Published in final edited form as:

Biochemistry. 2011 October 18; 50(41): 8914–8925. doi:10.1021/bi201050u.

Identification, Tissue Distribution and Molecular Modeling of Novel Human Isoforms of the Key Enzyme in Sialic Acid Synthesis, UDP-GlcNAc 2-epimerase/ManNAc Kinase†

Tal Yardeni^{‡,§}, Tsering Choekyil^{||}, Katherine Jacobs[‡], Carla Ciccone[‡], Katherine Patzel[‡], Yair Anikster[§], William A. Gahl[‡], Natalya Kurochkina^{||}, and Marjan Huizing^{‡,*}

[‡]Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, United States

[§]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, 69978, Israel

^{||}The School of Theoretical Modeling, Chevy Chase, MD 20825, United States

Abstract

UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE) catalyzes the first two committed steps in sialic acid synthesis. In addition to the previously described 3 human GNE isoforms (hGNE1-hGNE3), our database and PCR analysis yielded an additional 5 human isoforms (hGNE4-hGNE8). hGNE1 is the ubiquitously expressed major isoform, while the hGNE2-8 isoforms are differentially expressed and may act as tissue-specific regulators of sialylation. hGNE2 and hGNE7 display a 31-residue N-terminal extension compared to hGNE1. Based on similarities to kinases and helicases, this extension does not seem to hinder the epimerase enzymatic active site. hGNE3 and hGNE8 contain a 55 residue N-terminal deletion, and a 50-residue N-terminal extension compared to hGNE1. The size and secondary structures of these fragments are similar, and modeling predicted that these modifications do not affect the overall fold compared to hGNE1. However, epimerase enzymatic activity of GNE3 and GNE8 is likely absent, since the deleted fragment contains important substrate binding residues in homologous bacterial epimerases. hGNE5-hGNE8 have a 53-residue deletion, which was assigned a role in substrate (UDP-GlcNAc) binding. Deletion of this fragment likely eliminates epimerase enzymatic activity. Our findings imply that GNE is subject to evolutionary mechanisms to increase cellular functions, without increasing the number of genes. Our expression and modeling data contribute to elucidation of the complex functional and regulatory mechanisms of human GNE, and may contribute to further elucidating the pathology and treatment strategies of the human GNE-opathies sialuria and hereditary inclusion body myopathy.

The bifunctional enzyme uridine diphosphate (UDP)¹-N-acetylglucosamine (GlcNAc) 2-epimerase/N-acetylmannosamine (ManNAc) kinase (GNE), encoded by the *GNE* gene, catalyzes the first two committed, rate-limiting steps in the biosynthesis of N-acetylneuraminic acid (Neu5Ac) (1, 2). Neu5Ac is the most abundant mammalian sialic acid

[†]This study was supported by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States (T.Y., K.J., C.C., K.P., W.A.G., and M.H.) and Research Funds of The School of Theoretical Modeling, Chevy Chase, Maryland, United States (T.C. and N.K.). This work was performed in partial fulfillment of the requirements for a PhD degree of T.Y., Sackler Faculty of Medicine, Tel Aviv University, Israel.

*Corresponding author: Marjan Huizing, NHGRI, NIH, 10 Center Drive, Bld 10, Rm 10C103, Bethesda, MD 20892-1851, USA. Phone: (301) 402-2797. Fax: (301) 480-7825. mhuizing@mail.nih.gov.

SUPPORTING INFORMATION

This material is available free of charge via the Internet at <http://pubs.acs.org>.

and the precursor of most naturally existing sialic acids (3). Sialic acids are negatively charged, terminal residues on glycoconjugates, and assist in many cellular functions including cell-cell interactions, proliferation, and viral or bacterial infections (3, 4). The GNE enzyme consists of two enzymatic domains. The N-terminal domain carries out UDP-GlcNAc epimerase function, whereas the C-terminal domain is responsible for ManNAc kinase activity. In mammals, the end product of sialic acid synthesis, CMP-Neu5Ac, feedback-inhibits UDP-GlcNAc 2-epimerase activity of GNE by binding to its allosteric site (5).

Two distinct human disorders, sialuria (OMIM 269921) and hereditary inclusion body myopathy (HIBM; OMIM 600737), are associated with predominantly missense mutations in *GNE*. Sialuria is an autosomal dominant disorder characterized by coarse facies, variable developmental delay, hepatomegaly and recurrent infections. To date, only seven sialuria patients are described worldwide. All patients have a heterozygous missense mutation affecting the allosteric site of *GNE*, leading to loss of feedback-inhibition of GNE-epimerase activity by CMP-Neu5Ac, resulting in excessive sialic acid production (6, 7). HIBM and its allelic Japanese disorder, distal myopathy with rimmed vacuoles, or DMRV (OMIM 605820), is an autosomal recessive neuromuscular disorder of adult onset, characterized by slowly progressive muscle weakness and atrophy. More than 500 HIBM\DMRV patients exist worldwide, harboring over 60 *GNE* mutations. HIBM\DMRV patients have recessive (predominantly missense) mutations in either enzymatic domain of *GNE*, leading to decreased enzyme activity and, presumably, decreased sialic acid production (2, 8, 9). Whether hyposialylation is the main cause of the neuromuscular symptoms in HIBM\DMRV patients remains unknown.

In prokaryotes, GNE epimerase and kinase functions are carried out by two separate enzymes, and prokaryotic 2-epimerases have no allosteric feedback inhibition. In mammals, a bifunctional enzyme might have evolved by gene fusion of the two independent enzymes responsible for epimerase/kinase activity. Similarities between mammalian GNE N-terminal regions with prokaryotic UDP-GlcNAc 2-epimerases and mammalian GNE C-terminal regions with members of the sugar kinase superfamily previously assisted in identifying characteristic motifs of the GNE epimerase and kinase enzymatic domains (10, 11). Bacterial 2-epimerases are allosterically regulated by its substrate UDP-GlcNAc. A structural basis for allosteric activation was demonstrated by a crystallographic analysis of the *B. subtilis* ² 2-epimerase in complex with the reaction intermediate UDP (12). In addition, the crystallographic structures of the *E. coli* GNE enzyme unbound and in complex with UDP-N-acetylglucosamine (pdb code 1f6d, 1vgv), and the *V. cholera* (pdb code 1dzc) and *B. anthracis* (pdb code 3beo) enzymes in complex with UDP-N-acetylglucosamine were solved. Similarity of the N-terminal domain of the *H. sapiens* GNE to *V. cholera* (27% homology), *E. coli* (20% homology) and *B. anthracis* (18% homology) 2-epimerases was

¹Abbreviations: ADP, Adenosine diphosphate; ASKHA, Acetate and Sugar Kinase/Hsp70/Actin; ATP, Adenosine triphosphate; B2M, β 2 microglobulin; BLAST, Basic Local Alignment Search Tool; CDD1, cytidine deaminase 1; CDM, Consensus Data Mining; CMP, cytidine monophosphate; DMRV, distal myopathy with rimmed vacuoles; FDM, Fragment Database Mining; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GNE, uridine diphospho-N-acetylglucosamine (UDP-GlcNAc) 2-epimerase/N-acetylmannosamine (ManNAc) kinase; GOR, Garnier-Osguthorpe-Robson; HIBM, hereditary inclusion body myopathy; LDH, lactate dehydrogenase; MDH, malate dehydrogenase; NCBI, National Center for Biotechnology Information; Neu5Ac, N-acetylneuraminic acid; OMIM, Online Mendelian Inheritance in Man; qPCR, quantitative real-time polymerase chain reaction; pdb, protein databank; PSIPRED, Protein Structure Initiative Prediction; SGK1, glucocorticoid-induced kinase 1.

²Abbreviations of bacterial names: *A. avenae*: *Acidovorax avenae*; *A. orientalis*: *Amycolatopsis orientalis*; *B. anthracis*: *Bacillus anthracis*; *B. subtilis*: *Bacillus subtilis*; *E. coli*: *Escherichia coli*; *H. arsenicoxydan*: *Herminiimonas arsenicoxydans*; *H. pylori*: *Helicobacter pylori*; *H. sapiens*: *Homo sapiens*; *M. musculus*: *Mus musculus*; *P. aeruginosa*: *Pseudomonas aeruginosa*; *P. abyssi*: *Pyrococcus abyssi*; *P. carotovorum*: *Pectobacterium carotovorum*; *P. citrinum*: *Penicillium citrinum*; *P. denitrificans*: *Paracoccus denitrificans*; *P. falciparum*: *Plasmodium falciparum*; *R. rattus*: *Rattus rattus*; *S. achanthias*: *Squalus achanthias*; *S. aureus*: *Staphylococcus aureus*; *S. cerevisiae*: *Saccharomyces cerevisiae*; *T. thermophilus*: *Thermus thermophilus*; *T. reesei*: *Trichoderma reesei*; *V. cholera*: *Vibrio cholera*.

used to model its three-dimensional structure. In previous studies, structural elements and important ATP, ADP, Mg^{2+} and substrate-binding amino acids were assigned on the basis of these similarities (10, 11). The N-terminal epimerase domain of the human GNE enzyme contains two α/β domains (domains I and II) that form a cleft at the domain interface harboring the active site. Topology of both domains is similar to the Rossmann dinucleotide binding fold (13). Rossmann fold domains are conserved among mammalian and bacterial 2-epimerases. The human N-terminal GNE epimerase domain has a 7-stranded parallel β -sheet sandwiched between a total of 7 α -helices. The C-terminus of the GNE epimerase domain contains a 6-stranded β -sheet surrounded by a total of 7 α -helices (11). Other carriers of the Rossmann fold, including glycosyltransferases and the epimerase domains of 2-epimerases, have similar N-terminal and C-terminal domains (14, 15). The crystallographic structure of the ManNAc kinase domain of human GNE is solved at 2.84 Å resolution (pdb code 3eo3) (16). Residues 409–431 of the mammalian GNE ManNAc kinase domain showed similarities with the phosphate 1 motif of the ATP-binding domain of eukaryotic hexokinases (10). Similar to hexokinases (17), mammalian GNE ManNAc kinase contains a 5-stranded β -sheet $\beta 3\beta 2\beta 1\beta 4\beta 5$ with $\beta 2$ being anti-parallel to four other parallel strands with a pair of parallel α -helices located on each side of the β -sheet (Domain I). Another 5-stranded β -sheet $\beta 8\beta 7\beta 6\beta 9\beta 1$ with $\beta 7$ being anti-parallel to four other parallel strands is surrounded by a pair of parallel α -helices on each side (Domain II). The structure of two similar domains involved in ATP binding is a common feature of the ASKHA (Acetate and Sugar Kinase/Hsp70/Actin) superfamily, described in detail for the bacterial poly(P)/ATP-glucomannokinase (18, 19).

Recently, different human *GNE* mRNA splice variants and three predicted translated proteins, hGNE1, hGNE2 and hGNE3 were described (20, 21). Subsequently, two different mouse *Gne* mRNA splice variants were described, *Gne1* and *Gne2*, together with their expression in selected tissues (22). In the current study we identified additional human isoforms hGNE4-8, and demonstrate expression of hGNE isoform transcripts in a wide variety of tissues. It is unknown which role these isoforms play in GNE regulation, or GNE-related disease pathology. Based on our previous modeling results of the hGNE1 isoform (11), we now analyze and compare the structural features, with respect to catalytic activity, ligand binding and allosteric regulation, of all eight human GNE isoforms.

EXPERIMENTAL PROCEDURES

Bioinformatics

For nucleotide sequence homology searches, the National Center for Biotechnology Information (NCBI) BLAST searches were employed (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), as well as the Blat searches provided by the University of California Santa Cruz (UCSC) genome browser (<http://genome.ucsc.edu/cgi-bin/hgBlat>). For identification of GNE homologous protein structures, sequences were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), including the human GNE1 isoform (hGNE1: GenBank accession number NM_005476), hGNE2 (NM_001128227), hGNE3 (NM_001190388), hGNE4 (NM_001190383), and hGNE5 (NM_001190384). Atomic coordinates of proteins determined by X-ray crystallography were extracted from the Brookhaven Protein Databank (23). Protein sequences were aligned using the BLAST procedure. For a cut-off of homology, the expected threshold E-value of 0.1 was used. This value is the statistical significance threshold for reporting matches against database sequences and is comparable with a P-value. Matches of 50 or more proteins were made. These sequence alignments were combined with structural alignments to identify structurally identical residues responsible for similar enzymatic functions. Sequences of the additional N-terminal fragments of hGNE2 and hGNE7 (Table 2) and hGNE3 and hGNE8

(Table 3) were compared with sequences of proteins (and their homologs) whose structures are known and deposited in the Protein Data Bank.

RNA and cDNA

Tissue-specific RNA was purchased from Clontech (human Total RNA Master Panels). cDNA was created from all RNA using a High Capacity RNA-to-cDNA Reverse Transcription Kit (Applied Biosystems). Human multiple tissue cDNA panels (MTC panels I and II, Human Fetal MTC) were purchased from Clontech (Clontech Laboratories).

Transcript-Specific PCR

Primers were designed for transcript-specific PCR amplification of predicted human *GNE* splice variants (Figures 1 and 2, Supporting Information Table S1). All PCR reactions were performed with HotStarTaq polymerase, according to the manufacturer's instructions (Qiagen). PCR products were separated by agarose gel electrophoresis and selected bands were excised (QIAquick Gel Extraction Kit, Qiagen) and directly sequenced using BigDye® Terminator v3.1 chemistry and an ABI 3130xl Genetic Analyzer (Applied Biosystems).

Quantitative Real-Time PCR

TaqMan primers and probes were ordered as pre-manufactured assays-on demand (human *GNE* assay Hs01103402_m1 (exon 8–9)), or custom designed for splice variant specific sequences using the ABI Assay-by-Design service (Supporting Information Table S2) (Applied Biosystems). The housekeeping gene *GAPDH* (Glyceraldehyde 3-phosphate dehydrogenase; Hs99999905_m1) was used as internal control gene. All quantitative real-time PCR (qPCR) reactions and subsequent analyses were performed on an ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems). The pre-run thermal cycling conditions were 10 min at 95°C to activate the *Taq*DNA-polymerase, followed by 40 cycles of 95°C for 15 s and 60°C annealing/extension for 1 min. Each experiment was performed in triplicate. Within each experiment, reactions were run in triplicate. Relative gene expression levels were determined by the comparative threshold cycle method (ddCt) (24).

Modeling of the Human *GNE* Isoforms

Secondary structure predictions for the 31-residue N-terminal extension of the hGNE2 and hGNE7 isoforms and the 50-residue N-terminal extension of the hGNE3 and hGNE8 isoforms were performed by four methods: GOR V (Garnier- Osguthorpe-Robson) (25, 26), FDM (Fragment Database Mining) (27), CDM (Consensus Data Mining) (28), and PSIPRED (Protein Structure Initiative Prediction) (29). These extensions were aligned with fragments bearing similarities in amino acid sequence and assisted in determining three-dimensional structure (Figure 3B, Table 2, Table 3). Two sets of *GNE* epimerase domain deletions (either in *GNE* epimerase domain I involving hGNE3, hGNE5 and hGNE8 isoforms, or in the *GNE* epimerase domain II involving hGNE5, hGNE6, hGNE7 and hGNE8 isoforms) were analyzed on the presence of important catalytic residues and residues involved in ligand binding and allosteric regulation.

RESULTS

Human *GNE* Isoforms

The human *GNE* gene (*hGNE*) consists of 13 exons, which were previously numbered exons A1-12 (20). In this manuscript we introduce the more logical and useful numbering of *hGNE* exons 1–13 (Figure 1), a universal numbering also used by GenBank for the longest splice form of the *hGNE* gene (NM_001128227).

Three *GNE* mRNA splice variants were previously described, resulting from alternative splicing of 5' exons 1, 2 and 3 (20, 21), creating 3 protein open reading frames, i.e., hGNE1, hGNE2 and hGNE3 (Figure 1).

hGNE1 (GenBank NM_005476) is the isoform described in all previous biochemical and mutation analysis studies (2, 30, 31); it encodes 722 amino acids with its translation start codon in exon 3, considered nucleotide 1 in this study and previous *GNE* mutation reports. Isoform hGNE2 (NM_001128227) has a 31-amino acid extension at the N-terminus, yielding a 753 amino acids protein with its translation start codon in exon 1 at nucleotide -93 (Figure 1). hGNE3 (NM_001190388) is extended and deleted compared to hGNE1; it has 717 amino acids with its start codon in exon 1 at nucleotide -191 (Figure 1). Note that the start codons in exon 1 for hGNE2 and hGNE3 are at different positions (21, 22). Bioinformatic database searches revealed two additional *hGNE* transcript variants, encoding hGNE4 (NM_001190383) and hGNE5 (NM_001190384), skipping the in-frame exons 10 and 5, respectively.

Tissue-specific PCR amplification (Figure 1 and Supporting Information Table S1) of human cDNA tissue panels revealed hGNE1 encoding transcript expression in all tissues tested (Figure 2A and Table 1). Transcripts encoding isoforms hGNE2 and hGNE3 were expressed in the same pattern in a subset of tissues. These hGNE1-hGNE3 expression results resembled previous data (21), with two exceptions. In brain, we found only mRNA transcripts encoding hGNE1, and not hGNE2 or hGNE3; GNE2 was previously described to express in brain. In pancreas, we found expression of hGNE1, hGNE2 and hGNE3 transcripts; hGNE3 expression was not previously reported in pancreas.

Even though a transcript encoding isoform hGNE4 is listed in GenBank (NM_001190384), we were unable to detect expression of this transcript in the tested tissues (Figure 2A). The transcript encoding hGNE5 was minimally expressed in a subset of tissues (Figure 2A). Verification of the splice variant transcripts for hGNE4 and hGNE5 revealed three additional novel splice transcripts (Figures 1 and 2) whose proteins products we named hGNE6, hGNE7, and hGNE8. The hGNE6 encoded transcript is expressed in all tissues, similar to hGNE1. Likewise, transcripts for hGNE7 and hGNE8 followed the same expression pattern as those for hGNE2 and hGNE3 (Figures 1 and 2, Table 1).

We then quantified *hGNE* expression levels by qPCR in 10 selected human tissues with isoform-specific TaqMan primer-probe sets (Figure 2B, Table 1, Supporting Information Table S2). Expression of total *hGNE* mRNA differed among tissues, with the highest expression in placenta and liver. The transcripts encoding isoforms hGNE2-8 showed decreased expression compared to the hGNE1 transcript, had absent expression in some tissues (Figure 2B, Table 1), and exhibited a quantitative distribution pattern similar to that of the hGNE1 transcript in other tissues (i.e., liver, lung, placenta).

Modeling of hGNE Isoforms

Comparison of the isoforms hGNE1 - hGNE8 showed that the GNE-epimerase encoding domain (amino acids 1–378) exhibits differences among all isoforms, while only one isoform, hGNE4, showed differences in the kinase encoding domain (amino acids 410–722) (10, 32). Since we were unable to detect hGNE4 coding mRNA transcripts in tested tissues (Figure 2A), and further analysis showed that this variant was submitted as a mutation identified in a distal myopathy with rimmed vacuoles (DMRV) patient (GenBank EU093084) instead of a novel isoform, we did not further analyze hGNE4 secondary structure. In previous studies, structural elements and domains were assigned to the major hGNE1 isoform (10, 11). Here we align the secondary structure elements of the N- and C terminal domains of the human GNE epimerase domain of all human GNE isoforms (Figure

3 and Figure 4). There are two sites of differences between the isoforms hGNE1-hGNE8 in the epimerase domain. The first site is at the N-terminus of the (N-terminal) domain I, and the second site is at the N-terminus of the (C-terminal) domain II (gray shaded regions in Figure 4 and Figure 5).

Modeling of the N-Terminal Extensions of hGNE2 and hGNE7

The hGNE2 and hGNE7 isoforms contain all of the hGNE1 isoform and, in addition, have a 31-residue N-terminal extension compared to hGNE1 (Figure 3A). Secondary structure prediction algorithms and our modeling of this 31 residue N-terminal extension indicated possible presence of one or two additional α -helices (Figure 3B). Alpha-helical extension at the N-terminus of the nucleotide binding domains occurs in some other proteins, like in *S. achanthias* muscle lactate dehydrogenase (LDH) where it is also involved in the formation of intersubunit contacts (33). Extension of the Rossmann fold as N-terminal extra helix is present in *P. aeruginosa* UDP-N-acetylglucosamine 4-epimerase but not in *E. coli* UDP-galactose 4-epimerase (34).

The hGNE2 and hGNE7 N-terminal extension has high similarity to α -helices in several other proteins including the human serum and glucocorticoid-induced kinase 1 (SGK1) and *B. subtilis* putative transcriptional regulator YFHH (Table 2). SGK1 contains an intermolecular disulfide bond (35) in the region homologous to the hGNE2 and hGNE7 N-terminal extension. In SGK1, this region is predicted to come in close contact with an activation loop of the neighboring subunit and with its phosphorylation site (35). Interestingly, a cysteine residue is present in the 31 amino acid N-terminal extension of hGNE2 and hGNE7 (at residue 12), as well as in the homologous SGK1 sequence and in several homologous histidine kinases (Figure 3B and Table 2). In addition, SGK1 (residues 175–240) harbors another 25% homologous region to GNE (residues 370–471 of hGNE1). This GNE region is located between the epimerase and kinase enzymatic domains and contains a cysteine at residue 388. The cysteine residues C12 and C388 may be candidates for disulfide bonds, whether intermolecular (as in SGK1) or intramolecular (possibly between GNE regions) remains to be determined. However, as a cytosolic enzyme disulfide bonds may be unlikely for GNE due to the reducing milieu. Whether the high homology between SGK1 and GNE also provides data for similarities in structural or binding properties is a subject of future studies.

Modeling of the N-Terminal Extensions of hGNE3 and hGNE8

Compared to the major hGNE1 isoform, hGNE3 and hGNE8 contain a deletion of the first 55 N-terminal residues, including β 1, α 1, β 2, and part of α 2 (encoded by exon 3 in hGNE1). In addition, hGNE3 and hGNE8 contain a novel 50-residue N-terminal extension (encoded by exon 1). Secondary structure prediction algorithms and our modeling of the 50-residue N-terminal extension indicate possible presence of alternating α -helices and β -strands (Figure 3B), which may serve as a replacement of the deleted fragment (β 1, α 1, β 2, and part of α 2) in the three-dimensional structure of the protein. In bacterial epimerase enzymes, the region homologous to the N-terminal portion of hGNE3 and hGNE8, including deleted β 1, α 1, β 2, and part of α 2 (Figure 3A, Figure 4), contain an active site and an allosteric site. In the *B. anthracis* enzyme, this region is involved in contacts with the bacterial allosteric activator UDP-N-acetylglucosamine (mainly its phosphate and N-acetyl groups).

Alignment of the novel 50-residue fragment with the *A. orientalis* glycosyltransferase, also carrying a Rossmann fold (pdb code 1rrv, Table 3), suggests that it is likely that including the 50-residue extension may not change the overall fold of the protein. The N-terminal part of the 50-residue extension has lower homology to other proteins compared to the middle and C-terminal parts (Table 3). The middle part of the fragment shows homology to the N-

acetyl-D-glucosamine binding region of α -1,2-mannosidase, which possesses a $(\alpha\alpha)_7$ fold and includes short β -strands (pdb codes 1hcu, 2r18). The C-terminal part of the fragment shows the highest similarity with *S. cerevisiae* cytidine deaminase CDD1 (pdb code 1r5t), which is involved in RNA-editing activity (36). Among the deaminase CDD1 active site residues, W79 and C86 align with the hGNE3 and hGNE8 N-terminal extension in the region conserved between the two enzymes (Table 3). Amino acid W79 (W80 and W86 in the homologous deaminases APOBEC1 and AID, respectively) forms a van der Waals contact with the cytosine edited by the deaminase activity while C86 (C87 in AID) is involved in zinc coordination and nucleotide binding (36). The tertiary structure of the deaminase CDD1 consists of a five-stranded β -sheet surrounded by three α -helices on both sides, similar to other deaminases and to the loops followed by a β -strand in the homologous N-terminal regions of hGNE3 and hGNE8. The *M. musculus* deaminase (pdb code 2fr5, 36% homology to *S. cerevisiae* deaminase), binds the inhibitor tetrahydrouridine (pdb code 2fr5) (37) in the proximity of the region homologous to the GNE 50-residue extension.

The deleted portion of the N-terminus of hGNE3 and hGNE8 compared to hGNE1 (β 1, α 1, β 2, and part of α 2; Figure 3A), contains several amino acids important for enzymatic function. Helix α 1 [amino acids 18–35 of hGNE1] contains R19, D21, K24 (Figure 3A, Figure 4) corresponding to the α 1 helix [amino acids 9–26] and R10, E12, K15 of the *E. coli* enzyme, which are located in the vicinity of the active site (K15 in *E. coli*) and are involved in stabilization of the structure (E12) and binding of the UDP portion of the substrate (R10). Of the amino acids located in the β 2 α 2 loop and helix α 2 of hGNE1, which are deleted in hGNE3 and hGNE8 (Figure 3A), in *B. anthracis* 2-epimerase Q43, H44, and Q46 were found to compose an allosteric site and are conserved in bacterial enzymes. These residues, except the histidine, are not conserved in the human enzyme. The conserved histidine corresponds to H49 of the *R. rattus* and H49 of the *H. sapiens* 2-epimerase. This residue was shown to be directly involved in the epimerization process in rat GNE, which is 98% homologous to hGNE1. The point mutation H49A (erroneously indicated as H45A in Effertz et al. 1999 (ref. 10)) in the rat 2-epimerase resulted in loss of GNE epimerase activity but not in kinase activity and did not appear to perturb the enzyme oligomeric state (10).

Modeling of the Exon 5 Deletion in hGNE5-8

The hGNE5, hGNE6, hGNE7, and hGNE8 isoforms are derivatives of the hGNE1 isoform with deletion of a 53 amino acid fragment (encoded by exon 5) which includes the β 8, α 9, β 9, and part of α 10 secondary structure elements (Figure 3A, Figure 4). The deleted fragment contains a proposed catalytic residue, H220, which corresponds to H213 of the *E. coli* 2-epimerase (Figure 5B). Two amino acids from the deleted fragment form a bacterial allosteric site in the *B. anthracis* enzyme (R210 and H242) that are conserved in *V. cholera* and *E. coli* but not in *H. sapiens* 2-epimerase. Residue R210 of the *B. anthracis* 2-epimerase carries a role in anchoring UDP and UDP-GlcNAc and is conserved only in non-hydrolyzing enzymes. Residue H242 forms hydrogen bonds with β -phosphate oxygen and a hydroxyl group of UDP-GlcNAc upon conformational change. It is assigned a role in the binding of UDP-GlcNAc at the allosteric site and stabilization of the UDP intermediate.

DISCUSSION

In the current study we explored GNE isoform expression in human tissues and compared structural differences between hGNE1 and six other human GNE isoforms by molecular modeling. In addition to the previously described 3 human GNE isoforms, database and PCR analysis yielded an additional 5 hGNE isoforms, summarized in Figure 1. Since the commercially available GNE antibodies do not work well on Western blotting, we were unable to confirm expression of each transcript at the protein level. In our predictions below, we assumed each isoform transcript produces a protein.

Tissue analysis showed differential expression of the 8 hGNE isoform encoding transcripts. hGNE1 appeared to be the major isoform, with mRNA expression in all tissues tested. hGNE6 was the only other isoform whose mRNA was expressed in all tested tissues (although at low levels). hGNE6 likely only has kinase activity, since it lacks part of the epimerase functional domain (skipping of in-frame exon 5 compared to hGNE1). We identified 4 other hGNE transcripts (hGNE5, hGNE7, hGNE8) that lacked exon 5 and, therefore, are predicted to display only kinase activity. Expression of hGNE5 was very low with conventional PCR (Figure 2A), and no hGNE5-specific Taqman assay could be designed, limiting our further pursuit of hGNE5 expression. hGNE7 and hGNE8 followed a similar tissue-distribution pattern and were expressed at low levels in selected human tissues (Figure 2, Table 1). Interestingly, expression of hGNE2 and hGNE3 coding transcripts followed the same tissue-distribution pattern as hGNE7 and hGNE8, indicating that they may be regulated together. We did not detect hGNE4 coding mRNA transcripts in tested tissues, and further analysis showed that this variant was submitted as a mutation identified in a DMRV patient (GenBank EU093084), instead of a novel isoform. We list this variant because a GenBank accession number is ascribed to this isoform, listed in GenBank as variant 5, but we did not further analyze hGNE4 by molecular modeling.

Our cDNA analysis of a variety of mouse tissues identified transcripts encoding the two previously described mouse isoforms mGne1 and mGne2, which have high homologies of 98.5% and 96.8% to hGNE1 and hGNE2, respectively (21). Orthologs of the other human isoforms do not appear to exist in mouse tissues, perhaps due to an evolutionary mechanism.

The highest hGNE mRNA expression levels occurred in the liver and placenta. Liver is the major organ of sialic synthesis and was shown to display high GNE enzymatic activity (31). High hGNE expression in placenta may be related to an essential role of sialic acid during development (31). The lowest hGNE mRNA expression was found in skeletal muscle, as previously described (31), which may explain why skeletal muscle is the only affected tissue in HIBM\DMRV patients. A slight decrease in GNE enzymatic activities in HIBM\DMRV muscle may have a larger effect on sialic acid supply in muscle than in other tissues that have higher natural GNE expression and enzyme activity levels. The effect of HIBM\DMRV patient mutations on hGNE isoform expression remains to be determined.

GNE epimerase and kinase enzymatic activities were reported for the hGNE-1,-2, and -3 isoforms (22). Recombinantly expressed hGNE1 contained the highest epimerase activity (~1100 mU/mg) and existed in the tetrameric (full enzyme activity) and dimeric (kinase activity only) states. hGNE2 displayed an 80% reduction in epimerase activity compared to hGNE1 (~200 mU/mg) explained by its existence predominantly in the dimeric state. hGNE3 had no epimerase activity, likely because a part of the epimerase domain encoded by exon 3 is not expressed in this isoform. The oligomeric state of hGNE3 could not be determined due to low expression levels. hGNE-1,-2, and -3 all displayed high kinase activities (ranging from 2500–3500 mU/mg). Enzyme activities for the isoforms hGNE-5,-6,-7, and -8 are not known, but it is likely that they have no epimerase activity, since they lack exon 5, encoding part of the epimerase domain. In fact, our modeling studies of these isoforms, also predict diminished or absent epimerase activity. hGNE-5,-6,-7, and -8 likely have residual kinase activity, since they express the kinase domain encoding C-terminal exons. These results suggest that hGNE1 (with high epimerase activity) has a major role in sialic acid production. The other human isoforms (with primarily kinase activity) may assist in the crucial regulation of cellular sialic acid levels (22).

For molecular modeling all considered human isoforms were analyzed in 3 groups. First, the 31-residue N-terminal extension compared to the hGNE1 isoform, occurring in hGNE2 and hGNE7 was analyzed. This extension produces α -helices which extend the Rossmann fold.

Since similar extended Rossman folds N-terminal of the nucleotide binding domain are found in *S. achanthias* muscle lactate dehydrogenase, *P. aeruginosa* UDP-N-acetylglucosamine 4-epimerase and other proteins, we assume the additional N-terminal fragments of hGNE2 and hGNE7 do not hinder the active sites of the GNE enzyme and will likely not eliminate epimerase activity. In fact, this was confirmed by enzyme activity measurements in recombinantly expressed hGNE2, which displayed 80% epimerase activity and similar ManNAc kinase activity of that of GNE1 (22). The 31-residue extension at N-termini of hGNE2 and hGNE7 showed amino acid similarities with ATP-binding fragments of some kinases and helicases (Table 2). Based on homology to SGK1 and other kinases, the cysteine C12 residue in the 31-residue extension may be a candidate for disulfide bond formation. It was previously suggested that the 31-residue extension in GNE2 may have a regulatory role in fine-tuning sialic acid synthesis (21, 22).

Second, we considered the 55 residue N-terminal deletion, and a novel 50-residue extension in the hGNE3 and hGNE8 isoforms. Interestingly, the size and secondary structure (alternating α -helices and β -strands; Figure 3) of the deleted and extension fragments are similar, indicating that the extension may serve as a replacement for the deletion. In addition, the novel 50-residue extension has similarity to other proteins containing the Rossmann fold such as deaminases, glycosyltransferases, α -1,2-mannosidase, ATPase, and guardian (Table 3). These findings indicate that the N-terminal modification of hGNE3 and hGNE8 may not affect the overall fold of the protein, compared to hGNE1. However, epimerase function of the hGNE3 and hGNE8 isoforms is predicted to be diminished or absent, since the deleted 55 residue fragment contains several amino acids important for enzymatic function, including substrate binding. In homologous bacterial epimerase enzymes, this region is part of an active and allosteric site. In addition, it was shown that deletion of the N-terminal 39 amino acids of hGNE1 leads to complete loss of epimerase activity with 22% of the kinase activity retained (32). Indeed, activity measurements of the recombinant hGNE3 enzyme confirmed loss of epimerase enzymatic activity while ManNAc kinase activity was retained, likely due to presence of the unaffected kinase domain (21, 22).

Lastly, in addition to a N-terminal modification, a 53-residue deletion occurring in hGNE5, hGNE6, hGNE7, and hGNE8 was considered. The deleted region contains a predicted catalytic residue and is assigned a role in the binding of UDP-GlcNAc and stabilization of the UDP intermediate (displayed in Figure 5A). Deletion of this fragment likely results in abolished epimerase activity, while ManNAc kinase activity may remain present. No enzymatic studies have been performed for these isoforms. The hGNE5 isoform contains the above mentioned 53-residue deletion (encoded by exon 5) as well as a 110-residue N-terminal deletion (and no extension) compared to hGNE1. Therefore hGNE5 is also predicted to have abolished epimerase activity with residual kinase activity, which remains to be confirmed by enzymatic activity measurements.

We previously established that the major GNE1 isoform consists of 7-stranded parallel β -sheets are sandwiched between a total of 7 α -helices in the N-terminal domain and 6-stranded β -sheet surrounded by a total of 7 α -helices in the C-terminal domain (11). The epimerase encoding region of GNE consists of two domains, domain I (N-terminal) and domain II (C-terminal) (Figure 4). These two domains show structural similarities and both have a topology similar to the Rossman nucleotide binding fold. Interestingly, the two major deleted fragments in the hGNE isoforms, one 55-residue deletion at the N-terminus of the domain I in the hGNE3 and hGNE8 isoforms, and the other a 53-residue deletion at the N-terminus of the domain II in the hGNE5, hGNE6, hGNE7 and hGNE8 isoforms (Figure 2, Figure 3), also show structural similarities between each other. Both deleted fragments consist of $\beta\alpha\beta\alpha$ secondary structure elements at their N-termini, and both contain similarly

located conserved active site residue(s) and non-conserved bacterial allosteric site amino acid residues. It is possible that evolutionary involvement of the different hGNE isoforms illustrates a process of the elimination of the bacterial allosteric site from the enzyme, as well as further evolution of the mammalian bifunctional enzyme from two separate bacterial epimerase and kinase enzymes.

Evolution displays several mechanisms by which the number of cellular functions can be increased, without increasing the number of genes (38). These include expression of multiple mRNAs from one gene (by alternative mRNA transcription or alternative splicing), multiple translation products from one gene (by translation initiation from alternative in-frame start codons), tissue-specific mRNA or protein expression, protein targeting to different cellular compartments, post-translational protein modification (through phosphorylation, glycosylation, etc.), and oligomer formation, among others. It appears as if the *GNE* gene is an excellent example of being subject to some of the above mechanisms. First, the *GNE* gene displays an increased number of mRNA transcripts through evolution; while the mouse *Gne* gene only has 2 mRNA transcripts, the human *GNE* gene produces at least 7 different transcripts (formed by either alternative splicing or alternative transcription, or both). Second, *GNE* mRNA transcripts are tissue-specific expressed, possibly due to tissue-specific demands for sialylation or other regulatory processes independent of sialic acid production, such as sialyltransferase expression, ganglioside production and modulation of proliferation and apoptosis (39). Third, the different hGNE isoforms (if translated) may have different localization within the cell. hGNE1 can be localized to the nucleus, cytoplasm, or the Golgi-region, perhaps targeted by a putative leucine rich nuclear export signal, located within amino acids 121–140 of hGNE1 (40). This region is highly conserved and is not deleted or modified in any of the other human isoforms. Fourth, hGNE has several potential phosphorylation sites that may affect function (41). Some predicted phosphorylation sites (by NetPhos 2.0) are deleted (such as Thr34, Tyr22, Tyr54 of hGNE1 are deleted in hGNE3, hGNE5 and hGNE8) or added (such as Ser11 and Tyr4 of hGNE2 and hGNE7 are not present in hGNE1) in the different hGNE isoforms compared to hGNE1. And finally, GNE exists in two major oligomeric states, tetramers and dimers, which engage in dynamic interplay with monomers and higher aggregates. As a monomer, GNE has no enzymatic activity, its dimer exhibits only ManNAc kinase activity, and the tetrameric state displays both UDP-GlcNAc 2-epimerase and ManNAc kinase activities (42). In addition, the fully functional tetrameric state of GNE is stabilized by ligands of the epimerase domain, UDP-GlcNAc and CMP-Neu5Ac. Our molecular modeling studies may contribute to predictions of isoform involvement in oligomer formation and enzyme activity of oligomeric complexes.

In summary, our human GNE isoform studies revealed that hGNE1 exist as ubiquitously expressed, major isoform, while the hGNE2-8 isoforms are differentially expressed and may act as tissue-specific regulators of sialylation. Genetic analysis of the novel coding regions of hGNE exons 1–3 are warranted in patients with HIBM\DMRV, especially those in which only one *GNE* mutation could be identified. These expression and structural prediction data may contribute to elucidation of the complex functional and regulatory mechanisms of GNE, as well as to further elucidation of the pathology and treatment strategies of the human GNE-opathies sialuria and HIBM\DMRV, as well as of other disorders of hyposialylation.

Table S1 shows primer sequences for PCR amplification of the human *GNE* transcripts (*hGNE*). Table S2 shows TaqMan primer and probe sequences for qPCR to determine expression of *hGNE* transcripts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We greatly appreciate the expert laboratory work of Katherine Berger and Adrian Astiz- Martinez.

References

1. Hinderlich S, Stasche R, Zeitler R, Reutter W. A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Purification and characterization of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. *J Biol Chem.* 1997; 272:24313–24318. [PubMed: 9305887]
2. Eisenberg I, Avidan N, Potikha T, Hochner H, Chen M, Olender T, Barash M, Shemesh M, Sadeh M, Grabov-Nardini G, Shmilevich I, Friedmann A, Karpati G, Bradley WG, Baumbach L, Lancet D, Asher EB, Beckmann JS, Argov Z, Mitrani-Rosenbaum S. The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy. *Nat Genet.* 2001; 29:83–87. [PubMed: 11528398]
3. Chen X, Varki A. Advances in the biology and chemistry of sialic acids. *ACS Chem Biol.* 2010; 5:163–176. [PubMed: 20020717]
4. Varki NM, Varki A. Diversity in cell surface sialic acid presentations: implications for biology and disease. *Lab Invest.* 2007; 87:851–857. [PubMed: 17632542]
5. Kornfeld S, Kornfeld R, Neufeld EF, O'Brien PJ. The Feedback Control of Sugar Nucleotide Biosynthesis in Liver. *Proc Natl Acad Sci U S A.* 1964; 52:371–379. [PubMed: 14206604]
6. Seppala R, Lehto VP, Gahl WA. Mutations in the human UDP-N-acetylglucosamine 2-epimerase gene define the disease sialuria and the allosteric site of the enzyme. *Am J Hum Genet.* 1999; 64:1563–1569. [PubMed: 10330343]
7. Leroy JG, Seppala R, Huizing M, Dacremont G, De Simpel H, Van Coster RN, Orvisky E, Krasnewich DM, Gahl WA. Dominant inheritance of sialuria, an inborn error of feedback inhibition. *Am J Hum Genet.* 2001; 68:1419–1427. [PubMed: 11326336]
8. Huizing M, Krasnewich DM. Hereditary inclusion body myopathy: a decade of progress. *Biochim Biophys Acta.* 2009; 1792:881–887. [PubMed: 19596068]
9. Noguchi S, Keira Y, Murayama K, Ogawa M, Fujita M, Kawahara G, Oya Y, Imazawa M, Goto Y, Hayashi YK, Nonaka I, Nishino I. Reduction of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase activity and sialylation in distal myopathy with rimmed vacuoles. *J Biol Chem.* 2004; 279:11402–11407. [PubMed: 14707127]
10. Effertz K, Hinderlich S, Reutter W. Selective loss of either the epimerase or kinase activity of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase due to site-directed mutagenesis based on sequence alignments. *J Biol Chem.* 1999; 274:28771–28778. [PubMed: 10497249]
11. Kurochkina N, Yardeni T, Huizing M. Molecular modeling of the bifunctional enzyme UDP-GlcNAc 2-epimerase/ManNAc kinase and predictions of structural effects of mutations associated with HIBM and sialuria. *Glycobiology.* 2010; 20:322–337. [PubMed: 19917666]
12. Velloso LM, Bhaskaran SS, Schuch R, Fischetti VA, Stebbins CE. A structural basis for the allosteric regulation of non-hydrolysing UDP-GlcNAc 2-epimerases. *EMBO Rep.* 2008; 9:199–205. [PubMed: 18188181]
13. Rao ST, Rossmann MG. Comparison of super-secondary structures in proteins. *J Mol Biol.* 1973; 76:241–256. [PubMed: 4737475]
14. Ha S, Walker D, Shi Y, Walker S. The 1.9 Å crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.* 2000; 9:1045–1052. [PubMed: 10892798]

15. Campbell RE, Mosimann SC, Tanner ME, Strynadka NC. The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases. *Biochemistry*. 2000; 39:14993–15001. [PubMed: 11106477]
16. Tong Y, Tempel W, Nedyalkova L, Mackenzie F, Park HW. Crystal structure of the N-acetylmannosamine kinase domain of GNE. *PLoS One*. 2009; 4:e7165. [PubMed: 19841673]
17. Nishimasu H, Fushinobu S, Shoun H, Wakagi T. Crystal structures of an ATP-dependent hexokinase with broad substrate specificity from the hyperthermophilic archaeon *Sulfolobus tokodaii*. *J Biol Chem*. 2007; 282:9923–9931. [PubMed: 17229727]
18. Bork P, Sander C, Valencia A. An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc Natl Acad Sci U S A*. 1992; 89:7290–7294. [PubMed: 1323828]
19. Mukai T, Kawai S, Mori S, Mikami B, Murata K. Crystal structure of bacterial inorganic polyphosphate/ATP-glucosaminokinase. Insights into kinase evolution. *J Biol Chem*. 2004; 279:50591–50600. [PubMed: 15377666]
20. Watts GD, Thorne M, Kovach MJ, Pestronk A, Kimonis VE. Clinical and genetic heterogeneity in chromosome 9p associated hereditary inclusion body myopathy: exclusion of GNE and three other candidate genes. *Neuromuscul Disord*. 2003; 13:559–567. [PubMed: 12921793]
21. Reinke SO, Hinderlich S. Prediction of three different isoforms of the human UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. *FEBS Lett*. 2007; 581:3327–3331. [PubMed: 17597614]
22. Reinke SO, Eidenschink C, Jay CM, Hinderlich S. Biochemical characterization of human and murine isoforms of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase (GNE). *Glycoconj J*. 2009; 26:415–422. [PubMed: 18815882]
23. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol*. 1977; 112:535–542. [PubMed: 875032]
24. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008; 3:1101–1108. [PubMed: 18546601]
25. Garnier J, Osguthorpe DJ, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol*. 1978; 120:97–120. [PubMed: 642007]
26. Kloczkowski A, Ting KL, Jernigan RL, Garnier J. Combining the GOR V algorithm with evolutionary information for protein secondary structure prediction from amino acid sequence. *Proteins*. 2002; 49:154–166. [PubMed: 12210997]
27. Cheng H, Sen TZ, Kloczkowski A, Margaritis D, Jernigan RL. Prediction of protein secondary structure by mining structural fragment database. *Polymer (Guildf)*. 2005; 46:4314–4321. [PubMed: 19081746]
28. Sen TZ, Cheng H, Kloczkowski A, Jernigan RL. A Consensus Data Mining secondary structure prediction by combining GOR V and Fragment Database Mining. *Protein Sci*. 2006; 15:2499–2506. [PubMed: 17001039]
29. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT. Protein structure prediction servers at University College London. *Nucleic Acids Res*. 2005; 33:W36–38. [PubMed: 15980489]
30. Eisenberg I, Grabov-Nardini G, Hochner H, Korner M, Sadeh M, Bertorini T, Bushby K, Castellan C, Felice K, Mendell J, Merlini L, Shilling C, Wirguin I, Argov Z, Mitrani-Rosenbaum S. Mutations spectrum of GNE in hereditary inclusion body myopathy sparing the quadriceps. *Hum Mutat*. 2003; 21:99. [PubMed: 12497639]
31. Lucka L, Krause M, Danker K, Reutter W, Horstkorte R. Primary structure and expression analysis of human UDP-N-acetyl-glucosamine-2-epimerase/N-acetylmannosamine kinase, the bifunctional enzyme in neuraminic acid biosynthesis. *FEBS Lett*. 1999; 454:341–344. [PubMed: 10431835]
32. Blume A, Weidemann W, Stelzl U, Wanker EE, Lucka L, Donner P, Reutter W, Horstkorte R, Hinderlich S. Domain-specific characteristics of the bifunctional key enzyme of sialic acid biosynthesis, UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. *Biochem J*. 2004; 384:599–607. [PubMed: 15330759]

33. Abad-Zapatero C, Griffith JP, Sussman JL, Rossmann MG. Refined crystal structure of dogfish M4 apo-lactate dehydrogenase. *J Mol Biol.* 1987; 198:445–467. [PubMed: 3430615]
34. Ishiyama N, Creuzenet C, Lam JS, Berghuis AM. Crystal structure of WbpP, a genuine UDP-N-acetylglucosamine 4-epimerase from *Pseudomonas aeruginosa*: substrate specificity in udp-hexose 4-epimerases. *J Biol Chem.* 2004; 279:22635–22642. [PubMed: 15016816]
35. Zhao B, Lehr R, Smallwood AM, Ho TF, Maley K, Randall T, Head MS, Koretke KK, Schnackenberg CG. Crystal structure of the kinase domain of serum and glucocorticoid-regulated kinase 1 in complex with AMP PNP. *Protein Sci.* 2007; 16:2761–2769. [PubMed: 17965184]
36. Xie K, Sowden MP, Dance GS, Torelli AT, Smith HC, Wedekind JE. The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. *Proc Natl Acad Sci U S A.* 2004; 101:8114–8119. [PubMed: 15148397]
37. Teh AH, Kimura M, Yamamoto M, Tanaka N, Yamaguchi I, Kumasaka T. The 1.48 Å resolution crystal structure of the homotetrameric cytidine deaminase from mouse. *Biochemistry.* 2006; 45:7825–7833. [PubMed: 16784234]
38. Yogev O, Pines O. Dual targeting of mitochondrial proteins: mechanism, regulation and function. *Biochim Biophys Acta.* 2011; 1808:1012–1020. [PubMed: 20637721]
39. Wang Z, Sun Z, Li AV, Yarema KJ. Roles for UDP-GlcNAc 2- epimerase/ManNAc 6-kinase outside of sialic acid biosynthesis: modulation of sialyltransferase and BiP expression, GM3 and GD3 biosynthesis, proliferation, and apoptosis, and ERK1/2 phosphorylation. *J Biol Chem.* 2006; 281:27016–27028. [PubMed: 16847058]
40. Krause S, Hinderlich S, Amsili S, Horstkorte R, Wiendl H, Argov Z, Mitrani- Rosenbaum S, Lochmuller H. Localization of UDP-GlcNAc 2- epimerase/ManAc kinase (GNE) in the Golgi complex and the nucleus of mammalian cells. *Exp Cell Res.* 2005; 304:365–379. [PubMed: 15748884]
41. Horstkorte R, Nohring S, Danker K, Effertz K, Reutter W, Lucka L. Protein kinase C phosphorylates and regulates UDP-N-acetylglucosamine-2-epimerase/Nacetylmannosamine kinase. *FEBS Lett.* 2000; 470:315–318. [PubMed: 10745088]
42. Ghaderi D, Strauss HM, Reinke S, Cirak S, Reutter W, Lucka L, Hinderlich S. Evidence for dynamic interplay of different oligomeric states of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase by biophysical methods. *J Mol Biol.* 2007; 369:746–758. [PubMed: 17448495]

Figure 1.

The eight predicted human *GNE* transcripts.

Exon (boxes)-intron (lines) structures of the eight predicted human *GNE* mRNA transcripts encoding 8 predicted *GNE* isoforms (hGNE1-hGNE8). GenBank Accession numbers (when available) and predicted number of translated amino acids (aa) are indicated. Black boxes represent the open reading frame, gray boxes the untranslated mRNA regions, and white dotted lined boxes the skipped exons of each transcript. Primer locations for transcript-specific PCR amplifications (Figure 2A) are indicated as arrowed lines below each isoform. Note that the forward primers F2 and F3 are localized across exon-exon boundaries.

Figure 2.

Tissue expression of the eight human *GNE* transcripts.

(A) PCR amplification of tissue-specific cDNA from selected human tissues. Primers used for amplification and amplified isoform-transcripts are indicated.

(B) qPCR results of isoform-specific *hGNE* mRNA expression in a variety of human tissues. Transcript-specific Taqman primer-probe assays were used (see Supporting Information Table S2), as indicated above each panel. Displayed values represent the relative quantification (RQ) normalized to *GAPDH* with expression in colon set to 1 for each Taqman probe. Total *hGNE* (upper left panel) illustrates all *hGNE* isoforms combined. (The Taqman probe was located in the *hGNE* C-terminal part, common to all isoforms). Results are summarized in Table 1.

A.

N-Terminal region of domain I of the hGNE epimerase domain, $\beta 1\alpha 1\beta 2\alpha 2$:

```

hGNE1      MEKNGNNRRLRVVATCNADYSKLAPIMFGIKTEPEFFELDVVVLGSMIDDYGNTYRMIEQDDFDI
hGNE2      METYGYLQRESCFQGPHELYFKNLSKRKNQIMEKNGNNRRLRVVATCNADYSKLAPIMFGIKTEPEFFELDVVVLGSMIDDYGNTYRMIEQDDFDI
hGNE3      MPIGDCSVAAPRKQLLCSLFQTTLGYPARASGWKPMVICRGSHAFKDLIMEKNGNNRRLRVVATCNADYSKLAPIMFGIKTEPEFFELDVVVLGSMIDDYGNTYRMIEQDDFDI
hGNE4      MEKNGNNRRLRVVATCNADYSKLAPIMFGIKTEPEFFELDVVVLGSMIDDYGNTYRMIEQDDFDI
hGNE5      MEKNGNNRRLRVVATCNADYSKLAPIMFGIKTEPEFFELDVVVLGSMIDDYGNTYRMIEQDDFDI
hGNE6      MEKNGNNRRLRVVATCNADYSKLAPIMFGIKTEPEFFELDVVVLGSMIDDYGNTYRMIEQDDFDI
hGNE7      METYGYLQRESCFQGPHELYFKNLSKRKNQIMEKNGNNRRLRVVATCNADYSKLAPIMFGIKTEPEFFELDVVVLGSMIDDYGNTYRMIEQDDFDI
hGNE8      MPIGDCSVAAPRKQLLCSLFQTTLGYPARASGWKPMVICRGSHAFKDLIMEKNGNNRRLRVVATCNADYSKLAPIMFGIKTEPEFFELDVVVLGSMIDDYGNTYRMIEQDDFDI

B.a.      GPVDMTER-LKVTITFGTRPEAKMAPLVLELQKHPEKIESIVVTQAQHEQMLDQVLSIFGITPDFDL
LDH      TALDKDLGLHLSQSEPRSYNKITVVG-----DAVGHADAISVLMKDLADEVALVDVMEKLGEMMDLEHGSFLHT

```

N-terminal region of domain II of the hGNE epimerase domain, $\beta 8\alpha 9\beta 9\beta 10$:

```

hGNE1      NSIIRNWLGDVDSKDYIVALQHPVTDDIKHSIKM-FELTLDALISFNRKTLVLFP---NIDAGSKENVRYNRK
hGNE2      NSIIRNWLGDVDSKDYIVALQHPVTDDIKHSIKM-FELTLDALISFNRKTLVLFP---NIDAGSKENVRYNRK
hGNE3      NSIIRNWLGDVDSKDYIVALQHPVTDDIKHSIKM-FELTLDALISFNRKTLVLFP---NIDAGSKENVRYNRK
hGNE4      NSIIRNWLGDVDSKDYIVALQHPVTDDIKHSIKM-FELTLDALISFNRKTLVLFP---NIDAGSKENVRYNRK
hGNE5      NSIIRNWLGDVDSKDYIVALQHPVTDDIKHSIKM-FELTLDALISFNRKTLVLFP---NIDAGSKENVRYNRK
hGNE6      NSIIRNWLGDVDSKDYIVALQHPVTDDIKHSIKM-FELTLDALISFNRKTLVLFP---NIDAGSKENVRYNRK
hGNE7      NSIIRNWLGDVDSKDYIVALQHPVTDDIKHSIKM-FELTLDALISFNRKTLVLFP---NIDAGSKENVRYNRK
hGNE8      NSIIRNWLGDVDSKDYIVALQHPVTDDIKHSIKM-FELTLDALISFNRKTLVLFP---NIDAGSKENVRYNRK

B.a.      SHPVLEK-----LGNRLVLNTAHGRENLEGPMMNFRAIKRLVDKH--EDVQVQVYVHMNPVVRETAMDILG

```

B.

GNE2 and GNE7 (N-terminal extension):

```

          ETYGYLQRESCFQGPHELYFKNLSKRKNQIMEKNGNNR
PSIPRED   CCCCCCCCCCCCCCHHHHEHHCCCHHHHHHHHCCCCC
GOR V     CCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHCCCCC
FDM       CHHHCCCCCCCCCCCCCHHHHCCCHHHHHHHHHCCCCC
CDM       CHHHCCCCCCCCCCCCCHHHHCCCHHHHHHHHHCCCCC

```

GNE3 and GNE8 (N-terminal extension):

```

          MPIGDCSVAAPRKQLLCSLFQTTLGYPARASGWKPMVICRGSHAFKDLINTYRMIEQ
PSIPRED   CCCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCEEEEEECHHHHHHHHHHHCCCCC
GOR V     CCCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCCCCCEEECCCCCHHHHHHHHHHHCC
FDM       CCEEECCCCCCCCCHHHHHHHHCCCCCCCCCCCCCCCCCEEECCCCCCCCCCCCCHHHHHHH
CDM       CCEEECCCCCCCCCHHHHHHHHCCCCCCCCCCCCCCCCCEEECCCCCCCCCCCCCHHHHHHH

```

Figure 3.

Sequence alignments of the human GNE isoforms.

(A) Sequence alignments of the hGNE1-hGNE8 isoforms, *B. anthracis* 2-epimerase (B.a.), and *S. achantias* lactate dehydrogenase (LDH). Two fragments are shown: the N-terminal region of domain I of the hGNE epimerase domain, $\beta 1\alpha 1\beta 2\alpha 2$ (top), and the N-terminal region of domain II of the hGNE epimerase domain, $\beta 8\alpha 9\beta 9\beta 10$ (bottom). Strikethrough: amino acids deleted compared to hGNE1 isoform; green: α -helices; blue: β -sheets; bold: amino acids that form an active site in the *B. anthracis* 2-epimerase and a putative active site in the human GNE isoforms; #: Allosteric site in *B. anthracis* 2-epimerase; °: glucose binding residues in *B. anthracis* 2-epimerase.

(B) Secondary structure predictions with 4 prediction methods (PSIPRED, GOR V, FDM, and CDM) of the N-terminal 31-residue extension of the hGNE2 and hGNE7 isoforms and the 50-residue extension of the hGNE3 and hGNE8 isoforms. C, coil; H, α -helices; E, β -strands; black, novel N-terminal extensions; red, overlap with hGNE1 amino acids.

Figure 4.

Secondary structure elements of the epimerase domain of the hGNE enzyme. Secondary structure elements, including domain, α -helices and β -sheets structures (11), of the hGNE1 isoform are shown. Elements deleted in domain I in the hGNE3, hGNE5, and hGNE8 isoforms (light grey shaded) and in domain II in the hGNE5, hGNE6, hGNE7 and hGNE8 isoforms (dark grey shaded) are indicated. Numbers relate to hGNE1 amino acid numbering. N, N-terminus; C, C-terminus of the GNE epimerase domain.

Figure 5.

Tertiary structures of the major hGNE1 isoform and the active site of the GNE epimerase domain.

(A) Model of the hGNE1 isoform. α -helices (green) and β -sheets (blue) are indicated. Fragments deleted from domain I (right) in hGNE3, hGNE5 or hGNE8 and fragments deleted from domain II (left) in hGNE5, hGNE6, hGNE7, and hGNE8 are indicated in gray. In bacterial 2-epimerases these deleted fragments contain an allosteric site where UDP-GlcNAc binds (shown as balls) and an active site where the UDP substrate binds (shown as sticks).

(B) Superposition of the three-dimensional structure of the *E. coli* 2-epimerase in complex with UDP-*N*-acetylglucosamine (pdb code 1vgv) and the structure of the *H. sapiens* 2-epimerase domain of the hGNE1 isoform. Side chains of the active site residues (wire models), binding site residues of UDP-*N*-acetylglucosamine in *E. coli* 2-epimerase (upper lines) and the corresponding residues in the human 2-epimerase domain of hGNE1 (lower lines) are shown. Residues deleted in hGNE3, hGNE5, and hGNE8 are light gray shaded (hGNE1 Arg19, Asp21, Lys24, and His49). The hGNE1 His220 residue, deleted in hGNE5-hGNE8, is dark gray shaded. Red, oxygen; blue, nitrogen; gray, carbon; yellow, phosphorus.

Table 1

GNE isoform expression in human tissues.^a

	<i>hGNE1</i>	<i>hGNE2</i>	<i>hGNE3</i>	<i>hGNE6</i>	<i>hGNE7</i>	<i>hGNE8</i>
Liver	+	+	+	+	+	+
Spleen	+	-	-	+	-	-
Leukocyte	+	-	-	+	-	-
Lung	+	+	+	+	+	+
Skeletal muscle	+	-	-	+	-	-
Pancreas	+	+	+	+	+	+
Ovary	+	-	-	+	-	-
Heart	+	-	-	+	-	-
Thymus	+	-	-	+	-	-
Colon	+	+	+	+	+	+
Placenta	+	+	+	+	+	+
Small intestine	+	-	-	+	-	-
Prostate	+	-	-	+	-	-
Kidney	+	+	+	+	+	+
Brain	+	-	-	+	-	-
Testis	+	-	-	+	-	-
Salivary gland	+	+	+	+	+	+
Trachea	+	+	+	+	+	+
Adrenal	+	-	-	+	-	-
Thyroid	+	+	+	+	+	+
Uterus	+	+	+	+	+	+
Fibroblast	+	-	-	+	-	-

^a *hGNE4* was not detected in the tested human tissues. No *hGNE5*-specific Taqman assay could be designed. *hGNE5*-specific amplification could only be detected by conventional PCR (with very low expression in selected tissues, see Figure 2A).

+, expressed; -, not expressed; light gray highlighted, tissues that express all tested *hGNE* isoform transcripts.

Table 2

Amino acid similarities between the N-terminal 31 amino acid extension of the hGNE2 and hGNE7 isoforms and other homologous proteins

Protein/source/pdb code/	aa # ^a	Sequence alignment ^b
hGNE2 + hGNE7 extension	1-31	ME <u>TYGYLQ</u> RESCFQGPHELYFKNLSKRNKQI
SGK1/ <i>H. sapiens</i> /2r5t/	183-202	<u>ELFYHLQ</u> RERCFLEPRARFY
His kinase/ <i>A. avenae</i> /	198-213	<u>FLQ</u> RERCFAGDASHEL
His kinase/ <i>P. carotovorum</i> /	194-219	DE <u>LERFLQ</u> RERCFVSDASHELRTPLA
His kinase/ <i>H. arsenicoxydans</i> /	193-216	AELQOFLARERFFTGDVSHELRTP
YFHH/ <i>B. subtilis</i> /1sf9/	8-31	SSGVDLGT----ENLYFQSNAMKRYSQ
Leucyl-tRNA synthetase/ <i>T. thermophilus</i> /1obh /	703-733	LE <u>TS</u> GVFQAEALEGKDRELYGKLHETLK <u>KVT</u>
Ubiquitin thioesterase/ <i>H. sapiens</i> /2zfy/	177-207	LL <u>TS</u> GYLQRESKFFEHFIEGGRTVKEFCQ <u>QE</u>
Interleukin-5/ <i>H. sapiens</i> /1hul/	49-79	FQIGITLESQTVQGGTVRLFKNLSLIK <u>KYI</u>
Cell division control protein/ <i>P. falciparum</i> /1v0p/	238-268	FTVYEPLPWESFLKGLDESGIDLLSKML <u>KLD</u>
Ras-related C3 botulinum toxin/ <i>H. sapiens</i> /1ryf/	62-92	QEDYDRLRPLSYPTVGETY <u>GK</u> ITSRG <u>KDK</u>
Cdc42-interacting protein/ <i>H. sapiens</i> /2efk/	71-101	QSFVQILQEVNDEAGQRELVAENLSVRVCLE
Hypothetical protein HP0495/ <i>H. pylori</i> /2h9z/	25-55	KD <u>IS</u> TLKELLETYQRPFKLEFKNTSKNAKFY
Ecor124i restriction enzyme/ <i>E. coli</i> /2w00/	192-222	NSLEFKYLQLFVISNGTDTRYFANTTKRDKNS
ABC transporter/ <i>P. abyssi</i> /3bk7/	309-339	EELQGYLKDENVRFRPYEIRFTKLSE <u>RDVE</u>

^a aa #: amino acid number.

^b Shaded: homologous fragments; double underlined: α -helices; underlined: β -sheets.

Table 3

Amino acid similarities between the N-terminal 50 amino acid extension of the hGNE3 and hGNE8 isoforms and other homologous proteins

Protein Source /pdb code/	aa # ^a	Sequence alignment ^b
hGNE3 + hGNE8 extension	1-50	MPIGDCSVAAKPRKQLLCSLFQTTLGYRARASGWKPMVICRGSHAFKDLI
Deaminase CDD1 <i>S. cerevisiae</i> /1r5t/	77-89	SGWKCMVICGDS
<i>M. musculus</i> /2fr5/	81-91	D-FRAIATSSDL
Glycosyl transferase <i>A. orientalis</i> /1rrv/	346-394	IGVAHDGPTETTESLSAALTIVLATRARAEAVACMVLTDGAAAADLV
Alpha-1,2-mannosidase <i>T. reesei</i> /1hcu/	88-103	TTNQGSSEETNIRY
<i>P. citrinum</i> /2r18/	1113-1128	KTSDTVSLEETIRY
Alkylsulfatase <i>P. aeruginosa</i> /2cg3/	496-506	LGYQAENAGWR
ATP12 ATPase <i>P. denitrificans</i> /2zd2/	1114-1124	RAQAEGWDELI
Polycomb protein <i>H. sapiens</i> /2p0k/	112-142	PIGNCEKNG-----GMLQPPLGFRLNASSW-FMFLK
Alpha-14 giardin <i>E. coli</i> /3chj/	80-92	KPRAQLLELIR
Histidine-protein kinase <i>E. coli</i> /3iqi/	865-700	GSCGYSGVPRXKNLCOLIEQQLRSGTKEEDLEELL
Galactoside O-acetyl- transferase <i>S. aureus</i> /3ftu/	42-58	ATNKRKELIDQLFOTT

^a aa #: amino acid number;

^b Shaded: homologous fragments; double underlined: α -helices; underlined: β -sheets.