Primary structure and expression analysis of human UDP-*N*-acetyl-glucosamine-2-epimerase/*N*-acetylmannosamine kinase, the bifunctional enzyme in neuraminic acid biosynthesis[1]

Lothar Lucka[2,*], Matthias Krause[2], Kerstin Danker, Werner Reutter, Rüdiger Horstkorte

Institut für Molekularbiologie und Biochemie, Universitätsklinikum Benjamin Franklin, Freie Universität Berlin, Arnimallee 22, D-14195 Berlin-Dahlem, Germany

Received 7 May 1999; received in revised form 8 June 1999

Abstract ***N*-Acetylneuraminic acid is a main constituent of glycoproteins and gangliosides. In many membrane-bound receptors it is the target for external stimuli. The key enzyme for its biosynthesis is the bifunctional enzyme UDP-*N*-acetyl-glucosamine-2-epimerase/*N*-acetylmannosamine kinase, catalysing the first two steps of the biosynthesis in the cytosol. The rat enzyme was previously isolated and characterised. In this report we present the corresponding human cDNA sequence, compare it with the primary structure of the rodent enzyme, and report the analysis of its expression in different human tissues and cell lines.**

Key words: Neuraminic acid biosynthesis; UDP-*N*-acetylglucosamine-2-epimerase/*N*-acetylmannosamine kinase; Tissue expression

1. Introduction

In eukaryotic organisms *N*-acetylneuraminic acid is a terminal component of glycan structures bound to proteins and gangliosides [1,2]. It is considered a molecular determinant of specific biological functions [3] such as cellular adhesion [4–6], formation or masking of recognition determinants [7–10] and stabilisation of glycoprotein structures [11,12]. Recently it was shown that the two enzymes initiating the biosynthesis of Neu5Ac, UDP-*N*-acetylglucosamine-2-epimerase (UDP-GlcNAc-2-epimerase) and *N*-acetylmannosamine kinase (ManNAc kinase) (EC 5.1.3.14/EC 2.7.1.60) are parts of one bifunctional enzyme. This 79 kDa protein assembles to a dimer or hexamer. The hexameric structure represents both activities, the formation of *N*-acetylmannosamine and UDP from UDP-*N*-acetylglucosamine and the subsequent phosphorylation of *N*-acetylmannosamine (ManNAc), whereas the dimer catalyses only the phosphorylation of *N*-acetylmannosamine [13]. In rodents, the highest levels of expression and activity are found in liver, salivary gland and intestinal mucosa. In other organs the enzyme is expressed in smaller amounts paralleled by lower total enzyme activities [14] The UDP-GlcNAc-2-epimerase/ManNAc kinase is expressed at early stages during development [15], indicating its importance during organogenesis. The biological significance is further illustrated by the observation that in hepatoma activity of the UDP-GlcNAc-2-epimerase is dramatically reduced [16]. Recently, the 2-epimerase was found to be a major determinant of cell surface sialylation in haematopoietic cell lines [17]. The clinical relevance of the enzyme was demonstrated by the observation that the basic defect in sialuria has been identified as the loss of feedback control of UDP-GlcNAc-2-epimerase by CMP-*N*-acetylneuraminic acid with resultant overproduction of sialic acid [18,19]. Therefore characterisation of the human UDP-GlcNAc-2-epimerase/ManAc kinase could give new insights into this or possibly other human diseases. In this study we isolated, cloned and sequenced the cDNA from human liver. The deduced primary structure revealed high sequence similarity to the rodent enzymes, indicating high evolutionary conservation. Differential tissue and cell line expression was demonstrated by mRNA analysis.

2. Materials and methods

2.1. Materials

Unless otherwise stated, chemicals and reagents were obtained from Sigma-Aldrich, Deisenhofen, Germany. Enzymes were obtained from Life Technologies, Karlsruhe, Germany. *Escherichia coli* HB 101 and XL-1 blue were obtained from Stratagene, La Jolla, CA, USA and *E. coli* INVαF′ from Invitrogen, Leek, The Netherlands.

2.2. RT-PCR, cDNA library screening and 5′ RACE

DNA and RNA were prepared and analysed according to standard procedures [20]. Isolation of cDNA encoding the human UDP-GlcNAc-2-epimerase was done using a combination of RT-PCR, cDNA library screening and 5′ RACE.

Poly(A)$^+$ RNA from human liver (Clontech, Heidelberg, Germany) was used for RT-PCR. It was transcribed to cDNA using oligo(dt) primer and superscript reverse transcriptase (Life Technology, Karlsruhe, Germany) at 37°C for 60 min according to the manufacturer's instructions. The PCR reaction was performed under standard conditions [21]. For that UDP-GlcNAc-2-epimerase-specific primers according to the published rat sequence [14] were annealed to the human liver cDNA and incubated with Taq polymerase (Perkin Elmer, Branchburg, USA) over 25 cycles (30 s 94°C; 60 s 50°C; 90 s 72°C). Amplification products were analysed by horizontal agarose gel electrophoresis, isolated from the gel and cloned into the pCR2.1 vector (Invitrogen, Leek, The Netherlands) and sequenced.

A human liver cDNA library (oligo(dt)- and random-primed cDNA ligated into Uni (λZAP XR, Stratagene, La Jolla, CA, USA) was screened by hybridisation with a rat coding region probe according to standard procedures. Labelling of the rat DNA fragment was done by [α-^{32}P]dATP incorporation using a random priming labelling kit (Amersham-Buchler, Braunschweig, Germany). A total of 5×10^5 plaques were plated, lifted to nitrocellulose filters and hybridised to the labelled probe under stringent conditions overnight at 42°C. The phagemid pBluescript II SK$^-$ was excised from positive phages by in vivo excision using the helper phage R408 (Stratagene, La Jolla, CA, USA) according to the manufacturer's protocol. cDNA inserts were analysed in horizontal agarose gel electrophoresis and sequenced as described in Section 2.3.

*Corresponding author. Fax: (49) (30) 8 38 21 41.

[1] The novel nucleic acid sequence data reported here have been submitted to the EMBL DataBase, European Bioinformatics Institute, accession number AJ238764.

[2] These authors contributed equally to the results of this work.

0014-5793/99/$20.00
PII: S0014-5793(99)00837-6

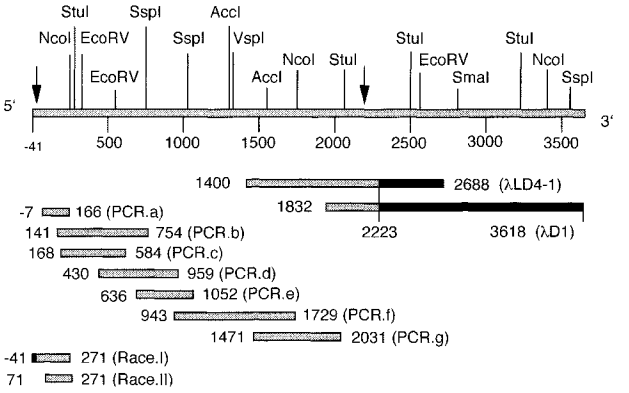

Fig. 1. Schematic presentation of cDNA fragments obtained by PCR and library screening. In the upper part of the figure the whole cDNA encoding the human UDP-GlcNAc-2-epimerase/ManNAc kinase is shown with indicated restriction enzyme cleavage sites. The start and stop codons are marked by arrows. In the lower part of the figure the obtained lambda cDNA clones (λLD4-1, λD1), the PCR fragments using primer specific for the corresponding rat sequence (PCR.a-g) and the PCR fragments from 5′ RACE priming (Race.I, II) are shown. The position relative to +1 of the coding region is indicated. All fragments were sequenced as described in Section 2. Parts of the bars indicated in grey show high similarity to the rat sequence.

Elongation of the 5′ cDNA terminus was performed using a 5′RACE kit (Life Technologies, Karlsruhe, Germany) according to the manufacturer's manual. As starting material for RT-PCR, 5 μg human liver poly(A)$^+$ RNA (Clontech, Heidelberg, Germany) were primed with human UDP-GlcNAc-2-epimerase-specific primers. Obtained cDNA was tailed with poly-dCTP using the Terminal deoxynucleotide Transferase (TdT). Subsequent amplification was done with an anchor primer containing an oligo(dG) sequence and a second internal UDP-GlcNAc-2-epimerase-specific primer. PCR products of the polymerase chain reactions were analysed in horizontal agarose gel electrophoresis, cloned into the pCR2.1 vector (Invitrogen, Leek, The Netherlands) and sequenced as described in Section 2.3.

2.3. Cloning and sequencing

Using the TA-cloning kit (Invitrogen, San Diego, CA, USA), approximately 50 ng of each reaction product were ligated directly into the pCR II vector (Invitrogen, San Diego, CA, USA). Ligated DNA was transformed into *E. coli* HB101 or INVαF′. Positive clones were identified by restriction analysis and sequencing. Sequencing according to Sanger et al. [22] was performed either with double-stranded plasmid DNA using [α-^{35}S]dATP (Hartmann Analytic, Braunschweig,

```
                                   AA CTC TAT TTT AAG AAC CTC TCA AAA CGA AAC AAG CAA ATC   -1
Met Glu Lys Asn Gly Asn Asn Arg Lys Leu Arg Val Cys Val Ala Thr Cys Asn Arg Ala   20
ATG GAG AAG AAT GGA AAT AAC CGA AAG CTG CGG GTT TGT GTT GCT ACT TGT AAC CGT GCA   60
Asp Tyr Ser Lys Leu Ala Pro Ile Met Phe Gly Ile Lys Thr Glu Pro Glu Phe Phe Glu   40
GAT TAT TCT AAA CTT GCC CCG ATC ATG TTT GGC ATT AAA ACC GAA CCT GAG TTC TTT GAA  120
Leu Asp Val Val Val Leu Gly Ser His Leu Ile Asp Asp Tyr Gly Asn Thr Tyr Arg Met   60
CTT GAT GTT GTG GTA CTT GGC TCT CAC CTG ATA GAT GAC TAT GGA AAT ACA TAT CGA ATG  180
Ile Glu Gln Asp Asp Phe Asp Ile Asn Thr Arg Leu His Thr Ile Val Arg Gly Glu Asp   80
ATT GAA CAA GAT GAC TTT GAC ATT AAC ACC AGG CTA CAC ACA ATT GTG AGG GGA GAA GAT  240
Glu Ala Ala Met Val Glu Ser Val Gly Leu Ala Leu Val Lys Leu Pro Asp Val Leu Asn  100
GAG GCA GCC ATG GTG GAG TCA GTA GGC CTG GCC CTA GTG AAG CTG CCA GAT GTC CTT AAT  300
Arg Leu Lys Pro Asp Ile Met Ile Val His Gly Asp Arg Phe Asp Ala Leu Ala Leu Ala  120
CGC CTG AAG CCT GAT ATC ATG ATT GTT CAT GGA GAC AGG TTT GAT GCC CTG GCT CTG GCC  360
Thr Ser Ala Ala Leu Met Asn Ile Arg Ile Leu His Ile Glu Gly Gly Glu Val Ser Gly  140
ACA TCT GCT GCC TTG ATG AAC ATC CGA ATC CTT CAC ATT GAA GGT GGG GAA GTC AGT GGG  420
Thr Ile Asp Asp Ser Ile Arg His Ala Ile Thr Lys Leu Ala His Tyr His Val Cys Cys  160
ACC ATT GAT GAC TCT ATC AGA CAT GCC ATA ACA AAA CTG GCT CAT TAT CAT GTG TGC TGC  480
Thr Arg Ser Ala Glu Gln His Leu Ile Ser Met Cys Glu Asp His Asp Arg Ile Leu Leu  180
ACC CGC AGT GCA GAG CAG CAC CTG ATA TCC ATG TGT GAG GAC CAT GAT CGC ATC CTT TTG  540
Ala Gly Cys Pro Ser Tyr Asp Lys Leu Leu Ser Ala Lys Asn Lys Asp Tyr Met Ser Ile  200
GCA GGC TGC CCT TCC TAT GAC AAA CTT CTC TCA GCC AAG AAC AAA GAC TAC ATG AGC ATC  600
Ile Arg Met Trp Leu Gly Asp Asp Val Lys Ser Lys Asp Tyr Ile Val Ala Leu Gln His  220
ATT CGC ATG TGG CTA GGT GAT GAT GTA AAA TCT AAA GAT TAC ATT GTT GCA CTA CAG CAC  660
Pro Val Thr Thr Asp Ile Lys His Ser Ile Lys Met Phe Glu Leu Thr Leu Asp Ala Leu  240
CCT GTG ACC ACT GAC ATT AAG CAT TCC ATA AAA ATG TTT GAA TTA ACA TTG GAT GCA CTT  720
Ile Ser Phe Asn Lys Arg Thr Leu Val Leu Phe Pro Asn Ile Asp Ala Gly Ser Lys Glu  260
ATC TCA TTT AAC AAG CGG ACC CTA GTC CTG TTT CCA AAT ATT GAC GCA GGG AGC AAA GAG  780
Met Val Arg Val Met Arg Lys Lys Gly Ile Glu His His Pro Asn Phe Arg Ala Val Lys  280
ATG GTT CGA GTG ATG CGG AAG AAG GGC ATT GAG CAT CAT CCC AAC TTT CGT GCA GTT AAA  840
His Val Pro Phe Asp Gln Phe Ile Gln Leu Val Ala His Ala Gly Cys Met Ile Gly Asn  300
CAC GTC CCA TTT GAC CAG TTT ATA CAG TTG GTT GCC CAT GCT GGC TGT ATG ATT GGG AAC  900
Ser Ser Cys Gly Val Arg Glu Val Gly Ala Phe Gly Thr Pro Val Ile Asn Leu Gly Thr  320
AGC AGC TGT GGG GTT CGA GAA GTT GGA GCT TTT GGA ACA CCT GTG ATC AAC CTG GGA ACA  960
Arg Gln Ile Gly Arg Glu Thr Gly Glu Asn Val Leu His Val Arg Asp Ala Asp Thr Gln  340
CGT CAG ATT GGA AGA GAA ACA GGG GAG AAT GTT CTT CAT GTC CGG GAT GCT GAC ACC CAA 1020
Asp Lys Ile Leu Gln Ala Leu His Leu Gln Phe Gly Lys Gln Tyr Pro Cys Ser Lys Ile  360
GAC AAA ATA TTG CAA GCA CTG CAC CTT CAG TTT GGT AAA CAG TAC CCT TGT TCA AAG ATA 1080
Tyr Gly Asp Gly Asn Ala Val Pro Arg Ile Leu Lys Phe Leu Lys Ser Ile Asp Leu Gln  380
TAT GGG GAT GGA AAT GCT GTT CCA AGG ATT TTG AAG TTT CTC AAA TCT ATC GAT CTT CAA 1140
Glu Pro Leu Gln Lys Lys Phe Cys Phe Pro Pro Val Lys Glu Asn Ile Ser Gln Asp Ile  400
GAG CCA CTG CAA AAG AAA TTC TGC TTT CCT CCT GTG AAG GAG AAT ATC TCT CAA GAT ATT 1200
Asp His Ile Leu Glu Thr Leu Ser Ala Leu Ala Val Asp Leu Gly Gly Thr Asn Leu Arg  420
GAC CAT ATT CTT GAA ACT CTA AGT GCC TTG GCC GTT GAT CTT GGC GGG ACG AAC CTC CGA 1260
Val Ala Ile Val Ser Met Lys Gly Glu Ile Val Lys Lys Tyr Thr Gln Phe Asn Pro Lys  440
GTT GCA ATA GTC AGC ATG AAG GGT GAA ATA GTT AAG AAG TAT ACT CAG TTC AAT CCT AAA 1320
Thr Tyr Glu Glu Arg Ile Asn Leu Ile Leu Gln Met Cys Val Glu Ala Ala Ala Glu Ala  460
ACC TAT GAA GAG AGG ATT AAT TTA ATC CTA CAG ATG TGT GTG GAA GCT GCA GCA GAA GCT 1380
Val Lys Leu Asn Cys Arg Ile Leu Gly Val Gly Ile Ser Thr Gly Gly Arg Val Asn Pro  480
GTA AAA CTG AAC TGC AGA ATT TTG GGA GTA GGC ATT TCC ACA GGT GGC CGT GTA AAT CCT 1440
Arg Glu Gly Ile Val Leu His Ser Thr Lys Leu Ile Gln Glu Trp Asn Ser Val Asp Leu  500
CGG GAA GGA ATT GTG CTG CAT TCA ACC AAA CTG ATC CAA GAG TGG AAC TCT GTG GAC CTT 1500
Arg Thr Pro Leu Ser Asp Thr Leu His Leu Pro Val Trp Val Asp Asn Asp Gly Asn Cys  520
AGG ACC CCC CTT TCT GAC ACT TTG CAT CTC CCT GTG TGG GTA GAC AAT GAT GGC AAC TGT 1560
```

```
Ala Ala Leu Ala Glu Arg Lys Phe Gly Gln Gly Lys Gly Leu Glu Asn Phe Val Thr Leu   540
GCT GCC CTG GCG GAA AGG AAA TTT GGC CAA GGA AAG GGA CTG GAA AAC TTT GTT ACA CTT 1620
Ile Thr Gly Thr Gly Ile Gly Gly Gly Ile Ile His Gln His Glu Leu Ile His Gly Ser   560
ATC ACA GGC ACA GGA ATC GGT GGT GGA ATT ATC CAT CAG CAT GAA TTG ATC CAC GGA AGC 1680
Ser Phe Cys Ala Ala Glu Leu Gly His Leu Val Val Ser Leu Asp Gly Pro Asp Cys Ser   580
TCC TTC TGT GCT GCA GAA CTG GGC CAC CTT GTT GTG TCT CTG GAT GGG CCT GAT TGT TCC 1740
Cys Gly Ser His Gly Cys Ile Glu Ala Tyr Ala Ser Gly Met Ala Leu Gln Arg Glu Ala   600
TGT GGA AGC CAT GGG TGC ATT GAA GCA TAC GCC TCT GGA ATG GCC TTG CAG AGG GAG GCA 1800
Lys Lys Leu His Asp Glu Asp Leu Leu Leu Val Glu Gly Met Ser Val Pro Lys Asp Glu   620
AAA AAG CTC CAT GAT GAG GAC CTG CTC TTG GTG GAA GGG ATG TCA GTG CCA AAA GAT GAG 1860
Ala Val Gly Ala Leu His Leu Ile Gln Ala Ala Lys Leu Gly Asn Ala Lys Ala Gln Ser   640
GCT GTG GGT GCG CTC CAT CTC ATC CAA GCT GCG AAA CTT GGC AAT GCG AAG GCC CAG AGC 1920
Ile Leu Arg Thr Ala Gly Thr Ala Leu Gly Leu Gly Val Val Asn Ile Leu His Thr Met   660
ATC CTA AGA ACA GCT GGA ACA GCT TTG GGT CTT GGG GTT GTG AAC ATC CTC CAT ACC ATG 1980
Asn Pro Ser Leu Val Ile Leu Ser Gly Val Leu Ala Ser His Tyr Ile His Ile Val Lys   680
AAT CCC TCC CTT GTG ATC CTC TCC GGA GTC CTG GCC AGT CAC TAT ATC CAC ATT GTC AAA 2040
Asp Val Ile Arg Gln Gln Ala Leu Ser Ser Val Gln Asp Val Asp Val Val Val Ser Asp   700
GAC GTC ATT CGC CAG CAG GCC TTG TCC TCC GTG CAG GAC GTG GAT GTG GTG GTT TCG GAT 2100
Leu Val Asp Pro Ala Leu Leu Gly Ala Ala Ser Met Val Leu Asp Tyr Thr Thr Arg Arg   720
TTG GTT GAC CCC GCC CTG CTG GGT GCT GCC AGC ATG GTT CTG GAC TAC ACA ACA CGC AGG 2160
Ile Tyr Ter
ATC TAC TAG ACC TCC AGG AAC AGA CAT GGA CCT TCT CTC CAG AGC TCC TGA GTG GAA TCA 2220
AGT TCT TGT CTT TAG GAT GAC CGT TTC TTA ACA ATC AAA TCT GGT ATT GAA CTG CAG GTG 2280
ACT TTG GCA GAG AAA TGT TTT CAC TTT TGG TCT CCT CTT CCA GAG TCA CCT TTC CCC ACT 2340
CCT ATT TTT GTA GAT GCT ATT CTT TCT GAT GTC TTC TTA CTA GGG GTC ATT TTA GCT CAA 2400
ACC CTG TAA GTT ACA GTC ACA ATT TTC TGT GCC AAA GCA GCT ACA ATA ATA GAG AGG AAG 2460
CCT TCT TAG AAC TCT GCT TAC TAA TGT ATT AAT ACC ACT GAG ACC TTC AGG CCT TGC TGG 2520
GAT ATC ACT TCA TCC TGA AGT TTG CAT TAA TAA TCC TTC CAG GCC GGG CAC AGT GGC TCA 2580
CGC CTG TAA TCC CAG CAC TTT GGG AGG CCG AGG CGG GCG GAT CAC GAG GTC AGG AGA TCG 2640
AGA CCG CCC TGG CTA ACA TGG TGA AAC ATG GTG AAA CCC CGT CTC TAC TAA AAA TAC AAA 2700
AAA TTA GCT GGG TGT GGT GGC GGG TCC AGC TAC TCG GGA GGC TGA GGC AGG AGA ATG GCA 2760
TGA ACC CCA GGC TGG AGT GCA GTG GCT CAA TGC AAC CTC TGC CCC CCG GGT TCG GTG ATT 2820
CTC CTG CCT CAG CCT CTT GAG TGG CTG GGA TTG CGG GCA CAT GCA CCA CGC CCG GCT GAT 2880
TTT TTT TTG TAT TTT TAG TGG ACA GGG TTT CAC CAT GTT GGC CAG GCT GGT CTT GAA CTC 2940
CTG ACC TCA GGT GAT CCG CCC GCC TCT GCC TCC CAA GGT GCT GGA TTA CAG GTG TAA GCC 3000
ATC ACA CCT GGC CCT AGT GAC AGG TTT TTA TGG GTA CTT TTA GAT GAT CTA AGA AAT CAT 3060
GTG CAT ATA TCT TTC AGA TTT CTA TTT TGG AAA ATG AAG GTT TCT ACA ACA TAT TGT TTC 3120
AGT GTT CAA ATA AAC TGA AGG ACT CAA CAT TAC ATT TGA ACT ATA TCC TTC CTA GTG GGT 3180
TAG TGT GAA AAA GAG TTT GGC TGA TTC CTA AAA CTC TGC CAG CCC TGC AGT AAT CTC CAG 3240
GCC TGG TTA TTG TTC AGA CAT TCC ATG GTG ATT CCT GGG AAG GAA GCT TGG CTG CTC AGT 3300
TTC TGA GTC TGG GGT GAG ATA ATG TTC TGG AAG GAC ATC TGT TCT TTG GTG TAA TCT CTC 3360
ATG GTG AAA TCT GCT CTG TAC ATC AGA CAA TTG CAT TGC TAC CAA GTT TCA TAC CAA ATA 3420
TTT GAA AGG ATG GTA TTG AAT CTA AAC CAA ATA TTA GGT TTT TAT TAA ACT CAT GGG AAG 3480
GCT AAT ATA TTC CAA CGT AAA TTA TTA CAT ATG GTT AAG TAA TTG CAT GTT AAT TTA TTT 3540
TAA TGT AAA TAT TTT TGT TAC TGT TCT GAG CCA AAT TCT AAA GAA AAA ATA AAT ACA TTT 3600
CCT TGT TGA AAA AAA AAA
```

Fig. 2. cDNA and deduced amino acid sequence of the human UDP-GlcNAc-2-epimerase/ManNAc kinase. The nucleotide position relative to +1 of the coding region and the numbered positions of the amino acid sequence are indicated on the right.

Germany) and the T7 sequencing kit (Pharmacia, Uppsala, Sweden), or with cycle sequencing and a LI-COR 4200 automatic sequencer (MWG-Biotech, Ebersberg, Germany). Both strands of all cDNA molecules were sequenced at least twice and the results analysed using the MacMolly Tetra software [23]. Nucleotides misincorporated by the Taq polymerase were identified by comparison of several independently obtained sequences.

2.4. Northern blotting

Commercially available Multiple Tissue Northern (MTN) blots (Clontech, Heidelberg, Germany) were used for expression analysis. The tissue blot carried 2 μg poly(A)$^+$ RNA from different human tissues: heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. The cell line blot carried 2 μg poly(A)$^+$ RNA from different human cancer cell lines: promyelocytic leukaemia HL-60, HeLa S3, chronic myelogenous leukaemia K-562, lymphoblastic leukaemia MOLT-4, Burkitt's lymphoma Raji, colorectal adenocarcinoma SW480, lung carcinoma A549 and melanoma G361. Blot membranes were prehybridised in ExpressHyb solution (Clontech) for 3h at 42°C and hybridised with a ^{32}P-labelled probe. A random priming kit (Life Technologies, Karlsruhe, Germany) was used for labelling. Hybridisation was performed using ExpressHyb solution at 42°C. Filters were washed several times in 0.2% SSC, 0.1% SDS and subjected to phosphor imaging.

```
                                        *                       #
MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKTEPEFFELDVVVLGSHLIDDYGNTYRM   60
""""""""""""""""""""""""""""""""""""""""""A""""""""""""""""""""""
""""""""""""""""""""""""""""""""""""""""""A""""""""""""""""""""""

IEQDDFDINTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALA   120
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

                    *   #                   *
TSAALMNIRILHIEGGEVSGTIDDSIRHAITKLAHYHVCCTRSAEQHLISMCEDHDRILL   180
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

          #                                     #
AGCPSYDKLLSAKNKDYMSIIRMWLGDDVKSKDYIVALQHPVTTDIKHSIKMFELTLDAL   240
""""""""""""""""""""""""""""""""C"""""""""""""""""""""""""""""""
""""""""""""""""""""""""""""""""C"""""""""""""""""""""""""""""""

ISFNKRTLVLFPNIDAGSKEMVRVMRKKGIEHHPNFRAVKHVPFDQFIQLVAHAGCMIGN   300
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

SSCGVREVGAFGTPVINLGTRQIGRETGENVLHVRDADTQDKILQALHLQFGKQYPCSKI   360
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

YGDGNAVPRILKFLKSIDLQEPLQKKFCFPPVKENISQDIDHILETLSALAVDLGGTNLR   420
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

   #          Δ      #
VAIVSMKGEIVKKYTQFNPKTYEERINLILQMCVEAAAEAVKLNCRILGVGISTGGRVNP   480
""""""""""""""""""""""""""S""""""""""""""""""""""""""""""""""""
""""""""""""""""""""""""""S""""""""""""""""""""""""""""""""""""

       #
REGIVLHSTKLIQEWNSVDLRTPLSDTLHLPVWVDNDGNCAALAERKFGQGKGLENFVTL   540
Q""V""""""""""""""""""""""""""""""""""""M""""""""""Q""""""
Q""V""""""""""""""""""""""""""""""""""""M""""""""""Q""""""

ITGTGIGGGIIHQHELIHGSSFCAAELGHLVVSLDGPDCSCGSHGCIEAYASGMALQREA   600
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
""""""""""""""""""""""""DV""""""""""""""""""""""""""""""""""""

KKLHDEDLLLVEGMSVPKDEAVGALHLIQAAKLGNAKAQSILRTAGTALGLGVVNILHTM   660
""""""""""""""""""""""""""""""V"""""""""""""""""""""""""""""""
""""""""""""""""""""""""""""""V"""""""""""""""""""""""""""""""

                                *                       ##
NPSLVILSGVLASHYIHIVKDVIRQQALSSVQDVDVVVSDLVDPALLGAASMVLDYTTRRIY 722
""""""""""""""""""R"""""""""""""""""""""""""""""""""""""""""""H
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""H
```

Fig. 3. Comparison of human, rat and mouse UDP-GlcNAc-2-epimerase/ManNAc kinase with indicated protein consensus sequences. The amino acid position is indicated on the right. Identical amino acids are marked by ". Putatively phosphorylated amino acids are underlined and specified: protein consensus sequences for #: protein kinase C, *: casein kinase II, Δ: cNMP-dependent kinase. Upper lane: human; middle lane: rat; lower lane: mouse.

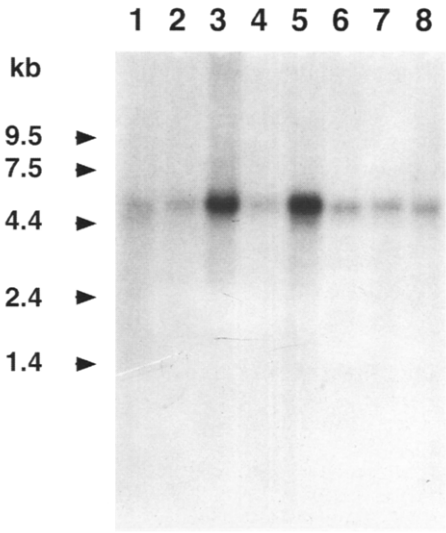

Fig. 4. Tissue-specific expression analysis of human UDP-GlcNAc-2-epimerase/ManNAc kinase by Northern blot analysis. A ^{32}P-labelled 500 bp coding region probe was hybridised to a MTN blot (Clontech, Heidelberg, Germany). Each lane carried 2 μg mRNA from various human tissues: (1) heart, (2) brain, (3) placenta, (4) lung, (5) liver, (6) skeletal muscle, (7) kidney, (8) pancreas. Hybridisation was done as described in Section 2.4.

3. Results and discussion

3.1. Nucleotide and amino acid sequence of the human UDP-GalNAc-2-epimerase/ManNAc kinase

To isolate the cDNA for the human UDP-GlcNAc-2-epimerase/ManNAc kinase we used a combination of RT-PCR, 5′ RACE and cDNA phage library screening. Based on the previously published sequence for the rat UDP-GlcNAc-2-epimerase/ManNAc kinase we generated primer pairs to amplify cDNA fragments encoding the human homologous enzyme. These were cloned into the vector pCR 2.1. At least three independently obtained clones were sequenced. By this technique we generated seven overlapping sequences by encoding the region +141 to +2031 bp. By the 5′ RACE method we obtained two overlapping clones encoding the 5′ end. Sequencing revealed that these fragments encode the −47 to 271 bp region. By screening a human liver cDNA library we obtained two clones encoding the region from +1400 to 3618 and from +1932 to 3618, respectively. The strategy for cDNA isolation and cloning is shown in Fig. 1.

The whole sequence for the UDP-GlcNAc-2-epimerase/ManAc kinase is shown in Fig. 2. The open reading frame encodes 722 amino acids with a predicted molecular weight of 79.4 kDa. Comparison of the human and rat coding sequences revealed 82% sequence identity based on nucleotides and 98% amino acid identity (Fig. 3). In the 3′ untranslated region there is no significant similarity between human and rat. At the nucleotide position +3588 a polyadenylation consensus sequence was identified. Within the polypeptide there are several protein consensus sequences for protein kinases, indicating that the human UDP-GlcNAc-2-epimerase/ManNAc kinase might be phosphorylated on serine/threonine residues; no conserved consensus sequence for tyrosine phosphorylation could be identified.

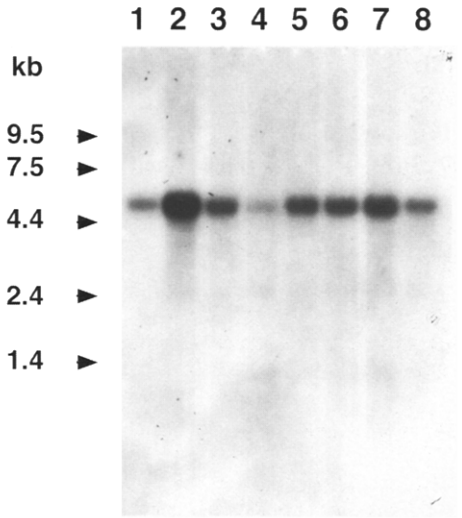

Fig. 5. Expression analysis of human UDP-GlcNAc-2-epimerase/ManNAc kinase in human cancer cell lines by Northern blot analysis. A ^{32}P-labelled 500 bp coding region probe was hybridised to a MTN blot (Clontech, Heidelberg, Germany). Each lane carried 2 μg mRNA from various human cell lines: (1) promyelocytic leukaemia HL-60, (2) HeLa S3, (3) chronic myelogenous leukaemia K-562, (4) lymphoblastic leukaemia MOLT-4, (5) Burkitt's lymphoma Raji, (6) colorectal adenocarcinoma SW480, (7) lung carcinoma A549, (8) melanoma G361. Hybridisation was done as described in Section 2.4.

3.2. *Expression of the human UDP-GlcNAc-2-epimerase/ ManNAc kinase*

The tissue-specific expression of the human UDP-GlcNAc-2-epimerase/ManAc kinase was examined by Northern blot analysis using poly(A)$^+$ RNA from different tissues (Fig. 4). Hybridisation with a coding region probe indicated that the corresponding mRNA migrates as a single 5.2 kb band. In contrast, the corresponding main band of rodent mRNA is ~3 kb in size. An additional mRNA population with a size of ~6 kb is found in rodent [14] but is missing in human mRNA. This indicates that humans possess one splice variant. The mRNA was detected in all human tissues tested, and the highest expression occurred in liver and placenta. The high mRNA level in liver is consistent with the high enzyme activity in this organ. The high expression in placenta indicates its essential role during development. This is supported by the detection of the enzyme in the early stages of mouse embryogenesis [15]. By analysing mRNA we could show that the UDP-GlcNAc-2-epimerase/ManNAc kinase is differentially expressed by human cancer cell lines. The enzyme is strongly expressed in cervix carcinoma-derived HeLa cells and weakly expressed in lymphoblastic leukaemia MOLT-4 cells (Fig. 5). Recently, the important role of the UDP-GalNAc-2-epimerase as a regulator of cell surface glycoconjugate sialylation in haematopoietic cell lines was demonstrated [17]. Whether the enzyme is functionally relevant for the progression of carcinogenesis remains to be clarified in a separate study.

Acknowledgements: The authors gratefully acknowledge the excellent technical assistance of Katrin Büttner and Sabine Nöhring, and Dr. Thomas Scott, Leeds, for the critical reading of the manuscript. This work was supported by the Bundesministerium für Bildung und Forschung, the Deutsche Forschungsgemeinschaft (Ho1959/1-1 and SFB 366), the Sonnenfeld-Stiftung, Berlin, and the Fonds der Chemischen Industrie, Frankfurt/Main.

References

[1] Schauer, R., Kelm, S. and Reuter, G. (1995) in: Biology of Sialic Acids (Rosenberg, A., Ed.), pp. 7–67, Plenum Press, New York.
[2] Tettamanti, G. and Riboni, L. (1993) Adv. Lipid Res. 25, 235–267.
[3] Reutter, W., Stäsche, R., Stehling, P. and Baum, O. (1997) in: Glycosciences: Status and Perspectives (Gabius, H.-J. and Gabius, S., Eds.), pp. 245–259, Chapman and Hall, Weinheim.
[4] Edelman, G.M. and Crossin, K.L. (1991) Annu. Rev. Biochem. 60, 155–190.
[5] Foxall, C., Watson, S.R., Dowbenko, D., Fennie, C., Lasky, L.A., Kiso, M., Hasegawa, A., Asa, D. and Brandley, B.K. (1992) J. Cell Biol. 117, 895–902.
[6] Kelm, S., Schauer, R. and Crocker, P.R. (1996) Glycoconjug. J. 13, 913–926.
[7] Yang, P., Major, D. and Rutishauser, U. (1994) J. Biol. Chem. 269, 23039–23044.
[8] Karlsson, K.A. (1995) Curr. Opin. Struct. Biol. 5, 622–635.
[9] Horstkorte, R., Lessner, N., Gerardy-Schahn, R., Lucka, L., Danker, K. and Reutter, W. (1999) Exp. Cell Res. 246, 122–128.
[10] Keppler, O.T., Stehling, P., Herrmann, M., Kayser, H., Grunow, D., Reutter, W. and Pawlita, M. (1995) J. Biol. Chem. 270, 1308–13014.
[11] Rens-Domiano, S. and Reisine, T. (1991) J. Biol. Chem. 266, 20094–20102.
[12] Pirie-Shepherd, S.R., Jett, E.A., Andon, N.L. and Pizzo, S.V. (1995) J. Biol. Chem. 270, 5877–5881.
[13] Hinderlich, S., Stäsche, R., Zeitler, R. and Reutter, W. (1997) J. Biol. Chem. 272, 24313–24318.
[14] Stäsche, R., Hinderlich, S., Weise, C., Effertz, K., Lucka, L., Moormann, P. and Reutter, W. (1997) J. Biol. Chem. 272, 24319–24324.
[15] Horstkorte, R., Nöhring, S., Wiechens, N., Schwarzkopf, M., Danker, K., Reutter, W. and Lucka, L. (1999) Eur. J. Biochem. 260, 923–927.
[16] Harms, E., Kreisel, W., Morris, H.P. and Reutter, W. (1973) Eur. J. Biochem. 32, 254–262.
[17] Keppler, O., Hinderlich, S., Langner, J., Schwartz-Albiez, R., Reutter, W. and Pawlita, M., Science 284, 1372–1375.
[18] Weiss, P., Tietze, F., Gahl, W.A., Seppala, R. and Ashwell, G. (1989) J. Biol. Chem. 264, 17635–17636.
[19] Seppala, R., Tietze, F., Krasnewich, D., Weiss, P., Ashwell, G., Barsh, G., Thomas, G.H., Packman, S. and Gahl, W.A. (1991) J. Biol. Chem. 266, 7456–7461.
[20] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) in: Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[21] Mullis, K.B. and Faloona, F.A. (1987) Methods Enzymol. 155, 335–350.
[22] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[23] Schöneberg, U., Vahrson, W., Priedemuth, U. and Wittig, B. (1993) in: Analysis and Interpretation of DNA and Protein Sequences using MacMolly® Tetra, 2nd edn., pp 1–105, KAROI-Verlag, Bielefeld.