

# *N*-glycolyl groups of nonhuman chondroitin sulfates survive in ancient fossils

Anne K. Bergfeld<sup>a,b</sup>, Roger Lawrence<sup>a,b</sup>, Sandra L. Diaz<sup>a,b</sup>, Oliver M. T. Pearce<sup>a,b</sup>, Darius Ghaderi<sup>a,b</sup>, Pascal Gagneux<sup>a,b,c</sup>, Meave G. Leakey<sup>d,e</sup>, and Ajit Varki<sup>a,b,1</sup>

<sup>a</sup>Glycobiology Research and Training Center, University of California, San Diego, La Jolla, CA 92093; <sup>b</sup>Center for Academic Research and Training in Anthropogeny, University of California, San Diego, La Jolla, CA 92093; <sup>c</sup>Department of Pathology, University of California, San Diego, La Jolla, CA 92093; <sup>d</sup>Turkana Basin Institute, 00502 Nairobi, Kenya; and <sup>e</sup>Department of Anthropology, Stony Brook University, Stony Brook, NY 11794

Edited by Chi-Huey Wong, Academia Sinica, Taipei, Taiwan, and approved July 25, 2017 (received for review April 19, 2017)

**Biosynthesis of the common mammalian sialic acid *N*-glycolylneuraminic acid (Neu5Gc) was lost during human evolution due to inactivation of the *CMAH* gene, possibly expediting divergence of the *Homo* lineage, due to a partial fertility barrier. Neu5Gc catabolism generates *N*-glycolylhexosamines, which are potential precursors for glycoconjugate biosynthesis. We carried out metabolic labeling experiments and studies of mice with human-like Neu5Gc deficiency to show that Neu5Gc degradation is the metabolic source of UDP-GlcNGc and UDP-GalNGc and the latter allows an unexpectedly selective incorporation of *N*-glycolyl groups into chondroitin sulfate (CS) over other potential glycoconjugate products. Partially *N*-glycolylated-CS was chemically synthesized as a standard for mass spectrometry to confirm its natural occurrence. Much lower amounts of GalNGc in human CS can apparently be derived from Neu5Gc-containing foods, a finding confirmed by feeding Neu5Gc-rich chow to human-like Neu5Gc-deficient mice. Unlike the case with Neu5Gc, *N*-glycolyl-CS was also stable enough to be detectable in animal fossils as old as 4 My. This work opens the door for investigating the biological and immunological significance of this glycosaminoglycan modification and for an “ancient glycans” approach to dating of Neu5Gc loss during the evolution of *Homo*.**

glycobiology | sialic acid | evolution

All vertebrate cells are covered with a complex array of glycoconjugates, with sialic acids (Sias) typically occupying terminal positions of many glycan chains (1). The two most common Sias in most mammals are *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc), which differ by a single oxygen atom (1). The only known metabolic pathway for Neu5Gc biosynthesis is hydroxylation of its precursor Neu5Ac, catalyzed by the enzyme CMP-Neu5Ac hydroxylase (*CMAH*) (2–4). A loss-of-function mutation in the corresponding single-copy *CMAH* gene is fixed in humans, who can no longer biosynthesize Neu5Gc (3, 5, 6). Multiple methods of genomic analysis estimated that the loss of *CMAH* function occurred ~2–3 Mya in the hominin lineage (7, 8).

There are many known and possible biological consequences of human Neu5Gc loss, discussed in detail elsewhere (9). Following loss of Neu5Gc biosynthesis, the human immune system also recognizes Neu5Gc-bearing glycans as foreign and antigenic molecules (10, 11). In human-like *Cmah*<sup>-/-</sup> Neu5Gc-deficient mice, the resulting anti-Neu5Gc antibodies could mediate female intrauterine immune selection against sperm from *Cmah*-positive WT males (12, 13). Models of frequency-dependent selection regimes showed that such a reproductive incompatibility by female immunity could drive this loss-of-function allele to fixation after it reached moderate initial frequencies. We therefore postulated that fixation of the *CMAH* null state in the hominin lineage leading to our species could have expedited divergence of the genus *Homo* ~2–3 Mya (12). However, there is as yet no direct proof for this hypothesis.

Notably, human cellular pathways still allow metabolic incorporation of scavenged Neu5Gc into endogenous cellular glycans (4, 10, 14). This process can even occur in intact humans,

by metabolically incorporating trace amounts of Neu5Gc from mammal-derived food products, particularly red meats (beef, lamb, and pork). Incorporation of trace amounts of such exogenous Neu5Gc into human cell-surface structures in the face of an anti-Neu5Gc antibody response makes Neu5Gc the first known “xeno-autoantigen” in humans (15–17). However, incorporation levels are far lower than endogenous levels in WT (*Cmah* intact) mice or in chimpanzees, our closest living evolutionary relatives (18).

As an alternative to activation to CMP-Neu5Gc and incorporation into glycoconjugates, Neu5Gc can also follow a degradative metabolic route, resulting in loss of the *N*-glycolyl group and formation of glucosamine 6-phosphate (4). One intermediate on this multistep pathway was found to be *N*-glycolylglucosamine (GlcNGc) (4) (Fig. 1A). Additionally, mammalian cells cultured in synthetic GlcNGc were found to biosynthesize UDP-GlcNGc, which was incorporated as *O*-GlcNGc modifications of proteins (19) and was also detected in *O*-glycans, likely as *N*-glycolylgalactosamine (GalNGc) (20, 21). Existence of HexNGcs in *O*-glycans and *N*-glycans was also suggested after analyzing cells supplemented with synthetic ManNGc (22). In addition, we found that culturing mammalian cells in synthetic GalNGc gave rise to UDP-GalNGc and UDP-GlcNGc, which serve as sugar nucleotide donors for incorporation into cellular *O*-glycans, gangliosides, heparan sulfates

## Significance

**We identified a glycan modification called *N*-glycolylated chondroitin sulfate (Gc-CS), derived from metabolic turnover of the nonhuman sialic acid *N*-glycolylneuraminic acid (Neu5Gc). The presence of Gc-CS could be demonstrated in species rich in Neu5Gc using chemically synthesized Gc-CS as a standard for mass spectrometry. Although humans cannot synthesize Neu5Gc due to a loss-of-function mutation, trace amounts of Gc-CS were found in humans, apparently derived from Neu5Gc-containing foods. Gc-CS was more easily detectable in animal fossils as old as 4 My, allowing indirect fossil evidence of Neu5Gc expression. These findings enable future studies to date the loss of Neu5Gc biosynthesis during human evolution and investigate this glycosaminoglycan modification in humans who consume Neu5Gc-rich foods (red meats).**

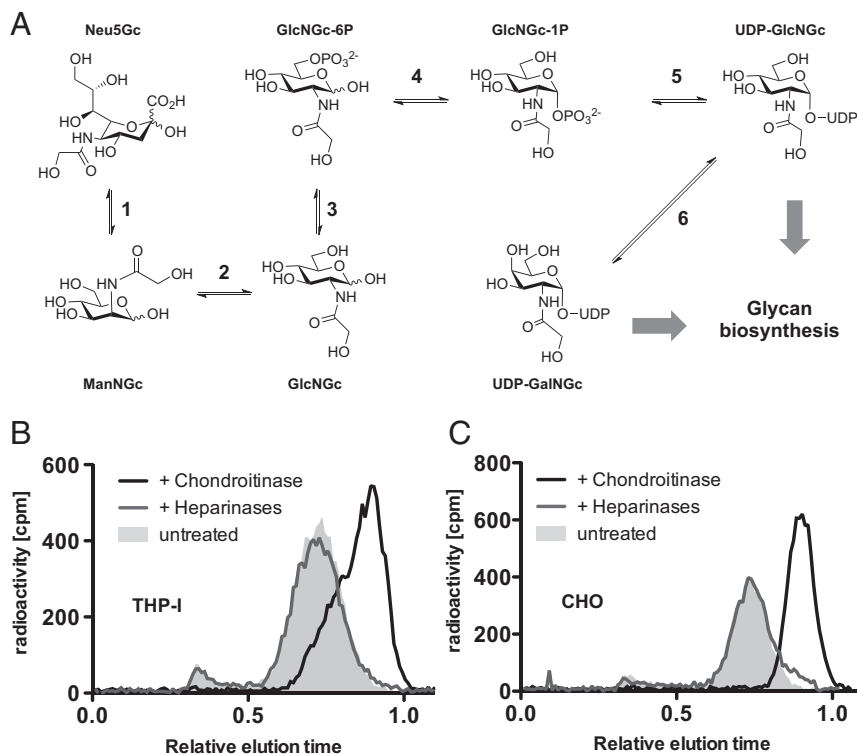
Author contributions: A.K.B., R.L., S.L.D., O.M.T.P., D.G., P.G., M.G.L., and A.V. designed research; A.K.B., R.L., S.L.D., O.M.T.P., and D.G. performed research; P.G. and M.G.L. contributed new reagents/analytic tools; A.K.B. and R.L. analyzed data; and A.K.B. and A.V. wrote the paper.

Conflict of interest statement: A.V. is cofounder of and has equity interest in SiaMab Therapeutics, Inc., which has licensed University of California, San Diego technologies related to Neu5Gc biology. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. However, there are no direct conflicts with the present study.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. Email: avarki@ucsd.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706306114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706306114/-DCSupplemental).



**Fig. 1.** Possible underlying mechanism for the natural occurrence of HexNGc in animal glycoconjugates. (A) The single known source for *N*-glycolyl groups in animals is the conversion of the *N*-acetyl group of CMP-Neu5Ac to an *N*-glycolyl group in CMP-Neu5Gc, which is catalyzed by Cmah (EC 1.14.18.2) (49). Therewith, *N*-glycolylhexosamines are to be Neu5Gc derivatives. Based on the well-studied *N*-acetylhexosamine pathways in animals we suggest a metabolic route to result in glycoconjugates comprising GlcNGc and GalNGc in nature. (i) Conversion of Neu5Gc into ManNGc is catalyzed by the *N*-acetylneuraminase lyase (EC 4.1.3.3) (4, 50, 51). (ii) Epimerization of ManNGc to GlcNGc can be achieved by GlcNAc-2'-epimerase (EC 5.1.3.8) (4). (iii) Phosphorylation of GlcNGc in the 6' position to result in GlcNGc-6P's beings catalyzed by GlcNAc kinase (EC 2.7.1.59) (4). (iv) Conversion of GlcNGc 6-P to GlcNGc 1-P might be catalyzed by GlcNAc 6-P phosphomutase (EC 5.4.2.3) (52). (v) GlcNGc 1-P would thereafter react with UTP to form UDP-GlcNAc, a reaction potentially catalyzed by UDP-*N*-acetylglucosamine diphosphorylase (EC 2.7.7.23) (53). (vi) Epimerization of UDP-GlcNAc to UDP-GalNAc is catalyzed by UDP-GlcNAc 4-epimerase (EC 5.1.3.7) (54). UDP-GlcNAc and UDP-GalNAc then serve as precursors for glycan assembly. (B) Human THP-I cells and (C) CHO LEC29.lec32 cells were cultured in the presence of [<sup>3</sup>H-glycolyl]Neu5Gc. Desalted GAGs were divided into three samples, from which one sample remained untreated (filled gray), one sample was treated with chondroitinase ABC (black line), and the last sample was treated with heparinases (gray line). The disaccharides were separated from the intact GAG chains by gel filtration chromatography.

(HS), and chondroitin sulfates (CS) (23). The latter two glycan classes are major components of extracellular matrix and bone (24). While cellular pathways allowed low-level incorporation of artificially provided GalNGc or GlcNGc into most major glycan classes, we noted that the *N*-glycolyl group was most prominently found in CS (23). Here we explore the possibility that metabolic turnover of naturally occurring and/or food-derived Neu5Gc might result in a so-far-unnoticed subset of cellular glycans that naturally contain HexNGc residues. We also apply these findings to challenging questions regarding indirect fossil evidence of Neu5Gc expression over a time span of millions of years.

## Results

**Possible Metabolic Pathway for Natural Occurrence of *N*-Glycolylhexosamines in Animal Glycans.** Mammalian cells cultured with synthetic *N*-glycolylhexosamines (HexNGc) incorporate GalNGc and GlcNGc into cellular glycans (19, 22, 23). Meanwhile, ManNGc and GlcNGc were found as natural products generated during metabolic degradation of Neu5Gc (4). Taken together, these findings raised the interesting question of whether cellular turnover of Neu5Gc naturally produces HexNGc that are subsequently incorporated into animal glycans. The proposed metabolic pathway involves the conversion of Neu5Gc via ManNGc and GlcNGc toward GlcNGc-6P (Fig. 1A). As an alternative to loss of the *N*-glycolyl group (generating glycolate and glucosamine-6-phosphate) (4), GlcNGc-1P might also be formed. Further metabolism toward

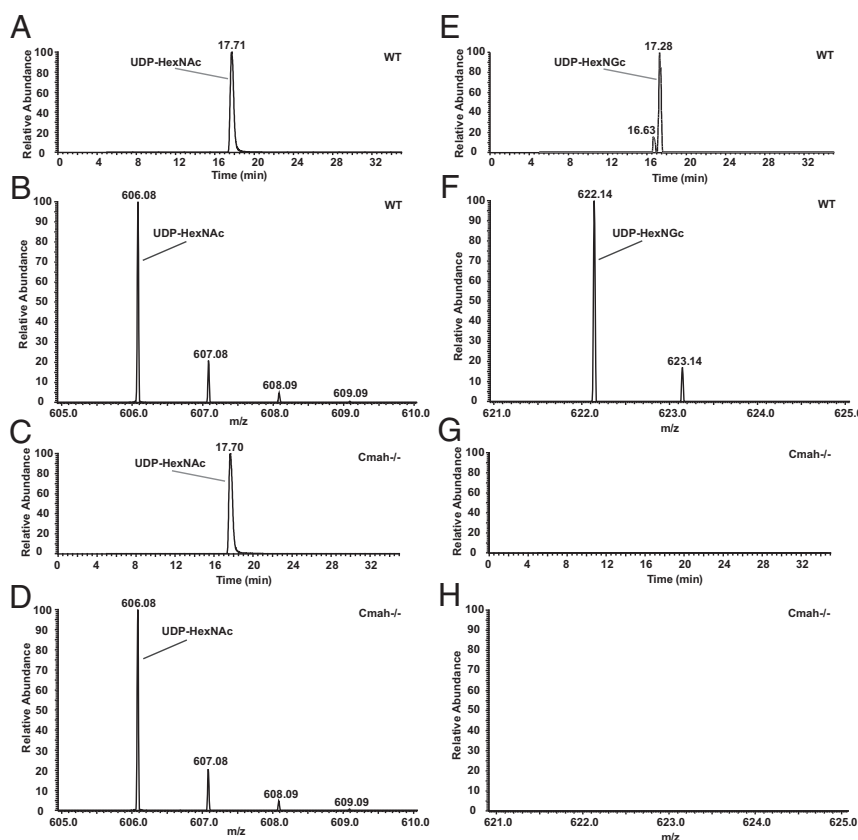
UDP-GlcNGc and subsequent epimerization to UDP-GalNGc could then provide the precursors for glycan assembly (Fig. 1A). To test this hypothesis, human THP-I cells were cultured with [<sup>3</sup>H-glycolyl]Neu5Gc followed by isolation of radiolabeled cellular glycosaminoglycans (GAGs). GAGs were selected because earlier studies indicated that incorporation of synthetic HexNGc in mammalian cells was highest into this class of glycans (23). The GAG fraction contained 2.5% of total cellular radioactivity and was divided into three aliquots, which were treated either with chondroitinase ABC or a mixture of heparin lyases-I/II/III or incubated as a nontreated control. Reaction products were analyzed by size-exclusion chromatography to separate large GAG polymers from disaccharide digestion products. Radioactivity in nontreated GAG chains (shaded gray) eluted early due to their polymeric nature (Fig. 1B). No differences were observed for the heparin lyase-treated sample (gray line), indicating that the glycolyl group derived from [<sup>3</sup>H-glycolyl]Neu5Gc was not incorporated into HS. In contrast, analysis of the chondroitinase-treated sample (black line) shows a peak eluting later, as expected for CS disaccharides (Fig. 1B), suggesting that the *N*-[<sup>3</sup>H]glycolyl group was incorporated as GalNGc into cellular CS. The same experiment was performed with CHO lec32.LEC29 cells, which lack the ability to activate Sias (25) and might therefore show enhanced metabolic flux of supplemented [<sup>3</sup>H-glycolyl]Neu5Gc toward HexNGc. Indeed, 8% of cellular radioactivity was now detected in the GAG fraction. Analysis of isolated GAGs was performed as described for

the human cells (Fig. 1C), and labeled, nontreated polymeric GAGs eluted early as expected (shaded gray). Again, no peak shift was observed after heparin lyase treatment (gray line) and all radioactivity eluted in the disaccharide peak after treatment with chondroitinase. Thus, cells selectively incorporated [ $^3\text{H}$ -glycolyl]GalNGc into CS, a likely location for naturally occurring HexNGc in animal glycans. Notably, despite the fact that UDP-[ $^3\text{H}$ -glycolyl]GalNGc would be the most distal product of [ $^3\text{H}$ -glycolyl]Neu5Gc metabolism (six steps away, see Fig. 1A) it was still preferentially used to synthesize [ $^3\text{H}$ -glycolyl]CS, versus the more proximal product UDP-[ $^3\text{H}$ -glycolyl]GlcNAc, which could have been used for biosynthesis of other glycan classes, such as *N*-glycans or HS.

**Natural Occurrence of UDP-HexNGc in Livers of WT but Not Neu5Gc-Deficient Mice.** To confirm that Neu5Gc metabolism is indeed the source of naturally occurring HexNGc in animal glycans we compared WT to *Cmah*<sup>-/-</sup> mice, which carry the human-like mutation in *Cmah* and are therefore devoid of any detectable Neu5Gc, making them an excellent negative control (4, 26, 27). Incorporation of HexNGc into animal glycans depends on the generation of its activated UDP-HexNGc precursor. Therefore, nucleotide sugars isolated from *Cmah*<sup>-/-</sup> and WT mouse livers were first subjected to liquid chromatography (LC)-MS analyses. In WT mouse liver samples a peak eluting at 17.71 min (Fig. 2A) with an underlying mass of 606 Da (Fig. 2B) represents UDP-HexNAc sugars. As expected, the same peak was found in corresponding *Cmah*<sup>-/-</sup> samples (Fig. 2C and D) at comparable

intensity. Samples from WT mice also showed a peak eluting at 17.28 min (Fig. 2E) with an underlying mass of 622 Da (Fig. 2F), which exactly fits the mass of UDP-HexNGc. Complete absence of this peak in samples from Neu5Gc-deficient *Cmah*<sup>-/-</sup> mice (Fig. 2G and H) demonstrates that UDP-HexNGc is a natural metabolite derived only from Neu5Gc turnover.

**Chemical Synthesis of *N*-Glycolyl-CS.** The above data suggest that CS disaccharides from Neu5Gc-containing animal tissues might contain naturally occurring GalNGc. To definitively identify the predicted CS disaccharides containing GalNGc, a synthetic standard was prepared by partial de-*N*-acetylation and subsequent re-*N*-glycolylation of commercially available CS from shark cartilage. Shark samples do not have endogenous Neu5Gc (16) and, as expected, commercial shark cartilage CS disaccharides showed no signs of *N*-glycolyl groups, with product ion spectra of CS disaccharides D0a4 and D0a6 being consistent with published data (for nomenclature, see the references in refs. 23 and 28). Shark CS was partially de-*N*-acetylated by treatment with hydrazine followed by subsequent re-*N*-glycolylation with acetoxyacetyl chloride (Fig. S1A). The final product was treated with chondroitinase ABC and resulting disaccharides analyzed by glycan reductive isotope labeling (GRIL)-LC/MS as described (23, 28). The extracted ion current is shown for CS disaccharides D0a4 and D0a6 (Fig. S1B) as well as corresponding D0q4 and D0q6 (Fig. S1C) of the partially *N*-glycolylated product. The nomenclature of *N*-glycolylated GAGs has been described previously



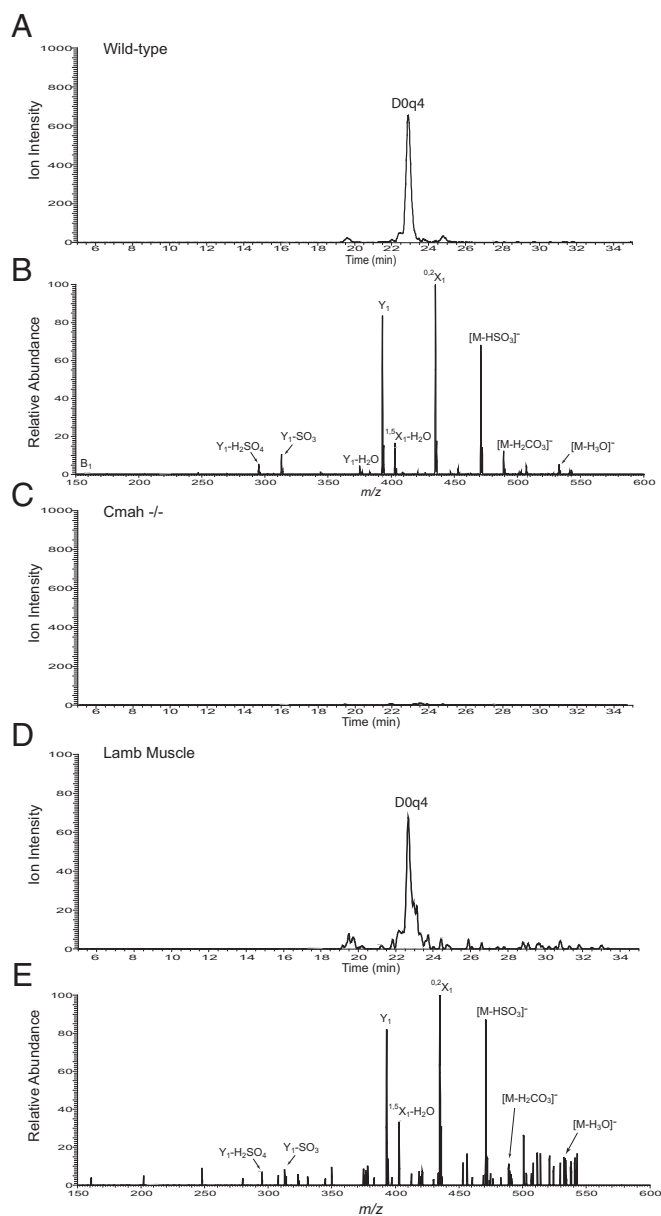
**Fig. 2.** Natural occurrence of UDP-*N*-glycolylhexosamines in mouse liver. Nucleoside diphospho sugars were isolated from deproteinized WT and *Cmah*<sup>-/-</sup> mouse liver homogenates via C-18 and charcoal columns. Samples were further treated with phosphatase following DEAE sephacel cleanup before LC-MS analysis. (A) Chromatogram of WT mouse liver nucleoside diphospho sugars with an underlying mass of UDP-HexNAc (606 Da) and (B) MS analysis of the respective peak fraction. (C) Chromatogram of *Cmah*<sup>-/-</sup> mouse liver nucleoside diphospho sugars with an underlying mass of UDP-HexNAc (606 Da) and (D) MS analysis of the respective peak fraction. (E) Chromatogram of WT mouse liver nucleoside diphospho sugars with an underlying mass of UDP-HexNGc (622 Da) and (F) MS analysis of the respective peak fraction. (G) chromatogram of *Cmah*<sup>-/-</sup> mouse liver nucleoside diphospho sugars with an underlying mass of UDP-HexNGc (622 Da) and (H) MS analysis of the theoretical peak fraction for the elution time at which UDP-HexNGc would elute if present.

(23). The product ion spectra of the *N*-glycosylated species D0q4 and D0q6 in Fig. S1 D and E are consistent with published data (23). The ratio of GalNGc:GalNAc in synthesized, partially *N*-glycosylated CS was estimated to be ~1:10 by MS of disaccharides, assuming similar signal intensities for both substituents. Complete absence of GalNGc in the starting material makes it a suitable negative control, and the ability to generate *N*-glycosylated CS provides not only the required standard but also the future opportunity to investigate the possible impact and function of *N*-glycosyl-CS in mammals.

**Natural Occurrence of GalNGc in Animal CS.** Liver GAGs of WT and *Cmah*<sup>-/-</sup> mice were isolated and depolymerized by chondroitinase ABC. Released disaccharides were analyzed by GRIL-LC/MS and differentially stable isotope-labeled *N*-acetylated CS-disaccharides were added as internal standards to all samples (23, 28). Besides the expected *N*-acetylated CS-disaccharides, samples from WT mouse liver also contained a peak at 22.9 min (35 min run), which fits the *N*-glycosylated CS-disaccharide D0q4 [*m/z* = 435 (551 with the aniline tag), Fig. 3A]. MS-MS analysis of the respective ion by collision-induced dissociation (CID) fits the expected fragmentation pattern for GalNGc-containing D0q4 disaccharide (Fig. 3B) (23). No D0q4-containing CS was detectable in *Cmah*<sup>-/-</sup> mouse livers (Fig. 3C). Similarly, *N*-glycosyl-CS was detectable in WT mouse kidney samples but absent from *Cmah*<sup>-/-</sup> mice (Table S1).

After confirming naturally occurring GalNGc using the well-controlled mouse model, samples from additional animal species were analyzed. Given that Neu5Gc was earlier found to be rich in “red meats” (10, 16, 29), we hypothesized that chances to detect its breakdown products are highest in such samples. Therefore, CS from lamb (ovine muscle), beef (bovine muscle), and pork (porcine muscle) were investigated next. The characteristic D0q4 CS-disaccharide peak (*m/z* = 551) of the lamb sample eluted at 22.8 min (35 min run; Fig. 3D) and its nature was identified by CID analysis, yielding the expected fragmentation pattern of D0q4 as mentioned above (Fig. 3E). Similarly, *N*-glycosyl-CS was detectable in porcine and bovine muscle (Table S1). In addition, CS purified from FCS and commercially available purified CS from porcine intestinal mucosa and bovine trachea were analyzed. As predicted, *N*-glycosyl-CS was detectable in all such animal-derived samples (Table S1). Based on the observed ion intensities, it is possible that *N*-glycosyl-CS exists at levels approaching 0.5%. Such estimates for rare residues in both CS and HS assume similar ionization efficiencies and matrix effects on ion suppression. There is no evidence that a significant ionization deficit exists in going from *N*-acetyl to *N*-glycosyl. However, more accurate quantitative work will require the synthesis of higher-purity standards.

**Neu5Gc Ingestion Generates Much Smaller Amounts of GalNGc, Detectable in Human CS.** Given that humans cannot biosynthesize Neu5Gc (5, 6), no GalNGc should be detectable in human CS samples, whereas related chimpanzee samples should contain *N*-glycosyl-CS. As expected, GalNGc-containing CS-disaccharides are present in chimpanzee serum as analyzed by selective reaction monitoring (SRM) for the dominant D0q4 structure eluting at 22.8 min (35 min run; Fig. 4A). The identity of the respective ion was confirmed by CID analysis (Fig. 4B). Interestingly, purified CS from commercial human serum actually showed trace amounts of *N*-glycosyl-CS, close to the detection limit of the instrument. The predominant D0q4 structure was eluting at 22.8 min (35 min run; Fig. 4C) and the identity of the ion was verified by MS/MS fragmentation analyses (Fig. 4D). To confirm this finding, multiple additional sera from healthy human volunteers were analyzed individually and trace amounts of GalNGc-containing CS were detectable in the majority of samples (Table S1). Given the complete absence of *N*-glycosyl groups in samples from *Cmah*<sup>-/-</sup> mice fed vegan, soy-based diets, the only conceivable

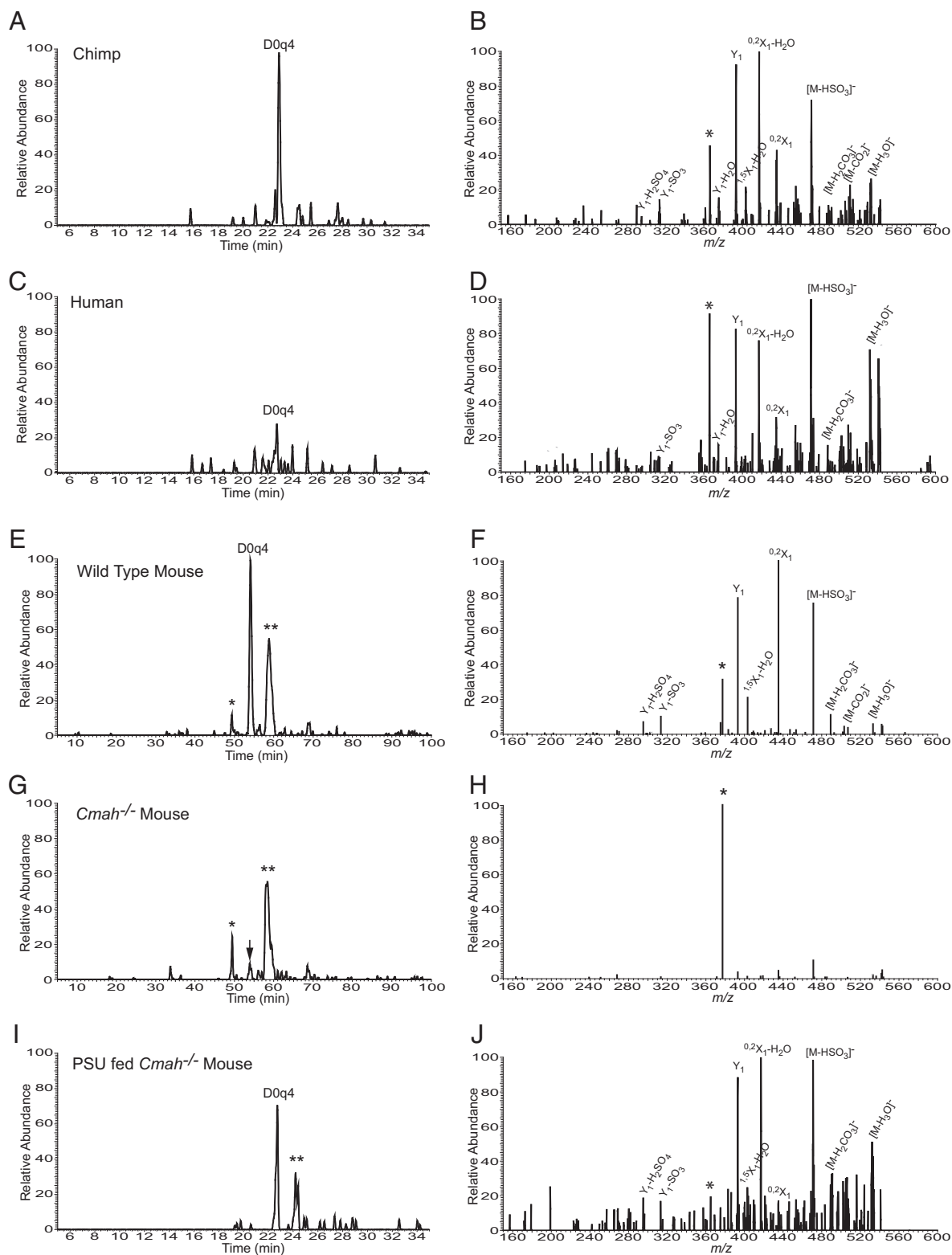


**Fig. 3.** Detection of *N*-glycosyl groups in CS isolated from WT and *Cmah*<sup>-/-</sup> liver and lamb muscle. (A) SRM for <sup>0.2</sup>X<sub>1</sub> product ion from D0q4 (*m/z* = 435) present in WT mouse liver. The y axis represents ion intensity, 0–1,000 arbitrary units. (B) CID mass spectrum of D0q4 detected in WT mouse liver. (C) SRM for <sup>0.2</sup>X<sub>1</sub> product ion from D0q4 present in *Cmah*<sup>-/-</sup> mouse liver. (D) SRM for <sup>0.2</sup>X<sub>1</sub> product ion from D0q4 (*m/z* = 435) present in normal lamb muscle. The y axis represents ion intensity, 0–100 arbitrary units. (E) CID mass spectrum of D0q4 present in lamb muscle.

source for GalNGc in humans would be uptake and metabolism of exogenous, diet-derived Neu5Gc.

To substantiate this hypothesis, Neu5Gc-deficient *Cmah*<sup>-/-</sup> mice were switched from a vegan soy chow diet onto a soy chow containing Neu5Gc at levels comparable to a human Western diet as described previously (16, 30). As expected, *N*-glycosyl-CS is readily detectable in WT mouse serum (Fig. 4 E and F) and absent in the sera of vegan *Cmah*<sup>-/-</sup> knockout mice (Fig. 4 G and H). As predicted, trace amounts of *N*-glycosyl-CS were consistently detectable only in *Cmah*<sup>-/-</sup> mice on the Neu5Gc-containing diet (Fig. 4 I and J). It has been concluded earlier that humans take up and incorporate the nonhuman, immunogenic Sia Neu5Gc





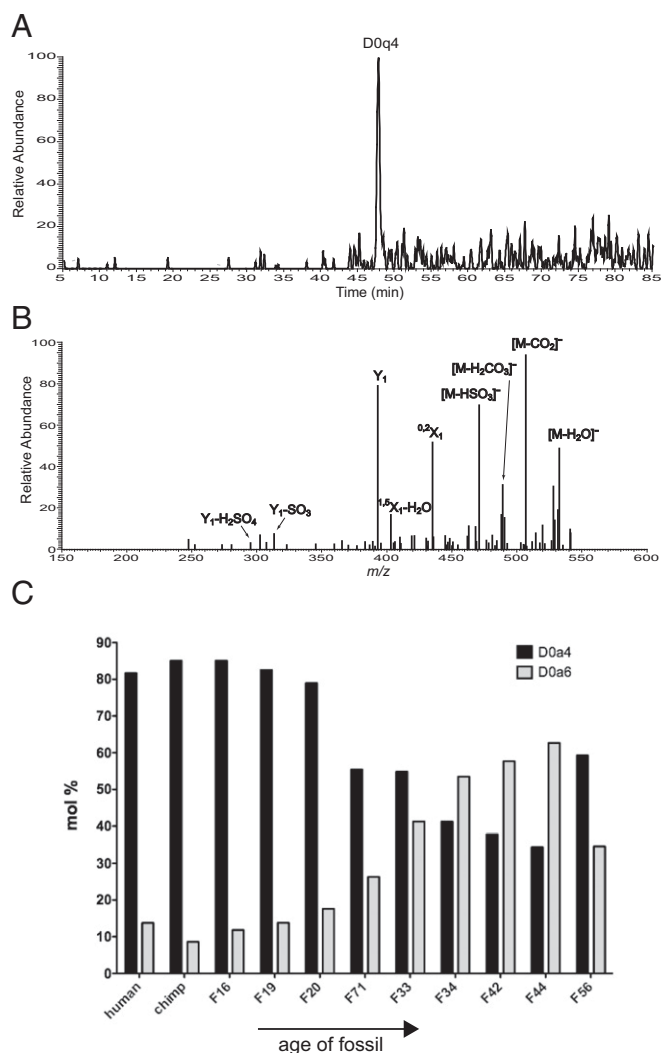
**Fig. 4.** LC/MS analysis of serum from primates and mice. (A and B) The extracted ion current (XIC) for D0q4 in chimpanzee serum identified by retention time and tandem MS (CID spectrum) compared with D0q4 standard. A 35-min gradient was used for this sample. (C and D) The XIC and corresponding CID spectrum for small trace amounts of D0q4 found in human serum. A 35-min gradient was used for this sample. (E and F) The XIC and corresponding CID spectrum for D0q4 found in a WT mouse. A 100-min gradient with higher resolution was used for this sample. (G and H) The XIC and corresponding CID spectrum for D0q4 detected in the time range where D0q4 normally elutes (arrow) in the 100-min gradient used for this sample. (I and J) The XIC and CID spectrum for D0q4 detected in *Cmah*<sup>-/-</sup> mice fed a Neu5Gc-containing diet. A 35-min gradient was used for this sample. The asterisks and double asterisks in the mouse XIC traces are isobaric species that are detected along with D0q4 but do not coelute with D0q4 standard. The asterisks in the CID spectra denote a daughter ion that does not correspond to a product ion from D0q4 standard.

from animal-derived food products (10, 16, 30). The present study suggests that humans also metabolize Neu5Gc toward GalNGc and incorporate this additional nonhuman monosaccharide into their glycoconjugates. However, it is present in trace amounts, much lower than that seen in species with intact CMAH and the capacity to synthesize endogenous Neu5Gc.

**Detection of *N*-Glycolyl-CS in Fossils.** Genetic analyses estimate the human loss of Neu5Gc biosynthesis to have occurred  $\sim 2\text{--}3$  Mya (7, 8). Taken together with mouse fertility studies and models of selection, we could posit that CMAH loss of function may have expedited divergence of the genus *Homo*, shortly afterward (12). To prove this hypothesis, fossil specimens of early bipedal hominins in the time period 4–2 Mya would ideally need to be analyzed for their Sia composition. Unfortunately, Sias are too unstable to remain detectable in fossil specimens from (sub)tropical regions where most of the more ancient bipedal hominins have been discovered (7). As CS molecules are significantly more stable and also abundant in bones, detection of the Neu5Gc-derivative GalNGc in CS of hominin fossils may allow dating of the human loss of Neu5Gc. For method establishment, CS was isolated from contemporary humans and chimpanzees starting with  $\sim 100$  mg of powdered bone material. CS was purified from both samples and treated with chondroitinase ABC before analysis by GRIL-LC/MS as described above (Table S2). *N*-glycolyl-CS was easily detected in chimpanzee bone and much smaller amounts were also noted in human bone, which is in line with the detection of *N*-glycolyl-CS traces in human sera described above. Thereafter, mammalian fossils ( $\sim 12,000\text{--}80,000$  y old from mammoth, elk, and cave bear) were investigated and CS as well as *N*-glycolyl-CS was detectable (Table S2). Based on these encouraging findings, we studied 100 mg of bone powder obtained from a  $\sim 1.65\text{-My}$ -old “Java Man” *Homo erectus* fossil (7, 31). Although CS could be successfully isolated from this significantly older fossil specimen, the total amount of CS in the sample was too low to allow detection of the minor *N*-glycolylated disaccharide species even if it was present (Table S2).

As the  $\sim 1.6\text{-My}$ -old Java man fossil is very precious and might even be too young to help date the human loss of Neu5Gc, we next focused on 4-My-old Kenyan bovid fossil samples, which were excavated from a bone bed in Area 261-1 at Allia Bay on the eastern shores of Lake Turkana. Using significantly larger amounts of starting material from these nonprecious samples accompanying famous hominin fossils (5 g each), the presence of *N*-glycolylated disaccharide residues in all animal fossil samples was evaluated using multiple reaction monitoring for  $Y_1$ ,  $^{0,2}X_1$ , and M-HSO<sub>3</sub> product ions. Except for sample F33, all ancient animal fossils exhibited detectable amounts of *N*-glycolyl-CS, as summarized in Table S2. An example of the data is shown for the  $\sim 3.4\text{-My}$ -old animal fossil F44. The *N*-glycolylated D0q4 disaccharide elutes at 48 min (100 min run) (Fig. 5A) and MS/MS fragmentation analysis confirms the identity of the predicted structure (Fig. 5B).

Interestingly, the disaccharide composition of CS extracted from fossil bones differs significantly from those of contemporary animal bones, including human bone. Whereas contemporary bone samples consistently contain  $\sim 75\text{--}85\%$  D0a4 disaccharide, this structure is less prominent in all fossil bones studied (e.g., only 34% in F44). In contrast, D0a6 disaccharide only makes up for  $\sim 5\text{--}15\%$  of CS in contemporary bone samples, whereas it can even be the predominant CS disaccharide in fossils (e.g., 63% in F44) as shown in Fig. 5C. Complete CS disaccharide compositional analysis is summarized in Fig. S2. Based on these findings, some samples were repurified to isolate HS and study the disaccharide composition found for HS. Similarly to CS, significantly altered disaccharide compositions were also found when analyzing HS (Fig. S3).



**Fig. 5.** Detection of *N*-glycolyl groups in CS isolated from partially mineralized fossil samples. (A) The accumulative extracted ion current chromatogram for  $Y_1$ ,  $^{0,2}X_1$ , and M-HSO<sub>3</sub> daughter ions consistent with *N*-glycolylated disaccharide residues ( $m/z = 393$ , 435 and 471, respectively) for fossil F44. (B) The product ion mass spectrum of D0q4 ( $m/z$  value = 551). (C) CS compositional analysis comparing D0a4 and D0a6 abundance of bone and fossil samples expressed as molar percent (mol %).

The altered disaccharide composition in fossil bones likely originates from geochemical processes (biochemical diagenesis) (32) and may be influenced by the individual conditions at excavation sites. Regardless, the unusually different disaccharide composition helps to exclude the possibility that CS or HS detected in fossils results from human contamination during handling and sample preparation. Overall, we successfully demonstrated the presence of CS as well as *N*-glycolyl-CS in animal fossils as old as 4 My. This method would theoretically allow dating the human loss of Neu5Gc when studying hominin fossils from this era. However, given the extreme rarity of older hominin fossil specimens and the currently needed starting material of 5 g, this can only be followed up when MS instruments reach a higher level of sensitivity.

## Discussion

We have demonstrated the natural occurrence of the monosaccharide GalNGc and its activated form UDP-GalNGc, which can act as a donor for animal CS biosynthesis, and proven an

unexpected cellular pathway for metabolism of the *N*-glycolyl group from the Sia Neu5Gc via six metabolic steps, leading to UDP-GalNGc. Conclusive proof of this pathway comes from the absence of UDP-HexNGc and *N*-glycolyl-CS in a mouse strain with human-like lack of endogenous Neu5Gc due to inactivation of the *Cmah* gene. Further supporting this finding, the presence of GalNGc in CS was shown in tissues and sera of other animals known to be rich in Neu5Gc, including commercially available CS from bovine trachea and porcine intestinal mucosa. In contrast, cartilage CS from shark (another group of species so far found devoid of Neu5Gc) was also devoid of GalNGc in CS. As expected, GalNGc was readily detectable in the chimpanzee specimens.

Considering the multiple metabolic steps involved in generating *N*-glycolyl-CS from Neu5Gc, it is surprising that natural occurrence of *N*-glycolyl groups is largely limited to CS *in vivo*. In this regard, it is notable that six metabolic steps tolerate the change from the usual hydrophobic *N*-acetyl group to a hydrophilic *N*-glycolyl group and thus allow conversion of Neu5Gc to UDP-GalNGc. This apparent “channeling” of the *N*-glycolyl group toward CS either means that this low-abundance monosaccharide disturbs cellular pathways least when incorporated in CS or that *N*-glycolylated CS has a specific biological function. Other rare modifications of GAGs such as 3-*O*-sulfation of HS are critical for interaction with HS-binding proteins (33, 34). The highly selective enrichment of *N*-glycolyl groups only in CS suggests the possibility that *N*-glycolyl-CS may have some as-yet-unknown function in mammals that was then also lost in humans. One possible explanation could be the specificity of the metabolic enzymes involved toward HexNAc versus HexNGc. Another possibility would be a shorter lifecycle or possible shedding from cells of other HexNGc-containing glycoconjugates. Regardless, availability of synthetic *N*-glycolyl-CS will allow future exploration of the functional question, for example by searching for endogenous proteins that selectively bind this GAG.

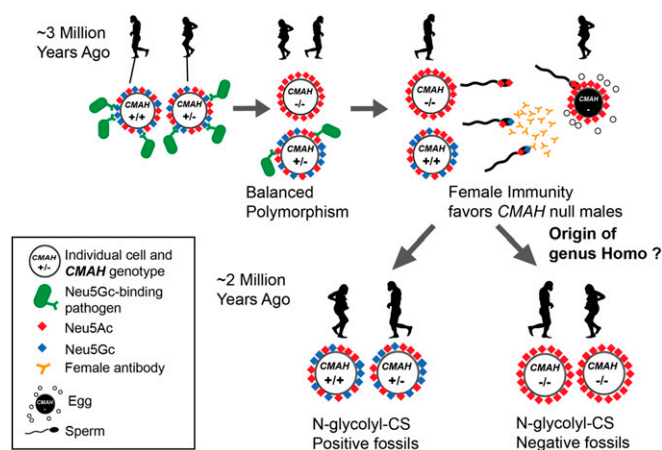
Unexpectedly, much smaller amounts of GalNGc were also detectable in CS from 8 out of 10 human serum samples as well as in human bone (Table S1). Taken together with the absence in *Cmah*<sup>-/-</sup> mice, the only explanation is that these trace amounts of GalNGc in human CS result from uptake and metabolism of Neu5Gc in animal-derived foods. Substantiating this hypothesis, *Cmah*<sup>-/-</sup> mice fed a Neu5Gc-containing diet showed trace amounts of GalNGc in the CS as well. Direct comparison of human meat eaters to a strictly vegan cohort of people in the future would further substantiate this finding. In this regard it is known that Neu5Gc represents a diet-derived xeno-autoantigen, with all humans having circulating antibodies against this epitope (11). Thus, it is possible that HexNGc-containing glycans such as GalNGc-containing CS represent additional diet-acquired xeno-autoantigens in humans. Future studies are needed to evaluate this possibility.

In the present study the chemical synthesis of partially *N*-glycolylated CS was developed, established, and fully described. The synthesis is based on commercially available shark cartilage, which is devoid of *N*-glycolylated CS and can therefore serve as a negative control. Following partial de-*N*-acetylation, partially *N*-glycolylated CS could be achieved using acetoxyacetyl chloride. The identity of the product was confirmed by MS. This material provides the starting point for comprehensive *in vitro* studies such as screening for potential binding partners to *N*-glycolyl-CS or screening for anti-HexNGc xeno-autoantibodies in the future.

While contemporary humans only have trace amounts of GalNGc in CS (derived from food sources), it is much more easily detectable in chimpanzee samples, as it is derived from the much larger amounts of endogenous Neu5Gc. Thus, analysis of CS from hominin fossils may allow dating the human loss of Neu5Gc biosynthesis. Previous attempts to analyze Sias directly in European Neanderthal specimens were successful (7), but attempts to do so in much older hominin fossils and faunal samples from Africa failed, likely because these sugars are too unstable in specimens from tropical and semitropical regions,

where most of the early bipedal hominins have been discovered (35). However, CS is significantly more stable than Sias and provides an alternative option to detect *N*-glycolyl groups in ancient fossils. Indeed, GalNGc-containing CS was successfully detected in various faunal fossil samples including a ~4-My-old long bone collected right next to *Australopithecus anamensis* in Allia bay, Kenya (Table S2). As the human loss of Neu5Gc is estimated to have occurred ~2–3 Mya (7) (Fig. 6), detection of GalNGc in hominin fossils from the relevant era should be possible. However, while *N*-glycolyl-CS is easily detectable in extracts from 100 mg of contemporary bones, a similar-sized sample obtained from an *H. erectus* in Java had very little CS and was below the detection limit for *N*-glycolyl-CS. At the current level of sensitivity, we estimate that ~5 g of starting material are required for such an analysis, making the study currently unfeasible given the rarity and precious nature of ancient hominin fossils.

However, improved sensitivity of MS and further optimization of CS isolation should render this study possible in the future. At that point it should be possible to access small samples of the relevant hominin fossils, spanning the time period from 1–4 Mya (35, 36), beginning possibly with definitive *Homo* samples for which larger amounts of material are available [e.g., *Homo georgicus* (37) and *Homo antecessor* (38, 39)]. If indeed such samples are negative for Gc-CS, it would then become important to request access to *Australopithecus* and *Paranthropus* samples (35) predicted to be positive, as well as undated samples of lineages with controversial connections to the *Homo* clade such as *Homo naledi* (40, 41) and *Australopithecus sediba* (42, 43). Assuming clear-cut results for samples, the division into *N*-glycolyl-CS positive and negative should help clarify the phylogeny of hominin fossils (samples with only trace amounts of *N*-glycolyl-CS will be assumed to represent Neu5Gc-null individuals who acquired the traces from dietary Neu5Gc, and thus classified



**Fig. 6.** Potential scenario for the role of Neu5Gc loss and female anti-Neu5Gc immunity in the origin of the genus *Homo* via interplay of natural and sexual selection acting on cell-surface Sias. There are many known pathogens that recognize and exploit Neu5Gc (blue diamond) as a receptor on host target cells (9). Natural selection by such pathogens may have selected for rare *CMAH* null alleles that abolish Neu5Gc expression in homozygote individuals (12). Such individuals have only Neu5Ac and its derivatives on their cells (red diamonds) allowing an escape from pathogens, but at higher frequencies would be targeted by adapting pathogens, resulting in maintenance of a balanced polymorphism. *CMAH*<sup>-/-</sup> females with anti-Neu5Gc antibodies also present in their reproductive tract would favor sperm from *CMAH*<sup>+/+</sup> males due to anti-Neu5Gc antibody-mediated cryptic selection against *CMAH*<sup>+/-</sup> or *CMAH*<sup>+/+</sup> males expressing Neu5Gc on their sperm. Once the frequency of the *CMAH* null allele reaches a critical level, this process can drive fixation of the loss-of-function allele in a population by directional selection (12).



as *N*-glycolyl-CS negative for this purpose) (Fig. 6). Assuming sensitivity hurdles can be overcome, this approach may eventually classify putative *Homo* fossils into CMAH-negative and CMAH-positive. The former would be presumed ancestors of the lineage leading to humans, and the latter are likely to be “dead-end” lineages that did not persist. In this regard some of the younger putative *Homo* lineages in Southeast Asia would also be interesting to study, to ascertain whether their ancestors left Africa before the fixation of the CMAH-null state (44–46).

Notably, the overall CS disaccharide composition of ancient fossils was distinct from contemporary human bone. In particular, while D0a6 is a minor disaccharide in contemporary bone samples it was the major structure in ancient samples (Fig. 5C and Fig. S2). Based on this finding, HS were also investigated. Although no GlcNGc was detectable in HS, the disaccharide composition of older fossils was also significantly altered compared with contemporary human bones (Fig. S3), likely due to diagenetic processes. The exact CS and HS disaccharide compositions of fossils are likely impacted by climatic and geochemical factors and certainly do not allow one to date a fossil. However, in contrast to other molecules analyzed in fossils, the significant difference to present-day humans makes CS and HS analyses in general a field where contaminations of fossil samples by the excavator or analyst can be reasonably excluded. This allows for future exploration of such ancient glycans to answer evolutionary questions.

## Materials and Methods

**Materials.** [Glycolyl-<sup>3</sup>H]Neu5Gc was prepared as described previously (4). Chimpanzee blood samples were collected as extra tubes only during routine noninvasive health screens of chimpanzee subjects at the Yerkes National Primate Center, Emory University, Atlanta, GA (supported by NIH Base Grant ORIP/OD P51OD011132, routine collection covered under local IRB approval by Emory University). All collections were made before the September 15, 2015, designation of captive chimpanzees as endangered species. Human AB serum was purchased from Valley Biomedical. Individual human serum samples were collected from healthy human volunteers, who provided informed consent. All human samples were completely anonymized, and the only information available to the investigators was age and sex. All individuals who handled the primary human and primate samples received the required training regarding precautions for blood-borne pathogens. All health and safety issues related to handling of human and nonhuman primate samples are covered by an Institutional biosafety approval from the University of California, San Diego Environmental Health and Safety Committee. Commercially available isolated CS from bovine trachea, porcine intestinal mucosa, and shark cartilage were obtained from Sigma. *H. erectus* samples from Indonesia were kindly provided by Etty Indriati and the late Teuku Jacob, Gadjah Mada University, Yogyakarta, Indonesia, as described earlier (7). Sources of other fossils are indicated in the main text and/or in Table S2.

**Tissue Culture.** Human acute monocytic leukemia cell line THP-1 (47) was cultivated in DMEM (high glucose; Invitrogen) and CHO LEC29.lec32 cells (25) were cultured in alpha MEM. Media were supplemented with 5% human serum (heat-inactivated and sterile-filtered; Valley Biomedical Inc.) and 2 mM glutamine. Cells were cultivated in a humidified 5% CO<sub>2</sub> atmosphere at 37 °C.

**Isolation and Analysis of Radiolabeled GAGs from Cells.** CHO cells were cultured under the conditions described above in a P-150 dish with 20 mL of media. The culturing media was supplemented with 7,400 kBq of [glycolyl-<sup>3</sup>H]Neu5Gc 3 d before reaching confluence. Thereafter, the media was removed and cells were harvested using 20 mM EDTA in PBS. For THP-1 cells, 30 × 10<sup>6</sup> cells were cultured for 3 d in the presence of 7,400 kBq of [glycolyl-<sup>3</sup>H]Neu5Gc in a total volume of 20 mL. Cells were washed twice with PBS to remove excess radiolabel and stored frozen at –80 °C until processed further. Cells were lysed by adding 2.5 mL of 0.1 M NaOH following 15-min incubation at 37 °C. Thereafter, 20 mL of buffer A (50 mM sodium acetate trihydrate and 200 mM NaCl, pH 6.0) were added to each sample and the pH was adjusted to 8.0 using 1 M acetic acid. Pronase E (Sigma) was added at a final concentration of 100 µg/mL using a 20 mg/mL stock solution and the sample was incubated in a head-over-head shaker at 37 °C overnight. An additional 50 µg/mL pronase E were added to each sample following an-

other overnight incubation. Thereafter, samples were incubated at 99 °C in a waterbath for 10 min to inactivate the proteinase. Samples were vortexed, spun at 3,000 × *g* for 1 h in 50-mL conical tubes, and supernatants collected. Columns were packed using 2 mL of DEAE sephacel (50% suspension; Sigma) and preequilibrated with 20 mL buffer B (50 mM sodium acetate trihydrate, 200 mM NaCl, and 0.1% Triton X-100, pH 6.0). The supernatants were loaded and the columns allowed to run by gravity flow. Columns were washed with 30 mL of buffer A followed by elution of bound GAGs with 5 mL of buffer C (50 mM sodium acetate trihydrate and 1 M NaCl, pH 6.0). Eluates were desalted using PD-10 desalting columns (GE Healthcare) as described in the manufacturer's guidelines using 10% ethanol. Samples were lyophilized, resuspended in 600 µL water each, and split into 3 × 200 µL for chondroitinase treatment, heparin lyase treatment, and control treatment. A 200-µL sample was supplemented with 23 µL chondroitinase buffer (500 mM Tris-HCl and 500 mM NaCl, pH 7.9) and 7 µL chondroitinase ABC (10 mU/µL stock; Sigma). Another 200-µL sample was supplemented with 23 µL of buffer (40 mM ammonium acetate and 3.33 mM calcium acetate, pH 7.0), 4.5 µL heparin lyase-Mix (2 mU each of heparin lyases I, II, and III from IBEX), and 2.5 µL water. The last 200-µL sample was supplemented with 23 µL chondroitinase buffer and 7 µL water. All samples were incubated at 37 °C overnight. Thereafter, size-exclusion chromatography was performed as described previously (23). In brief, phenol red solution (sterile filtered) was added to a final volume of 0.5 mL to each sample. A gel filtration column (1.7 cm outer diameter by 42 cm, CL-6B; Sigma) was poured and equilibrated in buffer A. Samples were loaded and the column run with ~40 µL/min using buffer A. Fractions were collected (15 min each, ~120 fractions total), and radioactivity was determined by scintillation counting. The elution time of phenol red was set to 1, and relative elution times of all fractions were calculated accordingly.

**Isolation of Nucleoside Diphospho Sugars.** An adult mouse liver was homogenized in 2 mL of 0.1× PBS, pH 7.4 (BioPioneer), containing 1:100 protease inhibitor mixture III (Calbiochem) using a polytron. Thereafter, 500 µL of liver homogenate were diluted with 2 mL PBS, pH 7.4, mixed well, and spun at 60,000 × *g* for 1 h at 4 °C in the ultracentrifuge. The supernatant was transferred into a new tube and an equivalent volume of 1.2 M perchloric acid was added. The tube was kept on ice for 1 h to complete protein precipitation before the sample was spun at 2,862 × *g* and 4 °C for 15 min. The supernatant was transferred to a new tube and spun again at 15,000 × *g* and 4 °C for 15 min. The resulting supernatant was diluted with 10 volumes of MilliQ water (~25 mL final volume). A 1 mL, 50 mg hypersep C-18 column (Thermo Fisher) was equilibrated by subsequently passing over 4 column volume (CV) of 10% acetic acid, 4 CV of 50% methanol, 4 CV methanol, 4 CV 72% isopropanol + 27.9% methanol + 0.1% formic acid, 4 CV anhydrous ethyl acetate, 4 CV chloroform, 2 CV methanol, 2 CV 50% methanol, and 8 mL MilliQ water. A 1 mL, 25 mg hypersep hypercarb PGC cartridge (Thermo Fisher) was equilibrated by subsequently passing over 4 CV acetonitrile, 4 CV 50% acetonitrile, and 8 CV 0.1% trifluoroic acid. Samples were loaded onto equilibrated C-18 column and allowed to run by gravity flow. The flow through of the C-18 column was collected and loaded onto the porous graphitic carbon (PGC) cartridge. The PGC was subsequently washed with 2 mL of 50 mM triethylamine acetate (TEAA) buffer, pH 7.0. Nucleotide sugars were eluted with 1.6 mL of 50% acetonitrile containing 50 mM TEAA buffer, pH 7.0. Acetonitrile and TEAA were removed using nitrogen gas before samples were frozen and lyophilized. Dried material was reconstituted in 85 µL of MilliQ water before 10 µL of 10× NEB3 buffer and 5 µL (50 U) of calf intestinal phosphatase (both New England Biolabs) were added. Samples were incubated 45 min at 37 °C and thereafter diluted with 10 mM NH<sub>4</sub>HCO<sub>3</sub> pH 7.5 to a final volume of 1 mL. DEAE sephacel columns (1-mL bed volume; Sigma) were equilibrated with 5 CV of 10 mM ammonium carbonate before samples were loaded. Columns were washed with 5 CV of 10 mM ammonium carbonate before elution of nucleoside sugars with 2.25 mL of 250 mM ammonium carbonate (discarded the first 0.75 mL and collected the remaining 1.5 mL). Samples were frozen and lyophilized to remove ammonium carbonate; this step was repeated two times. Dried material was taken up in water and analyzed by LC-MS analysis as described below using the 35-min gradient (28).

**Blood Samples.** Chimpanzee blood samples were collected as extra tubes only during routine noninvasive health screens of chimpanzee subjects at the Yerkes National Primate Center, Emory University, Atlanta, GA (supported by NIH Base Grant ORIP/OD P51OD011132, routine collection covered under local IRB approval by Emory University). All collections were made before the September 15, 2015, designation of captive chimpanzees as endangered species. Chimpanzee blood samples were shipped overnight on ice to the University of California, San Diego. Human blood was collected at about the



same time into identical tubes from healthy volunteer donors (following informed consent, under the approval from the University of California, San Diego Human Subjects Institutional Review Board), and stored overnight on ice, to ensure similar treatment conditions before analysis. All health and safety issues related to handling of human and nonhuman primate samples are covered by an Institutional biosafety approval from the University of California, San Diego Environmental Health and Safety Committee. All individuals who handle the samples receive the required training regarding precautions for blood-borne pathogens.

**Mice and Chow.** *Cmah<sup>-/-</sup>* mice were described previously (26) and syngeneic WT C57BL/6 mice were purchased from Harlan Laboratories. Mice were fed vegan standard chow (110951/110751; Dyets, Inc.) or—only where specifically mentioned—Neu5Gc-containing chow (Dyets, Inc. custom order) as described (16). Mice had access to water for ad libitum consumption and were maintained on a 12-h light/dark cycle. Animal work was performed in accordance with The Association for Assessment and Accreditation of Laboratory Animal Care under protocol S01227 approved by The Institutional Animal Care and Use Committee of the University of California, San Diego.

**Isolation of GAGs.** An adult mouse liver and kidney were each homogenized in 2 mL and 1 mL of 0.1× PBS, pH 7.4 (BioPioneer) containing 1:100 protease inhibitor mixture III (Calbiochem) using a polytron, respectively. The resulting kidney homogenate was diluted with buffer A to a final volume of 20 mL, whereas only 500  $\mu$ L of the liver homogenate were used and diluted to 20 mL with buffer A. For purification of GAGs from animal muscle, 2 g of ground beef, lamb, and pork were each homogenized using a polytron in 1:10 diluted PBS, pH 7.4, reaching a final volume of 27.5 mL for each homogenate. For purification of GAGs from human and chimpanzee sera, 5 mL of each serum were diluted with buffer A to a final volume of 20 mL. GAGs from mouse sera were purified from 2 mL of starting material following addition of 18 mL buffer A and 10 mL of FCS were supplemented with 10 mL of buffer A. To prevent bacterial growth, all serum and tissue samples were supplemented with 0.01% sodium azide and 100  $\mu$ g/mL carbenicillin. Thereafter, 100  $\mu$ g/mL pronase E were added and samples were incubated in a head-over-head shaker at 37 °C overnight. The next day another 50  $\mu$ g/mL protease was added to each sample following an additional overnight incubation. All samples were heated for 10 min at 99 °C to inactivate the protease. GAGs were purified as described previously (23, 48). In brief, samples were spun for 1 h at 3,000  $\times$  g and the resulting supernatant was passed over a 5- $\mu$ m low-protein-binding syringe filter, which removed the floating fat and residual particles, especially from tissue samples. The flow-through was applied onto DEAE Sephacel columns (Sigma) with  $\sim$ 200  $\mu$ L bed volume pre-equilibrated in buffer B. Columns were washed with 20 mL of buffer A and bound GAGs were eluted with 5 mL of buffer C. Eluted GAGs were subsequently passed over PD-10 columns (GE Healthcare) using 10% ethanol and following the manufacturer's guidelines. Desalted GAGs were lyophilized and stored at  $-20$  °C.

**Release of GAGs from Fossils and Bones.** A rotary power tool (Dremel) with a brush attachment was used to carefully remove debris from the outside surface of the fossils. Various grinding/drilling attachments were used thereafter to generate 50–100 mg of powdered fossils or contemporary bones. The resulting powder was suspended in 5 mL of 0.5 M EDTA, pH 8.0, in a 50-mL conical tube. The suspension was incubated for 48 h at room temperature in a head-over-head shaker. Buffer A was added to a final volume of 20 mL. For larger-scale preparations (5 g of starting material), the powdered fossils were transferred into a 250-mL plastic Erlenmeyer flask (Corning) and supplemented with 50 mL of 0.5 M EDTA, pH 8.0, for 48 h at room temperature and 140 rpm. Samples were diluted with buffer A to a final volume of 100 mL. All samples were supplemented with Proteinase K (Invitrogen) and Pronase E (Sigma) at final concentrations of 100  $\mu$ g/mL each. Samples were incubated 24 h at 37 °C and 140 rpm. Thereafter, another 50  $\mu$ g/mL of both proteases were added to the samples following another overnight incubation at 37 °C and 140 rpm. Samples from large-scale preparations were transferred into 4  $\times$  50-mL conicals each and all samples were heat-inactivated at 99 °C for 10 min. Samples were vortexed following centrifugation at 3,000  $\times$  g for 1 h. Supernatants were collected and GAGs were purified via DEAE

sephacel ( $\sim$ 250  $\mu$ L bed volume) following desalting on PD-10 columns as described above. Eluted GAGs were lyophilized.

**Chondroitinase Treatment.** Nonradioactive, lyophilized GAGs were resuspended in 40  $\mu$ L MilliQ water before addition of 5  $\mu$ L 10 $\times$  chondroitinase buffer (500 mM Tris-HCl and 500 mM NaCl, pH 7.9) and 50 mU of chondroitinase ABC (Sigma). The final reaction volume was 50  $\mu$ L and samples were incubated at 37 °C overnight. Thereafter, samples were lyophilized again.

**Treatment of GAGs with Heparin Lyases.** Nonradioactive, lyophilized GAGs were analyzed as described previously (28).

**Aniline Tagging of CS and HS Disaccharides.** Nonradioactive CS disaccharides were coupled to [<sup>12</sup>C<sub>6</sub>]aniline as described previously (23, 28). In brief, 130 mg NaCNBH<sub>3</sub> were dissolved in 1.4 mL DMSO before adding 0.6 mL glacial acetic acid. Enzymatically depolymerized GAG samples were dried down after digestion and reconstituted in 17  $\mu$ L [<sup>12</sup>C<sub>6</sub>]aniline to which 17  $\mu$ L of the above-mentioned NaCNBH<sub>3</sub> solution were added. The reactions were incubated for 16 h at 37 °C (dry heat). Thereafter, samples were dried down using a speedvac and stored in the dark at 4 °C until analyzed by LC/MS.

**LC/MS Analysis of CS and HS.** LC/MS was performed as described previously (28). Deviating from the above, a shorter gradient was also found suitable for analysis: After the first 5 min at 0% B, there is a single linear gradient from 5 to 30 min with percent B rising from 0 to 100% with a final re-equilibration step at 0% B for 5 min, for a total of 35 min per run. CS derived from fossils were run with the longer gradient to maximize resolution. Additionally, *N*-glycolyl-CS-specific peaks in fossil samples were confirmed by spiking with chemically synthesized *N*-glycolyl-CS disaccharide standard.

**Synthesis of Partially *N*-Glycolylated CS.** In a glass tube, 50 mg of CS sodium salt (from shark cartilage; Sigma) were dissolved in 1 mL anhydrous hydrazine. The tube was closed with a glass marble and incubated at 100 °C for 2 h (hood, shielded) to partially de-*N*-acetylate the CS. Thereafter, the sample was dialyzed [molecular weight cutoff (MWCO) 3,500; Spectra/Por] three times against 2 L of water each to remove hydrazine, frozen, and lyophilized. Dried material was resuspended in 2.5 mL of water and 135 mg sodium bicarbonate (seven equivalents) was added. After dissolving completely, 135  $\mu$ L of acetoxyacetyl chloride was added while stirring vigorously over the course of 10 min (six equivalents). Sample was mixed at room temperature for 1 h to allow re-*N*-glycolylation of the CS. Thereafter, the sample was dialyzed again (MWCO 3,500; Spectra/Por) three times against 2 L of water each, frozen, and lyophilized. Dried material was resuspended in 2 mL of 0.1 M NaOH to remove potential *N*-acetyl groups attached to the hydroxyl group of the *N*-glycolyl function. The sample was incubated at 37 °C for 10 min and dialyzed again three times against 2 L of water each, frozen, and lyophilized. The dried compound was stored at  $-20$  °C. The presence of *N*-glycolyl groups was confirmed by MS after chondroitinase ABC treatment as described below.

**Generation of an *N*-Glycolyl-CS Disaccharide Standard.** Partially *N*-glycolylated CS (50  $\mu$ g) described above was incubated with 50 mM Tris-HCl, pH 7.9, 50 mM NaCl, and 20 mU of chondroitinase ABC (Sigma) in a final volume of 55  $\mu$ L at 37 °C overnight. The sample was lyophilized and CS disaccharides tagged with aniline as described above but using [<sup>13</sup>C<sub>6</sub>]aniline instead of [<sup>12</sup>C<sub>6</sub>]aniline.

**ACKNOWLEDGMENTS.** We thank Etty Indriati and the late Teuku Jacob, Gadjah Mada University, Yogyakarta, Indonesia for the Indonesian *Homo* samples; the Yerkes Primate Research Center for the chimpanzee serum samples; healthy volunteers for blood samples; David Vocado for the UDP-HexNGc standard used in previous related studies; Pamela Stanley for CHO lec29.LEC32 cells; Margaret Schoeninger from the Center for Academic Research and Training in Anthropogeny for the chimpanzee and human bones; and Jeffrey D. Esko for his critical review of the manuscript. This work was supported by NIH Grants R01GM32373 (to A.V.), GM33063, and HL107150 (J. Esko, Principal Investigator) and the Mathers Foundation of New York. Mass spectrometry analyses were performed at the Glycotechnology Core Resource at the Glycobiology Research and Training Center, University of California, San Diego, managed by Biswa Choudhury.

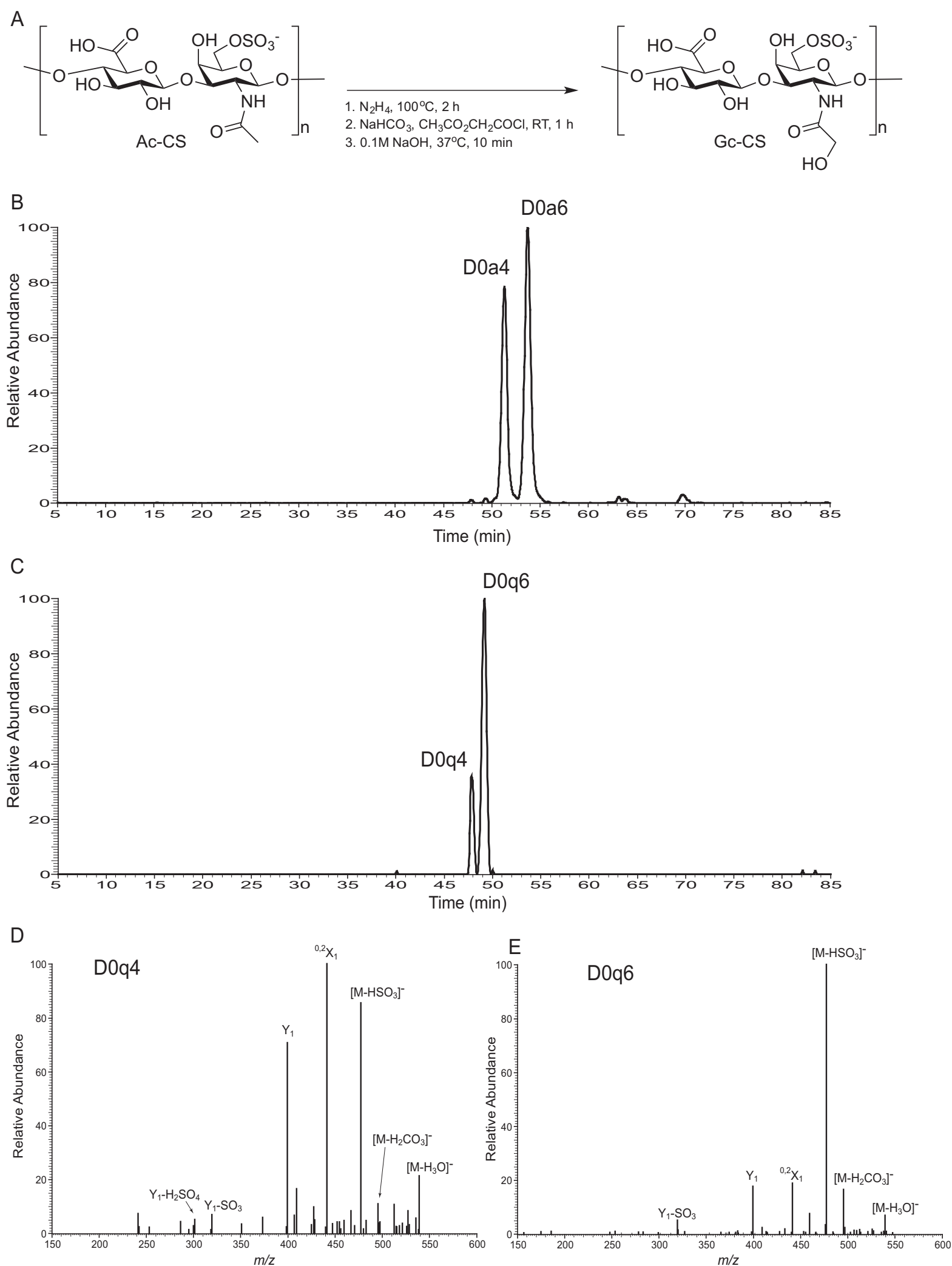
- Varki A, Schauer R (2009) *Essentials of Glycobiology*, eds Varki A, et al. (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 199–218.
- Shaw L, Schauer R (1989) Detection of CMP-N-acetylneuraminic acid hydroxylase activity in fractionated mouse liver. *Biochem J* 263:355–363.
- Hedlund M, Padler-Karavani V, Varki NM, Varki A (2008) Evidence for a human-specific mechanism for diet and antibody-mediated inflammation in carcinoma progression. *Proc Natl Acad Sci USA* 105:18936–18941.

- Bergfeld AK, Pearce OM, Diaz SL, Pham T, Varki A (2012) Metabolism of vertebrate amino sugars with *N*-glycolyl groups: Elucidating the intracellular fate of the non-human sialic acid *N*-glycolylneuraminic acid. *J Biol Chem* 287:28865–28881.
- Chou HH, et al. (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc Natl Acad Sci USA* 95:11751–11756.
- Irie A, Koyama S, Kozutsumi Y, Kawasaki T, Suzuki A (1998) The molecular basis for the absence of *N*-glycolylneuraminic acid in humans. *J Biol Chem* 273:15866–15871.

7. Chou HH, et al. (2002) Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc Natl Acad Sci USA* 99: 11736–11741.
8. Hayakawa T, Aki I, Varki A, Satta Y, Takahata N (2006) Fixation of the human-specific CMP-N-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. *Genetics* 172:1139–1146.
9. Okerblom J, Varki A (2017) Biochemical, cellular, physiological, and pathological consequences of human loss of N-glycolylneuraminic acid. *ChemBioChem* 18: 1155–1171.
10. Tangvoranuntakul P, et al. (2003) Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc Natl Acad Sci USA* 100:12045–12050.
11. Padler-Karavani V, et al. (2008) Diversity in specificity, abundance, and composition of anti-Neu5Gc antibodies in normal humans: Potential implications for disease. *Glycobiology* 18:818–830.
12. Ghaderi D, et al. (2011) Sexual selection by female immunity against paternal antigens can fix loss of function alleles. *Proc Natl Acad Sci USA* 108:17743–17748.
13. Ma F, et al. (2016) A mouse model for dietary xenosialitis: Antibodies to xenoglycan can reduce fertility. *J Biol Chem* 291:18222–18231.
14. Bardor M, Nguyen DH, Diaz S, Varki A (2005) Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. *J Biol Chem* 280:4228–4237.
15. Varki NM, Strobert E, Dick EJ, Jr, Benirschke K, Varki A (2011) Biomedical differences between human and nonhuman hominids: Potential roles for uniquely human aspects of sialic acid biology. *Annu Rev Pathol* 6:365–393.
16. Samraj AN, Läubl H, Varki N, Varki A (2014) Involvement of a non-human sialic acid in human cancer. *Front Oncol* 4:33.
17. Alisson-Silva F, Kawanishi K, Varki A (2016) Human risk of diseases associated with red meat intake: Analysis of current theories and proposed role for metabolic incorporation of a non-human sialic acid. *Mol Aspects Med* 51:16–30.
18. Muchmore EA, Diaz S, Varki A (1998) A structural difference between the cell surfaces of humans and the great apes. *Am J Phys Anthropol* 107:187–198.
19. Macauley MS, et al. (2012) Metabolism of vertebrate amino sugars with N-glycolyl groups: Intracellular  $\beta$ -O-linked N-glycolylglucosamine (GlcNGc), UDP-GlcNGc, and the biochemical and structural rationale for the substrate tolerance of  $\beta$ -O-linked  $\beta$ -N-acetylglucosaminidase. *J Biol Chem* 287:28882–28897.
20. Pouilly S, Bourgeois V, Piller F, Piller V (2012) Evaluation of GalNAc analogs as substrates for enzymes of the mammalian GalNAc salvage pathway. *ACS Chem Biol* 7: 753–760.
21. Pouilly S, Piller V, Piller F (2011) Metabolic glycoengineering through the mammalian GalNAc salvage pathway. *FEBS J* 279:586–598.
22. Bateman AC, et al. (2010) Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: The importance of NeuA $\alpha$ 2-6 glycans. *J Biol Chem* 285:34016–34026.
23. Bergfeld AK, et al. (2012) Metabolism of vertebrate amino sugars with N-glycolyl groups: Incorporation of N-glycolylhexosamines into mammalian glycans by feeding N-glycolylgalactosamine. *J Biol Chem* 287:28898–28916.
24. Esko JD, Kimata K, Lindahl U (2009) *Essentials of Glycobiology*, eds Varki A, et al. (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 229–248.
25. Potvin B, Raju TS, Stanley P (1995) Lec32 is a new mutation in Chinese hamster ovary cells that essentially abrogates CMP-N-acetylneuraminic acid synthetase activity. *J Biol Chem* 270:30415–30421.
26. Hedlund M, et al. (2007) N-glycolylneuraminic acid deficiency in mice: Implications for human biology and evolution. *Mol Cell Biol* 27:4340–4346.
27. Diaz SL, et al. (2009) Sensitive and specific detection of the non-human sialic acid N-glycolylneuraminic acid in human tissues and biotherapeutic products. *PLoS One* 4: e4241.
28. Lawrence R, et al. (2008) Evolutionary differences in glycosaminoglycan fine structure detected by quantitative glycan reductive isotope labeling. *J Biol Chem* 283: 33674–33684.
29. Li H, Fan X (2014) Quantitative analysis of sialic acids in Chinese conventional foods by HPLC-FLD. *Open J Prev Med* 4:57–63.
30. Banda K, Gregg CJ, Chow R, Varki NM, Varki A (2012) Metabolism of vertebrate amino sugars with N-glycolyl groups: Mechanisms underlying gastrointestinal incorporation of the non-human sialic acid xeno-autoantigen N-glycolylneuraminic acid. *J Biol Chem* 287:28852–28864.
31. Kaifu Y, et al. (2008) Cranial morphology of Javanese Homo erectus: New evidence for continuous evolution, specialization, and terminal extinction. *J Hum Evol* 55: 551–580.
32. Cowie GL, Hedges JI (1994) Biochemical indicators of diagenetic alteration in natural organic matter mixtures. *Nature* 369:304–307.
33. Thacker BE, Xu D, Lawrence R, Esko JD (2014) Heparan sulfate 3-O-sulfation: A rare modification in search of a function. *Matrix Biol* 35:60–72.
34. Xu D, Esko JD (2014) Demystifying heparan sulfate-protein interactions. *Annu Rev Biochem* 83:129–157.
35. Wood B, K Boyle E (2016) Hominin taxic diversity: Fact or fantasy? *Am J Phys Anthropol* 159:537–578.
36. Harcourt-Smith W (2016) Early hominin diversity and the emergence of the genus Homo. *J Anthropol Sci* 94:19–27.
37. Lordkipanidze D, et al. (2013) A complete skull from Dmanisi, Georgia, and the evolutionary biology of early Homo. *Science* 342:326–331.
38. Bermúdez de Castro JM, et al. (1997) A hominid from the lower Pleistocene of Atapuerca, Spain: Possible ancestor to Neandertals and modern humans. *Science* 276: 1392–1395.
39. Campaña I, et al. (2016) New interpretation of the Gran Dolina-TD6 bearing Homo antecessor deposits through sedimentological analysis. *Sci Rep* 6:34799.
40. Berger LR, et al. (2015) Homo naledi, a new species of the genus Homo from the Dinaledi Chamber, South Africa. *Elife* 4:e09560.
41. Dembo M, et al. (2016) The evolutionary relationships and age of Homo naledi: An assessment using dated Bayesian phylogenetic methods. *J Hum Evol* 97:17–26.
42. Berger LR, et al. (2010) Australopithecus sediba: A new species of Homo-like australopithecine from South Africa. *Science* 328:195–204.
43. Pickering R, et al. (2011) Australopithecus sediba at 1.977 Ma and implications for the origins of the genus Homo. *Science* 333:1421–1423.
44. Indriati E, et al. (2011) The age of the 20 meter Solo River terrace, Java, Indonesia and the survival of Homo erectus in Asia. *PLoS One* 6:e21562.
45. Callaway E, et al. (2014) The discovery of Homo floresiensis: Tales of the hobbit. *Nature* 514:422–426.
46. Culotta E (2016) HUMAN ORIGINS. Likely hobbit ancestors lived 600,000 years earlier. *Science* 352:1260–1261.
47. Auwerx J (1991) The human leukemia cell line, THP-1: A multifaceted model for the study of monocyte-macrophage differentiation. *Experientia* 47:22–31.
48. Esko JD (2001) Special considerations for proteoglycans and glycosaminoglycans and their purification. *Curr Protoc Mol Biol*, 10.1002/0471142727.mb1702s22.
49. Shaw L, Schauer R (1988) The biosynthesis of N-glycolylneuraminic acid occurs by hydroxylation of the CMP-glycoside of N-acetylneuraminic acid. *Biol Chem Hoppe Seyler* 369:477–486.
50. Bulai T, Bratosin D, Artenie V, Montreuil J (2002) Characterization of a sialate pyruvate-lyase in the cytosol of human erythrocytes. *Biochimie* 84:655–660.
51. Schauer R, Sommer U, Krüger D, van Unen H, Traving C (1999) The terminal enzymes of sialic acid metabolism: Acylneuraminase pyruvate-lyases. *Biosci Rep* 19:373–383.
52. Mio T, Yamada-Okabe T, Arisawa M, Yamada-Okabe H (2000) Functional cloning and mutational analysis of the human cDNA for phosphoacetylglucosamine mutase: Identification of the amino acid residues essential for the catalysis. *Biochim Biophys Acta* 1492:369–376.
53. Strominger JL, Smith MS (1959) Uridine diphosphoacetylglucosamine pyrophosphorylase. *J Biol Chem* 234:1822–1827.
54. Glowacka D, Zwierz K, Gindziński A, Gałasiński W (1978) The metabolism of UDP-N-acetyl-D-glucosamine in the human gastric mucous membrane. II. The activity of UDP-N-acetylglucosamine 4-epimerase (E.C.5.1.3.7.). *Biochem Med* 19:202–210.

# Supporting Information

Bergfeld et al. 10.1073/pnas.1706306114



**Fig. S1.** Synthesis of partially *N*-glycolylated CS. (A) Schematic representation of the procedure to generate partially *N*-glycolylated CS from commercial-grade shark cartilage CS. (B) The extracted ion current chromatograph for [ $^{13}\text{C}_6$ ]aniline-labeled nonmodified singly sulfated disaccharide residues present after chemical modification. The relative abundances of both D0a4 and D0a6 are consistent with unmodified starting material. (C) The extracted ion current chromatograph for the corresponding [ $^{13}\text{C}_6$ ]aniline-labeled glycolylated disaccharides D0q4 and D0q6. (D) The product ion mass spectrum for D0q4. (E) The product ion mass spectrum for D0q6.



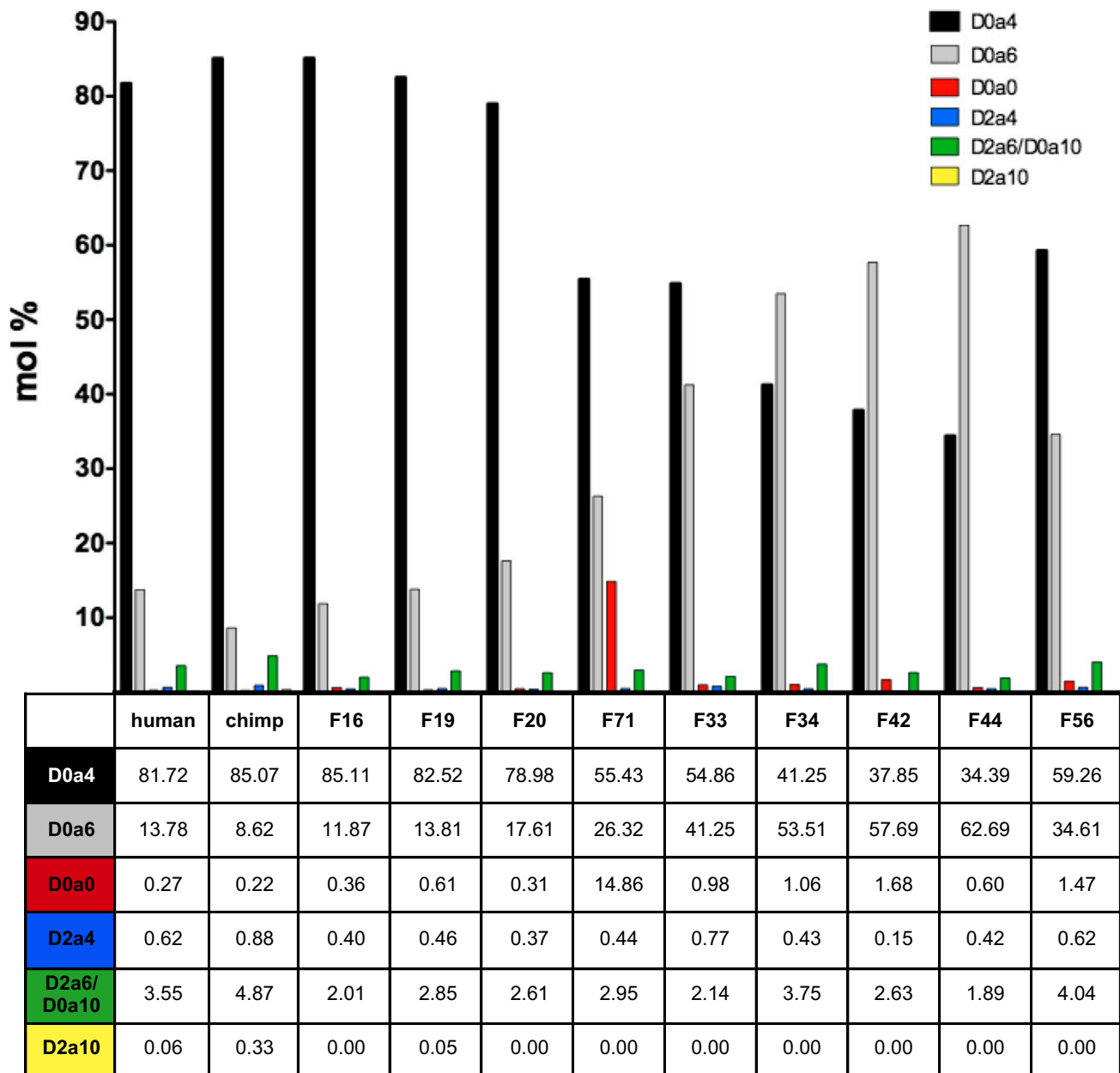
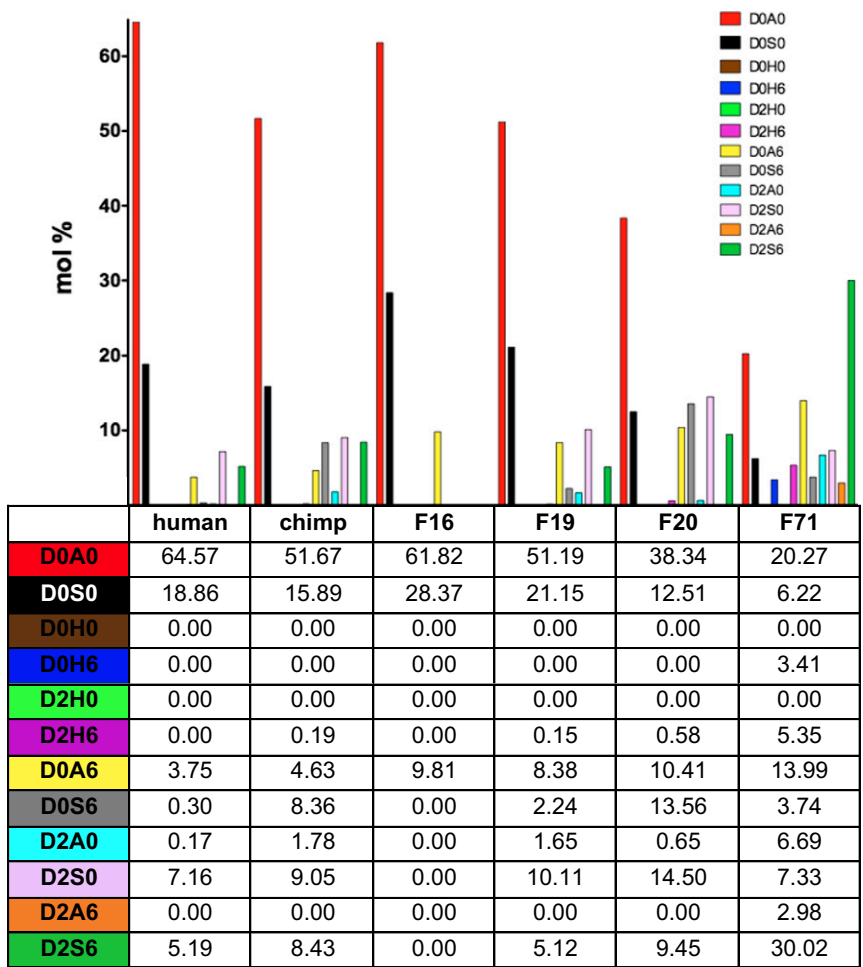


Fig. S2. Disaccharide compositional analyses of CS. Relative abundance of CS disaccharides in contemporary (human and chimpanzee bone) compared with fossil specimens expressed as molar percent (mol %).



**Fig. S3.** Disaccharide compositional analyses of HS. Relative abundance of HS disaccharides in contemporary (human and chimpanzee bone) compared with fossil specimens expressed as molar percent (mol %).

**Table S1. Detection of naturally occurring *N*-glycolylated CS**

CS purified from	Presence of Gc-CS
WT mouse liver	Yes (Fig. 3)
<i>Cmah</i> <sup>-/-</sup> mouse liver	No (Fig. 3)
WT mouse kidney	Yes
<i>Cmah</i> <sup>-/-</sup> mouse kidney	No
WT mouse serum	Yes (Fig. 4)
<i>Cmah</i> <sup>-/-</sup> mouse serum	No (Fig. 4)
Bovine muscle (beef)	Yes
Porcine muscle (pork)	Yes
Ovine muscle (lamb)	Yes (Fig. 3)
FCS	Yes
Bovine trachea (commercial CS)	Yes
Porcine intestinal mucosa (commercial CS)	Yes
Chimpanzee serum	Yes (Fig. 4)
Human AB serum	Trace (Fig. 4)
Human serum (S-8)	Trace
Human serum (S-19)	Trace
Human serum (S-30)	Trace
Human serum (S-34)	Trace
Human serum (S-39)	No
Human serum (S-46)	No
Human serum (S-54)	Trace
Human serum (S-55)	Trace
Human serum (S-56)	Trace
<i>Cmah</i> <sup>-/-</sup> mouse serum from Neu5Gc-fed mice	Trace (Fig. 4)
Shark cartilage (commercial CS)	No (Fig. S1)

Summary of samples analyzed for the presence of *N*-glycolylated CS.



Table S2. Detection of *N*-glycolyl-CS in fossil specimens

Name	Origin	Location	Estimated age	Description/site where collected	Total amount of CS, $\mu\text{g/g}$	Total amount of HS, $\mu\text{g/g}$	<i>N</i> -glycolyl-CS detectable?
Human	Human (humerus)	USA	Contemporary	Provided by USCD Anatomical Preparation	139.2	337.8	Yes (*)
Chimpanzee	Chimpanzee (vertebrae)	USA	Contemporary	Provided by CARTA	68.1	657.4	Yes (**)
F16	Mammoth (astragalus)	Dickinson Ct, Kansas, USA	~12,000 y	Forals Fossils (Alan Foral, 5417 Calvert, Lincoln, NE 68506)	2.6	11.1	Yes (*)
F19	Elk (antler)	Dickinson Ct, Kansas, USA	~12,000 y	Forals Fossils (Alan Foral, 5417 Calvert, Lincoln, NE 68506)	28.8	106.4	Yes (****)
F20	Cave bear (jaw bone)	Romania	40,000–80,000 y	Canada Fossils (John Issa, 3333 Eighth Street SE, Calgary, Alberta T2G3A4)	4.6	178.7	Yes (++)
F71	<i>H. Erectus</i> (Sangiran 27, skull)	Sangrian, Java, Indonesia	1.65 My	Teuku Jacob, Ety Indriati	0.2	86.1	n.d.
F33	Bovidae	West Turkana Kenya	1.47 My	Nariokotome 3 (NK3) next to excavation of KNM-WT 15,000, <i>H. erectus</i> skeleton	0.054	n.a.	n.d.
F34	Bovidae (mandible fragment, WT 4608)	West Turkana, Kenya	Between 0.7 and 1.4 My	Nachukui 1 (NC1) above Chari Tuff	0.089	n.a.	Yes (++++)
F42	Bovid or hippo (occipital condyle fragment)	West Turkana, Kenya	3.4 My	Lomekwi 4 (L04), level of Tulu Bor tuff	0.013	n.a.	Yes (+)
F44	Mammal (vertebra fragment)	West Turkana, Kenya	3.4 My	Lomekwi 4 (L04), level of the Tulu Bor tuff	0.026	n.a.	Yes (++)
F56	Mammal (long bone fragment)	Allia bay, Kenya	3.9 My	Collected in excavation 261-1 with <i>Australopithecus anamensis</i>	0.051	n.a.	Yes (+)

Total amounts of CS and HS were calculated by MS using a [<sup>13</sup>C]laniline-tagged internal standard. The presence of *N*-glycolyl-CS in all samples is provided with the number of crosses and asterisks indicating relative abundance of D0q4 among the two sample sets. n.a., not analyzed; n.d., not detectable.