ii) or (iii);

[0016] (v) a fragment of any of the nucleic acids of (i) to (iv), that hybridizes to a nucleic acid of (i).

[0017] Preferably, the above nucleic acids encode a polypeptide of the invention, preferably one having an enzymatic galactosyltransferase activity, more preferably one having a β-1,4-galactosyltransferase activity, preferably one with L-fucoside-, more preferably one with α-L-fucoside-, more preferably one with Fuc-α-1,6-GlcNAc— and most preferably one with $GnGnF^6$— (nomenclature according to Schachter, Biochem. Cell. Biol. 64(3), 163-181, 1986) containing poly/oligosaccharides or glycoconjugates as acceptor substrates.

[0018] Galactosyltransferase activity, as used herein, is meant to describe an enzymatic transfer of a galactose residue from an activated donor form (i.e. nucleotide-activated galactose, preferably UDP-Gal) to an acceptor. β-1,4-Galactosyltransferase activity, as used herein, is meant to describe the specificity of the galactosyltransferase activity, i.e the transfer of galactose in a beta 1,4-configuration onto an acceptor molecule. β-1,4-Galactosyltransferase activity on L-fucosides as acceptor substrate, as used herein, is meant to describe the specificity of the galactosyltransferase activity in a beta-linked 1,4-transfer onto L-fucosides as the acceptor substrate. L-fucosides, as meant herein, are meant to describe poly/oligosaccharides or glycoconjugates as acceptor substrates containing terminal L-fucose in alpha, most preferably in alpha-1,6 configuration, e.g. as part of MMF6 or $GnGnF^6$ (Schachter, Biochem. Cell. Biol. 64(3), 163-181, 1986).

[0019] In a most preferred embodiment, the encoded polypeptide comprises a polypeptide sequence selected from the group consisting of polypeptide sequences listed in SEQ ID NOs 2, 4, 6, 8 and 10, preferably SEQ ID NO: 2, or a functional fragment or functional derivative of any of these.

SEQ ID NO: 1 is the nucleic acid sequence coding for SEQ ID NO 2:
(also listed in NCBI as Ref Seq NM_072144.4 and in Wormbase as M03F8.4; coding for galactosyltransferase [referred to as GalT in the Examples section] from *Caenorhabditis elegans*)

ATGCCTCGAATCACCGCCAGTAAAATAGTTCTTCTAATTGCATTATCATTTTGTATTA

CTGTTATTTATCACTTTCCAATAGCAACGAGAAGCAGTAAGGAGTACGATGAATATG

-continued

GAAATGAATATGAAAACGTTGCATCGATAGAGTCGGATATAAAAAATGTACGTCGAT

TACTTGACGAGGTACCGGATCCCTCACAAAACCGTCTACAATTCCTGAAACTTGATG

AGCATGCTTTTGCATTCTCGGCCTACACAGACGATCGAAATGGAAATATGGGGTAC

AAATATGTCCGAGTCCTGATGTTTATCACGTCACAAGACAACTTTTCCTGTGAAATAA

ACGGGAGAAAGTCCACAGATGTATCACTTTACGAGTTCTCGGAAAATCACAAAATGA

AGTGGCAAATGTTTATTTTGAATTGTAAACTACCCGATGGTATAGATTTCAATAATGT

TAGCTCTGTAAAGGTCATAAGAAGCACAACCAAGCAGTTTGTTGATGTGCCGATTCG

GTATAGAATTCAAGATGAGAAAATAATTACGCCAGACGAATATGACTATAAAATGTCA

ATTTGTGTTCCAGCATTGTTIGGAAATGGATATGATGCAAAGCGAATTGTTGAGITTA

TTGAGCTGAATACTTTGCAAGGAATCGAGAAAATATACATTTACACTAATCAAAAAGA

GCTTGATGGATCCATGAAGAAAACGTTGAAATACTATTCGGATAATCACAAAATAAC

ATTAATTGATTACACATTACCATTCAGAGAGGATGGTGTTTGGTATCACGGGCAATT

GGCAACTGTTACTGATTGTTTACTGAGAAACACTGGAATCACAAAATACACATTTTTC

AATGATTTTGATGAGTTCTTCGTCCCCGTTATCAAAAGTCGGACTCTCTTTGAAACAA

TCAGTGGGCTTTTTGAAGATCCCACTATTGGATCGCAACGAACAGCTTTGAAGTATA

TAAATGCAAAAATCAAGAGCGCTCCGTATTCACTGAAAAATATTGTTTCCGAAAAAC

GAATTGAAACAAGATTCACGAAATGTGTAGTTCGACCGGAAATGGTTTTTGAACAGG

GTATTCATCATACGAGTAGAGTGATTCAAGACAACTATAAAACGGTTTCCCATGGCG

GATCCCTTCTACGGGTTTATCATTACAAGGATAAAAAGTATTGTTGCGAAGACGAGA

GCCTCTTGAAAAAACGGCATGGAGATCAACTTCGGGAAAAATTCGATTCAGTTGTTG

GTCTTTTAGACTTGTAG

SEQ ID NO: 2 (also listed in NCBI Ref Seq NP_504545.2)
MPRITASKIVLLIALSFCITVIYHFPIATRSSKEYDEYGNEYENVASIESDIKNVRRLLD

EVPDPSQNRLQFLKLDEHAFAFSAYTDDRNGNMGYKYVRVLMFITSQDNFSCEINGRK

STDVSLYEFSENHKMKWQMFILNCKLPDGIDFNNVSSVKVIRSTTKQFVDVPIRYRIQDE

KIITPDEYDYKMSICVPALFGNGYDAKRIVEFIELNTLQGIEKIYIYTNQKELDGSMKKTLK

YYSDNHKITLIDYTLPFREDGVWYHGQLATVTDCLLRNTGITKYTFFNDFDEFFVPVIKSR

TLFETISGLFEDPTIGSQRTALKYINAKIKSAPYSLKNIVSEKRIETRFTKCVVRPEMVFEQ

GIHHTSRVIQDNYKTVSHGGSLLRVYHYKDKKYCCEDESLLKKRHGDQLREKFDSVVG

LLDL

SEQ ID NO: 3 is the nucleic acid sequence coding for SEQ ID NO: 4:
(also listed in NCBI Ref Seq XM_001674213.1; coding for galactosyltransferase from *Caenorhabditis briggsae*)
ATGCCACGAA TAACGGCAAG CAAAATAGTG TTATTATCTG TATTATCCTT

ACTAACAGTT TTCTATCTGA ATACATTTTC GTCTATTAAA ATTGAAAACG

ATCTCGACGG GACTGATTAC GACTTGGATT ACATAGAATC TGATATCAAA

AAGACGCGTC GATTACTCAA TGAAATCCCT GATCCATCTC AAAACCGAGT

TCAATTTTTT AAACTCGATG ATAATGGATA TGCATTCTCA GCATATACAG

ATAATAGGAA AGGAAATATG GGTCACAAAT ATGTCAGAAT ATTAGTGTTC

CTAACTAAAT TTGATGATTT TTCTTGCGAA ATTAACTCGA AGAAATCCTA

TGTTGTTACA CTCTACGAGC TATCAGAAAA TCACAATATG AAGTGGAAAA

-continued

TGTATATTTT GAATTGTTTA CTTCCCGATG GAATCACTTT CAACGATGTG

AATTCTGTAA AAATATCTAG AAGTTCTTCA AAACTTTCAG TCCAAATCCC

GATCAGATAT AGAATTCAAG ATGAGAAAAT GATGACTCCA GATGAATACG

ATTATAAGTT GTCGATTTGT GTTCCTGCAC TTTTTGGAAA CGTTTATTAT

CCAAGGAGGA TTATTGAATT TGTGGAACTA AACAGCTTGC AAGACATCGA

CAAAATCTAC ATCTACTACA ATCCTTTAGA AATGACAGAT GAGGCCACAG

AAAGGACTTT GAAGTTTTAT TCCAATAATG GGAAAATCAA TTTAATAGAA

TTCATTCTCC CATTTTCTAC TCGAGATGTT TGGTATTATG GGCAATTGGC

CACCGTTACA GATTGTCTTC TCCGTAACAC TGGAATAACT CAATACACAT

TTTTCAATGA TTTGGATGAA TTTTTCGTGC CAGTACTGGA CAACCAAACT

CTCTCTGAAA CTGTGTCAGG ATTATTTGAA AATCGAAAAA TTGCCTCTCA

GAGAACGGCC TTGAAATTTA TTAGTACAAA AATCAATCGA TCTCCTGTAA

CTCTCAATAA TATTGTGTCT TCTAAAAATT TTGAAACGAG ATTCACAAAA

TGCGTCGTAC GGCCGGAAAT GGTTTTTGAG CAGGGCATTC ACCATACGAG

TAGAGTAATA CAAGACGACT ACGAAACCCC ATCCCATGAT GGATCACTTT

TGCGTGTGTA TCACTACAGA GAACCAAGAT ATTGCTGCGA AAACGAGAAT

CTTCTAAAAC AAAGATACGA TAAGAAGCTT CAAGAAGTTT TTGATGCTGT

AGTTCTTATA TTGCATGTCA CATTTGATGT ATGGATATAT CACCTGAAAA

ACACCCTCTA A

SEQ ID NO: 4 (also listed in NCBI Ref Seq XP_001674265.1)
MPRITASKIV LLSVLSLLTV FYLNTFSSIK IENDLDGTDY DLDYIESDIK KTRRLLNEIP

DPSQNRVQFF KLDDNGYAFS AYTDNRKGNM GHKYVRILVF LTKFDDFSCE

INSKKSYVVT LYELSENHNM KWKMYILNCL LPDGITFNDV NSVKISRSSS

KLSVQIPIRY RIQDEKMMTP DEYDYKLSIC VPALFGNVYY PRRIIEFVEL NSLQDIDKIY

IYYNPLEMTD EATERTLKFY SNNGKINLIE FILPFSTRDV WYYGQLATVT

DCLLRNTGIT QYTFFNDLDE FFVPVLDNQT LSETVSGLFE NRKIASQRTA

LKFISTKINR SPVTLNNIVS SKNFETRFTK CVVRPEMVFE QGIHHTSRVI

QDDYETPSHD GSLLRVYHYR EPRYCCENEN LLKQRYDKKL QEVFDAVVLI

LHVTFDVWIY HLKNTL

SEQ ID NO: 5 is the nucleic acid sequence coding for SEQ ID NO: 6 (1428 nucleic acids) followed by a stop codon and further 68 nucleotides: (also listed in NCBI Ref Seq XM_001629141.1; coding for galactosyltransferase from *Nematostella vectensis*)
ATGCGATGCT ATATTTACAA ATTGAGGTTG TCCGTTTGTC TGTTTGTAGT

GCTCTTCACA GCACTGCTTT TCATCACCTA TTTAAACCAC TCAGAGCTTG

AATCAGCAGA GAAAAGTAGC GGAAAAAGGA AGACGCGACA TCGTAAACGA

ACACGTTCAC GCAAACAACA CGAGAGCCAT TTTCAGAAAG CTCGACTACA

AGAAAGAGAA CTAGTATTAA GATCTACAGC GCCACCAACA TTACGAAGAG

AAGTACAAGC GCATCGATTA GGGCAGATCC GTGGCAAGAA CACGGACCAG

GGGATAACTG GAAAGTTCAC AGAGATCGCT AAAGACACGC ATATTTATTC

AGCGTTTTAC GACGATGCCA AGTCAAATCC ATTCATTCGT CTTATCATCC

-continued

TCTCGGGAAA ACACTACCAG CCTGGATTAT CTTGCCAATT TTGCGAACCT

TTGTCCGCCA GTTGTAGTTT TGCGGACTCT AAAGCTGAAT ACTACACGAC

CAACGAGAAC CATGGGAGAG TATTTGGCGG GTTCATTGCG AGTTGCCTCG

TGCCTGATGG ATTCAATGCA GTGCCATTGT TTGTTGACAT AACGGCCGAT

GTTAAGGGGG AGAAAAGCAA GGCACGGGTA CCTGTGGTGT CTAATGCACA

TCTCTACTAC CCTATTAAAT ACGCAATCTG CGTCCCACCC CTCCGATCAG

AGAAACTAAC AGCGAAAAGA CTCATAGAGT TTGTCGAGCT AACCAAACTT

TTAGGCGCTA ACCATTTTAC TTTTTATGAC TTCAAAACGG ACCCGGAAGT

CAATAACGTT TTAAGATATT ACCAGGAGAC ACAAGTAGCA AATGTTCTGC

CATGGAATCT ACCTTCAAAT TTGGTATCCA GGCCGAACGA TATTTGGTAC

TTTGGTCAGG TTTTGGCTAT TCTAGATTGC TTGTATCGCT ACAAGAACAG

GGCAAAATTT GTAGCCTTCA ATGACGTAGA TGAGTTTATC GTTCCGCTAA

GGAACAGCTC GATAGTGGAA ATACTAAACG CGTTTCACCG GCCATACCAC

TGTGGACATT GCTTTCAGAG CGTGGTGTTC AGCTCAAACG CGAGATTTCC

CAGGCAAAAA AGCGAGTTAG TTTCTCAGCG GTTCTTCCAC AGGACCCAGG

AAACCATCCC TCTCCTCTCG AAATGCATTG TGGATCCTTT GAGAGTGTTC

GAGATGGGGA TTCACCACAT AAGCAAGGCT ACAGGTCTGC GGTATTCCGT

CAACTCAGTA CACGAGAGTG ACGCGGTTAT CTTCCATTAC AGGACTTGCA

CTACGTCATT TGGTATACGT CATCAGTGCA TGAACCTAGT GCATGATGGG

ACCATGGCCA AATATGGAAA ACGACTTCAG AAAATGTTTA GAAAGGTTGT

AAATGATTTA AAACTTTTGG CACCAACGTA GCTATTTCGT AACACTTCAC

ACTTTCATTG TTATAACAGA ATACAGAATA AATTAATGAT TGTTGTGCC

SEQ ID NO: 6 (also listed in NCBI Ref Seq XP_001629191)
MRCYIYKLRL SVCLFVVLFT ALLFITYLNH SELESAEKSS GKRKTRHRKR

TRSRKQHESH FQKARLQERE LVLRSTAPPT LRREVQAHRL GQIRGKNTDQ

GITGKFTEIA KDTHIYSAFY DDAKSNPFIR LIILSGKHYQ PGLSCQFCEP

LSASCSFADS KAEYYTTNEN HGRVFGGFIA SCLVPDGFNA VPLFVDITAD

VKGEKSKARV PVVSNAHLYY PIKYAICVPP LRSEKLTAKR

LIEFVELTKL LGANHFTFYD FKTDPEVNNV LRYYQETQVA NVLPWNLPSN

LVSRPNDIWY FGQVLAILDC LYRYKNRAKF VAFNDVDEFI VPLRNSSIVE

ILNAPHRPYH CGHCFQSVVF SSNARFPRQK SELVSQRFFH RTQETIPLLS

KCIVDPLRVF EMGIHHISKA TGLRYSVNSV HESDAVIFHY RTCTTSFGIR

HQCMNLVHDG TMAKYGKRLQ KMFRKVVNDL KLLAPT

SEQ ID NO: 7 is the nucleic acid sequence coding for SEQ ID NO: 8:
(also listed in NCBI Ref Seq XM_002189335, coding for galactosyltransferase from *Taeniopygia guttata*)
ATGACTGTAA CTTTAATGCT TGTGGTTTCT TATCTGAGAT TACAGAGACT

TTCTCATCAG CCAAAAGTAA TTCAAGAAAG TAGAAGATGT AGAGGGAAAA

TTGCCCTTAG CACAATAACA GCATTGGAAG GTAACAAAAC TGATATTATA

TCCCCATACT TTGATGACAG AGAAAACAAA ATCACTCGTC TGATTGGGAT

TGTTCACCAT AAAGATGTAA AACAACTGTT CTGCTGGTTC TGCTGTCAAG

-continued

CCAATGGAAA GATATATGTA TCAAAAGCAG AAATAGATGT TCACTCGGAT

AGATTTGGAT TCCCTTATGG TGCAGCAGAT ATAATTTGTT TGGAACCTGA

AAACTGTGAT CCAACACATG TATCAATTCA TCAGTCTCCA TATGGAAATA

TTGACCAGCT GCCGAGGTTT GAAATTAAAA ATCGCAGGCC TGAGACCTTT

TCTGTTGACT TCACCGTGTG CATTTCTGCC ATGTTTGGAA ACTACAACAA

TGTCTTGCAG TTTGTACAGA GTATGGAAAT GTATAAGATT CTTGGAGTAC

AGAAAGTGGT GATCTATAAG AACAACTGCA GCCATCTGAT GGAGAAAGTC

TTGAAATTTT ATATAGAAGA AGGAACTGTT GAGGTAATTC CCTGGCCAAT

AGACTCACAC CTCAGGGTTT CTTCTAAATG GCGCTTCATG GAAGACGGGA

CACACATTGG CTACTATGGA CAAATCACAG CTCTAAATGA CTGTATATAC

CGCAACATGG AAAGGACCAA GTTTGTGGTC CTTAATGACG CTGATGAAAT

AATTCTTCCC CTTAAACACC CAGACTGGAA AACAATGATG AACAGTCTTC

AGGAGCAAAA CCCAGGGACT AGTGTTTTCC TTTTfGAGAA CCATATCTTC

CCAGAAACTG TATTTTCTCC CATGTTCAAC ATTTCATCTT GGAATACTGT

GCCAGGTGTT AACATATTGC AGCATGTGTA CAGAGAGCCT GACAGGAAAC

ATGTAATCAA TCCCAGGAAA ATGATAGTTG ATCCACGAAA GGTGATTCAG

ACTTCAGTCC ATTCTGTCCT ACGTGCTTAT GGGAAGAGCG TGAATGTTCC

CATGGAAGTT GCCCTCATTT ATCACTGTCG GAAGGCCCTT CAAGGAAACC

TTCCCAGAGA ATCTCTCATC AGGGATACAA CACTGTGGAG ATATAACTCA

TCATTAATCA TGAATGTTAA CAAGGTTCTA TCTCAAACCA TGCTGCAAAC

TCAAAATTGA

SEQ ID NO: 8 (also listed in NCBI Ref Seq XP_002189371)
MTVTLMLVVS YLRLQRLSHQ PKVIQESRRC RGKIALSTIT ALEGNKTDII

SPYFDDRENK ITRLIGIVHH KDVKQLFCWF CCQANGKIYV SKAEIDVHSD

RFGFPYGAAD IICLEPENCD PTHVSIHQSP YGNIDQLPRF EIKNRRPETF

SVDFTVCISA MFGNYNNVLQ FVQSMEMYKI LGVQKVVIYK NNCSHLMEKV

LKFYIEEGTV EVIPWPIDSH LRVSSKWRFM EDGTHIGYYG QITALNDCIY

RNMERTKFVV LNDADEIILP LKHPDWKTMM NSLQEQNPGT SVFLFENHIF

PETVFSPMFN ISSWNTVPGV NILQHVYREP DRKHVINPRK MIVDPRKVIQ

TSVHSVLRAY GKSVNVPMEV ALIYHCRKAL QGNLPRESLI RDTTLWRYNS

SLIMNVNKVL SQTMLQTQN

SEQ ID NO: 9 is the nucleic acid sequence coding for SEQ ID NO: 10:
(also listed in NCBI Ref Seq XM_626032, coding for galactosyltransferase from *Cryptosporidium parvum*)
ATGCAAAGTA AAGTCATTTT TAGGATCTTG GTATTGATCA TTTCGGTGAT

TGGATCCTTA TACTCAATAA TTCAATTAAT GCTAAAGGAG CTATCAAGTA

ACAAAAATAT TCAAGAGGTT AGTCATTCAA GGAGGCTAAT AAGTGAACCT

TACAGTGAAA GTATTAATGA ACAAAATGAT CAAGATTGGA AAGAACTAAA

GCTAATAATT CCAAATCATT CTCAAATTAA CCAGCAGGAA AAAAATGGTA

ATTTGATTGA GTTTAAAGTT TATATATACT CAGCATATTA TGATTGGAGA

-continued

ATAGATAGGA TACGAATAAA TTCACTTATC CCATCGAATT TTTATGATCG

AATAGAAATG GAATGTGCAA TAATCTTGGA CAAAAATATT TACACAGGAA

CTATTAAAAA AGTGATTCAT AAGGAGCACC ATAATAAAGA ATATGTATCA

TCGACTTTAC TCTGCGAAAT TGCAAAAAAT GAAATTAAAT TTGAGGATAT

TTCAAGGAAA GTTTTGATAA CAATTTTGGA AAATGGAAAC AGCACAAATA

AATCAGAAAT ATGGATAACT CTAAAAAAAA TTCCAAAAAA TAGCTCTAAT

AATCATGAGC TGACTGTTTG TGTGAGACCT TGGTGGGGAG AGCCAATAAA

GAATGGAAAC TTGGGAAATA AACAAAAATT TAACAATTCA GGGTTAATGC

TTGAATTTAT TAATTCATAT TTATTCTTAG GAGCAAATAA ATTTTATTTA

TATCAAAATT ACTTGGACAT TGACGAAGAT GTAAGAAATA TAATAAATTA

TTATTCTAAT ATCAAAAATG TTTTGGAAAT TATTCCATAC TCATTACCAA

TAATTCCATT TAAACAAGTT TGGGATTTCG CACAAACAAC AATGATACAG

GACTGCCTAC TAAGAAATAT TGGAAAAACA AAATACTTGT TATTCGTAGA

TACCGATGAA TTTGTATTTC CAAACTTGAA AAATTATAAC TTAATGGATT

TTTTAAATTT ATTAGAAGCC AACAATCCTT ATTATAAAAA CAAAGTCGGG

GCAATGTGGA TTCCAATGTA TTTTCATTTT TTAGAGTGGG AATCTGATAA

AAATAATTTG AAGAAATATT CAACAATTGA GAAAAAAATT AAGAAAAAGA

TGGCAAATAT TGAGTTTGTT CTATATCGTA AAACATGTAG AATGTTAAGT

TCTGGAACAA AAAAAAGTGA CAAGACGAGA AGAAAAGTTA TTATTAGACC

TGAAAGAGTT TTGTATATGG GTATACATGA AACAGAAGAG ATGCTAAGCA

AAAAATTTCA TTTCATTAGA GCTCCTGTAA TTAATGTGGG TGGAGGAAAC

GAACTAAGTA TATATTTACA TCATTATAGA AAAGCAAAAG GTATTGTAAA

CAATGATCCC AAACAAAGAG AACTTGTGAA TATGTATTTA GAAAATGTTT

GTTCAGATAA GCTGTTAGAT TCAGGGGGAG ATTCCATTCA AGATGGAGTA

ATTGTCGACA ATACTGTTTG GGAGATATTT GGAACACACT TATACCAGAT

AATTTTTGAG CATATTAAAG AAATCCAAGA TATGTACACA AATAAGGAAA

TAATTAATGG AAATAAAAAT TTAAGTGTTG AAGAATTACA TAATTAA

SEQ ID NO: 10 (also listed in NCBI Ref Seq XP_626032)

MQSKVIFRIL VLIISVIGSL YSIIQLMLKE LSSNKNIQEV SHSRRLISEP YSESINEQND

QDWKELKLII PNHSQINQQE KNGNLIEFKV YIYSAYYDWR IDRIRINSLI PSNFYDRIEM

ECAIILDKNI YTGTIKKVIH KEHHNKEYVS STLLCEIAKN EIKFEDISRK VLITILENGN

STNKSEIWIT LKKIPKNSSN NHELTVCVRP WWGEPIKNGN LGNKQKFNNS

GLMLEFINSY LFLGANKFYL YQNYLDIDED VRNIINYYSN IKNVLEIIPY SLPIIPFKQV

WDFAQTTMIQ DCLLRNIGKT KYLLFVDTDE FVFPNLKNYN LMDFLNLLEA

NNPYYKNKVG AMWIPMYFHF LEWESDKNNL KKYSTIEKKI KKKMANIEFV

LYRKTCRMLS SGTKKSDKTR RKVIIRPERV LYMGIHETEE MLSKKFHFIR

APVINVGGGN ELSIYLHHYR KAKGIVNNDP KQRELVNMYL ENVCSDKLLD

SGGDSIQDGV IVDNTVWEIF GTHLYQIIFE HIKEIQDMYT NKEIINGNKN

LSVEELHN

[0020] The term "nucleic acid encoding a polypeptide" as it is used in the context of the present invention is meant to include allelic variations and redundancies in the genetic code.

[0021] The term "% (percent) identity" as known to the skilled artisan and used herein indicates the degree of relatedness among two or more nucleic acid molecules that is determined by agreement among the sequences. The percentage of "identity" is the result of the percentage of identical regions in two or more sequences while taking into consideration the gaps and other sequence peculiarities.

[0022] The identity of related nucleic acid molecules can be determined with the assistance of known methods. In general, special computer programs are employed that use algorithms adapted to accommodate the specific needs of this task. Preferred methods for determining identity begin with the generation of the largest degree of identity among the sequences to be compared. Preferred computer programs for determining the identity among two nucleic acid sequences comprise, but are not limited to, BLASTN (Altschul et al., J. Mol. Biol., 215, 403-410, 1990) and LALIGN (Huang and Miller, Adv. Appl. Math., 12, 337-357, 1991). The BLAST programs can be obtained from the National Center for Biotechnology Information (NCBI) and from other sources (BLAST handbook, Altschul et al., NCB NLM NIH Bethesda, Md. 20894).

[0023] The nucleic acid molecules according to the invention may be prepared synthetically by methods well-known to the skilled person, but also may be isolated from suitable DNA libraries and other publicly available sources of nucleic acids and subsequently may optionally be mutated. The preparation of such libraries or mutations is well-known to the person skilled in the art.

[0024] In a preferred embodiment, the nucleic acid molecules of the invention are cDNA, genomic DNA, synthetic DNA, RNA or PNA, either double-stranded or single-stranded (i.e. either a sense or an anti-sense strand). The nucleic acid molecules and fragments thereof, which are encompassed within the scope of the invention, may be produced by, for example, polymerase chain reaction (PCR) or generated synthetically using DNA synthesis or by reverse transcription using mRNA from *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* or *Cryptosporidium parvum.*

[0025] In some instances the present invention also provides novel nucleic acids encoding the polypeptides of the present invention characterized in that they have the ability to hybridize to a specifically referenced nucleic acid sequence, preferably under stringent conditions. Next to common and/or standard protocols in the prior art for determining the ability to hybridize to a specifically referenced nucleic acid sequence under stringent conditions (e.g. Sambrook and Russell, Molecular cloning: A laboratory manual (3 volumes), 2001), it is preferred to analyze and determine the ability to hybridize to a specifically referenced nucleic acid sequence under stringent conditions by comparing the nucleotide sequences, which may be found in gene databases (e.g. http://www.-ncbi.nlm.nih.gov/entrez/query.fcgi?db=nucleotide) with alignment tools, such as e.g. the above-mentioned BLASTN (Altschul et al., J. Mol. Biol., 215, 403-410, 1990) and LALIGN alignment tools.

[0026] Most preferably the ability of a nucleic acid of the present invention to hybridize to a nucleic acid, e.g. those listed in any of SEQ ID NOs 1, 3, 5, 7 and/or 9, is confirmed in a Southern blot assay under the following conditions: 6× sodium chloride/sodium citrate (SSC) at 45° C. followed by a wash in 0.2×SSC, 0.1% SDS at 65° C.

[0027] The nucleic acid of the present invention is preferably operably linked to a promoter that governs expression in suitable vectors and/or host cells producing the polypeptides of the present invention in vitro or in vivo.

[0028] Suitable promoters for operable linkage to the isolated and purified nucleic acid are known in the art. In a preferred embodiment the nucleic acid of the present invention is one that is operably linked to a promoter selected from the group consisting of the *Pichia pastoris* AOX1 or GAP promoter (see for example *Pichia* Expression Kit Instruction Manual, Invitrogen Corporation, Carlsbad, Calif.), the *Saccharomyces cerevisiae* GAL1, ADH1, ADH2, MET25, GPD or TEF promoter (see for example Methods in Enzymology, 350, 248, 2002), the Baculovirus polyhedrin p10 or ie1 promoter (see for example Bac-to-Bac Expression Kit Handbook, Invitrogen Corporation, Carlsbad, Calif., and Novagen Insect Cell Expression Manual, Merck Chemicals Ltd., Nottingham, UK), the *E. coli* T7, araBAD, rhaP BAD, tetA, lac, trc, tac or pL promoter (see Applied Microbiology and Biotechnology, 72, 211, 2006), the plant CaMV35S, ocs, nos, Adh-1, Tet promoters (see e.g. Lau and Sun, Biotechnol Adv. 27, 1015-1022, 2009) or inducible promoters for mammalian cells as described in Sambrook and Russell (2001).

[0029] Preferably, the isolated and purified nucleic acid is in the form of a recombinant vector, such as an episomal or viral vector. The selection of a suitable vector and expression control sequences as well as vector construction are within the ordinary skill in the art. Preferably, the viral vector is a baculovirus vector (see for example Bac-to-Bac Expresssion Kit Handbook, Invitrogen Corporation, Carlsbad, Calif.). Vector construction, including the operable linkage of a coding sequence with a promoter and other expression control sequences, is within the ordinary skill in the art.

[0030] Hence and in a further aspect, the present invention relates to a recombinant vector, comprising a nucleic acid of the invention.

[0031] A further aspect of the present invention is directed to a host cell comprising a nucleic acid and/or a vector of the invention and preferably producing polypeptides of the invention. Preferred host cells for producing the polypeptide of the invention are selected from the group consisting of yeast cells, preferably *Saccharomyces cerevisiae* (see for example Methods in Enzmology, 350, 248, 2002), *Pichia pastoris* cells (see for example *Pichia* Expression Kit Instruction Manual, Invitrogen Corporation, Carlsbad, Calif.), *E. coli* cells (BL21(DE3), K-12 and derivatives) (see for example Applied Microbiology and Biotechnology, 72, 211, 2006), plant cells, preferably *Nicotiana tabacum* or *Physcomitrella patens* (see e.g. Lau and Sun, Biotechnol Adv. 27, 1015-1022, 2009), NIH-3T3 mammalian cells (see for example Sambrook and Russell, 2001) and insect cells, preferably sf9 insect cells (see for example Bac-to-Bac Expression Kit Handbook, Invitrogen Corporation, Carlsbad, Calif.)

[0032] Another important aspect of the invention is directed to an isolated and purified polypeptide selected from the group consisting of

[0033] (a) polypeptides having an amino acid sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8 and 10, preferably SEQ ID NO: 2,

[0034] (b) polypeptides encoded by a nucleic acid of the present invention,

[0035] (c) polypeptides having an amino acid sequence identity of at least 25, 30 or 40%, preferably at least 50 or 60%, more preferably at least 70 or 80%, most preferably at least 90 or 95% with the polypeptides of (a) and/or (b),

[0036] (d) a fragment and/or functional derivative of (a), (b) or (c).

[0037] The identity of related amino acid molecules can be determined with the assistance of known methods. In general, special computer programs are employed that use algorithms adapted to accommodate the specific needs of this task. Preferred methods for determining identity begin with the generation of the largest degree of identity among the sequences to be compared. Preferred computer programs for determining the identity among two amino acid sequences comprise, but are not limited to, TBLASTN, BLASTP, BLASTX or TBLASTX (Altschul et al., J. Mol. Biol., 215, 403-410, 1990). The BLAST programs can be obtained from the National Center for Biotechnology Information (NCBI) and from other sources (BLAST handbook, Altschul et al., NCB NLM NIH Bethesda, Md. 20894).

[0038] Preferably, said polypeptides are encoded by an above-mentioned nucleic acid of the invention.

[0039] In a preferred embodiment, the polypeptide, fragment and/or derivative of the invention is functional, i.e. has enzymatic galactosyltransferase activity, preferably an enzymatic β-1,4-galactosyltransferase activity, more preferably an enzymatic β-1,4-galactosyltrans-ferase activity, preferably with L-fucoside-, more preferably with α-L-fucoside-, more preferably with Fuc-α-1,6-GlcNAc— and most preferably with $GnGnF^6$— (nomenclature according to Schachter, *Biochem. Cell. Biol.* 64(3), 163-181, 1986) containing poly/oligosaccharides or glycoconjugates as acceptor substrates.

[0040] For example, a preferred assay for determining the functionality, i.e. enzymatic activity, of the polypeptides, fragments and derivatives thereof according to the present invention is provided in example 4 below.

[0041] The term "functional derivative" of a polypeptide of the present invention is meant to include any polypeptide or fragment thereof that has been chemically or genetically modified in its amino acid sequence, e.g. by addition, substitution and/or deletion of amino acid residue(s) and/or has been chemically modified in at least one of its atoms and/or functional chemical groups, e.g. by additions, deletions, rearrangement, oxidation, reduction, etc. as long as the derivative still has at least one of the above enzymatic activities to a measurable extent, e.g. of at least about 1 to 10% of the original unmodified polypeptide.

[0042] In this context a functional fragment of the invention is one that forms part of a polypeptide or derivative of the invention and still has at least one of the above enzymatic activities in a measurable extent, e.g. of at least about 1 to 10% of the complete protein.

[0043] The term "isolated and purified polypeptide" as used herein refers to a polypeptide or a peptide fragment which either has no naturally-occurring counterpart (e.g., a peptide-mimetic), or has been separated or purified from components which naturally accompany it, e.g. in *Caenorhabditis elegans* tissue or a fraction thereof. Preferably, a polypeptide is considered "isolated and purified" when it makes up for at least 60% (w/w) of a dry preparation, thus being free from most naturally-occurring polypeptides and/or organic molecules with which it is naturally associated. Preferably, a polypeptide of the invention makes up for at least 80%, more preferably at 90%, and most preferably at least 99% (w/w) of a dry preparation. More preferred are polypeptides according to the invention that make up for at least 80%, more preferably at least 90%, and most preferably at least 99% (w/w) of a dry polypeptide preparation. Chemically synthesized polypeptides are by nature "isolated and purified" within the above context.

[0044] An isolated polypeptide of the invention may be obtained, for example, by extraction from a natural source, e.g. *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* or *Cryptosporidium parvum*; by expression of a recombinant nucleic acid encoding the polypeptide in a host, preferably a heterologous host; or by chemical synthesis. A polypeptide that is produced in a cellular system being different from the source from which it naturally originates is "isolated and purified", because it is separated from components which naturally accompany it. The extent of isolation and/or purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, HPLC analysis, NMR spectroscopy, gas liquid chromatography, or mass spectrometry.

[0045] Furthermore, in one aspect the present invention relates to antibodies, functional fragments and functional derivatives thereof that specifically bind a polypeptide of the invention. These are routinely available by hybridoma technology (Kohler and Milstein, Nature, 256, 495-497, 1975), antibody phage display (Winter et al., Annu. Rev. Immunol. 12, 433-455, 1994), ribosome display (Schaffitzel et al., J. Immunol. Methods, 231, 119-135, 1999) and iterative colony filter screening (Giovannoni et al., Nucleic Acids Res. 29, E27, 2001) once the target antigen is available. Typical proteases for fragmenting anti-bodies into functional products are well-known. Other fragmentation techniques can be used as well as long as the resulting fragment has a specific high affinity and, preferably a dissociation constant in the micromolar to picomolar range.

[0046] A very convenient antibody fragment for targeting applications is the single-chain Fv fragment, in which a variable heavy and a variable light domain are joined together by a polypeptide linker. Other antibody fragments for identifying the polypeptide of the present invention include Fab fragments, Fab_2 fragments, miniantibodies (also called small immune proteins), tandem scFv-scFv fusions as well as scFv fusions with suitable domains (e.g. with the Fc portion of an immunoglobulin). For a review on certain antibody formats, see Holliger and Hudson, Biotechnol., 23(9), 1126-36, 2005.

[0047] The term "functional derivative" of an antibody for use in the present invention is meant to include any antibody or fragment thereof that has been chemically or genetically modified in its amino acid sequence, e.g. by addition, substitution and/or deletion of amino acid residue(s) and/or has been chemically modified in at least one of its atoms and/or functional chemical groups, e.g. by additions, deletions, rearrangement, oxidation, reduction, etc. as long as the derivative has substantially the same binding affinity as to its original antigen and, preferably, has a dissociation constant in the micro-, nano- or picomolar range.

[0048] In a preferred embodiment, the antibody, fragment or functional derivative thereof for use in the invention is one that is selected from the group consisting of polyclonal antibodies, monoclonal antibodies, chimeric antibodies, humanized antibodies, CDR-grafted antibodies, Fv-fragments, Fab-

fragments and Fab_2-fragments and antibody-like binding proteins, e.g. affilines, anticalines and aptamers.

[0049] For a review of antibody-like binding proteins see Binz et al. on engineering binding proteins from non-immunoglobulin domains in Nature Biotechnol., 23(10), 1257-1268, 2005. The term "aptamer" describes nucleic acids that bind to a polypeptide with high affinity. Aptamers can be isolated from a large pool of different single-stranded RNA molecules by selection methods such as SELEX (see, e.g., Jayasena, Clin. Chem., 45, 1628-1650, 1999; Klug and Famulok, M. Mol. Biol. Rep., 20, 97-107, 1994; U.S. Pat. No. 5,582,981). Aptamers can also be synthesized and selected in their mirror form, for example, as the L-ribonucleotide (Nolte et al., Nat. Biotechnol., 14, 1116-1119, 1996; Klussmann et al., Nat. Biotechnol., 14, 1112-1115, 1996). Forms isolated in this way have the advantage that they are not degraded by naturally occurring ribonucleases and, therefore, have a greater stability.

[0050] Another antibody-like binding protein and alternative to classical antibodies are the so-called "protein scaffolds", for example, anticalines, that are based on lipocaline (Beste et al., Proc. Natl. Acad. Sci. USA, 96, 1898-1903, 1999). The natural ligand binding sites of lipocalines, for example, of the retinol-binding protein or bilin-binding protein, can be changed, for example, by employing a "combinatorial protein design" approach, and in such a way that they bind selected haptens (Skerra, Biochem. Biophys. Acta, 1482, pp. 337-350, 2000). For other protein scaffolds it is also known that they are alternatives for antibodies (Skerra, J. Mol. Recognition, 13, 167-287, 2000; Hey, Trends in Biotechnology, 23, 514-522, 2005).

[0051] In summary, the term functional antibody derivative is meant to include the above protein-derived alternatives for antibodies, i.e. antibody-like binding proteins, e.g. affilines, anticalines and aptamers, that specifically recognize a polypeptide, fragment or derivative thereof.

[0052] A further aspect relates to a hybridoma cell line, expressing a monoclonal antibody according to the invention.

[0053] The nucleic acids, vectors, host cells, polypeptides and antibodies of the present invention have a number of new applications.

[0054] In one aspect the present invention relates to the use of a polypeptide, a cell extract comprising a polypeptide of the invention, preferably a nematode extract, more preferably an extract of *Caenrhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* or *Cryptosporidium parvum*, and/or a host cell of the present invention for producing galactoside-containing oligo/polysaccharides and/or glycoconjugates, preferably galactosyl-fucoside-containing oligo/polysaccharides and glycoconjugates, more preferably D-galactopyranosyl-β-1,4-L-fucopyranosyl-α-1,6-GlcNAc-containing oligo/polysaccharides and glycoconjugates, most preferably $GnGnF^6Gal$- or MMF^6Gal-containing oligo/polysaccharides and glycoconjugates.

[0055] It is understood that the term glycoconjugate, as used herein is non-limiting with respect to the nature of the non-sugar component. Preferably the non-sugar component of the glycoconjugate is a poly/oligopeptide.

[0056] The enzymatic synthesis of galactosyl-fucosyl-specific oligosaccharides and glycoconjugates is highly specific, controlled and environment-friendly and the products can serve as highly parasite-specific (this epitope is only known to also exist in octopus [Zhang et al., Glycobiology, 7, 1153-1158, 1997], squid [Takahashi et al., Eur. J. Biochem., 270, 2627-2632, 2003] and limpets [Wuhrer et al., Biochem. J., 378, 625-632, 2004]) vaccine components for the treatment and prevention of parasitic, preferably nematode and apicomplexa infections in a subject, such as a human or other mammal, in need thereof.

[0057] Exemplary and preferred galactosyl-fucosyl-specific oligosaccharides and glycoconjugates are selected from the group consisting of N-linked glycans, N-glycoproteins, glycolipids and lipid-linked oligosaccharides (LOS). The term "glycoconjugate" as used herein, is meant to include any type of conjugate, preferably but not necessarily a covalently bonded one, for example bonded by a covalent linker, of an oligosaccharide- and a non-saccharide component, e.g. a polypeptide or any other type of organic or inorganic carrier that is physiologically acceptable and might even have a desired physiological function, e.g. as an immune stimulating adjuvant, imparting nematode toxicity, etc.

[0058] For example, raw extracts of *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* or *Cryptosporidium parvum* or recombinant insect cells producing a polypeptide of the invention can produce Gal-Fuc-containing conjugates, e.g. free Gal-Fuc glycans, Gal-Fuc-peptides, Gal-Fuc-polypeptides, Gal-Fuc-folded proteins. Alpha-1,6-linked fucosides are strongly preferred over alpha-1,3-linked fucosides.

[0059] Another aspect of the present invention is directed to a method for producing galactosyl-fucosyl derivatives, comprising the following steps:

[0060] (i) providing at least one polypeptide of the invention,

[0061] (ii) providing at least one fucosylated acceptor substrate,

[0062] (iii) incubating (i) and (ii) in the presence of at least one suitable divalent metal cation cofactor, preferably selected from manganese (II), cobalt (II) and/or iron (II) ions, more preferably manganese (II), and at least one activated sugar substrate, preferably uridine diphosphate (UDP)-galactose under conditions suitable for enzymatic activity of the polypeptide of the invention,

[0063] (iv) optionally isolating the galactosyl-fucose derivatives.

[0064] The polypeptide of the invention may be provided as an isolated polypeptide, in dry or soluble form, in a buffer, a host cell, a cell extract or any other system that will sustain its enzymatic activity and allow access to its substrate and activated sugar substrate. The fucosylated acceptor substrate is any kind of fucosyl-containing substrate, optionally in isolated form or as a component of a system that can be enzymatically modified by the polypeptide of the invention. The activated sugar substrate is preferably UDP-galactose but can also be any other type of activated, preferably phosphate-activated galactosyl derivative that can be transferred to a fucosylated acceptor substrate. The method of the invention preferably leads to galactopyranosyl-β-1,4-L-fucopyranosyl-derivatives, more preferably D-galactopyranosyl-β-1,4-L-fucopyranosyl-α-1,6-βGlcNAc (Gal-Fuc) derivatives.

[0065] The polypeptides of the present invention have a broad substrate specificity as long as the substrate features a suitable fucosyl-moiety. Galactosyl-transferase activity was demonstrated for substrates such as, e.g. fucosyl-saccharides, fucosyl-peptides, fucosyl-polypeptides and even complex and folded fucosyl-polypeptides. For example, galactosyl-

transferase activity was demonstrated for human IgG1, a glycoprotein having $GnGnF^6$ carbohydrate structures as prevalent epitopes. These IgG1 glycans are known to be accessible for PNGaseF digest. Glycosylation of human IgG1 was demonstrated with the crude sf9 insect cell extract containing the core galactosyltransferase of *Caenorhabditis elegans*. Incubation of human IgG1 with radioactively labelled UDP-Gal in the presence of enzyme extract from *Caenorhabditis elegans* led to substrate galactosylation. In addition, galactosylation was demonstrated on remodelled human transferrin carrying $GnGnF^6$ carbohydrate structures as prevalent epitopes. For this purpose human apotransferrin was sequentially treated with sialidase (Iskratsch et al, Anal. Biochem., 368, 133-146, 2009), β1,4-galactosidase from *Aspergillus oryzae* and recombinant *Anopheles* core α1,6-FucT expressed in *Pichia pastoris* to produce a glycoprotein having $GnGnF^6$ carbohydrate structures as prevalent epitopes. Incubation with a crude sf9 insect cell extract containing the core galactosyltransferase of *Caenorhabditis elegans* led to galactosylation which was monitored by dot blotting with the fucose-specific *Aleuria aurantia* lectin and by MALDI-TOF MS of tryptic peptides of the various neoglycoforms.

[0066] It has very recently been shown that the serum content of core fucosylated alpha feto-protein (AFP) is highly specific for hepatocellular carcinomas (HCC), because benign liver diseases such as chronic hepatitis and liver cirrhosis do not lead to core-fucosylated AFP in mammals, in particular humans (see Tateno et al., Glycobiology, 19(5), 527-536. 2009).

[0067] Therefore, in a further aspect the polypeptides of the invention, host cells comprising polypeptides of the invention and/or cell extracts of *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* and/or *Cryptosporidium parvum* can be used for covalently binding galactosyl compounds to core-fucosylated alpha-fetoprotein (AFP), preferably for detecting and/or quantifying hepatocellular carcinoma (HCC) cells, preferably by selectively labelling core-fucosylated alpha-fetoprotein (AFP) from the blood of HCC patients, because core-fucosylated AFP is selectively suitable as an acceptor substrate for the polypeptides of the present invention.

[0068] Hence, the present invention relates to polypeptides of the invention, host cells comprising polypeptides of the invention and/or cell extracts of *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* and/or *Cryptosporidium parvum* for preparing diagnostic means for detecting core-fucosylated AFP, i.e. for detecting and/or quantifying hepatocellular carcinoma (HCC) cells.

[0069] Also, the polypeptides of the invention, host cells comprising polypeptides of the invention and/or cell extracts of *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* and/or *Cryptosporidium parvum* are useful for preparing diagnostic means for detecting further core-fucosylated marker glycoproteins whose appearance correlates with other types of carcinoma cells.

[0070] In a preferred embodiment, the invention relates to a method of diagnosis, comprising the following steps:

(i) providing blood or a fraction thereof, that comprises AFP, preferably serum,

(ii) incubating said blood or said fraction thereof with (a) a polypeptide of the invention, a host cell of the invention and/or cell extracts of *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* and/or *Cryptosporidium parvum* and (b) an activated galactosyl derivative, preferably a labelled galactosyl derivative, preferably labelled UDP-galactose, under conditions that allow for the galactosyltransfer of activated galactose to core-fucosylated AFP (AFP-L3),

(iii) and detecting the galactose-labelled and hence core-fucosylated AFP (AFP-L3).

[0071] Labels for activated galactosyl derivatives for practicing the above method are selected from the group consisting of isotopes e.g. ^{14}C, chemical modifications e.g. halogen substitutions and other selectively detectable modifications e.g. biotin, azide etc. Preferably, all of the steps (i) to (iii) are performed outside the living body, i.e. in vitro.

[0072] A further aspect of the invention is directed to the use of antibodies specifically binding a polypeptide of the invention, preferably a polypeptide having a sequence selected from any of SEQ ID NOs: 2, 4, 6, 8 and/or 10, for identifying and/or quantifying nematodes and apicomplexa, preferably *Caenorhabditis elegans, Caenorhabditis briggsae*, and *Cryptosporidium parvum*, respectively, in a sample of interest, for example a human or mammalian sample, preferably in a cell fraction or extract sample. The design and development of typical antibody assays, e.g. ELISAs, is within the ordinary skill in the art and need not be further elaborated.

[0073] The invention has been described with the emphasis upon preferred embodiments and illustrative examples. However, it will be obvious to those of ordinary skill in the art that variations of the preferred embodiments may be used and that it is intended that the invention may be practiced otherwise than as specifically described herein. Moreover, as the foregoing examples are included for purely illustrative purposes, they should not be constructed to limit the scope of the invention in any respect. Accordingly, this invention includes all modifications encompassed within the spirit and scope of the invention as defined by the claims appended hereto.

FIGURES

[0074] FIG. **1** is an anti-FLAG immunoblotting of baculovirus-infected sf9 whole cell extracts. Different clones of baculoviruses containing empty vector control (e.v.), N-terminally FLAG-tagged M03F8.4 (FLAG-GalT) and untagged M03F8.4 (GalT). Loading ca. 150 kcells/slot, SDS-PAGE 12%, α-FLAG (1:2000, SIGMA), α-mouse-HRP (1:2000, Santa Cruz Biotechnology), ECL (Pierce, 2 s exposure).

[0075] FIG. **2** is an SDS-PAGE analysis of baculovirus-infected sf9 whole cell extracts. Different clones of baculoviruses containing empty vector control (e.v.), N-terminally FLAG-tagged M03F8.4 (FLAG-GalT) and untagged M03F8.4 (GalT). Loading ca. 150 kcells/slot, SDS-PAGE 12%, detection by silver staining. (Protein is expressed in low amounts, not detectable by silver staining with respect to the empty vector construct in crude extracts.)

[0076] FIG. **3** is a column chart showing the galactosylation turnover of a $GnGnF^6$ acceptor substrate (dabsyl-GEN [$GnGnF^6$]R) in the presence of Mn^{2+}, Mg^{2+} and EDTA demonstrating metal ion dependency; MES, pH 6, r.t., 2.5 h, turnover determined by ratio of MALDI-MS peak intensity ([m/z 2369/(m/z 2207+m/z 2369)]*100) from crude reaction mixture.

[0077] FIG. **4** is a column chart showing the galactosylation of a GnGnF^6 acceptor substrate (dabsyl-GEN[GnGnF^6]R)—functionality of the tagged and non-tagged construct; MES, pH 6, r.t., 2.5 h, turnover determined by ratio of MALDI-MS peak intensity ([m/z 2369/(m/z 2207+m/z 2369)]*100) from crude reaction mixture.

[0078] FIG. **5** shows the galactosylation of a GnGnF^6 acceptor substrate (dabsyl-GEN[GnGnF^6]R)—functionality of the tagged and non-tagged construct (MES pH 6, r.t., 2.5 h) by way of MS analysis. Upper spectrum: reaction without UDP-Gal, central spectrum: with UDP-Gal, bottom spectrum: digest of the product from the central spectrum with *Aspergillus* β-galactosidase (citrate buffer, pH 5, r.t., 2 d). The enzyme clearly adds a galactose to this acceptor substrate which can be digested with (3-galactosidase, and therefore shows a β-linked Gal residue incorporated by the GalT. Additional GlcNAc removal takes place after prolonged reaction times (>2 d) due to presence of hexosaminidase in the insect cell crude extract.

[0079] FIG. **6** is a comparison of MS/MS spectra of acceptor (upper spectrum) and galactosylated reaction product (lower spectrum) of FIG. **5**. The MS/MS analysis clearly shows the galactose being linked to the core fucose, as observed from secondary ion 1272.61 corresponding to a Hex-dHex-HexNAc motif linked to the dabsylated GENR peptide.

[0080] FIG. **7** is a comparative analysis of the donor specificity of the galactosyl transferase (dansyl-N[GnGnF^6]ST, MES pH 6.5, Mn^{2+}, r.t., 13 h). The enzyme seems to have a high specificity for UDP-Gal, with a negligible residual activity on UDP-Glc.

[0081] FIG. **8** is column chart of an analysis of the acceptor specificity: *Caenorhabditis elegans* GalT galactosylates selectively α-1,6 linked over α-1,3-linked fucose; dabsyl-GEN-[MMF$^{6/3}$]R, MES pH 6.5, r.t., 2.5 h, turnover determined by ratio of MALDI-MS peak intensity ([m/z 1963/(m/z 1801+m/z 1963)]*100) from crude reaction mixture.

[0082] FIG. **9** shows the graphic determination of the K_m (app) of the untagged galactosyl transferase for UDP-Gal: K_m (app, UDP-Gal)=ca. 40 μM.

[0083] FIG. **10** is an analysis of the temperature dependency of the galactosyltransferase of the invention (dansyl-N[GnGnF^6]ST, UDP-Gal, MES pH 6.5, Mn^{2+}, 2.5 h).

[0084] FIG. **11** is a column chart demonstrating the glycosylation of human IgG1 (possessing GnGnF^6 epitopes) with the polypeptide of the invention, i.e. *Caenorhabditis elegans* core galactosyltransferase.

[0085] FIG. **12** is a MALDI-TOF MS spectrum demonstrating the glycosylation of remodelled human transferrin (possessing GnGnF^6 epitopes) with a polypeptide of the invention, i.e. *Caenorhabditis elegans* core galactosyltransferase. The indicated mk values correspond to peptide 622-642 carrying GnGn (3813), GnGnF^6 (3957) and GnGnF^{6}Gal (4119), respectively.

EXAMPLES

Experimental Procedures

Chemicals and Suppliers

[0086] UDP-Gal (VWR International and Sigma), UDP-Glc, UDP-GlcNAc, UDP-GalNAc (all SIGMA), UDP-^{14}C-Gal (GE Healthcare), GlcNAc-β-1,2-Man-α-1,6-[GlcNAc-β-1,2-Man-α-1,3-]-Man, Man-β-1,4-GlcNAc-β-1,4-[α-1,6-Fuc]-GlcNAc, MMF6, GnGnF^6 (all Dextra Laboratories, UK), Fuc-α-1,6-GlcNAc (Carbosynth Ltd., UK), dabsyl-GEN[GnGnF^6]R (Paschinger et al., Glycobiology, 15(5), 463-474, 2005), dabsyl-GEN[MMF6]R (Fabini et al., J. Biol. Chem. 276(30), 28058-28067, 2001), dabsyl-GEN[MMF3]R (Fabini et al., J. Biol. Chem. 276(30), 28058-28067, 2001) and dansyl-N[GnGnF^6]ST (Roitinger et al., Glycoconj. J., 15(1), 89-91, 1998) were obtained according to previously published methods.

Example 1

Isolation of *Caenorhabditis elegans* cDNA and Cloning of M03F8.4 into Expression Vectors

Nematode Strains:

[0087] Methods for culturing *Caenorhabditis elegans* are described in Brenner, S. (Genetics 77(1), 71-94, 1974). The wild type Bristol N2 strain was grown at 20° C. on standard NGM agar plates seeded with *Escherichia coli* OP50.

Isolation of *Caenorhabditis elegans* M03F8.4 cDNA:

A *Caenorhabditis elegans* mixed culture was harvested from one standard NGM agar plate and washed twice in sterile M9 buffer (22 mM KH_2PO_4, 42 mM Na_2HPO_4, 85 mM NaCl, 1 mM $MgSO_4$). Total RNA was extracted using the NucleoSpin® RNA II RNA isolation kit (MACHEREY-NAGEL AG). cDNA synthesis was performed with 0.5 μg total RNA using the First-strand cDNA synthesis step of the SuperScript™ III Platinum Two-Step qRT-PCR Kit (Invitrogen AG).

Construction of the pFastBac1 Donor Plasmid for Recombinant Gene Expression in sf9 Insect Cells:

[0088] M03F8.4 cDNA was isolated from a previously prepared cDNA library by PCR using Phusion High-Fidelity DNA Polymerase (Finnzymes) according to the manual supplied. For construction of an untagged version, the following forward and reverse primers, flanked with SalI and XbaI restrictions sites, respectively, were used: 5'-TTTGTCGACACTTCTGAATGCCTCG-3' (SEQ ID NO: 11) and 5'-TTTTCTAGACTACAAGTCTAA-AAGACCAAC-3' (SEQ ID NO: 12). The resulting fragment was digested with the appropriate restriction enzymes and cloned into the pFastBac1 donor plasmid (Invitrogen). For construction of an N-terminally FLAG tagged version, a forward primer lacking the start codon was used: 5'-TTTGTCGACCCTCGAATCACCGCC-3' (SEQ ID NO: 13). The resulting fragment was cloned into a pFastBac1 donor plasmid containing an N-terminal FLAG sequence (Muller et al., J. Biol. Chem. 277(36), 32417-32420, 2002) (both vectors kindly provided by Thierry Hennet, Institute of Physiology, University of Zurich).

Example 2

Expression of Recombinant Proteins

[0089] Recombinant baculoviruses containing the *Caenorhabditis elegans* core beta-1,4-GalT candidate cDNA (with and without N-terminal FLAG-tag) and an empty vector control were generated according to the manufacturers instructions (Invitrogen). After infection of 2×10^6 *S. frugiperda* (sf9) adherent insect cells with recombinant baculoviruses and incubation for 72 h at 28° C., the cells were lysed with shaking (4° C., 15 min) in 150 μL tris-buffered saline (pH 7.4) containing 2% (v/v) Triton-X100 and protease inhibitor cocktail (Roche, complete EDTA-free). The lysis

mixtures were centrifuged (2000×g, 5 min) and the post-nuclear supernatant was recovered and used for all further enzymatic studies.

Example 3

Denaturing Gel Electrophoretic Analysis and Immunoblotting

[0090] Infected sf9 cells (2×10^6 cells, see above) were vortexed in 200 μL Laemmli buffer and proteins denatured by heating (95° C., 5 min). After cooling to r.t. the samples were centrifuged (16 krpm, 5 min) and the supernatant was used for further analysis. The samples were separated by SDS-PAGE (12% acrylamide, 120 V). The resulting gels were either analyzed by silver-staining or by blotting onto a nitrocellulose membrane. After blocking the membrane (5% BSA in PBST) immuno-detection was performed by incubation with anti-FLAG antibody M2 (SIGMA, dilution 1:2000 in PBST+1% BSA) followed by anti-mouse-HRP (Santa Cruz Biotechnology, dilution 1:10000 in PBST+1% BSA) after extensive washing (PBST) and final detection using ECL (Pierce) and exposure to photographic film.

Example 4

Glycosyltransferase Assays

[0091] Enzymatic activity towards appropriate carbohydrates or glycoconjugates was assessed using 0.5 μL of raw extract of sf9 cells (containing either an empty vector control bacmid, a putative GalT expressing bacmid or a putative FLAG-tagged GalT expressing bacmid) in 2.5 μL final volume of MES buffer (pH 6.5, 40 μM) containing manganese (II) chloride (10 μM), UDP-galactose (1 mM) and the acceptor fucoside (glycan or glyco(poly)peptide, 40 μM). Glycosylation reactions were typically run for 2 h at room temperature, unless noted otherwise. For donor specificity analysis UDP-galactose was replaced by equal concentrations of UDP-Glc, UDP-GlcNAc or UDP-GalNAc (Sigma) respectively. For co-factor-specificity analysis $MnCl_2$ was replaced by equal concentrations of the various metal chlorides or Na_2EDTA. To quantify the incorporation of galactose into the acceptor glycans total UDP-Gal concentration was doped with 10% UDP-^{14}C-Gal (25 nCi, GE Healthcare). Excess radioactivity (UDP-^{14}C-Gal) was removed by loading the reaction mixture (quenched with 100 μL H_2O) onto a column of anion exchange resin (AG1-X8, Cl^- form, Bio-Rad Laboratories, 200 mg) and elution of the uncharged products (H_2O, 900 μL).

[0092] Glycosylation of human IgG1 (5 μL of 3 g/L, Calbiochem) was performed in 50 μL total volume using the same buffer, salt and enzyme conditions as described above, except the absence of non-radioactive UDP-Gal, which was replaced by UDP-^{14}C-Gal (75 nCi). The reaction was performed at r.t. over night. A suspension of sepharose-protein G beads (Amersham Biosciences, 10 μL) in PBS (200 μL) was added and binding of IgG1 to the beads was done with shaking (4° C., 1 h). The beads were washed with PBS (5×200 μL) and IgG1 was eluted with 20 mM aqueous HCl (3×100 μL). Analysis (vide infra) of the reaction products was performed either by direct MALDI-TOF mass spectrometry, HPLC analysis of fluorescently labelled glycopeptides for donor specificity or scintillation counting of radio-labelled assays.

[0093] Stepwise remodelling of human asialotransferrin N-glycans was performed as follows: Asialotransferrin (GalGal) was previously prepared by sialidase treatment of human apotransferrin (Iskratsch et al, Anal. Biochem., 368, 133-146, 2009).

[0094] To produce asialoagalactotransferrin (GnGn), β1,4-galactosidase (3U, from *Aspergillus oryzae*) was added to about 1 mg of GalGal and the sample was incubated for 48 hours at 37° C. (total volume 50 μl).

[0095] To obtain GnGnF^6, the sample was brought to a neutral pH with 0.5 μl 1M NaOH, before 50 nmol of GDP-fucose and 15 μl of a preparation of recombinant *Anopheles* core α1,6-FucT, expressed in *Pichia pastoris*, were added. The preparation was incubated overnight before another 50 nmol of GDP-fucose and a further 15 μl enzyme (FucT) were added and again incubated overnight at 37° C. In total, approximately 1 mg of GnGnF^6 was obtained.

[0096] To prepare GalFuc-transferrin, 1 μl of a preparation of recombinant *Caenorhabditis elegans* GalT, 0.2 mmol of $MnCl_2$ and 20 nmol of UDP-galctose were added to an aliquot of GnGnF^6 (300 μg) and incubated overnight at 30° C. Again, the desired glycan structure was boosted with a second incubation overnight after the addition of further substrate (UDP-galactose) and enzyme (GalT).

[0097] The degree of modification of the transferrin was monitored by dot blotting with the fucose-specific *Aleuria aurantia* lectin and by MALDI-TOF MS of tryptic peptides of the various neoglycoforms.

Example 5

Structural Analysis

[0098] After exposing dabsyl-GEN[GnGnF^6]R to galactosylation conditions, the resulting crude mixture was adjusted to 50 mM sodium citrate and pH 4.5, digested with *Aspergillus oryzae* β-galactosidase (27 mU) (see Gutternigg et al., *J. Biol. Chem.* 282(38), 27825-27840, 2007) for 2 days at 30° C. The samples were analyzed by MALDI-TOF mass spectrometry (vide infra).

[0099] HPLC Analysis:

[0100] Both, for analysis of donor specificity and the reaction rate dependence on donor concentration, the dansyl-N[GnGnF^6]ST acceptor substrate was separated from the galactosylated reaction product using an isocratic solvent system (0.7 mL/min, 9% MeCN (95%, (v/v)) in 0.05% aqueous TFA (v/v)) on a reversed phase Hypersil ODS C18 column (4×250 mm, 5 μm) and fluorescence detection (excitation at 315 nm, emission detected at 550 nm) at room temperature. The Shimadzu HPLC system consisted of a SCL-10A controller, two LC10AP pumps and a RF-10AXL fluorescence detector controlled by a personal computer using Class-VP software (V6.13SP2). Dansyl-N[GnGnF^6]ST eluted at a retention time of 9.09 min and the galactosylated reaction product at 8.06 min.

Mass Spectrometry:

[0101] Glycans were analyzed by MALDI-TOF mass spectrometry on a BRUKER Ultraflex TOF/TOF machine using a α-cyano-4-hydroxy cinnamic acid matrix. A peptide standard mixture (Bruker) was used for external calibration.

Scintillation Counting:

[0102] The eluates of the anion exchange resin column and protein G beads were thoroughly mixed with scintillation

fluid (Irga-Safe Plus, Packard, 4 mL) and measured with a Perkin Elmer Tri-Carb 2800TR.

Abbreviations for Carbohydrates:

[0103] Fuc—L-fucose, Gal—D-galactose, GalNAc—D-N-acetylgalactosamine, Glc—D-glucose, GlcNAc—D-N-acetylglucosamine, Man—D-mannose

[0104] Abbreviations for complex glycans (according to the Schachter nomenclature [Biochem Cell Biol 64(3), 163-181, 1986]):

[0105] GalGal Gal-β-1,4-GlcNAc-β-1,2-Man-α-1,6-[Gal-β-1,4-GlcNAc-β-1,2-Man-α-1,3-]-Man-β-1,4-GlcNAc-β-1,4-GlcNAc

[0106] GnGn GlcNAc-β-1,2-Man-α-1,6-[GlcNAc-β-1,2-Man-α-1,3-]-Man-β-1,4-GlcNAc-β-1,4-GlcNAc

[0107] GnGnF^{6} GlcNAc-β-1,2-Man-α-1,6-[GlcNAc-β-1,2-Man-α-1,3-]-Man-β-1,4-GlcNAc-β-1,4-[α-1,6-Fuc]-GlcNAc

[0108] GnGnF^{6}Gal GlcNAc-β-1,2-Man-α-1,6-[GlcNAc-β-1,2-Man-α-1,3-]-Man-β-1,4-GlcNAc-β-1,4-[Gal-β-1,4-Fuc-α-1,6]-GlcNAc

[0109] MMF^{6} Man-α-1,6-[Man-α-1,3-]-Man-β-1,4-GlcNAc-β-1,4-[α-1,6-Fuc]-GlcNAc

[0110] MMF^{6}Gal Man-α-1,6-[Man-α-1,3-]-Man-β-1,4-GlcNAc-β-1,4-[Gal-β-1,4-Fuc-α-1,6]-GlcNAc

[0111] MMF^{3} Man-α-1,6-[Man-α-1,3-]-Man-β-1,4-GlcNAc-β-1,4-[α-1,3-Fuc]-GlcNAc

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 13

<210> SEQ ID NO 1
<211> LENGTH: 1278
<212> TYPE: DNA
<213> ORGANISM: Caenorhabditis elegans

<400> SEQUENCE: 1

atgcctcgaa tcaccgccag taaaatagtt cttctaattg cattatcatt ttgtattact 60

gttatttatc actttccaat agcaacgaga agcagtaagg agtacgatga atatggaaat 120

gaatatgaaa acgttgcatc gatagagtcg gatataaaaa atgtacgtcg attacttgac 180

gaggtaccgg atccctcaca aaaccgtcta caattcctga aacttgatga gcatgctttt 240

gcattctcgg cctacacaga cgatcgaaat ggaaatatgg ggtacaaata tgtccgagtc 300

ctgatgttta tcacgtcaca agacaacttt tcctgtgaaa taaacgggag aaagtccaca 360

gatgtatcac tttacgagtt ctcggaaaat cacaaaatga agtggcaaat gtttattttg 420

aattgtaaac tacccgatgg tatagatttc aataatgtta gctctgtaaa ggtcataaga 480

agcacaacca agcagtttgt tgatgtgccg attcggtata gaattcaaga tgagaaaata 540

attacgccag acgaatatga ctataaaatg tcaatttgtg ttccagcatt gtttggaaat 600

ggatatgatg caaagcgaat tgttgagttt attgagctga atactttgca aggaatcgag 660

aaaatataca tttacactaa tcaaaaagag cttgatggat ccatgaagaa aacgttgaaa 720

tactattcgg ataatcacaa aataacatta attgattaca cattaccatt cagagaggat 780

ggtgtttggt atcacgggca attggcaact gttactgatt gtttactgag aaacactgga 840

atcacaaaat acacattttt caatgatttt gatgagttct tcgtccccgt tatcaaaagt 900

cggactctct ttgaaacaat cagtgggctt tttgaagatc ccactattgg atcgcaacga 960

acagctttga agtatataaa tgcaaaaatc aagagcgctc cgtattcact gaaaaatatt 1020

gtttccgaaa aacgaattga aacaagattc acgaaatgtg tagttcgacc ggaaatggtt 1080

tttgaacagg gtattcatca tacgagtaga gtgattcaag acaactataa aacggtttcc 1140

catggcggat cccttctacg ggtttatcat tacaaggata aaaagtattg ttgcgaagac 1200

gagagcctct tgaaaaaacg gcatggagat caacttcggg aaaaattcga ttcagttgtt 1260

ggtcttttag acttgtag 1278

<210> SEQ ID NO 2

-continued

<211> LENGTH: 425
<212> TYPE: PRT
<213> ORGANISM: Caenorhabditis elegans

<400> SEQUENCE: 2

Met Pro Arg Ile Thr Ala Ser Lys Ile Val Leu Leu Ile Ala Leu Ser
1 5 10 15

Phe Cys Ile Thr Val Ile Tyr His Phe Pro Ile Ala Thr Arg Ser Ser
20 25 30

Lys Glu Tyr Asp Glu Tyr Gly Asn Glu Tyr Glu Asn Val Ala Ser Ile
35 40 45

Glu Ser Asp Ile Lys Asn Val Arg Arg Leu Leu Asp Glu Val Pro Asp
50 55 60

Pro Ser Gln Asn Arg Leu Gln Phe Leu Lys Leu Asp Glu His Ala Phe
65 70 75 80

Ala Phe Ser Ala Tyr Thr Asp Asp Arg Asn Gly Asn Met Gly Tyr Lys
85 90 95

Tyr Val Arg Val Leu Met Phe Ile Thr Ser Gln Asp Asn Phe Ser Cys
100 105 110

Glu Ile Asn Gly Arg Lys Ser Thr Asp Val Ser Leu Tyr Glu Phe Ser
115 120 125

Glu Asn His Lys Met Lys Trp Gln Met Phe Ile Leu Asn Cys Lys Leu
130 135 140

Pro Asp Gly Ile Asp Phe Asn Asn Val Ser Ser Val Lys Val Ile Arg
145 150 155 160

Ser Thr Thr Lys Gln Phe Val Asp Val Pro Ile Arg Tyr Arg Ile Gln
165 170 175

Asp Glu Lys Ile Ile Thr Pro Asp Glu Tyr Asp Tyr Lys Met Ser Ile
180 185 190

Cys Val Pro Ala Leu Phe Gly Asn Gly Tyr Asp Ala Lys Arg Ile Val
195 200 205

Glu Phe Ile Glu Leu Asn Thr Leu Gln Gly Ile Glu Lys Ile Tyr Ile
210 215 220

Tyr Thr Asn Gln Lys Glu Leu Asp Gly Ser Met Lys Lys Thr Leu Lys
225 230 235 240

Tyr Tyr Ser Asp Asn His Lys Ile Thr Leu Ile Asp Tyr Thr Leu Pro
245 250 255

Phe Arg Glu Asp Gly Val Trp Tyr His Gly Gln Leu Ala Thr Val Thr
260 265 270

Asp Cys Leu Leu Arg Asn Thr Gly Ile Thr Lys Tyr Thr Phe Phe Asn
275 280 285

Asp Phe Asp Glu Phe Phe Val Pro Val Ile Lys Ser Arg Thr Leu Phe
290 295 300

Glu Thr Ile Ser Gly Leu Phe Glu Asp Pro Thr Ile Gly Ser Gln Arg
305 310 315 320

Thr Ala Leu Lys Tyr Ile Asn Ala Lys Ile Lys Ser Ala Pro Tyr Ser
325 330 335

Leu Lys Asn Ile Val Ser Glu Lys Arg Ile Glu Thr Arg Phe Thr Lys
340 345 350

Cys Val Val Arg Pro Glu Met Val Phe Glu Gln Gly Ile His His Thr
355 360 365

Ser Arg Val Ile Gln Asp Asn Tyr Lys Thr Val Ser His Gly Gly Ser

-continued

```
    370                 375                 380

Leu Leu Arg Val Tyr His Tyr Lys Asp Lys Lys Tyr Cys Cys Glu Asp
385                 390                 395                 400

Glu Ser Leu Leu Lys Lys Arg His Gly Asp Gln Leu Arg Glu Lys Phe
                405                 410                 415

Asp Ser Val Val Gly Leu Leu Asp Leu
            420                 425
```

<210> SEQ ID NO 3
<211> LENGTH: 1311
<212> TYPE: DNA
<213> ORGANISM: Caenorhabditis briggsae

<400> SEQUENCE: 3

```
atgccacgaa taacggcaag caaaatagtg ttattatctg tattatcctt actaacagtt     60

ttctatctga atacattttc gtctattaaa attgaaaacg atctcgacgg gactgattac    120

gacttggatt acatagaatc tgatatcaaa aagacgcgtc gattactcaa tgaaatccct    180

gatccatctc aaaaccgagt tcaatttttt aaactcgatg ataatggata tgcattctca    240

gcatatacag ataataggaa aggaaatatg ggtcacaaat atgtcagaat attagtgttc    300

ctaactaaat ttgatgattt ttcttgcgaa attaactcga agaaatccta tgttgttaca    360

ctctacgagc tatcagaaaa tcacaatatg aagtggaaaa tgtatatttt gaattgttta    420

cttcccgatg gaatcacttt caacgatgtg aattctgtaa aaatatctag aagttcttca    480

aaactttcag tccaaatccc gatcagatat agaattcaag atgagaaaat gatgactcca    540

gatgaatacg attataagtt gtcgatttgt gttcctgcac tttttggaaa cgtttattat    600

ccaaggagga ttattgaatt tgtggaacta aacagcttgc aagacatcga caaaatctac    660

atctactaca atcctttaga aatgacagat gaggccacag aaaggacttt gaagttttat    720

tccaataatg ggaaaatcaa tttaatagaa ttcattctcc cattttctac tcgagatgtt    780

tggtattatg ggcaattggc caccgttaca gattgtcttc tccgtaacac tggaataact    840

caatacacat ttttcaatga tttggatgaa tttttcgtgc cagtactgga caaccaaact    900

ctctctgaaa ctgtgtcagg attatttgaa aatcgaaaaa ttgcctctca gagaacggcc    960

ttgaaattta ttagtacaaa aatcaatcga tctcctgtaa ctctcaataa tattgtgtct   1020

tctaaaaatt ttgaaacgag attcacaaaa tgcgtcgtac ggccggaaat ggtttttgag   1080

cagggcattc accatacgag tagagtaata caagacgact acgaaacccc atcccatgat   1140

ggatcacttt tgcgtgtgta tcactacaga gaaccaagat attgctgcga aaacgagaat   1200

cttctaaaac aaagatacga taagaagctt caagaagttt ttgatgctgt agttcttata   1260

ttgcatgtca catttgatgt atggatatat cacctgaaaa acaccctcta a            1311
```

<210> SEQ ID NO 4
<211> LENGTH: 436
<212> TYPE: PRT
<213> ORGANISM: Caenorhabditis briggsae

<400> SEQUENCE: 4

```
Met Pro Arg Ile Thr Ala Ser Lys Ile Val Leu Leu Ser Val Leu Ser
1               5                   10                  15

Leu Leu Thr Val Phe Tyr Leu Asn Thr Phe Ser Ser Ile Lys Ile Glu
            20                  25                  30
```

-continued

```
Asn Asp Leu Asp Gly Thr Asp Tyr Asp Leu Asp Tyr Ile Glu Ser Asp
        35                  40                  45

Ile Lys Lys Thr Arg Arg Leu Leu Asn Glu Ile Pro Asp Pro Ser Gln
    50                  55                  60

Asn Arg Val Gln Phe Phe Lys Leu Asp Asp Asn Gly Tyr Ala Phe Ser
65                  70                  75                  80

Ala Tyr Thr Asp Asn Arg Lys Gly Asn Met Gly His Lys Tyr Val Arg
                85                  90                  95

Ile Leu Val Phe Leu Thr Lys Phe Asp Asp Phe Ser Cys Glu Ile Asn
            100                 105                 110

Ser Lys Lys Ser Tyr Val Val Thr Leu Tyr Glu Leu Ser Glu Asn His
        115                 120                 125

Asn Met Lys Trp Lys Met Tyr Ile Leu Asn Cys Leu Leu Pro Asp Gly
    130                 135                 140

Ile Thr Phe Asn Asp Val Asn Ser Val Lys Ile Ser Arg Ser Ser Ser
145                 150                 155                 160

Lys Leu Ser Val Gln Ile Pro Ile Arg Tyr Arg Ile Gln Asp Glu Lys
                165                 170                 175

Met Met Thr Pro Asp Glu Tyr Asp Tyr Lys Leu Ser Ile Cys Val Pro
            180                 185                 190

Ala Leu Phe Gly Asn Val Tyr Tyr Pro Arg Arg Ile Ile Glu Phe Val
        195                 200                 205

Glu Leu Asn Ser Leu Gln Asp Ile Asp Lys Ile Tyr Ile Tyr Tyr Asn
    210                 215                 220

Pro Leu Glu Met Thr Asp Glu Ala Thr Glu Arg Thr Leu Lys Phe Tyr
225                 230                 235                 240

Ser Asn Asn Gly Lys Ile Asn Leu Ile Glu Phe Ile Leu Pro Phe Ser
                245                 250                 255

Thr Arg Asp Val Trp Tyr Tyr Gly Gln Leu Ala Thr Val Thr Asp Cys
            260                 265                 270

Leu Leu Arg Asn Thr Gly Ile Thr Gln Tyr Thr Phe Phe Asn Asp Leu
        275                 280                 285

Asp Glu Phe Phe Val Pro Val Leu Asp Asn Gln Thr Leu Ser Glu Thr
    290                 295                 300

Val Ser Gly Leu Phe Glu Asn Arg Lys Ile Ala Ser Gln Arg Thr Ala
305                 310                 315                 320

Leu Lys Phe Ile Ser Thr Lys Ile Asn Arg Ser Pro Val Thr Leu Asn
                325                 330                 335

Asn Ile Val Ser Ser Lys Asn Phe Glu Thr Arg Phe Thr Lys Cys Val
            340                 345                 350

Val Arg Pro Glu Met Val Phe Glu Gln Gly Ile His His Thr Ser Arg
        355                 360                 365

Val Ile Gln Asp Asp Tyr Glu Thr Pro Ser His Asp Gly Ser Leu Leu
    370                 375                 380

Arg Val Tyr His Tyr Arg Glu Pro Arg Tyr Cys Cys Glu Asn Glu Asn
385                 390                 395                 400

Leu Leu Lys Gln Arg Tyr Asp Lys Lys Leu Gln Glu Val Phe Asp Ala
                405                 410                 415

Val Val Leu Ile Leu His Val Thr Phe Asp Val Trp Ile Tyr His Leu
            420                 425                 430
```

-continued

Lys Asn Thr Leu
435

<210> SEQ ID NO 5
<211> LENGTH: 1499
<212> TYPE: DNA
<213> ORGANISM: Nematostella vectensis

<400> SEQUENCE: 5

atgcgatgct atatttacaa attgaggttg tccgtttgtc tgtttgtagt gctcttcaca 60

gcactgcttt tcatcaccta tttaaaccac tcagagcttg aatcagcaga gaaaagtagc 120

ggaaaaagga agacgcgaca tcgtaaacga acacgttcac gcaaacaaca cgagagccat 180

tttcagaaag ctcgactaca agaaagagaa ctagtattaa gatctacagc gccaccaaca 240

ttacgaagag aagtacaagc gcatcgatta gggcagatcc gtggcaagaa cacggaccag 300

gggataactg gaaagttcac agagatcgct aaagacacgc atatttattc agcgttttac 360

gacgatgcca agtcaaatcc attcattcgt cttatcatcc tctcgggaaa acactaccag 420

cctggattat cttgccaatt ttgcgaacct ttgtccgcca gttgtagttt tgcggactct 480

aaagctgaat actacacgac caacgagaac catgggagag tatttggcgg gttcattgcg 540

agttgcctcg tgcctgatgg attcaatgca gtgccattgt ttgttgacat aacggccgat 600

gttaaggggg agaaaagcaa ggcacgggta cctgtggtgt ctaatgcaca tctctactac 660

cctattaaat acgcaatctg cgtcccaccc ctccgatcag agaaactaac agcgaaaaga 720

ctcatagagt ttgtcgagct aaccaaactt ttaggcgcta accattttac tttttatgac 780

ttcaaaacgg acccggaagt caataacgtt ttaagatatt accaggagac acaagtagca 840

aatgttctgc catggaatct accttcaaat ttggtatcca ggccgaacga tatttggtac 900

tttggtcagg ttttggctat tctagattgc ttgtatcgct acaagaacag ggcaaaattt 960

gtagccttca atgacgtaga tgagtttatc gttccgctaa ggaacagctc gatagtggaa 1020

atactaaacg cgtttcaccg gccataccac tgtggacatt gctttcagag cgtggtgttc 1080

agctcaaacg cgagatttcc caggcaaaaa agcgagttag tttctcagcg gttcttccac 1140

aggacccagg aaaccatccc tctcctctcg aaatgcattg tggatccttt gagagtgttc 1200

gagatgggga ttcaccacat aagcaaggct acaggtctgc ggtattccgt caactcagta 1260

cacgagagtg acgcggttat cttccattac aggacttgca ctacgtcatt tggtatacgt 1320

catcagtgca tgaacctagt gcatgatggg accatggcca aatatggaaa acgacttcag 1380

aaaatgttta gaaaggttgt aaatgattta aaacttttgg caccaacgta gctatttcgt 1440

aacacttcac actttcattg ttataacaga atacagaata aattaatgat tgttgtgcc 1499

<210> SEQ ID NO 6
<211> LENGTH: 476
<212> TYPE: PRT
<213> ORGANISM: Nematostella vectensis

<400> SEQUENCE: 6

Met Arg Cys Tyr Ile Tyr Lys Leu Arg Leu Ser Val Cys Leu Phe Val
1 5 10 15

Val Leu Phe Thr Ala Leu Leu Phe Ile Thr Tyr Leu Asn His Ser Glu
20 25 30

Leu Glu Ser Ala Glu Lys Ser Ser Gly Lys Arg Lys Thr Arg His Arg

-continued

```
        35                  40                  45

Lys Arg Thr Arg Ser Arg Lys Gln His Glu Ser His Phe Gln Lys Ala
    50                  55                  60

Arg Leu Gln Glu Arg Glu Leu Val Leu Arg Ser Thr Ala Pro Pro Thr
65                  70                  75                  80

Leu Arg Arg Glu Val Gln Ala His Arg Leu Gly Gln Ile Arg Gly Lys
                85                  90                  95

Asn Thr Asp Gln Gly Ile Thr Gly Lys Phe Thr Glu Ile Ala Lys Asp
            100                 105                 110

Thr His Ile Tyr Ser Ala Phe Tyr Asp Asp Ala Lys Ser Asn Pro Phe
        115                 120                 125

Ile Arg Leu Ile Ile Leu Ser Gly Lys His Tyr Gln Pro Gly Leu Ser
    130                 135                 140

Cys Gln Phe Cys Glu Pro Leu Ser Ala Ser Cys Ser Phe Ala Asp Ser
145                 150                 155                 160

Lys Ala Glu Tyr Tyr Thr Thr Asn Glu Asn His Gly Arg Val Phe Gly
                165                 170                 175

Gly Phe Ile Ala Ser Cys Leu Val Pro Asp Gly Phe Asn Ala Val Pro
            180                 185                 190

Leu Phe Val Asp Ile Thr Ala Asp Val Lys Gly Glu Lys Ser Lys Ala
        195                 200                 205

Arg Val Pro Val Val Ser Asn Ala His Leu Tyr Tyr Pro Ile Lys Tyr
    210                 215                 220

Ala Ile Cys Val Pro Pro Leu Arg Ser Glu Lys Leu Thr Ala Lys Arg
225                 230                 235                 240

Leu Ile Glu Phe Val Glu Leu Thr Lys Leu Leu Gly Ala Asn His Phe
                245                 250                 255

Thr Phe Tyr Asp Phe Lys Thr Asp Pro Glu Val Asn Asn Val Leu Arg
            260                 265                 270

Tyr Tyr Gln Glu Thr Gln Val Ala Asn Val Leu Pro Trp Asn Leu Pro
        275                 280                 285

Ser Asn Leu Val Ser Arg Pro Asn Asp Ile Trp Tyr Phe Gly Gln Val
    290                 295                 300

Leu Ala Ile Leu Asp Cys Leu Tyr Arg Tyr Lys Asn Arg Ala Lys Phe
305                 310                 315                 320

Val Ala Phe Asn Asp Val Asp Glu Phe Ile Val Pro Leu Arg Asn Ser
                325                 330                 335

Ser Ile Val Glu Ile Leu Asn Ala Phe His Arg Pro Tyr His Cys Gly
            340                 345                 350

His Cys Phe Gln Ser Val Val Phe Ser Ser Asn Ala Arg Phe Pro Arg
        355                 360                 365

Gln Lys Ser Glu Leu Val Ser Gln Arg Phe Phe His Arg Thr Gln Glu
    370                 375                 380

Thr Ile Pro Leu Leu Ser Lys Cys Ile Val Asp Pro Leu Arg Val Phe
385                 390                 395                 400

Glu Met Gly Ile His His Ile Ser Lys Ala Thr Gly Leu Arg Tyr Ser
                405                 410                 415

Val Asn Ser Val His Glu Ser Asp Ala Val Ile Phe His Tyr Arg Thr
            420                 425                 430

Cys Thr Thr Ser Phe Gly Ile Arg His Gln Cys Met Asn Leu Val His
        435                 440                 445
```

-continued

```
Asp Gly Thr Met Ala Lys Tyr Gly Lys Arg Leu Gln Lys Met Phe Arg
    450                 455                 460

Lys Val Val Asn Asp Leu Lys Leu Leu Ala Pro Thr
465                 470                 475


<210> SEQ ID NO 7
<211> LENGTH: 1260
<212> TYPE: DNA
<213> ORGANISM: Taeniopygia guttata

<400> SEQUENCE: 7

atgactgtaa ctttaatgct tgtggtttct tatctgagat tacagagact ttctcatcag     60

ccaaaagtaa ttcaagaaag tagaagatgt agagggaaaa ttgcccttag cacaataaca    120

gcattggaag gtaacaaaac tgatattata tccccatact ttgatgacag agaaaacaaa    180

atcactcgtc tgattgggat tgttcaccat aaagatgtaa aacaactgtt ctgctggttc    240

tgctgtcaag ccaatggaaa gatatatgta tcaaaagcag aaatagatgt tcactcggat    300

agatttggat tcccttatgg tgcagcagat ataatttgtt tggaacctga aaactgtgat    360

ccaacacatg tatcaattca tcagtctcca tatggaaata ttgaccagct gccgaggttt    420

gaaattaaaa atcgcaggcc tgagaccttt tctgttgact tcaccgtgtg catttctgcc    480

atgtttggaa actacaacaa tgtcttgcag tttgtacaga gtatggaaat gtataagatt    540

cttggagtac agaaagtggt gatctataag aacaactgca gccatctgat ggagaaagtc    600

ttgaaatttt atatagaaga aggaactgtt gaggtaattc cctggccaat agactcacac    660

ctcagggttt cttctaaatg gcgcttcatg gaagacggga cacacattgg ctactatgga    720

caaatcacag ctctaaatga ctgtatatac cgcaacatgg aaaggaccaa gtttgtggtc    780

cttaatgacg ctgatgaaat aattcttccc cttaaacacc cagactggaa aacaatgatg    840

aacagtcttc aggagcaaaa cccagggact agtgttttcc tttttgagaa ccatatcttc    900

ccagaaactg tattttctcc catgttcaac atttcatctt ggaatactgt gccaggtgtt    960

aacatattgc agcatgtgta cagagagcct gacaggaaac atgtaatcaa tcccaggaaa   1020

atgatagttg atccacgaaa ggtgattcag acttcagtcc attctgtcct acgtgcttat   1080

gggaagagcg tgaatgttcc catggaagtt gccctcattt atcactgtcg gaaggccctt   1140

caaggaaacc ttcccagaga atctctcatc agggatacaa cactgtggag atataactca   1200

tcattaatca tgaatgttaa caaggttcta tctcaaacca tgctgcaaac tcaaaattga   1260


<210> SEQ ID NO 8
<211> LENGTH: 419
<212> TYPE: PRT
<213> ORGANISM: Taeniopygia guttata

<400> SEQUENCE: 8

Met Thr Val Thr Leu Met Leu Val Val Ser Tyr Leu Arg Leu Gln Arg
1               5                   10                  15

Leu Ser His Gln Pro Lys Val Ile Gln Glu Ser Arg Arg Cys Arg Gly
            20                  25                  30

Lys Ile Ala Leu Ser Thr Ile Thr Ala Leu Glu Gly Asn Lys Thr Asp
        35                  40                  45

Ile Ile Ser Pro Tyr Phe Asp Asp Arg Glu Asn Lys Ile Thr Arg Leu
    50                  55                  60
```

-continued

```
Ile Gly Ile Val His His Lys Asp Val Lys Gln Leu Phe Cys Trp Phe
65                  70                  75                  80

Cys Cys Gln Ala Asn Gly Lys Ile Tyr Val Ser Lys Ala Glu Ile Asp
                85                  90                  95

Val His Ser Asp Arg Phe Gly Phe Pro Tyr Gly Ala Ala Asp Ile Ile
            100                 105                 110

Cys Leu Glu Pro Glu Asn Cys Asp Pro Thr His Val Ser Ile His Gln
        115                 120                 125

Ser Pro Tyr Gly Asn Ile Asp Gln Leu Pro Arg Phe Glu Ile Lys Asn
    130                 135                 140

Arg Arg Pro Glu Thr Phe Ser Val Asp Phe Thr Val Cys Ile Ser Ala
145                 150                 155                 160

Met Phe Gly Asn Tyr Asn Asn Val Leu Gln Phe Val Gln Ser Met Glu
                165                 170                 175

Met Tyr Lys Ile Leu Gly Val Gln Lys Val Val Ile Tyr Lys Asn Asn
            180                 185                 190

Cys Ser His Leu Met Glu Lys Val Leu Lys Phe Tyr Ile Glu Glu Gly
        195                 200                 205

Thr Val Glu Val Ile Pro Trp Pro Ile Asp Ser His Leu Arg Val Ser
    210                 215                 220

Ser Lys Trp Arg Phe Met Glu Asp Gly Thr His Ile Gly Tyr Tyr Gly
225                 230                 235                 240

Gln Ile Thr Ala Leu Asn Asp Cys Ile Tyr Arg Asn Met Glu Arg Thr
                245                 250                 255

Lys Phe Val Val Leu Asn Asp Ala Asp Glu Ile Ile Leu Pro Leu Lys
            260                 265                 270

His Pro Asp Trp Lys Thr Met Met Asn Ser Leu Gln Glu Gln Asn Pro
        275                 280                 285

Gly Thr Ser Val Phe Leu Phe Glu Asn His Ile Phe Pro Glu Thr Val
    290                 295                 300

Phe Ser Pro Met Phe Asn Ile Ser Ser Trp Asn Thr Val Pro Gly Val
305                 310                 315                 320

Asn Ile Leu Gln His Val Tyr Arg Glu Pro Asp Arg Lys His Val Ile
                325                 330                 335

Asn Pro Arg Lys Met Ile Val Asp Pro Arg Lys Val Ile Gln Thr Ser
            340                 345                 350

Val His Ser Val Leu Arg Ala Tyr Gly Lys Ser Val Asn Val Pro Met
        355                 360                 365

Glu Val Ala Leu Ile Tyr His Cys Arg Lys Ala Leu Gln Gly Asn Leu
    370                 375                 380

Pro Arg Glu Ser Leu Ile Arg Asp Thr Thr Leu Trp Arg Tyr Asn Ser
385                 390                 395                 400

Ser Leu Ile Met Asn Val Asn Lys Val Leu Ser Gln Thr Met Leu Gln
                405                 410                 415

Thr Gln Asn


<210> SEQ ID NO 9
<211> LENGTH: 1647
<212> TYPE: DNA
<213> ORGANISM: Cryptosporidium parvum

<400> SEQUENCE: 9
```

-continued

```
atgcaaagta aagtcatttt taggatcttg gtattgatca tttcggtgat tggatcctta     60

tactcaataa ttcaattaat gctaaaggag ctatcaagta acaaaaatat tcaagaggtt    120

agtcattcaa ggaggctaat aagtgaacct tacagtgaaa gtattaatga acaaaatgat    180

caagattgga aagaactaaa gctaataatt ccaaatcatt ctcaaattaa ccagcaggaa    240

aaaaatggta atttgattga gtttaaagtt tatatatact cagcatatta tgattggaga    300

atagatagga tacgaataaa ttcacttatc ccatcgaatt tttatgatcg aatagaaatg    360

gaatgtgcaa taatcttgga caaaaatatt tacacaggaa ctattaaaaa agtgattcat    420

aaggagcacc ataataaaga atatgtatca tcgactttac tctgcgaaat tgcaaaaaat    480

gaaattaaat ttgaggatat ttcaaggaaa gttttgataa caattttgga aaatggaaac    540

agcacaaata aatcagaaat atggataact ctaaaaaaaa ttccaaaaaa tagctctaat    600

aatcatgagc tgactgtttg tgtgagacct tggtggggag agccaataaa gaatggaaac    660

ttgggaaata aacaaaaatt taacaattca gggttaatgc ttgaatttat taattcatat    720

ttattcttag gagcaaataa attttattta tatcaaaatt acttggacat tgacgaagat    780

gtaagaaata taataaatta ttattctaat atcaaaaatg ttttggaaat tattccatac    840

tcattaccaa taattccatt taaacaagtt tgggatttcg cacaaacaac aatgatacag    900

gactgcctac taagaaatat tggaaaaaca aaatacttgt tattcgtaga taccgatgaa    960

tttgtatttc caaacttgaa aaattataac ttaatggatt ttttaaattt attagaagcc   1020

aacaatcctt attataaaaa caaagtcggg gcaatgtgga ttccaatgta ttttcatttt   1080

ttagagtggg aatctgataa aaataatttg aagaaatatt caacaattga gaaaaaaatt   1140

aagaaaaaga tggcaaatat tgagtttgtt ctatatcgta aaacatgtag aatgttaagt   1200

tctggaacaa aaaaaagtga caagacgaga agaaaagtta ttattagacc tgaaagagtt   1260

ttgtatatgg gtatacatga aacagaagag atgctaagca aaaaatttca tttcattaga   1320

gctcctgtaa ttaatgtggg tggaggaaac gaactaagta tatatttaca tcattataga   1380

aaagcaaaag gtattgtaaa caatgatccc aaacaaagag aacttgtgaa tatgtattta   1440

gaaaatgttt gttcagataa gctgttagat tcaggggggag attccattca agatggagta   1500

attgtcgaca atactgtttg ggagatattt ggaacacact tataccagat aatttttgag   1560

catattaaag aaatccaaga tatgtacaca aataaggaaa taattaatgg aaataaaaat   1620

ttaagtgttg aagaattaca taattaa                                       1647
```

```
<210> SEQ ID NO 10
<211> LENGTH: 548
<212> TYPE: PRT
<213> ORGANISM: Cryptosporidium parvum

<400> SEQUENCE: 10

Met Gln Ser Lys Val Ile Phe Arg Ile Leu Val Leu Ile Ile Ser Val
1               5                  10                  15

Ile Gly Ser Leu Tyr Ser Ile Ile Gln Leu Met Leu Lys Glu Leu Ser
            20                  25                  30

Ser Asn Lys Asn Ile Gln Glu Val Ser His Ser Arg Arg Leu Ile Ser
        35                  40                  45

Glu Pro Tyr Ser Glu Ser Ile Asn Glu Gln Asn Asp Gln Asp Trp Lys
    50                  55                  60
```

-continued

```
Glu Leu Lys Leu Ile Ile Pro Asn His Ser Gln Ile Asn Gln Gln Glu
65                  70                  75                  80

Lys Asn Gly Asn Leu Ile Glu Phe Lys Val Tyr Ile Tyr Ser Ala Tyr
                85                  90                  95

Tyr Asp Trp Arg Ile Asp Arg Ile Arg Ile Asn Ser Leu Ile Pro Ser
            100                 105                 110

Asn Phe Tyr Asp Arg Ile Glu Met Glu Cys Ala Ile Ile Leu Asp Lys
        115                 120                 125

Asn Ile Tyr Thr Gly Thr Ile Lys Lys Val Ile His Lys Glu His His
    130                 135                 140

Asn Lys Glu Tyr Val Ser Ser Thr Leu Leu Cys Glu Ile Ala Lys Asn
145                 150                 155                 160

Glu Ile Lys Phe Glu Asp Ile Ser Arg Lys Val Leu Ile Thr Ile Leu
                165                 170                 175

Glu Asn Gly Asn Ser Thr Asn Lys Ser Glu Ile Trp Ile Thr Leu Lys
            180                 185                 190

Lys Ile Pro Lys Asn Ser Ser Asn Asn His Glu Leu Thr Val Cys Val
        195                 200                 205

Arg Pro Trp Trp Gly Glu Pro Ile Lys Asn Gly Asn Leu Gly Asn Lys
    210                 215                 220

Gln Lys Phe Asn Asn Ser Gly Leu Met Leu Glu Phe Ile Asn Ser Tyr
225                 230                 235                 240

Leu Phe Leu Gly Ala Asn Lys Phe Tyr Leu Tyr Gln Asn Tyr Leu Asp
                245                 250                 255

Ile Asp Glu Asp Val Arg Asn Ile Ile Asn Tyr Tyr Ser Asn Ile Lys
            260                 265                 270

Asn Val Leu Glu Ile Ile Pro Tyr Ser Leu Pro Ile Ile Pro Phe Lys
        275                 280                 285

Gln Val Trp Asp Phe Ala Gln Thr Thr Met Ile Gln Asp Cys Leu Leu
    290                 295                 300

Arg Asn Ile Gly Lys Thr Lys Tyr Leu Leu Phe Val Asp Thr Asp Glu
305                 310                 315                 320

Phe Val Phe Pro Asn Leu Lys Asn Tyr Asn Leu Met Asp Phe Leu Asn
                325                 330                 335

Leu Leu Glu Ala Asn Asn Pro Tyr Tyr Lys Asn Lys Val Gly Ala Met
            340                 345                 350

Trp Ile Pro Met Tyr Phe His Phe Leu Glu Trp Glu Ser Asp Lys Asn
        355                 360                 365

Asn Leu Lys Lys Tyr Ser Thr Ile Glu Lys Lys Ile Lys Lys Lys Met
    370                 375                 380

Ala Asn Ile Glu Phe Val Leu Tyr Arg Lys Thr Cys Arg Met Leu Ser
385                 390                 395                 400

Ser Gly Thr Lys Lys Ser Asp Lys Thr Arg Arg Lys Val Ile Ile Arg
                405                 410                 415

Pro Glu Arg Val Leu Tyr Met Gly Ile His Glu Thr Glu Glu Met Leu
            420                 425                 430
```

-continued

```
Ser Lys Lys Phe His Phe Ile Arg Ala Pro Val Ile Asn Val Gly Gly
        435                 440                 445

Gly Asn Glu Leu Ser Ile Tyr Leu His His Tyr Arg Lys Ala Lys Gly
    450                 455                 460

Ile Val Asn Asn Asp Pro Lys Gln Arg Glu Leu Val Asn Met Tyr Leu
465                 470                 475                 480

Glu Asn Val Cys Ser Asp Lys Leu Leu Asp Ser Gly Gly Asp Ser Ile
                485                 490                 495

Gln Asp Gly Val Ile Val Asp Asn Thr Val Trp Glu Ile Phe Gly Thr
            500                 505                 510

His Leu Tyr Gln Ile Ile Phe Glu His Ile Lys Glu Ile Gln Asp Met
        515                 520                 525

Tyr Thr Asn Lys Glu Ile Ile Asn Gly Asn Lys Asn Leu Ser Val Glu
    530                 535                 540

Glu Leu His Asn
545


<210> SEQ ID NO 11
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 11

tttgtcgaca cttctgaatg cctcg                                          25


<210> SEQ ID NO 12
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 12

ttttctagac tacaagtcta aaagaccaac                                     30


<210> SEQ ID NO 13
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer

<400> SEQUENCE: 13

tttgtcgacc ctcgaatcac cgcc                                           24
```

1. An isolated and purified nucleic acid, wherein said nucleic acid is selected from the group consisting of:

(i) a nucleic acid comprising at least a nucleic acid sequence selected from the group consisting of nucleic acid sequences listed in SEQ ID NOs: 1, 3, 5, 7 and 9;

(ii) a nucleic acid having a sequence of at least 60 or 70% identity, preferably at least 80 or 90% identity, more preferred at least 95% identity, most preferred at least 98% identity with the nucleic acid sequence listed in SEQ ID NO 1;

(iii) a nucleic acid that hybridizes to a nucleic acid of (i) or (ii);

(iv) a nucleic acid, wherein said nucleic acid is derivable by substitution, addition and/or deletion of one of the nucleic acids of (i), (ii) oder (iii);

(v) a fragment of any of the nucleic acids of (i) to (iv), that hybridizes to a nucleic acid of (i).

2. The nucleic acid according to claim **1**, wherein said nucleic acid is a DNA, RNA or PNA, preferably DNA or PNA, more preferably DNA.

3. The nucleic acid according to claim **1**, wherein said nucleic acid encodes a protein having galactosyltransferase activity, preferably β-1,4-galactosyltransferase activity, pref-

erably with L-fucoside-, more preferably with α-L-fucoside-, more preferably with Fuc-α-1,6-GlcNAc— and most preferably with GnGnF^6— containing poly/oligosaccharides or glycoconjugates as acceptor substrates.

4. An isolated and purified polypeptide selected from the group consisting of:

(a) polypeptides having an amino acid sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8 and 10, preferably SEQ ID NO: 2,

(b) polypeptides encoded by a nucleic acid of claim **1**,

(c) polypeptides having an amino acid sequence identity of at least 25, 30 or 40%, preferably at least 50 or 60%, more preferably at least 70 or 80%, most preferably at least 90 or 95% with the polypeptides of (a) and/or (b),

(d) a fragment and/or functional derivative of (a), (b) or (c).

5. The polypeptide according to claim **4**, wherein said polypeptide has galactosyltransferase activity, preferably β-1,4-galactosyltransferase activity, preferably with L-fucoside-, more preferably with α-L-fucoside-, more preferably with Fuc-α-1,6-GlcNAc— and most preferably with GnGnF^6-containing poly/oligosaccharides or glycoconjugates as acceptor substrates.

6. A recombinant vector comprising a nucleic acid of claim **1**, preferably a viral or episomal vector, preferably a baculovirus vector.

7. A host cell comprising a nucleic acid claim **1**, preferably selected from the group consisting of yeast cells, preferably *Saccharomyces cerevisiae, Pichia pastoris* cells, *E. coli* cells, plant cells, preferably *Nicotiana tabacum* or *Physcomitrella patens* cells, NIH-3T3 mammalian cells and insect cells, more preferably sf9 insect cells.

8. An antibody that specifically binds a polypeptide of claim **4**.

9. An antibody according to claim **8**, wherein said antibody is monoclonal antibody.

10. Hybridoma cell line, expressing a monoclonal antibody that specifically binds a polypeptide according to claim **5**.

11. Use of a polypeptide of claim **4**, a cell extract comprising a polypeptide of claim **4**, preferably a *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* or *Cryptosporidium parvum* extract, and/or a host cell comprising a nucleic acid, selected from the group consisting of yeast cells, *E. Coli*, plant cells, NIH-3T3 mammalian cells and insect cells, for producing galactosyl-containing oligo/polysaccharides and/or glycoconjugates, preferably galactosyl-fucoside-containing oligo/polysaccharides and/or glycoconjugates, more preferably D-galactopyranosyl-β-1,4-L-fucopyranosyl-α-1,6-GlcNAc-containing oligo/polysaccharides and/or glycoconjugates, most preferably GnGnF^{6}Gal- and/or MMF^{6}Gal-containing oligosaccharides and glycoconjugates.

12. Method for producing galactosyl-fucosyl derivatives, comprising the following steps:

(i) providing at least one polypeptide of the invention,

(ii) providing at least one fucosylated acceptor substrate,

(iii) incubating (i) and (ii) in the presence of at least one suitable divalent metal cation cofactor, preferably selected from manganese (II), cobalt (II) and/or iron (II) ions, more preferably manganese (II), and at least one activated sugar substrate, preferably uridine diphosphate (UDP)-galactose under conditions suitable for enzymatic activity of the polypeptide of the invention,

(iv) optionally isolating the galactosyl-fucose derivatives.

13. Use of at least one polypeptide of claim **4**, a host cell comprising a nucleic acid, selected from the group consisting of yeast cells, *E. Coli*, plant cells, NIH-3T3 mammalian cells and insect cells,

and/or cell extracts of *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* and/or *Cryptosporidium parvum* for covalently binding galactosyl compounds to core-fucosylated alpha-fetoprotein (AFP), preferably for detecting and/or quantifying hepatocellular carcinoma (HCC).

14. A method of diagnosis, comprising the following steps:

(i) providing blood or a fraction thereof, that comprises AFP, preferably serum,

(ii) incubating said blood or said fraction thereof with (a) a polypeptide of claim **4**, a host cell comprising a nucleic acid, selected from the group consisting of yeast cells, *E. Coli*, plant cells, NIH-3T3 mammalian cells and insect cells, and/or cell extracts of *Caenorhabditis elegans, Caenorhabditis briggsae, Nematostella vectensis, Taeniopygia guttata* and/or *Cryptosporidium parvum* and (b) an activated galactosyl derivative, preferably a labelled galactosyl derivative, preferably labelled UDP-galactose, under conditions that allow for the galactosyl-transfer of activated galactose to core-fucosylated AFP (AFP-L3),

(iii) and detecting the galactose-labelled and hence core-fucosylated AFP (AFP-L3).

15. Use of antibodies according to claim **8** for identifying and/or quantifying nematodes or pathogens, preferably *Caenorhabditis elegans, Caenorhabditis briggsae* and/or *Cryptosporidium parvum* in a sample of interest, for example a human or mammalian sample, preferably in a cell fraction or extract sample.

* * * * *

专利名称(译)	N-聚糖核心β-半乳糖基转移酶及其用途		
公开(公告)号	US20120064541A1	公开(公告)日	2012-03-15
申请号	US13/322505	申请日	2010-05-28
[标]申请(专利权)人(译)	UNIV FUR BODENKULTUR WIEN 苏黎世大学		
申请(专利权)人(译)	苏黎世联邦理工学院 Universität大学FUR BODENKULTUR WIEN Universität大学ZURICH		
当前申请(专利权)人(译)	苏黎世联邦理工学院 Universität大学ZURICH Universität大学FUR BODENKULTUR WIEN		
[标]发明人	KUNZLER MARKUS AEBI MARKUS WILSON LAIN TITZ ALEXANDER WALTER HENGARTNER MICHAEL BUTSCHI ALEX		
发明人	KUNZLER, MARKUS AEBI, MARKUS WILSON, LAIN TITZ, ALEXANDER WALTER HENGARTNER, MICHAEL BUTSCHI, ALEX		
IPC分类号	G01N33/53 C07K14/00 C07H21/02 C12N9/96 C12P19/00 C12N1/19 C12N1/21 C12N5/10 C07K16/40 C12N5/12 C07H21/04 C12N15/63		
CPC分类号	C12N9/1051		
优先权	2009007139 2009-05-28 EP		
外部链接	Espacenet USPTO		

摘要(译)

本发明涉及新的半乳糖基转移酶，编码它们的核酸，以及重组载体，宿主细胞，抗体，用途和与其相关的方法。