



HAL
open science

Genetic and enzymatic characterization of 3-O-sulfotransferase SNPs associated with *Plasmodium falciparum* parasitaemia

Ngoc Thy Nguyen, Romain R Vivès, Magali Iché-Torres, Vincent Delauzun,
Els Saesen, Veronique Roig-Zamboni, Hugues Lortat-Jacob, Pascal Rihet,
Yves Bourne

► **To cite this version:**

Ngoc Thy Nguyen, Romain R Vivès, Magali Iché-Torres, Vincent Delauzun, Els Saesen, et al.. Genetic and enzymatic characterization of 3-O-sulfotransferase SNPs associated with *Plasmodium falciparum* parasitaemia. *Glycobiology*, 2018, 28 (7), pp.534-541. 10.1093/glycob/cwy038 . hal-02045171

HAL Id: hal-02045171

<https://amu.hal.science/hal-02045171>

Submitted on 22 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Genetic and enzymatic characterization of 3-O-sulfotransferase SNPs associated with *Plasmodium falciparum* parasitaemia

Ngoc Thy Nguyen^{1,2¶}, Romain R. Vivès³, Magali Torres¹, Vincent Delauzun², Els Saesen³, Véronique Roig-Zamboni², Hugues Lortat-Jacob³, Pascal Rihet^{1*}, Yves Bourne^{2*}

¹ Aix-Marseille Univ, INSERM, TAGC, Marseille, France

² Aix-Marseille Univ, CNRS, AFMB, Marseille, France

³ Univ. Grenoble Alpes, CNRS, CEA, IBS, Grenoble, France

* **co-corresponding authors:** yves.bourne@afmb.univ-mrs.fr, Tel: +33(0)491825566 or pascal.rihet@univ-amu.fr, Tel: +33(0)491828723

¶ **Present address:** University of Science and Technology of Hanoi (USTH), Vietnam Academy of Science and Technology (VAST), Hanoi, Vietnam

Running title: Human 3-OST-3 SNPs effect on promoter and catalytic activities

Keywords: 3-O-sulfotransferase, heparin sulfate biosynthesis, malaria infection, genetic association, enzymatic activity

Abstract (179 words)

The *HS3ST3A1/B1* genes encode two homologous 3-O-sulfotransferases involved in the late modification step during heparan sulfate (HS) biosynthesis. In addition to the SNPs rs28470223 (C>T) in the promoter region of both *HS3ST3A1* and rs62636623 (Gly/Arg) in the stem region of *HS3ST3B1*, three missense mutations (rs62056073, rs61729712 and rs9906590) located within the catalytic sulfotransferase domain of 3-OST-B1 are linked and associated to *P. falciparum* parasitaemia. To ascertain the functional effects of these SNP associations, we investigated the regulatory effect of rs28470223 and characterized the enzymatic activity of the missense SNP rs61729712 (Ser279Asn) localized at proximity of the substrate binding cleft. The SNP rs28470223 results in decreased promoter activity of *HS3ST3A1* in K562 cells, suggesting a reduced *in vivo* transcription activity of the target gene. A comparative kinetic analysis of *wt* HS3ST3B1 and the Ser269Asn variant (rs61729712) using a HS-derived oligosaccharide substrate reveals a slightly higher catalytic activity for the SNP variant. These genetic and enzymatic studies suggest that genetic variations in enzymes responsible of HS 3-O sulfation can modulate their promoter and/or enzymatic activities and may influence *P. falciparum* parasitaemia.

Introduction

Heparan sulfate (HS) is a family of highly N- and O-sulfated glycosaminoglycans that is commonly found on the mammalian cell surface and in the extracellular matrix. HS is attached to a variety of different core proteins defined as HS proteoglycans (HSPGs) and participates in a wide range of physiological and pathophysiological functions, including embryonic development, inflammatory responses, blood coagulation, and assisting viral/bacterial infections (Bishop et al. 2007). HS consists of a disaccharide repeating unit with glucuronic acid (GlcA) or iduronic acid (IdoA) and glucosamine, each of which is capable of carrying sulfate groups and contributes to dictate HS function (Bishop et al. 2007). The HS maturation process involves a series of biosynthetic enzymes, including bifunctional enzymes (NSDT, N-deacetylase/N-sulfotransferase-1), a critical step for formation of N-sulfated domains, a C5-epimerase and various O-sulfotransferases (Fu et al. 2016). The three classes of O-sulfotransferases transfer sulfate groups at various positions, including C2 of IdoA, C6 and C3 of the GlcN units, leading to clusters of highly sulfated regions (NS domain) separated by long stretches of unsulfated GlcNAc-containing disaccharides (NA domain). The sulfate-rich regions are likely to generate highly flexible regions because of their high content of the conformationally versatile IdoA and IdoA2S residues (Turnbull and Gallagher 1991). In addition, tissue specific expression of different enzyme isoforms fine-tunes synthesis of specific HS structures, which are predominantly determined by a controlled positioning of N-, 2-, 6-, and 3-O-sulfate groups along HS chains, to confer an important regulatory role in various biological processes (Esko and Selleck 2002).

The 3-O-sulfation of GlcN unit is a relatively rare modification, present in only a limited number of HS polymers or absent entirely, and occur at a late step during the biosynthetic pathway (Thacker et al. 2014). Conversely, the HS 3-O-sulfotransferases (3-OST-3 that are encoded by *HS3ST3* genes) represent the largest gene family among all HS sulfotransferases and vertebrates generally possess seven isozymes divided into two subgroups according to sequence homology of their sulfotransferase domain (Liu and Pedersen 2007). Based on the large diversity of 3-OST enzymes and the fact that they probably act at a late stage, one might assume that they show selectivity for the sulfation pattern in their target sites. One group (3-OST-2,-3A,-3B,-4 and -6) is often referred to as “gD-type” HS3STs, because members of the subfamily can generate binding sites for glycoprotein gD of type I herpes simplex virus (HSV-1). Members of the second group (3-OST-1, -5) have in common the capacity to generate a binding site for antithrombin and are thus designed “AT-type” HS3STs. Hence, 3-OST-1

preferentially modifies GlcN sites in which GlcA at position +1 is devoid of 2-O-sulfate groups (Mochizuki et al. 2008). In contrast, 3-OST-2, -3, -4 and -6 (gD-type) preferentially modify GlcN sites in which position -1 is IdoA-2S (Meissen et al. 2009). Unlike 3-OST-1, the 3-OST-2, -3, -3A and -3B enzymes are predicted to have type II integral membrane architecture (Shworak et al. 1999). Both mouse and human forms of 3-OST-3B and a human form of 3-OST-3A can modify the HS of HSV-1-resistant cells at specific sites to generate cells susceptible to HSV-1 infection (Shukla et al. 1999).

In addition to their functional implications in various cell events, HS proteoglycans can play an important role in infectious diseases, such as malaria caused by the parasite *Plasmodium*, involving the HS chains of both the mammalian host and the vector. Several studies suggest that the outcome of malaria infection may be influenced by variation in HS composition, owing to genetic variations within genes encoding the biosynthesis enzymes (Coppi et al. 2007; Sinnis et al. 2007; Armistead et al. 2011). The genes encoding the 3-O sulfotransferases are located in different chromosomal regions except for *HS3ST3A1* and *HS3ST3B1* which encode the 3-OST-3A and 3-OST-3B sulfotransferases, respectively, and target the same disaccharide (Liu, Shriver, et al. 1999). In fact, those two genes are located within chromosome 17p12 that showed a genome-wide significant linkage with *P. falciparum* parasitaemia in a population living in Burkina Faso (Brisebarre et al. 2014). Moreover, genetic variations within *HS3ST3A1* and *HS3ST3B1* are linked to *P. falciparum* parasitaemia in humans, as found in a population living in an endemic area in Burkina Faso (Atkinson et al. 2012). Alteration of HS 3-O-sulfation may affect i) the binding of *P. falciparum* antigen on host cells and/or ii) the pro-inflammatory response. To date, several single nucleotide polymorphisms (SNPs) have been identified in the promoter region of both *HS3ST3A1/B1*, and one of them (rs28470223, C/T), which is located within the promoter of *HS3ST3A1*, was linked and associated with *P. falciparum* parasitaemia. Moreover, four missense SNPs (rs62636623 G>C Gly83Arg, rs62056073 A>G Ile196Val, rs61729712 G>A Ser269Asn and rs9906590 G>A Glu363Lys) are located within *HS3ST3B1*, of which the later three SNPs are located within the sulfotransferase domain. These missense SNPs were genetically linked to parasitaemia and were associated with parasitaemia in combination with other SNPs, suggesting that they may alter 3-OST-3 functionality. While epidemiological correlation studies relate naturally occurring variations in 3-O-sulfotransferases to the susceptibility for infectious diseases in these populations, no functional data are yet available for most of these genetic variants.

To ascertain the functional role of the SNPs localized either in the promotor region or within the sulfotransferase domain of the HS 3-OST-3 modification enzyme, we investigated the effect of the SNP rs28470223 on the expression level of the *HS3ST3A1* gene and the effect of the Ser269Asn (rs61729712) variant on the enzymatic activity of human 3-OST-3B1. Bioinformatics analyses reveal that seven SNPs in *HS3ST3A1/B1* shows a predicted deleterious effect, confirming that genetic variations within the genes encoding 3-O-sulfotransferases may affect the susceptibility to infectious diseases, such as malaria. The SNP rs28470223 is associated with a significantly altered promoter activity that may affect both the transcriptional activity and *P. falciparum* parasitaemia in humans. Complementary enzyme kinetic analyses of *wt* human 3-OST-3 and the Ser269Asn (rs61729712) variant, which is located at the entry of the substrate binding cleft based on the crystal structure of the homologous 3-OST-3A1 isoform, reveal a slightly higher catalytic activity compared to the *wt* enzyme.

Results and discussion

Bioinformatics analysis of the phenotypic effect of HS3ST3A1 and HS3ST3B1 genetic variants

All SNPs previously identified in the chromosomal region containing *HS3ST3A1* and *HS3ST3B1* from a population living in an endemic area in Burkina Faso (Atkinson et al. 2012) were investigated using the bioinformatics PredictSNP2 tool, which integrates binary predictions and uniform confidence values computed by six different tools. The comparative analysis, which provides a consensus score based on the five best-performing tools, reveals that seven SNPs show a predicted deleterious effect, including three (rs62056073 A>G Ile196Val, rs61729712 G>A Ser269Asn and rs9906590 G>A Glu363Lys) out of the four missense variants (**Table I**). In contrast, rs28470223 (C>T) could not be predicted for deleterious effects using the PredictSNP2 tool in contrast to the alternative FunSeq2 and GWAVA predicting tools.

SNP rs28470223 regulates transcriptional promoter activity of HS3ST3A1

We further used bioinformatics to investigate the regulatory effect of the SNP rs28470223 (C>T) associated with a malaria phenotype (Atkinson et al. 2012). The location of rs28470223 within the promoter suggested that this SNP is a cis-regulatory variant controlling *HS3ST3A1* gene expression. We first map the peaks from ChIP-seq experiments covering a total of 485 transcription factors using the Remap atlas (Griffon et al. 2015; Chèneby et al. 2018) and found 56 transcription factors that bind to rs28470223-containing genomic sequences (**Table S1**), indicating a specific regulatory region. We further assess a possible regulatory functional effect of the SNP rs28470223 on the binding sites of transcription factors using the perfectos-ape tool (E. Vorontsov et al. 2015) based on the hocomoco v1.1 motif database (Kulakovskiy et al. 2018). The SNP rs28470223 alters binding of several transcription factors (**Table S2**), predicting a functional effect of this SNP in modifying transcription factor binding affinities to rs28470223-containing genomic sequences. The P value decreases more than 10-fold for the allele T compared to the allele C for the AP2D and KAISO transcription factors whereas it increases more than 10-fold for the BRAC and TCFP2 transcription factors (**Table S2**). Finally, the SNP rs28470223 was found to be an expression quantitative trait locus (eQTL) ($P = 3.06 \cdot 10^{-6}$) in the Genotype-Tissue Expression database (Consortium 2013). Overall, the bioinformatics analysis supports a cis-regulatory effect of the SNP rs28470223.

We next perform 6 gene reporter assays to evaluate the functional effect of the SNP rs28470223 on the *HS3ST3A1* promoter activity in K562 cells. As a result, the minor allele (rs28470223-T) significantly decreases the promoter activity by ~54% compared to the wild type allele (rs28470223-C) ($P=6 \times 10^{-5}$) (**Figure 1**), confirming the regulatory function of rs28470223 on the promoter transcriptional activity of *HS3ST3A1*.

A missense SNP is located within the catalytic domain of 3-OST-3 on the reducing end of the large open cleft

The catalytic domain of human 3OST-3B1 shares 99% sequence identity with that of 3-OST-3A1 for which the structure has been determined, arguing that this functionally important region is actively maintained by gene conversion between the *3-OST-3A* and *3-OST-3B* loci; each gene being duplicated in human (Shworak et al. 1999). The catalytic domain of 3-OST-3 displays an overall α/β motif typically found in sulfotransferases with a large open cleft running across the active site. Central to this structural motif are the phosphosulfate binding loop (P-loop like motif, Lys162-Arg166) of the high-energy sulfate donor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) and a conserved glutamate acting as the catalytic base, a binding site well conserved in sulfotransferases (Liu et al. 2012). The acceptor site for sulfonation, referred as the acceptor sugar, lies at the center of the cleft and corresponds to the position of a GlcNS6S residue (T-2 in 3-OST-3/H-3 in 3-OST-1) (**Figure 2**) (Moon et al. 2004; Moon et al. 2012).

Given that substrate specificity in HS O-sulfotransferases could be determined by residues distal to the catalytic site, we hypothesized that the three naturally occurring 3-OST-3 SNP isoforms would display differential 3-O-sulfotransferase activities compared to the wild-type enzyme. To investigate potential functional effects of the coding SNPs on the 3-OST-3 catalytic activity, we mapped the three missense SNPs onto the crystal structure of 3-OST-3 in ternary complex with PAP and a tetrasaccharide (Moon et al. 2004). This structural comparison was extended to the crystal structure of 3-OST isoform 1 in ternary complex with PAP and an heptasaccharide (Moon et al. 2012). The three missense SNPs are located in loop regions at the surface and hardly alter enzyme functionality (**Figure 2**). Yet, the Ser269Asn SNP is located at the tip of helix $\alpha 8$ and faces the entry of the large substrate binding at site T-4/H-5 on the reducing end at proximity of the GlcNS6S N-sulfate group (tetrasaccharide complex in 3-OST-3) or 6-O-sulfate group (heptasaccharide complex in 3-OST-1). Sequence analysis of a subset of the 3-O-sulfotransferase family members reveals that Ser269 is conserved across mammalian

3-OST-3 orthologs but variability is tolerated at this position in other 3-OST-3 isoforms (Asn in 3-OST-2 and Lys in 3-OST-1/5).

The Ser269Asn SNP variant exhibits a slightly higher catalytic activity

We expressed and purified human *wt* 3-OST-3B1 and the Ser269Asn variant using a mammalian cell expression system and anti-FLAG affinity purification tag. Expression of *wt* 3-OST-3B1 and the Ser269Asn variant clearly shows a band at 37 kDa by Western Blot analysis corresponding to the recombinant proteins. We used an HS-derived octasaccharide substrate (HS8) as a mimic of a natural HS substrate to identify with a HPLC-based assay any differences in enzyme activities between the wildtype and the Ser269Asn variant. HS8 was prepared by depolymerisation of HS with heparinase III, an enzyme that specifically cleaves GlcNAc-containing disaccharide units, thereby releasing intact NS-domain-like oligosaccharides. Digestion products were purified using combined size-exclusion and anion-exchange chromatographies to yield size and charge homogenous oligosaccharide species (Pye et al. 1998). Amongst these, HS8 was selected for the 3-OST enzymatic assay as this oligosaccharide displays an average sulfation content suitable for 3-OST-3B1 activity (**Figure S1**). HS8 was incubated with *wt* 3-OST-3B1 or the Ser269Asn variant and reaction was monitored by measuring the disaccharide composition using a dedicated HPLC-based assay (see Experimental procedures). HPLC analysis of disaccharides present in untreated HS8 yields four major peaks, enabling determination of the disaccharide composition after integration and normalization of the peak areas (**Table S3**). In contrast, a similar disaccharide analysis of 3-OST-3B1-treated HS8 showed the appearance and time-dependent increase of an unknown, late-eluting saccharide species (**Figure 3**). This species most likely corresponds to a heparinase-resistant 3-O sulfated tetrasaccharide, as previously reported (Yamada et al. 1993). Concomitantly, the 3-OST-3B1 activity decreases the Δ HexA2S,GlcNS disaccharide content while the pattern of other HS8 disaccharides (i.e. Δ HexA,GlcNAc, Δ HexA,GlcNS and Δ HexA2S,GlcNS6S) remains unchanged (**Figure 3**). These data clearly indicate that the 3-OST-3B1-catalyzed 3-O-sulfation event on HS8 is non-random and may preferentially target IdoA2S-GlcNS units, in agreement with previous studies and on the fact that the 3-OST-3 modification must precede the 6-O-sulfation step (Liu, Shworak, et al. 1999; Thacker et al. 2014; Wang, P.H. Hsieh, et al. 2017). Instead, the 3-OST-1 modification occurs only after 6-O-sulfation to generate the -GlcA,GlcNS3S6S- disaccharide unit (Wang, P.-H.H. Hsieh, et al. 2017).

A kinetic analysis of the Ser269Asn variant activity compared to *wt* 3-OST-3B1 reveals similar elution profiles, with a slightly faster increase of the peak * area associated to a concomitant faster decrease of the Δ HexA2S,GlcNS disaccharide content, as monitored by a time course treatment (**Figure 3**). This result points to a slightly higher catalytic activity for the Ser269Asn variant compared to the *wt* enzyme. In fact, structural analysis suggests that a bulkier Asn in the 3-OST-3B1 variant could provide additional polar interactions with the N- or 6-O-sulfate group of GlcNS6S at site T4 near the reducing end compared to a shorter Ser residue in *wt* 3-OST3. Site-directed mutagenesis of the adjacent Trp268 residue in 3-OST-3, which contributes to a sulfate-binding pocket near the reducing end of the substrate, resulted in a 41-fold reduction in catalytic efficiency, arguing for a functional role of the contiguous Ser269 in substrate binding (Moon et al. 2012). In addition to Ser269, other residues remote from the active site but located at the reducing end, including Thr256 in 3-OST-3 and Arg268 in 3-OST-1, were shown to play important roles in substrate binding and specificity between 3-OST isoforms (Moon et al. 2012).

Overall, this first genetic and biochemical analyses of two genetic variants in the *HS3ST3A1/B1* genes associated with *P. falciparum* parasitaemia offer first insights into the modulation of 3-OST-3 gene transcription and catalytic activity by SNPs. Whether the SNPs affected HS proteoglycan level and structure in the population living in an endemic area in Burkina Faso is unknown and merits further investigation.

Materials and methods:

In silico analysis of the SNP regulatory effect

We explored ChIP-seq results by using the ReMap atlas covering 485 transcription factors to identify transcription factors that bind DNA sequence containing the studied SNPs (Chèneby et al. 2018). To this aim, we used the ReMap tool to intersect the genomic coordinates of SNPs with those of the ChIP-seq peaks. Furthermore, we used the PERFECTOS-APE software to predict the effect of the SNPs on the binding of transcription factors (E. Vorontsov et al. 2015). Briefly, the extracted SNPs sequences and their alleles together with their flanking regions were scanned against a recent collection of motifs, which contains binding models of 680 human transcription factors (Kulakovskiy et al. 2018). The site score was obtained from position-specific scoring matrices on sequences containing each allele, allowing us to assess the P value for each allele and therefore the P value ratio. The candidate transcription factors was retained on the basis of the best P value obtained for one of the allele; it should be lower than 10^{-3} . The calculated P value ratio reflects the predicted effect of the SNP on the binding of the transcription factor on the SNP-containing DNA sequence.

Transient transfection and dual-luciferase reporter assay

K-562 cells (ATCC CLL-243) were grown in Gibco RPMI 1640 medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% fetal bovine serum. A 1200bp DNA fragment upstream the *HS3ST3A1* translation start site (chromosome 17: 13504472-13505672 according to the hg38 assembly) was cloned by gene synthesis (GeneCust Custom Services, Luxembourg). This fragment was cloned into the MluI-XhoI sites of the pGL3-basic vector (Promega, Madison, WI, USA), which contains the firefly luciferase coding sequence. Initially, the pGL3 construct contained the rs28470223 T-allele which represents the minor allele in Africa, whereas the rs28470223-C is the major allele in Africa and is considered as the wild type allele. To generate the pGL3 construct containing the rs28470223-C allele, site-directed mutagenesis was performed with the Q5 Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich, MA, USA) using primers (5'-AGGTTCTCTCGGCCAAGGAGC-3', 3'-GAACTCGGGCGGAGAGAA-5') designed by NEBaseChanger tool and provided by the supplier, and an annealing temperature of 63°C. K562 transfection was performed with the Neon™ Transfection System (Invitrogen) following manufacturer's instructions. For each of the 6 tests performed, 10^6 cells were co-transfected with 1 µg of the empty pGL3-basic control vector, or with positive pGL3-promoter control vector, or with 1µg of pGL3-basic vector

containing the rs28470223-T (p-T) or rs28470223-C (p-C) *HS3ST3A1* promoter sequence together with 200ng of the pRL-SV40 plasmid encoding Renilla luciferase (Promega, Madison, WI, USA), which was used as a transfection efficiency control. Transfected cells were maintained at 37°C in 5% CO₂ during 24h. Firefly and Renilla luciferase activities were monitored using a TriStar LB 941 Multimode Microplate Reader (Berthold technologies, Thermo Fisher Scientific, Waltham, MA, USA) by analyzing 20µl of cell lysate according to the manufacturer's instructions provided in the Dual Luciferase Kit (Promega). The firefly luciferase activity of each sample was normalized by the Renilla luciferase activity and adjusted by the pGL3-basic mean activity.

Statistical analysis

To assess the effect of SNPs on the promoter activity, Student's t-test was used after checking the normality of the distributions by using the Shapiro-Wilk method and their variance equality by using a Fisher test. All analyses were performed by using either R or SPSS software (SPSS, Boulogne, France). Only terms significant at the 5% level were retained.

Expression and purification of human 3-OST-3B1 and the Ser269Asn variant

Following unsuccessful attempts to express a soluble form of the human 3-OST-3B1 catalytic domain (Ile125-Asp390) in *E. coli*, either as a 6xHis- or thioredoxine-tagged form, a cDNA fragment encoding the catalytic domain of *wt* 3-OST-3B1 was cloned into the pYD7 vector suitable for expression in human embryonic kidney HEK293 EBNA cell line. The recombinant pYD7-3-OST-3B1 vector consists of a Kozak consensus sequence, a SEAP peptide signal and a Flag tag upstream of the catalytic domain for 3-OST-3B1 detection and purification. The pYD7 plasmid carrying the 3-OST-3B1 Ser269Asn variant (rs61729712) was amplified by PCR from the wild type construct using the forward 5'-TCGACACGTCGTGGAACGCCATCCAGATCGG-3' and reverse 5'-CCGATCTGGATGGCGTTCCACGACGTGTCGA-3' oligonucleotides. The PCR product was cloned into the linearized PYD7 plasmid DNA and the sequence was verified by sequencing. Plasmid DNA (2 µg) was transfected with 4 µg of linear polyethylenimine (Polysciences) into HEK293 EBNA cells in a 6-well culture plate. Transfected cells cultured in DMEM 10 % FBS were selected with blasticidin (5 µg/ml) for stable cell lines establishment. Higher expression and secretion levels of the protein produced by the cell lines were achieved in DMEM 2 % FBS, 1.25 mM sodium valproate (Sigma-Aldrich) and 0.5 % Tryptone N1

(Organotechnie). Protein productions were performed in multilayer culture flasks at 32°C instead 37°C in order to reduce protein aggregation. Protein secretion was analysed by Western Blot using an anti-FLAG tag antibody. Before purification, culture medium containing secreted protein was harvested after 7 days of culture and dialysed overnight at 4°C against 20 mM HEPES pH 8.0 and 100 mM NaCl. FLAG-tagged *wt* 3-OST-3B1 was purified by anti-FLAG M2 affinity (Sigma-Aldrich) and eluted in a buffer containing 20 mM HEPES pH 8.0, 100 mM NaCl and 100 µg/ml FLAG tag peptide (Sigma-Aldrich). Fractions containing 3-OST-3B1 were pooled and further purified by size exclusion chromatography on a Superdex 200 26/60 column (GE-Healthcare) equilibrated with 20 mM Tris pH 8.0 and 200 mM NaCl. The Ser269Asn variant was produced and purified as for *wt* 3-OST-3B1. Protein purity and integrity of *wt* 3-OST-3B1 and the Ser269Asn variant were analysed by SDS-PAGE electrophoresis and matrix-assisted laser desorption/ionization (MALDI-TOF) mass spectrometry, and concentrated by ultrafiltration to 1 mg/ml in 20 mM Tris-HCl pH 8.0, 100 mM NaCl and 1 mM PAP (Sigma-Aldrich).

Preparation of HS-derived octasaccharides

Porcine mucosal HS (Celsius) was extensively digested with 25 mU/ml of heparinase III (Grampian Enzyme) in 5 mM Tris-HCl, 2 mM CaCl₂, 0.1 mg/mL BSA, 50 mM NaCl, pH 7.5 for 72 h at 30°C. The resulting mixture was resolved onto a 1500 mm x 44 mm Biogel P10 column (Bio-Rad) in 200 mM NaCl at 1 mL/min, leading to a graded series of size-uniform oligosaccharides from disaccharide to octadecasaccharide. Samples were dialyzed against distilled water, freeze-dried and quantified. The octasaccharides were further purified by strong-anion-exchange HPLC, on a 9 x 250 mm preparative ProPac PA1 column (Dionex). After equilibration in mobile phase (distilled water adjusted to pH 3.5 with HCl) at 1 mL/min, samples (7.5 mg) were injected and eluted with a gradient of NaCl (0 to 0.5 M over 10 min, then 0.5 to 1 M over 80 min) in the same mobile phase. The eluate was monitored on-line for UV absorbance at 232 nm, and 23 different fractions were collected, dialyzed against distilled water, freeze-dried, quantified and analyzed as described below. Fraction HS8 was used as a substrate to monitor the 3-OST-3B1 catalytic activity.

Digestion of HS8 oligosaccharide with 3-OST-3B1

Lyophilized HS8 was resuspended in 50mM MES pH 7.0, 5mM MgCl₂, 200µM PAPS and incubated with 1µg of 3-OST-3B1 (wild type or variant) for 24h at 37°C. For time-course digestion, aliquots were taken off after 0, 1, 3, 6 and 24 hours. Samples were then boiled for 5

minutes to inactivate the enzyme, then stored at -20°C.

Disaccharide analysis

Disaccharide analysis of the HS8 samples was performed as previously described (Henriet et al. 2017). Briefly, samples in 100 mM sodium acetate, 0.5 mM calcium acetate, pH 7.1 were digested into disaccharides by incubation with a cocktail of heparinase I, II and III (10 mU each) overnight at 37°C. Disaccharide composition was determined by RPIP-HPLC, by injection on a Luna 5µm C18 reversed phase column (4.6 x 300 mm, Phenomenex, Le Pecq, France) equilibrated at 0.5 mL/min in 1.2 mM tetra-N-butylammonium hydrogen sulfate (TBA) in 8.5% acetonitrile. Disaccharides were then resolved using a multi-step NaCl gradient (0–30 mM in 1 min, 30–90 mM in 39 min, 90–228 mM in 2 min, 228 mM for 4 min, 228-300 mM in 2 min, 300 mM for 4 min) calibrated with HS disaccharide standards (Iduron, Alderley Edge, UK). On-line post-column disaccharide derivatization was achieved by addition of 2-cyanoacetamide (0.25%) in NaOH (0.5%) at a flow rate of 0.16 mL/min, followed by fluorescence detection (excitation 346 nm, emission 410 nm).

Structural analysis

A structural analysis was performed to map the three 3-OST-3B1 missense SNPs on the crystal structure of 3-OST-3 bound to PAP and a tetrasaccharide substrate (accession code 1T8U) or 3-OST-1 bound to PAP and heptasaccharide substrate (3UAN). The rmsd value between the two structures is 0.95 Å for 244 Cα atoms (44% sequence identity).

Acknowledgements:

TNN was supported by a PhD fellowship from the Vietnamese government. We thank Melanie Daligault (TAGC, Marseille) for helpful assistance with bioinformatics analyses, Ahmad Ali-Ahmad and Pascale Marchot (AFMB, Marseille) for helpful assistance with biochemical analyses. This work was supported in part by the CNRS, INSERM and Aix-Marseille University, the GDR GAG (GDR 3739), the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01 and the “Investissements d’avenir” program Glyco@Alps (ANR-15-IDEX-02).

References:

- Armistead JS, Wilson IBH, van Kuppevelt TH, Dinglasan RR. 2011. A role for heparan sulfate proteoglycans in *Plasmodium falciparum* sporozoite invasion of anopheline mosquito salivary glands. *Biochem J* 438:475–483. doi:10.1042/BJ20110694.
- Atkinson A, Garnier S, Afridi S, Fumoux F, Rihet P. 2012. Genetic variations in genes involved in heparan sulphate biosynthesis are associated with *Plasmodium falciparum* parasitaemia: a familial study in Burkina Faso. *Malar J* 11:108. doi:10.1186/1475-2875-11-108.
- Bishop JR, Schuksz M, Esko JD. 2007. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* 446:1030–1037. doi:10.1038/nature05817.
- Brisebarre A, Kumulungui B, Sawadogo S, Atkinson A, Garnier S, Fumoux F, Rihet P. 2014. A genome scan for *Plasmodium falciparum* malaria identifies quantitative trait loci on chromosomes 5q31, 6p21.3, 17p12, and 19p13. *Malar J* 13:198. doi:10.1186/1475-2875-13-198.
- Chèneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. 2018. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res* 46:D267–D275. doi:10.1093/nar/gkx1092.
- Consortium Gte. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45:580–585. doi:10.1038/ng.2653.
- Coppi A, Tewari R, Bishop JR, Bennett BL, Lawrence R, Esko JD, Billker O, Sinnis P. 2007. Heparan sulfate proteoglycans provide a signal to Plasmodium sporozoites to stop migrating and productively invade host cells. *Cell Host Microbe* 2:316–327. doi:10.1016/j.chom.2007.10.002.
- E. Vorontsov I, V. Kulakovskiy I, Khimulya G, D. Nikolaeva D, J. Makeev V. 2015. PERFECTOS-APE - Predicting Regulatory Functional Effect of SNPs by Approximate P-value Estimation. In: *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms*. Vol. 1. BIOSTEC 2015. (BIOINFORMATICS). p. 102–108.
- Esko JD, Selleck SB. 2002. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu Rev Biochem* 71:435–471. doi:10.1146/annurev.biochem.71.110601.135458.
- Fu L, Suflita M, Linhardt RJ. 2016. Bioengineered heparins and heparan sulfates. *Adv Drug Deliv Rev* 97:237–249. doi:10.1016/j.addr.2015.11.002.
- Griffon A, Barbier Q, Dalino J, van Helden J, Spicuglia S, Ballester B. 2015. Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res* 43:e27. doi:10.1093/nar/gku1280.
- Henriet E, Jäger S, Tran C, Bastien P, Michelet J-FF, Minondo A-MM, Formanek F, Dalko-Csiba M, Lortat-Jacob H, Breton L, et al. 2017. A jasmonic acid derivative improves skin healing and induces changes in proteoglycan expression and glycosaminoglycan structure. *Biochim Biophys Acta* 1861:2250–2260. doi:10.1016/j.bbagen.2017.06.006.
- Kulakovskiy I V, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, Medvedeva YA, Magana-Mora A, Bajic VB, Papatsenko DA, et al. 2018. HOCOMOCO: towards a complete

- collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* 46:D252–D259. doi:10.1093/nar/gkx1106.
- Liu J, Moon AF, Sheng J, Pedersen LC. 2012. Understanding the substrate specificity of the heparan sulfate sulfotransferases by an integrated biosynthetic and crystallographic approach. *Curr Opin Struct Biol* 22:550–557. doi:10.1016/j.sbi.2012.07.004.
- Liu J, Pedersen LC. 2007. Anticoagulant heparan sulfate: structural specificity and biosynthesis. *Appl Microbiol Biotechnol* 74:263–272. doi:10.1007/s00253-006-0722-x.
- Liu J, Shriver Z, Blaiklock P, Yoshida K, Sasisekharan R, Rosenberg RD. 1999. Heparan sulfate D-glucosaminyl 3-O-sulfotransferase-3A sulfates N-unsubstituted glucosamine residues. *J Biol Chem* 274:38155–38162.
- Liu J, Shworak NW, Sinaÿ P, Schwartz JJ, Zhang L, Fritze LM, Rosenberg RD. 1999. Expression of heparan sulfate D-glucosaminyl 3-O-sulfotransferase isoforms reveals novel substrate specificities. *J Biol Chem* 274:5185–5192.
- Meissen JK, Sweeney MD, Girardi M, Lawrence R, Esko JD, Leary JA. 2009. Differentiation of 3-O-sulfated heparin disaccharide isomers: identification of structural aspects of the heparin CCL2 binding motif. *J Am Soc Mass Spectrom* 20:652–657. doi:10.1016/j.jasms.2008.12.002.
- Mochizuki H, Yoshida K, Shibata Y, Kimata K. 2008. Tetrasulfated disaccharide unit in heparan sulfate: enzymatic formation and tissue distribution. *J Biol Chem* 283:31237–31245. doi:10.1074/jbc.M801586200.
- Moon AF, Edavettal SC, Krahn JM, Munoz EM, Negishi M, Linhardt RJ, Liu J, Pedersen LC. 2004. Structural analysis of the sulfotransferase (3-o-sulfotransferase isoform 3) involved in the biosynthesis of an entry receptor for herpes simplex virus 1. *J Biol Chem* 279:45185–45193. doi:10.1074/jbc.M405013200.
- Moon AF, Xu Y, Woody SM, Krahn JM, Linhardt RJ, Liu J, Pedersen LC. 2012. Dissecting the substrate recognition of 3-O-sulfotransferase for the biosynthesis of anticoagulant heparin. *Proc Natl Acad Sci U S A* 109:5265–5270. doi:10.1073/pnas.1117923109.
- Pye DA, Vives RR, Turnbull JE, Hyde P, Gallagher JT. 1998. Heparan sulfate oligosaccharides require 6-O-sulfation for promotion of basic fibroblast growth factor mitogenic activity. *J Biol Chem* 273:22936–22942. doi:9722514.
- Saesen E, Sarrazin S, Laguri C, Sadir R, Maurin D, Thomas A, Imberty A, Lortat-Jacob H. 2013. Insights into the mechanism by which interferon- γ basic amino acid clusters mediate protein binding to heparan sulfate. *J. Am. Chem. Soc.* 135:9384–9390. doi:10.1021/ja4000867.
- Shukla D, Liu J, Blaiklock P, Shworak NW, Bai X, Esko JD, Cohen GH, Eisenberg RJ, Rosenberg RD, Spear PG. 1999. A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell* 99:13–22.
- Shworak NW, Liu J, Petros LM, Zhang L, Kobayashi M, Copeland NG, Jenkins NA, Rosenberg RD. 1999. Multiple isoforms of heparan sulfate D-glucosaminyl 3-O-sulfotransferase. Isolation,

- characterization, and expression of human cdnas and identification of distinct genomic loci. *J Biol Chem* 274:5170–5184.
- Sinnis P, Coppi A, Toida T, Toyoda H, Kinoshita-Toyoda A, Xie J, Kemp MM, Linhardt RJ. 2007. Mosquito heparan sulfate and its potential role in malaria infection and transmission. *J Biol Chem* 282:25376–25384. doi:10.1074/jbc.M704698200.
- 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. doi:10.1016/j.matbio.2013.12.001.
- Turnbull JE, Gallagher JT. 1991. Distribution of iduronate 2-sulphate residues in heparan sulphate. Evidence for an ordered polymeric structure. *Biochem J* 273 (Pt 3:553–559. doi:1996955.
- Wang Z, Hsieh P-HH, Xu Y, Thieker D, Chai EJE, Xie S, Cooley B, Woods RJ, Chi L, Liu J. 2017. Synthesis of 3-O-Sulfated Oligosaccharides to Understand the Relationship between Structures and Functions of Heparan Sulfate. *J Am Chem Soc.* doi:10.1021/jacs.7b01923.
- Wang Z, Hsieh PH, Xu Y, Thieker D, Chai EJE, Xie S, Cooley B, Woods RJ, Chi L, Liu J. 2017. Synthesis of 3-O-Sulfated Oligosaccharides to Understand the Relationship between Structures and Functions of Heparan Sulfate. *J. Am. Chem. Soc.* 139:5249–5256. doi:10.1021/jacs.7b01923.
- Yamada S, Yoshida K, Sugiura M, Sugahara K, Khoo KH, Morris HR, Dell A. 1993. Structural studies on the bacterial lyase-resistant tetrasaccharides derived from the antithrombin III-binding site of porcine intestinal heparin. *J Biol Chem* 268:4780–4787. doi:8444855.

Legends to figures:

Figure 1: rs28470223 polymorphism alters *HS3ST3A1* promoter activity. Six luciferase reporter gene assays with constructs containing the rs28470223-C or rs28470223-T *HS3ST3A1* promoter in K562 cell line were performed. All constructs were co-transfected with pRL-SV40 to standardize transfection efficiency and luciferase activity for each sample was adjusted by the empty pGL3-Enhancer. Statistical analysis was performed using two-tailed Student's t-test after controlling the normality of the data and the equality of their variance. Mean and standard error mean of relative luciferase activity are shown for each allele.

Figure 2: Mapping residues corresponding to missense SNP polymorphism on the crystal structure of 3-OST-3A1. (A) Overall view of the catalytic domain of the homologous 3-OST-3 dimer with each subunit colored in cyan and yellow. Mutated residues (orange) that severely affect the catalytic activity are mapped onto the solvent-accessible surface of human 3-OST-3 (accession code 1T8U) in complex with PAP (orange with magenta phosphate) and a tetrasaccharide (blue with green sulfate); those that correspond to the three 3-OST-3B1 SNPs are shown in red with the Ile196Val rs62056073 variant located on the back face. The Ile288Ala mutant, which is closely located to the Ser269Asn variant and less severely affects the catalytic activity, is shown in violet. (B) Close-up of boxed area, oriented as in panel A, of the active site of 3-OST-3 with bound PAP and tetrasaccharide.

Figure 3: Analysis of 3-OST-3B1-catalyzed 3-O-sulfation of HS8 oligosaccharide (A) RPIP-HPLC-elution profiles of untreated (gray trace) and 3-OST-3B1-treated (black trace) HS8 oligosaccharides. The peak (marked *) eluted at the highest NaCl concentration corresponds to a new 3-OST-3B1-catalyzed 3-O-sulfated species. (B) Time course of 3-O-sulfation of HS8 by *wt* 3-OST-3B1 (open circles) and the Ser269Asn mutant (closed circles) showing the evolution of the disaccharide content in Δ HexA,GlcNAc (left), unknown species (peak *, middle) and Δ HexA2S,GlcNS (right).

Table I. Predicted effect for *HS3ST3A1* and *HS3ST3B1* variants

<i>SNP_ID</i>	<i>GRCh37 location (bp)</i>	<i>Variant</i>	<i>Gene location</i>	<i>Type of change</i>	<i>PredictSNP 2</i>	<i>CADD</i>	<i>DANN</i>	<i>FATHMM</i>	<i>FunSeq2</i>	<i>GWAVA</i>
rs62057033	13399616	A>T	HS3ST3A1	Exon 2 - Synonymous	96% Neutral	73% Neutral	90% Neutral	77% Neutral	93% Neutral	56% Neutral
rs61732181	13399928	C>T	HS3ST3A1	Exon 2 - Synonymous	88% Neutral	58% deleterious	90% Neutral	93% deleterious	93% Neutral	56% Neutral
rs61744056	13400057	T>C	HS3ST3A1	Exon 2 - Synonymous	91% Neutral	94% Neutral	97% Neutral	90% Neutral	93% Neutral	62% deleterious
rs8080565	13400153	G>A	HS3ST3A1	Intron	77% Neutral	86% Neutral	54% deleterious	85% Neutral	80% Neutral	58% Neutral
rs3744337	13504665	A>G	HS3ST3A1	5'UTR	74% Neutral	67% deleterious	62% Neutral	93% Neutral	83% Neutral	64% deleterious
rs3744335	13504884	T>G	HS3ST3A1	5'UTR	88% Neutral	86% Neutral	80% Neutral	93% Neutral	–	53% Neutral
rs28470223	13505023	A>G	HS3ST3A1	5'UTR	88% Neutral	76% Neutral	86% Neutral	91% Neutral	67% deleterious	64% deleterious
rs78863672	13505237	C>A	HS3ST3A1	5'UTR	91% deleterious	86% deleterious	62% Neutral	86% deleterious	67% deleterious	84% deleterious
rs2072243	14204380	C>T	HS3ST3B1	5'UTR	97% deleterious	76% deleterious	89% deleterious	–	68% deleterious	80% deleterious
rs2072242	14204410	T>C	HS3ST3B1	5'UTR	97% deleterious	98% deleterious	89% deleterious	88% deleterious	62% deleterious	78% deleterious
rs115229628	14204423	G>A	HS3ST3B1	5'UTR	97% deleterious	98% deleterious	100% deleterious	91% deleterious	67% deleterious	76% deleterious
rs62636623	14205082	G>C	HS3ST3B1	Exon 1 - Missense Gly>Arg	89% Neutral	94% Neutral	90% Neutral	77% Neutral	84% Neutral	–
rs62636622	14205168	G>A	HS3ST3B1	Exon 1 - Synonymous	96% Neutral	92% Neutral	90% Neutral	87% Neutral	93% Neutral	54% Neutral
rs62056073	14248376	A>G	HS3ST3B1	Exon 2 - Missense Ile>Val	87% deleterious	58% Neutral	60% deleterious	59% deleterious	62% Neutral	51% deleterious
rs9906855	14248423	T>C	HS3ST3B1	Exon 2 - Synonymous	96% Neutral	94% Neutral	97% Neutral	90% Neutral	93% Neutral	62% deleterious
rs61729712	14248596	G>A	HS3ST3B1	Exon 2 - Missense Ser>Asn	87% deleterious	53% deleterious	60% deleterious	82% deleterious	62% Neutral	51% deleterious
rs55688668	14248702	G>A	HS3ST3B1	Exon 2 - Synonymous	93% Neutral	83% Neutral	90% Neutral	57% deleterious	93% Neutral	56% Neutral
rs9906590	14248877	G>A	HS3ST3B1	Exon 2 - Missense Glu>Lys	87% deleterious	53% deleterious	60% deleterious	65% deleterious	62% Neutral	50% deleterious
rs3785655	14249167	C>T	HS3ST3B1	3'UTR	88% Neutral	86% Neutral	68% Neutral	75% Neutral	83% Neutral	53% Neutral
rs73979332	14249433	C>T	HS3ST3B1	3'UTR	88% Neutral	80% Neutral	62% Neutral	89% Neutral	81% Neutral	71% Neutral

Figure 1

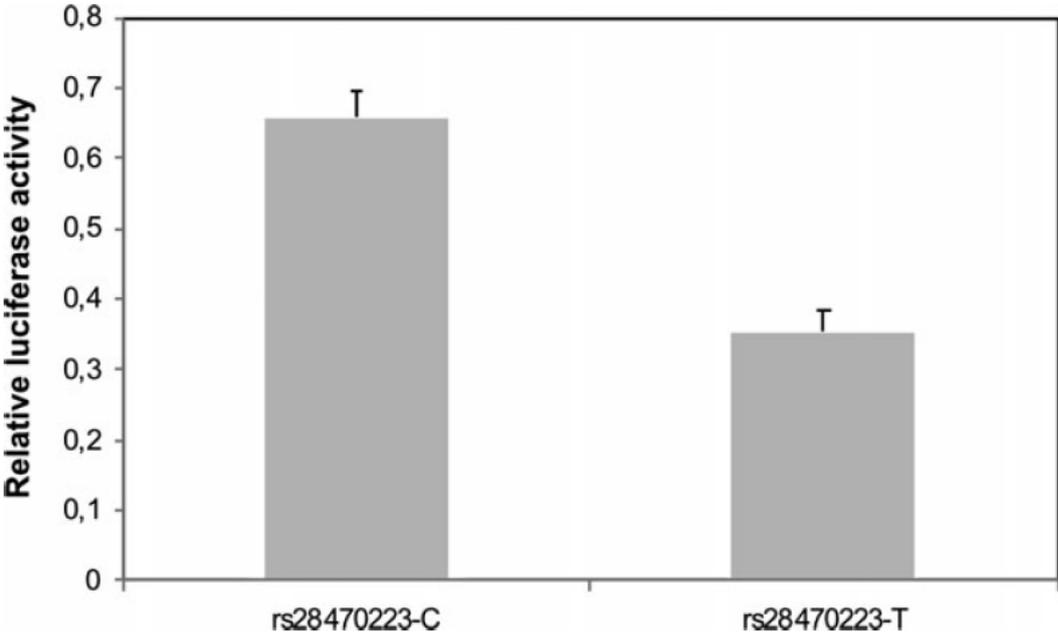


Figure 2

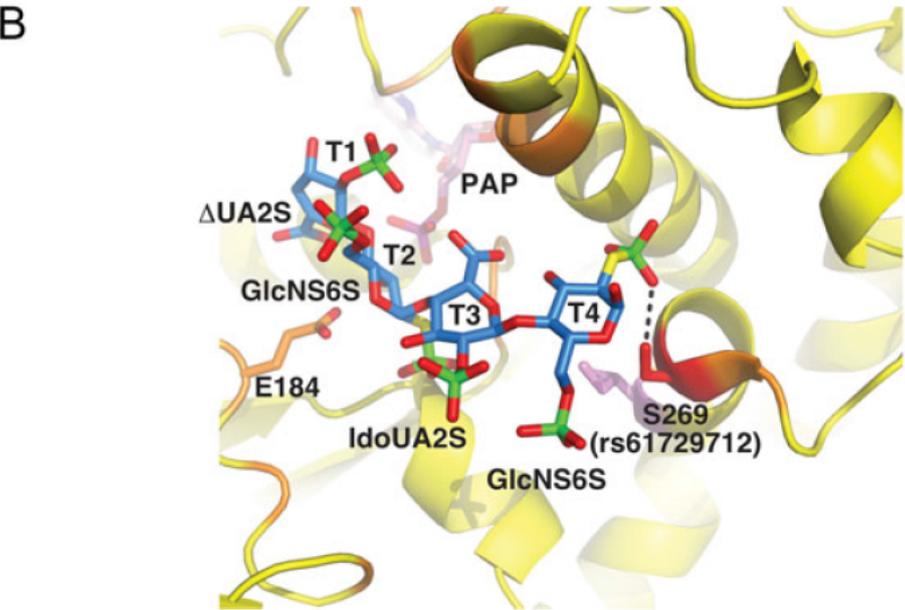
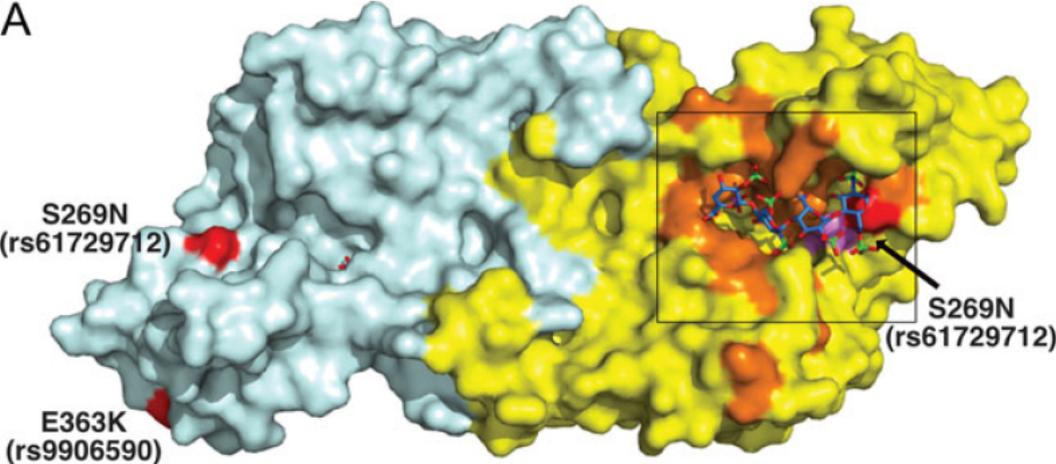


Figure 3

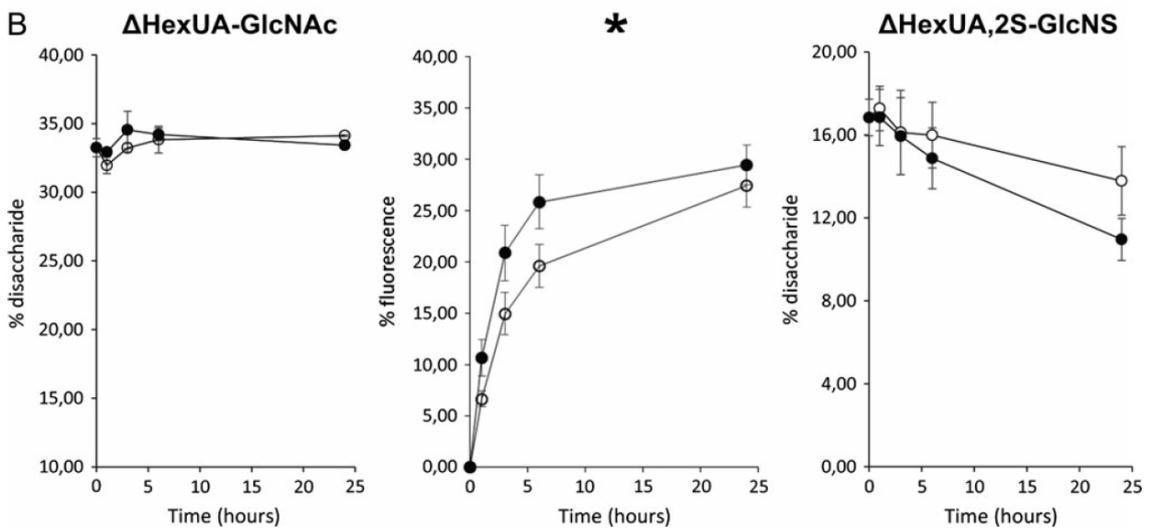
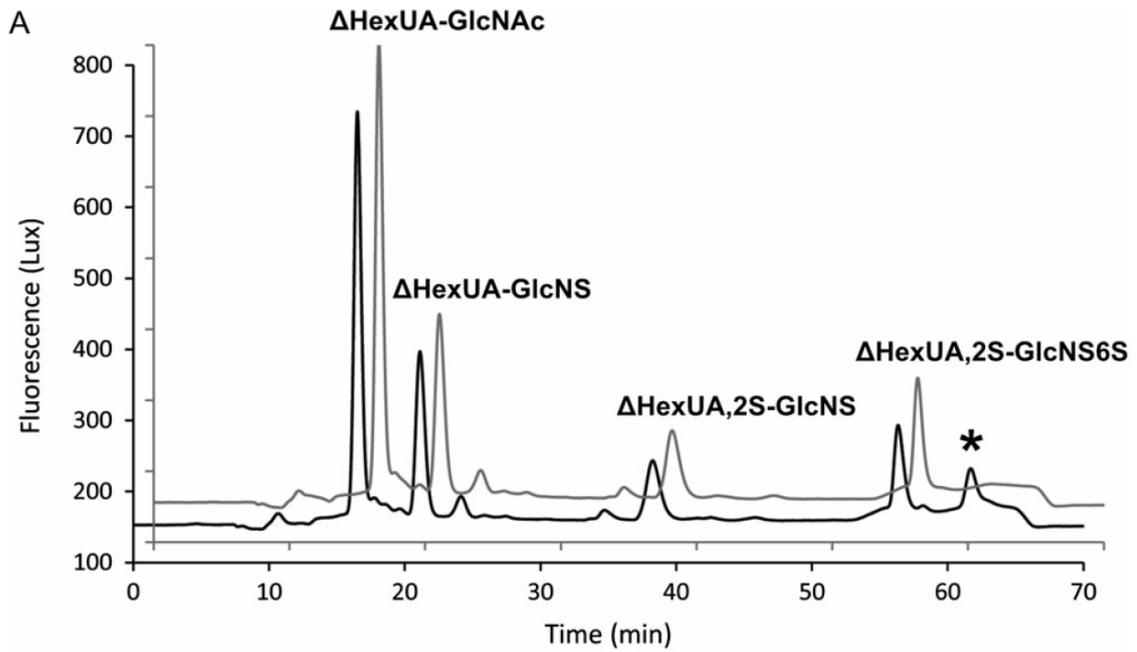


Table S1. Transcription factors that bind sequence containing rs28470223

List of transcription factors^a

atf1	med1
atf7	mllt1
bach1	mnt
brd2	mta2
brd4	mta3
cbfb	mx1
cdk9	myc
chd1	mycn
chd8	ncor
crem	nrf1
ctcf	phf8
ctcfl	pou5f1
cxxc4	rad21
e2f6	rbbp5
eed	runx2
ep300	sap30
erg	sin3a
ets1	smc3
ezh2	snai2
hdac1	stag1
hdac2	stat3
hexim1	suz12
hsf1	tbl1x
jarid2	tp53
kdm4a	trim24
kdm5b	twist1
max	ubtf
maz	zfn263

^aResults of the integrative analysis of transcriptional regulators CHIP-seq experiments using the ReMap tool

Table S2. *In silico* analysis of the effect of rs28470223

Motif	Allele		P-value 1	P-value 2	Fold change	Fold change
	1/allele	2			T/C	C/T
AP2D_HUMAN.H11MO.0.D	T/C		2,44E-03	2,11E-04	11,567	0,086
BRAC_HUMAN.H11MO.1.B	T/C		4,27E-04	8,54E-03	0,050	19,990
KAISO_HUMAN.H11MO.1.A	T/C		1,28E-03	9,70E-05	13,170	0,076
NFIA_HUMAN.H11MO.1.D	T/C		1,50E-04	8,66E-04	0,174	5,760
NFIB_HUMAN.H11MO.0.D	T/C		7,66E-05	7,51E-04	0,102	9,811
PAX5_HUMAN.H11MO.0.A	T/C		1,62E-03	3,11E-04	5,216	0,192
TFCP2_HUMAN.H11MO.0.D	T/C		3,31E-05	5,11E-04	0,065	15,431
ZN121_HUMAN.H11MO.0.C	T/C		1,43E-04	2,08E-05	6,892	0,145

Table S3. Disaccharide composition of HS8-derived oligosaccharides

Disaccharide structure	Disaccharide content
Δ HexUA-GlcNAc	33.3%
Δ HexUA-GlcNS	26.4%
Δ HexUA-GlcNAc,6S	2.8%
Δ HexUA,2S-GlcNAc	<i>nd</i>
Δ HexUA-GlcNS6S	2.6%
Δ HexUA,2S-GlcNS	17.3%
Δ HexUA,2S-GlcNAc,6S	0.5%
Δ HexUA,2S-GlcNS6S	17%
<i>N-sulfation</i>	63.3%
<i>2-O-sulfation</i>	34.8%
<i>6-O-sulfation</i>	22.9%
<i>Sulfate/dp2</i>	1.21

Data indicate the relative proportion of the detected oligosaccharides. *nd* : not detected