bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A cooperative network of molecular "hot spots" highlights the complexity of LH3 collagen glycosyltransferase activities

Antonella Chiapparino[a,*,1], Francesca De Giorgi[a,*], Luigi Scietti[a,*], Silvia Faravelli[a], Tony Roscioli[b,c,d], Federico Forneris[a]

[a] *The Armenise-Harvard Laboratory of Structural Biology, Dept. Biology and Biotechnology - University of Pavia, via Ferrata 9, 27100 Pavia, Italy, http://fornerislab.unipv.it;* [b] *NSW Health Pathology East laboratory, Prince of Wales Hospital, Randwick, NSW, Australia;* [c] *Centre for Clinical Genetics, Sydney Children's Hospital, Randwick, Australia;* [d] *Neuroscience Research Australia (NeuRA), Prince of Wales Clinical School, University of New South Wales, Sydney, Australia*

* These authors contributed equally to this work.

[1] Present address: Heidelberg University Biochemistry Center (BZH), INF 328, D-69120 Heidelberg, Germany.

Correspondence to: Federico Forneris (federico.forneris@unipv.it)

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Hydroxylysine glycosylations are collagen-specific post-translational modifications essential for maturation and homeostasis of fibrillar as well as non-fibrillar collagen molecules. Lysyl hydroxylase 3 (LH3) is the only human enzyme capable of performing two chemically-distinct collagen glycosyltransferase reactions using the same catalytic site: inverting beta-1,O-galactosylation of hydroxylysines and retaining alpha-1,2-glycosylation of galactosyl hydroxylysines.

Here, we used structure-based mutagenesis to show that both glycosyltransferase activities are strongly dependent on a broad cooperative network of amino acid side chains which includes the first-shell environment of the glycosyltransferase catalytic site and shares features with both retaining and inverting enzymes. We identified critical "hot spots" leading to selective loss of inverting activity without affecting the retaining reaction. Finally, we present molecular structures of LH3 in complex with UDP-sugar analogs which provide the first structural templates for LH3 glycosyltransferase inhibitor development.

Collectively, our data provide a comprehensive overview of the complex network of shapes, charges and interactions that enable LH3 glycosyltransferase activities, expanding the molecular framework for the manipulation of glycosyltransferase functions in biomedical applications.

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SIGNIFICANCE STATEMENT

Multifunctional collagen lysyl hydroxylase-glycosyltransferase LH3 is essential for biosynthesis and maturation of collagen molecules and has strong implications in rare developmental diseases as well as in metastatic spreading of solid tumors. LH3 is the only human enzyme capable of carrying out two different glycosyltransferase reaction mechanisms within the same catalytic site to produce glucosyl-galactosyl-hydroxylysine, the most ubiquitous glycosylation pattern found in nature. This work elucidates the molecular networks enabling the LH3 glycosyltransferase reactions, highlighting a highly cooperative network of amino acid residues that concertedly ensure collagen glycosylation, and suggesting that the LH3 molecular organization may be the progenitor of more specialized glycosyltransferases. In addition, this work presents the first two small molecule scaffolds capable of modulating LH3 glycosyltransferase activities.

KEYWORDS

Collagen biosynthesis; extracellular matrix; post-translational modifications; collagen glycosylations; glycosyltransferase; lysyl hydroxylase.

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

\body

1. INTRODUCTION

Collagens are the most abundant proteins in the human body and are highly conserved from sponges to mammals (1, 2). The different oligomeric architectures and roles of collagen molecules strongly depend on a variety of post-translational modifications, including proline and lysine hydroxylations, as well highly specific glycosylations of hydroxylated lysines (HyK) (1, 3). The disaccharide present on HyK contains a highly conserved glucosyl(α−1,2)-galactosyl(β−1,O) glycan moiety, whose identity was discovered in the late sixties (4-6). Monosaccaridic galactosyl-(β−1,O) HyK have been identified as well, as a result of catabolic reactions carried out by the collagen α-glucosidase, an enzyme highly specific for the disaccharide present on collagenous domains. The role for this enzyme is to localize collagen in the glomerular basement membrane (7-9).

The spread of glycosylation largely depends on collagen type (4, 10, 11), on the functional area inside tissues (12-14), on the developmental stage (15, 16) and on disease states (17-19). Although extensively studied, the precise mechanisms of collagen glycosylation and their biological relevance in collagen homeostasis have remained poorly understood.

The identity and exquisite stereochemistry of the Glc(α−1,2)-Gal(β−1,O)-HyK-linked carbohydrate supports the idea that at least two distinct enzyme types are needed to fully incorporate these complex post-translational modifications on collagen molecules (20). The first reaction indeed requires an inverting-type galactosyltransferase (GalT) acting on HyK, whereas the subsequent glucosylation is catalyzed by a retaining-type glucosylgalactosyltransferase (GlcT). Multifunctional lysyl hydroxylase 3 (LH3) was the only enzyme identified as possessing lysyl hydroxylase activity as well as GalT and GlcT activities *in vitro* (21). The glycosyltransferase activities are specific of LH3, as highly homologous LH1 and LH2a/b are not capable of catalyzing these reactions (22). Conversely, *in vivo*

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

studies have demonstrated that decreased LH3 protein levels and/or pathogenic mutations in LH3 GT domain, exclusively impair the GlcT activity (18, 23, 24). This occurs secondarily to the LH3 p.Asn223Ser, which introduces an additional glycosylation site within the enzyme's GT domain leading to an osteogenesis imperfecta-like phenotype (18); and in the recently identified LH3 p.Pro270Leu, which results in a Stickler-like syndrome with vascular complications and variable features typical of Ehlers-Danlos syndrome and Epidermolysis Bullosa (24). Mouse studies have also shown that only the LH3 GlcT activity is indispensable for the biosynthesis of collagen IV and formation of the basement membrane during embryonic development (15, 25), consistent with the presence of additional collagen galactosyltransferases. Two genes encoding for O-galactosyltransferases (*GLT25D1* and *GLT25D2*) were recently identified (26, 27). It is of interest that GLT25D1 and LH3 were proposed to act in concert on collagen molecules (26, 28, 29). Studies on osteosarcoma cell lines which produce large amounts of fibrillar collagens, showed that the simultaneous deletion of *GLT25D1* and *GLT25D2* resulted in growth arrest due to lack of glycosylation, further indicating that the GalT activity of LH3 might not be as essential as its GlcT activity (30).

These data support the existence of a specific and highly conserved machinery for collagen O-glycosylation, and sustain the hypothesis that *in vivo* the entire collagen glycosylation machinery may involve distinct proteins and protein complexes for GalT and GlcT reactions. This raises the intriguing question of how this highly conserved process is spatiotemporally regulated at the molecular level. Our current understanding of collagen glycosyltransferases is however restricted to three-dimensional structures of human LH3 in complex with UDP-sugar donor substrates (31) and to few mutagenesis studies focusing on the main hallmarks of Mn^{2+}-dependent glycosyltransferase catalysis (21, 32).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Here, we combine site-directed mutagenesis scanning with biochemistry and structural biology to characterize the mechanisms of the two distinct human LH3 glycosyltransferase activities. Our data highlight an overall distribution of "hot spots" around the extended glycosyltransferase cavity of LH3, critically involved in both GalT and GlcT functions, but also amino acid residues capable of selectively abolishing transfer of galactose to HyK without affecting GlcT activity. Finally, we also identify and characterize UDP-sugar substrate analogs acting as inhibitors of LH3 glycosyltransferase activities.

Together, our results provide a detailed molecular blueprint of LH3 glycosyltransferase activities and a structural framework for the development of the first collagen GalT/GlcT inhibitors. These insights will assist with the manipulation of LH3 protein functions and donor substrate specificity in biomedical applications.

RESULTS

Features and roles of the non-conserved LH3 "glycoloop"

The LH3 N-terminal GT domain shares its fold with Mn^{2+}-dependent GT-A glycosyltransferases, encompassing a UDP-donor substrate binding cavity stretched towards a GalT/GlcT catalytic pocket (31). We firstly inspected the structural organization of the UDP binding cavity to identify signatures for LH3 glycosyltransferase activity by comparing the amino acid sequences for the residues surrounding the UDP-sugar donor substrate with those of GalT/GlcT-inactive LH1 and LH2a/b isoforms. We found that nearly all residues involved in Mn^{2+} and UDP-sugar binding are conserved in orthologs, with the exception of Val80, becoming Lys68 in LH1 and Gly80 in LH2a/b (Figure 1). The presence of a different amino acid side chain surrounding the donor substrate cavity led us to consider whether this could be a discriminating functional feature among the GT domains in LH enzymes. In LH3, Val80 is located in the middle of a flexible "glycoloop" (Gly72-Gly87), not visible in the electron

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

density of the ligand-free LH3 structure and stabilized upon UDP-substrate binding (31). Within the glycoloop, residue Val80 is in close proximity to the ribose ring of the UDP-sugar donor substrates. We hypothesized that introduction of a large, positively charged residue such as Lys in LH1, or alterations due to complete removal of side chain steric hindrance such as Gly in LH2 could lead to inability of binding donor substrates. We therefore generated the LH3 Val80Lys mutant. This enzyme variant was found to be folded based on analytical gel filtration and differential scanning fluorimetry (DSF), and showed lysine hydroxylation activity comparable to wild-type LH3 (SI Appendix, Fig. S1). Conversely, the mutation resulted in a strong impairment of both GalT and GlcT activities, with 30% and 50% residual activity, respectively when the reaction was carried out in presence of both donor and acceptor (gelatin) substrates (Figure 2, SI Appendix, Table S1). Considering that wild-type LH3 is also capable of activating donor UDP-sugar substrates and release UDP in absence of the acceptor collagen substrate (31), we investigated the impact of the Val80Lys mutation in absence of acceptor substrates. The experiments yielded comparable enzymatic activities between wild-type and mutant LH3 (Figure 2, SI Appendix, Table S1), indicating that this residue might be involved in the productive positioning of the donor substrate during transfer of the glycan moiety to the acceptor substrate, rather than stabilizing the UDP moiety in the catalytic pocket.

To further rationalize the implications of LH3 Val80 in GalT and GlcT activities, we crystallized and solved the 3.0-Å resolution structure of this mutant in complex with Mn^{2+}, and also obtained the 2.3-Å resolution structure of LH3 Val80Lys mutant in presence of both Mn^{2+} and the UDP-Glc donor substrate (SI Appendix, Table S2). Overall, both structures superimpose almost perfectly with wild-type LH3 for all domains (SI Appendix, Fig. S2). The structure of the LH3 Val80Lys mutant bound to Mn^{2+} appeared identical to that of wild-type LH3. In both structures, the glycoloop containing Val/Lys80 could not be modelled in the

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

electron density due to its high flexibility (SI Appendix, Fig. S2B). On the other hand, the side chain of the Val80Lys residue could be modelled unambiguously in the experimental electron density of the UDP-donor substrate bound structure (Figure 3A). Despite the increased steric hindrance, the mutated Lys80 residue adopted a conformation compatible with the simultaneous presence of the UDP-Glc in the catalytic cavity. However, similar to what observed for wild-type LH3, the glycan moiety of UDP-Glc was completely flexible and therefore not visible in the electron density (Figure 3A). Collectively, these data are consistent with the alteration introduced by the Val80Lys mutation impacting partially on the LH3 glycosyltransferase catalytic activities

The glycoloop is a structural feature found exclusively in the GT domains of LH enzymes. It incorporates Trp75, a residue whose aromatic side chain stabilizes the uridine moiety of the donor substrate and, together with residue Tyr114 of the DxxD motif (a distinguishing feature of LH3 GT domain, shared among LH enzymes, (31)) “sandwiches” the donor substrate in an aromatic stacking environment (Figure 1A). Both residues are critical for the LH3 GalT and GlcT enzymatic activities (31). The conformation adopted by the LH3 glycoloop in the presence of UDP-donor substrates is however not accompanied by other significant structural changes in surrounding amino acids, with the exception of minor rearrangements of distant residue Trp92 (not conserved in other LH isoenzymes (Figure 1A-B)), whose bulky side chain rearranges pointing towards the aromatic ring of Trp75. Prompted by this observation, we mutated this residue to alanine and found that the presence of this variant did not alter the folding of the enzyme (SI Appendix, Fig. S1A-B) nor its LH enzymatic activity (SI Appendix, Fig. S1C). Conversely, the mutant showed 40% decrease for both GalT and GlcT activities in presence of donor and acceptor substrate compared to LH3 wild-type (Figure 2, SI Appendix, Table S1); the impact of the Trp92Ala mutation on reactions in absence of acceptor substrate seemed to affect both activities at similar levels, with slightly stronger

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reduction in GalT (70%) compared to GlcT (60%) (Figure 2, SI Appendix, Table S1). These findings suggest that LH3-specific long-range interactions in the GT domain may contribute to the productive conformations of the glycoloop in donor substrate-bound states.

In UDP-sugar bound structures, the glycoloop contacts a poly-Asp sequence (Asp188-Asp191, Figure 1A); this sequence is partially conserved in LH isoforms lacking glycosyltransferase activities (Figure 1B). Mutations of Asp190 and Asp191 were reported to affect the glycosyltransferase enzymatic activities of LH3 (32). Based on LH3 crystal structures, such behaviour is expected, since residues Asp190 and Asp191 point towards the GalT/GlcT catalytic cavity. Interestingly, superposition of LH3 molecular structures with GT-A fold glycosyltransferases show that Asp191 caps the N-terminal end of an α-helix in a highly conserved position, with functional relevance in both retaining (33, 34) and inverting enzymes (35, 36) (SI Appendix, Fig. S3, Table 1).

Collectively, these data point towards an involvement of the poly-Asp repeat, and in particular Asp190 and Asp191, in the positioning and recognition of donor or acceptor substrates.

LH3 shares features with both retaining and inverting glycosyltransferases

After investigating the amino acid residues involved in stabilization of the UDP moiety of donor substrates, we focused on another group of residues within the GT catalytic pocket, opposite to the putative position of the flexible sugar rings of the same substrates (Figure 1A). Many LH3 residues shaping this part of the glycosyltransferase catalytic cavity matched catalytic amino acids found in other GT-A type glycosyltransferases (Figure 1A, SI Appendix, Fig. S3) (37, 38). In particular, LH3 Trp145, a residue located in one of the loops of the GT domain uniquely found in LH3, was previously suggested as a possible candidate for modulation of LH3 GalT and GlcT activities. This residue was found to adopt different side chain conformations in substrate-free and substrate-bound structures, affecting the shape

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and steric hindrance of the enzyme's catalytic cavity (31); interestingly, nearly identical conformational changes were observed when comparing substrate-free and substrate bound LH3 Val80Lys structures (Figure 3B). Mutating the LH3 Trp145 residue into alanine strongly reduced both GalT (6% residual) and GlcT (10% residual) enzymatic activities (Figure 2, SI Appendix, Table S1), without affecting other enzyme's properties (SI Appendix, Fig. S1). This supports previous hypotheses of a "gating" role for Trp145 in the GT catalytic cavity, assisting the productive positioning of sugar moieties of donor substrates for effective transfer during catalysis. Comparison with molecular structures of other glycosyltransferases (including distant homologs) highlighted that most structurally-related enzymes position aromatic side chains from different structural elements of their fold in their catalytic cavities in a structural arrangement reminiscent to that of Trp145 in LH3. In particular, similar aromatic residues were found in other glycosyltransferases such as Tyr186 in LgtC from *Neisseria meningitidis*, Trp314 in the N-acetyllactosaminide α-1,3-galactosyl transferase GGTA1, Trp300 in the histo-blood group ABO system transferase, and Trp243 and Phe245 in the two glucoronyltransferases B3GAT3 and B3GAT1, respectively (Table 1, SI Appendix, Fig. S3). This is surprising, considering that the loop encompassing residues 142-163 is a unique structural feature of LH3 (31). This supports glycosyltransferases being highly versatile enzymes, displaying an impressive structural plasticity to carry out reactions characterized by a very similar mechanism on a large variety of specific donor and acceptor substrates.

Our previous structural comparisons of ligand-free and substrate-bound LH3 highlighted the additional possibility of a concerted mechanism involving conformational changes of a non-conserved aromatic residue located on the LH3 surface (Trp148, Figure 1), together with Trp145 (31). To investigate such possibility, we mutated Trp148 into alanine. The mutant enzyme was also found to behave like wild-type LH3 in folding and LH activity (Suppl. Fig.

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1). Glycosyltransferase assays showed that this variant had reduced GalT (30% residual) and GlcT (50% residual) activities compared to the wild-type, in presence of both donor and acceptor substrates (Figure 2, SI Appendix, Table S1). Despite the less pronounced alterations than those observed when mutating Trp145, these data are consistent with possible synergistic mechanisms between long-range substrate recognition on the enzyme's surface and conformational rearrangements upon substrate binding in the enzyme's catalytic site.

Molecular structures of LH3 in complex with UDP-Glc and Mn^{2+}, showed weak electron density near the UDP pyrophosphate group partially compatible with glycan donor substrates, likely representative of multiple conformations simultaneously trapped in the substrate binding cavity (31). We explored the LH3 catalytic cavity in its proximity, looking for additional amino acids potentially critical for catalysis. In particular, we searched for residues carrying carboxylic or amide side chains, thus capable of acting as candidate catalytic nucleophiles for the formation of a (covalent) glycosyl-enzyme intermediate prior to glycosylation of the acceptor substrate (39).

In retaining-type glycosyltransferases belonging to the GT-6 family, a conserved glutamate has been found to act as a nucleophile (40-42). In LH3 structures, we noticed that residue Glu141 points towards the cavity that accommodates the glycan moiety (Figure 1A). Removal of this carboxylate side chain resulted in a folded enzyme (SI Appendix, Fig. S1) completely deprived of both GalT and GlcT activities (Figure 2, SI Appendix, Table S1). These data are consistent with the catalytic role of the glutamate proposed to act as a nucleophile (40), suggesting that LH3 may indeed act through similar mechanisms of donor substrate processing, despite the fact that in LH3 Glu141 occupies a different position within the catalytic cavity when compared to GT-6 glycosyltransferases (SI Appendix, Fig. S3) (40, 43, 44).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Proximate to Glu141 in the GalT/GlcT cavity, residue Asn255 is the closest amino acid to the UDP phosphate-sugar bond. Despite being fully conserved in LH isoforms (Figure 1B), it is not found in any GT-A-type glycosyltransferases with known structures to date. The side chain of Asn255 consistently points to a direction opposite to the donor substrate in all LH3 structures (Figure 3B, SI Appendix, Fig. S2B). We wondered whether the side chain amide group might be involved in catalysis, possibly through recognition of acceptor substrates given the conformation displayed by this side chain. Surprisingly, LH3 Asn255Ala mutants showed that their GalT activity was completely abolished, whereas the GlcT enzymatic activity was reduced to 50% (Figure 2, SI Appendix, Table S1); the protein was properly folded and showed LH activity comparable to LH3 wild-type (SI Appendix, Fig. S1). These results identify Asn255 as a possible critical discriminating residue for the two glycosyltransferase activities of LH3, and rule out possible functions for this residue as catalytic nucleophile for retaining-type glycosyltransferase mechanisms, given its major impact restricted to the GalT catalytic activity.

Pathogenic LH3 mutations in the LH3 GT domain affect protein folding

Recently a new pathogenic LH3 mutation, Pro270Leu, has been identified and mapped at the interface of the AC and GT domain (24). This residue localizes in a loop which is critical for shaping the GT cavity, although given its position it is unlikely to play direct roles in catalysis. To better understand the impact of this pathogenic mutation on LH3 enzymatic activity, we generated a Pro270Leu mutant. In this case, due to very low expression levels, we could not reliably carry out any *in vitro* investigations. Considering the high reproducibility associated to recombinant production of a large variety of LH3 point mutants, this result may indicate that this mutation is likely to severely impact the overall enzyme stability rather than

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

its enzymatic activity, resulting in extremely low protein expression levels *in vitro* and likely also *in vivo*.

Molecular structures of LH3 in complex with UDP-sugar analogs provide insights on how glycan moieties are processed inside the LH3 catalytic cavity

A frequent limitation associated to molecular characterizations of glycosyltransferases is the high flexibility of the donor substrate glycan moiety within the catalytic cavity. Such limitation becomes even more relevant when the enzyme is capable of processing UDP-sugar molecules in absence of acceptor substrates, such as in the case of LH3. Considering our previous (31) and current co-crystallization results, we wondered whether free UDP, the product of the enzymatic reaction, could remain bound in the LH3 GT domain with the same efficiency as physiological donor substrates even after processing. We therefore compared the binding of free UDP and donor UDP-sugars using DSF and detected a thermal shift of 3.5 °C for free UDP, compared to a 2-2.5 °C shift using UDP-sugar substrates (Figure 4A). These results suggested that free UDP may bind to LH3, likely with even higher affinity than UDP-glycan substrates, and that the GalT and GlcT reactions may therefore be affected by product inhibition. To our surprise, the increase in thermal stability did not correlate with efficient trapping of the reaction product in LH3 molecular structures. Independently from the UDP concentration used in co-crystallization and soaking experiments, we never observed any electron density for free UDP, yielding LH3 structures completely identical to ligand-free enzyme (SI Appendix, Fig. S4A).

A report from Kivirikko and colleagues (45) suggested that UDP-glucuronic acid (UDP-GlcA) could act as competitive inhibitor of collagen glycosyltransferases. Based on that, UDP-GlcA was used to isolate LH3 from chicken embryos preparation (21, 46). However, no follow-up biochemical studies could be found in the literature. We used DSF and

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

luminescence-based GalT and GlcT activity assays to investigate whether and how UDP-GlcA could affect LH3 enzymatic activity. DSF showed that UDP-GlcA indeed binds weakly to LH3, resulting in a thermal shift of 1-1.5 °C (Figure 4A), highlighting limited stabilization compared to UDP-glycan substrates and free UDP. Enzymatic assays also confirmed the competitive inhibition displayed by this molecule (Figure 4B), with IC_{50} values in the millimolar range (SI Appendix, Table S1). We also successfully co-crystallized and determined the 2.2-Å resolution crystal structure of wild-type LH3 in complex with Mn^{2+} and UDP-GlcA (Suppl. Table 2), and found that the inhibitor could efficiently replace UDP-sugar donor substrates in the substrate cavity (Figure 4C). We observed additional electron density for the glucuronic acid moiety of the inhibitor in the enzyme's catalytic cavity, however this density could not be interpreted with a single inhibitor conformation. Nevertheless, analysis of the experimental electron density for the glucuronic acid moiety unambiguously showed that the inhibitor adopts a bent conformation: the glycan moiety is deeply buried in the enzyme's catalytic cavity proximate to residues Lys89, Asp190, Asp191, but distant from the residues found critical for catalysis, including Trp145, Asn255 and Glu141 (Figure 4C), thereby leaving the remaining space in the cavity for accommodating acceptor substrates.

Considering the possible conformations adopted by the glucuronic acid moiety based on analysis of the electron density and the close proximity of the glucuronic acid moiety to LH3 Val80, we wondered whether the LH3 Val80Lys mutation could interfere with inhibitor binding. We therefore co-crystallized and solved the 2.7-Å resolution structure of LH3 Val80Lys mutant in complex with UDP-GlcA (SI Appendix, Table S2), and surprisingly observed partial displacement of the glycoloop, for which we could not observe the typical well defined electron density present in UDP-sugar-bound wild-type LH3 structures (Figure 4D). At the same time, we could not observe improvements in the quality of the electron density for the glucuronic acid moiety, resulting even poorer than what observed in wild-type

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

LH3 (Figure 4D). This suggests that the intrinsic flexibility of the sugar-like moiety is not influenced by specific conformations of the glycoloop, but rather by lack of specific protein-ligand interactions that could provide stabilization of the sugar ring in a unique structural arrangement.

The lack of a precise conformation for the glucuronic acid moiety observed crystal structures prompted us for a further investigation of another UDP-sugar substrate analog, characterized by lack of the carboxylic moiety of UDP-GlcA: UDP-Xylose (UDP-Xyl). Similar to UDP-GlcA, UDP-Xyl resulted to be a weak inhibitor of LH3 GalT and GlcT activities (Figure 4A-B), with IC_{50} in the high micromolar range (SI Appendix, Table S1). The 2.0-Å resolution structure of LH3 in complex with Mn^{2+} and UDP-Xyl also showed the inhibitor bound inside the enzyme's catalytic cavity, with weak electron density associated to the sugar moiety suggesting multiple conformations of the xylose moiety attached to UDP, similar to what observed for UDP-GlcA (Figure 4E). Taken together, these results suggest that the LH3 GT binding cavity is capable of hosting a variety of UDP-sugar substrates, and that inhibition likely depends on the reduced flexibility (and therefore increased stabilization) of the ligand within the cavity. In this respect, we could expect that UDP-sugar analogs strongly interacting with side chains proximate to the glycan moieties of UDP-GlcA and UDP-Xyl, may have the potential to become powerful inhibitors of LH3 glycosyltransferase activities.

DISCUSSION

Glycosyltransferases are highly versatile, yet very specific enzymes. If carefully inspected, they reveal a series of recurrent features that allow their comparative characterization even in presence of very low sequence/structure conservation. LH3 has been known for long time as a promiscuous enzyme able to exploit both an inverting and a retaining catalytic mechanism *in vitro* for the specific transfer of different sugars to at least two different acceptor substrates:

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the HyK and the GalHyK of collagens (1). Our *in vitro* investigations highlight multiple hot spots important for LH3 glycosyltransferase activity: the glycoloop, the poly-Asp helix, the acceptor substrate cavity, and the region proximate to the UDP-sugar donor substrate (Figure 1A).

LH3 is the only isoform of its family found capable of glycosyltransferase activities, however the strong sequence conservation among human LH isoforms cannot be used to elucidate the key determinants for this additional function. The GT domain has features (such as the DxxD motif, the poly-Asp region, the glycoloop) not found in other glycosyltransferases, yet highly conserved within the LH family (Figure 1B). Computational homology models of homologous LH1 and LH2a/b (47) (SI Appendix, Fig. S5) support the possibility of UDP-sugar donor binding and processing. To fully investigate LH enzymes' GT domains, we focused on the only non-matching residue present within the whole amino acid sequence directly surrounding the UDP-donor substrate. This residue (Val80 in LH3, corresponding to Lys68 in LH1 and Gly80 in LH2a/b) is located in the middle of the glycoloop, in close proximity to the ribose ring of the UDP-sugar donor substrate(s). Our data were consistent with this residue being important for catalysis, as it assists the positioning of the bound donor substrate and in particular the glycan moiety. However, the results also emphasize how sequence alterations at this site are not sufficient to justify the lack of GalT/GlcT enzymatic activities in homologous LH1/2. We therefore expanded our investigation to the second-shell environment surrounding the donor substrate, and found that non-conserved LH3 Trp92 (Leu80 in LH1, Leu92 in LH2a/b) positions its aromatic side chain in a conformation that stabilizes the entire glycoloop to facilitate the enzymatic reactions. Again, mutating this residue did not lead to full loss of the glycosyltransferase activities. This is consistent with the loss of GalT/GlcT functions in LH1 and LH2a/b being associated to a broad set of subtle

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

alterations, possibly involving residues distant from the actual enzyme's catalytic site, likely essential for recognition of collagen acceptor substrates.

Prompted by these observations, we expanded our investigation to the LH3 catalytic cavity expected to host the glycan moieties of the UDP-sugar donor substrates and the acceptor molecules. Residues Asp190 and Asp191 of the characteristic LH3 poly-Asp helix lie in a conserved position compared to other retaining and inverting type GT-A (SI Appendix, Fig. S3, Table 1) and their carboxylate moieties are critical for catalysis (32). Although residues matching these positions have been proposed to act as catalytic nucleophiles in retaining type GT-A (32-34), and as catalytic bases in inverting type GT-A (35, 36), their distances and the relative orientations with respect to UDP-sugar donor substrates in the LH3 GT domain (Figure 1A) do not support this hypothesis. Nevertheless, these residues play critical roles in recognizing and assisting the proper positioning of donor UDP-sugar substrates, as shown by their proximity to glycan moieties in LH3 co-crystal structures with UDP-sugar analogs (Figure 4C-E), a feature observed also in mutagenesis studies on other GT-A retaining type glycosyltransferases (48).

On the opposite site of the UDP-binding pocket, LH3 exhibits a non-conserved loop shaping the GT catalytic cavity bearing two aromatic residues, the Trp145 and the Trp148 that seem to act in a concerted way during catalysis, as suggested by comparisons between substrate-free and substrate-bound LH3 molecular structures (Figure 1A). Trp145 is indispensable for both GalT/GlcT activities: its conformational changes seem to respond to the presence and conformational positioning of the donor substrate inside the catalytic cavity. Although located in a loop that is uniquely found in LH3, the Trp145 side chain matches a site frequently occupied by bulky aromatic residues in other GT-A glycosyltransferases that shape a portion of the GT cavity to facilitate donor substrate processing and catalysis. This further highlights the versatility of glycosyltransferases, in which many different structural features have

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

evolved to specifically recognize distinct donor and acceptor substrates, while preserving the ability to carry out the same catalytic reaction. The implication in catalysis of the less conserved Trp148 on the surface of the GT domain is even more intriguing: this residue seems to coordinately respond to the rearrangements of its counterpart Trp145 in the catalytic site, suggesting involvement in recognition of the acceptor substrate prior to its access into the GT cavity. Alternatively, Trp148 may contribute to long-range stabilizing interactions with collagen molecules, while they dock their HyK or GalHyK residues in the acceptor substrate site during the GalT or GlcT reactions, respectively.

Catalytic nucleophiles have been clearly identified so far only in the retaining-type glycosyltransferases belonging to the GT-6 family (41, 42), such as the α–1,3 galactosyltransferase (GGTA1) where a conserved glutamate is found positioned on the β–face of the donor sugar (40, 43, 44) (SI Appendix, Fig. S3). Conversely, extensive structural comparisons and mutagenesis experiments have been performed in the O-galactosyltransferase LgtC from *Neisseria menengitidis,* focusing on matching residue, Gln189 (48). However, the role of this residue as catalytic nucleophile was ruled out. This site is occupied by Gln192 in LH3. This residue is next to the poly-Asp helix, distant from the sites occupied by donor substrates and in an arrangement that is not compatible with a direct role in catalysis. In spite of that, we identified the carboxylate side chain of LH3 residue Glu141 pointing directly towards the glycan moiety (Figure 1A), and found that this residue is essential for both GalT and GlcT activities (Figure 2, SI Appendix, Table S1). Glu141 adopts a conformation corresponding to Asp130 in the O-galactosyltransferase LgtC from *Neisseria meningitidis*, Asp125 in the O-glucosyltransferase GYG1 from rabbit, and Gln247 in the O-glucosyltransferase GGTA1 from *Bos taurus* (SI Appendix, Fig. S3, Table 1). As the removal of this carboxylate side chain completely abolishes LH3 glycosyltransferase activities (Figure 2, SI Appendix, Table S1), we suggest that the site corresponding to Glu141 might be a

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

critical "hot spot" during catalysis in those glycosyltransferases that present in the same position carboxylate or amide side chains.

Side chain amides are also good candidate nucleophiles. In LH3, residue Asn255 is the closest amino acid to the UDP phosphate-sugar bond, but in crystal structures its side chain consistently points to a direction opposite to the donor substrate (Figure 1A). When inspecting the molecular structures of other GT-A glycosyltransferases, we noticed that this residue is not conserved (SI Appendix, Fig. S3). Strikingly, the LH3 Asn255Ala mutant showed a complete loss of GalT activity, but partially preserved the GlcT activity. Although the significance of *in vivo* LH3 GalT activity is uncertain, such activity is clearly detectable *in vitro* (21, 31). The ability of Asn255 to selectively abolish only LH3 GalT activity highlights how LH3 can promiscuously accept and recognize very different acceptor substrates (i.e., collagen HyK *versus* GalHyK) within the same catalytic site and carry out two glycosyltransferase reactions with different mechanisms (i.e., inverting GalT and retaining GlcT). Presently, LH3 is the only known glycosyltransferase capable of such promiscuity, and our data show that it shares numerous features with both retaining and inverting glycosyltransferases. Collectively, these results suggest the intriguing possibility that LH3 is not just the ancestor of the whole LH family, but may preserve in its sequence features belonging to the evolutionary precursors of both retaining and inverting glycosyltransferases. In this respect, the carbohydrate-active enzyme (CAZy) classification (42) requires for enzymes belonging to the same family to carry out reactions with the same stereochemical outcome. Given the distinguishing features and dual mode of action observed and now characterized for LH3, we suggest this enzyme as the funding member of a new family of GT-A glycosyltransferases, the GT-X family.

Our work provides a set of 3D structures of LH3 in complex with UDP-sugar analogs, which work as mild inhibitors (Figure 4). Despite the high flexibility observed for the glycan

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

moieties of the bound molecules, the new molecular structures presented provide valuable insights for structure-based drug development of inhibitors of LH3 GalT/GlcT enzymatic activities. These molecules may give the spark to innovative therapeutic strategies against pathological conditions characterized by excess collagen glycosylations, such as osteogenesis imperfecta (49).

Together with the mutagenesis scanning of the entire GT catalytic site, our work provides a comprehensive overview of the complex network of shapes, charges and interactions that enable LH3 GalT and GlcT activities.

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MATERIALS AND METHODS

5.1 Chemicals

All chemicals were purchased from Sigma-Aldrich (Germany) unless otherwise specified.

5.2 Site-directed mutagenesis

The LH3 coding sequence (GenBank accession number BC011674.2 - Source Bioscience), devoid of the N-terminal signal peptide was amplified using oligonucleotides containing in-frame 5′-BamHI and 3′-NotI (SI Appendix, Table S3) and cloned in a pCR8 vector, that was used as a template for subsequent mutagenesis experiments. All LH3 mutants were generated using the Phusion Site Directed Mutagenesis Kit (ThermoFisher Scientific). The entire plasmid was amplified using phosphorylated primers. For all mutants the forward primer introduced the mutation of interest (SI Appendix, Table S3). The linear mutagenized plasmid was phosphorylated prior to ligation. All plasmids were checked by Sanger sequencing prior to cloning into the expression vector.

5.3 Recombinant protein expression and purification

Wild-type and mutant LH3 coding sequences were cloned into the pUPE.106.08 expression vector (U-protein Express BV, The Netherlands) in frame with a 6xHis-tag followed by a Tobacco Etch Virus (TEV) protease cleavage site. Suspension growing HEK293F cells (Life Technologies, UK) were transfected at a confluence of 10^6 cells ml^{-1}, using 1 μg of plasmid DNA and 3 μg of linear polyethyleneimine (PEI; Polysciences, Germany). Cells were harvested 6 days after transfection by centrifuging the medium for 15 minutes at 1000 x *g*. The clarified medium was filtered using a 0.2 mm syringe filter and the pH was adjusted to 8.0 prior to affinity purification as previously described (31). All proteins were isolated from the medium exploiting the affinity of the 6xHis tag for the HisTrap Excel (GE Healthcare,

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

USA) affinity column. The purified protein was further polished using a Superdex 200 10/300 GL (GE Healthcare) equilibrated in 25 mM HEPES/NaOH, 200 mM NaCl, pH 8.0, to obtain a homogenous protein sample; peak fractions containing the protein of interest were pooled and concentrated to 1 mg mL^{-1}.

5.4 Crystallization, data collection, structure determination and refinement

Wild-type and mutant LH3 co-crystallization experiments were performed using the hanging-drop vapor-diffusion method protocols as described in (31), by mixing 0.5 µL of enzyme concentrated at 3.5 mg mL^{-1} with 0.5 µL of reservoir solutions composed of 600 mM sodium formate, 12% poly-glutamic-acid (PGA-LM, Molecular Dimensions), 100 mM HEPES/NaOH, pH 7.8, 500 µM $FeCl_2$, 500 µM $MnCl_2$, supplemented with 1 mM of the appropriate UDP-sugar analogs (UDP, UDP-glucose, UDP-glucuronic acid, UDP-xylose). Crystals were cryo-protected with the mother liquor supplemented with 20% ethylene glycol, harvested using MicroMounts Loops (Mitegen), flash-cooled and stored in liquid nitrogen prior to data acquisition. X-ray diffraction data were collected at various beamlines of the European Synchrotron Radiation Facility, Grenoble, France and at the Swiss Light Source, Villigen, Switzerland. Data were indexed and integrated using *XDS* (50) and scaled using *Aimless* (51). Data collection statistics are summarized in Suppl. Table 2. The data showed strong anisotropy and therefore underwent anisotropic cut-off using STARANISO (52) prior to structure determination and refinement. The structures were solved by molecular replacement using the structure of wild-type LH3 in complex with Fe^{2+}, 2-OG and Mn^{2+} (pdb: 6FXM) (31) as search model using *PHASER* (53). The structure was refined with successive steps of manual building in *COOT* (54) and automated refinement with *phenix.refine* (55). Model validation was performed with *MolProbity* (56). Refinement statistics for the final models are reported in Suppl. Table 2.

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5.5 Evaluation of LH3 GalT/GlcT enzymatic activity

LH, GalT and GlcT activities were tested using luminescence-based enzymatic assays with a Glomax Discover (Promega, USA) as described in *Scietti et al.* (31). Minor modifications have been done for GalT/GlcT competitive inhibition assays, where 1 µL of a miture of 250 µM $MnCl_2$, 500 µM UDP-galactose (GalT assay) or UDP-glucose (GlcT assay) and increasing concentrations of either UDP-GlcA or UDP-Xyl were initially added to the enzyme and gelatin substrate to start the reactions. All experiments were performed in triplicates. Control experiments were performed in the same conditions by selectively removing LH3. Data were analyzed and plotted using the GraphPad Prism 7 (Graphpad Software, USA)

5.6 Differential Scanning Fluorimetry (DSF)

DSF assays were performed on LH3 wt and mutants using a Tycho NT.6 instrument (NanoTemper Technologies GmbH, Germany). LH3 samples at a concentration of 1 mg/mL in a buffer composed of 25 mM HEPES, 500 mM NaCl, pH 8. Binding assays were performed by incubating LH3 with 50 uM $MnCl_2$ and 5 mM free UDP or UDP-sugar donor substrates or their analogs. Data were analyzed and plotted using the GraphPad Prism 7 (Graphpad Software, USA).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGEMENTS

We thank the European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS) for the provision of synchrotron radiation facilities. We thank Dr. M. Campioni for useful discussions and M. Miao for support in crystallization experiments. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Curie grant agreement COTETHERS - n. 745934 to AC, the Italian Association for Cancer Research (AIRC, "My First AIRC Grant" id. 20075 to FF), by Fondazione Giovanni Armenise-Harvard (CDA2013 to FF), and by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022, to the Dept. of Biology and Biotechnology "L. Spallanzani", University of Pavia). None of the funding sources had roles in study design, collection, analysis and interpretation of data, in the writing of the report and in the decision to submit this article for publication.

AUTHOR CONTRIBUTIONS

AC generated and purified LH3 mutants with support from SF, carried out biochemical investigations and crystallized LH3 mutants; LS purified and crystallized wild-type LH3 in complex with UDP-sugar analogs; FDG performed LH3 enzymatic activity assays; LS, FDG and FF carried out structural refinement and analyses; TR was part of the study group focusing on LH3 pathogenic variants. FF designed the study with help from AC, FDG and LS. AC and FF wrote the paper, with contributions from all authors.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: List of glycosyltransferase enzymes used for comparisons with human LH3. The list includes the indication of the catalytic bases and nucleophile residues as proposed in the original papers describing the various glycosyltransferases, with the corresponding residue number in human LH3 based on structural superpositions.

Protein Name	Type	PDB ID	catalytic base residue	nucleophile acceptor residue	corresp. LH3 residue	reference paper
LgtC - GALACTOSYL TRANSFERASE LGTC *(N. meningitidis)*	retaining	1GA8				Persson et al., 2001 10.1038/84168
GYG1 - Glycogenin *(O. cuniculus)*	retaining	1LL2		Asp163	Asp191	Gibbons et al., 2002 10.1016/S0022-2836(02)00305-4
mgs - Mannosylglycerate synthase (*R. marinus)*	retaining	2BO8				Flint et al., 2005 10.1038/nsmb950
GALNT10 - Polypeptide N-acetylgalactosaminyl transferase 10 (*H. sapiens)*	retaining	2D7R		Gln346	Asp190	Kubota et al., 2006 10.1016/j.jmb.2006.03.061
GGTA1 - N-Acetyllactosaminide α-1,3-galactosyl transferase (R365K) *(B. taurus)*	retaining	5NRB		Glu317	Gln192	Albesa-Jove et al., 2017 10.1002/anie.201707922
ABO - Histo-blood group ABO system transferase *(H. sapiens)*	retaining	1LZI		Glu303	Gln192	Patenaude et al., 2002 10.1038/nsb832
spsA - PROTEIN (SPORE COAT POLYSACCHARIDE BIOSYNTHESIS PROTEIN SPSA)	inverting	1QGQ	Asp191		Asp191	Charnock et al., 1999 10.1128/JB.183.1.77-85.2001
MGAT1 - N-acetylglucosaminyl transferase I *(O. cuniculus)*	inverting	1FOA	Asp 291		Asp191	Unligil et al., 2000 10.1093/emboj/19.20.5269
Mfng - Manic Fringe glycosyltransferase *(M. musculus)*	inverting	2J0B	Asp 232		Asp191	Jinek et al., 2006 10.1038/nsmb1144
B3GAT3 - GLUCURONYLTRANSFERASE I	inverting	1FGG	Glu281		Asp190	Pedersen et al., 2000 10.1074/jbc.M007399200
B3GAT1 – Galactosylgalactosylxylosyl protein 3-beta-glucuronosyltransferase 1		1V84				Kakuda et al., 2004 10.1074/jbc.M400622200

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE LEGENDS

Figure 1: Features of the LH3 glycosyltransferase (GT) domain. (A) Cartoon representation of the LH3 GT domain (PDB ID: 6FXR) showing the key residues shaping the catalytic site as sticks. The PolyAsp motif (brown) and the Glycoloop (cyan) involved in binding of UDP-sugar donor substrates are shown. The residues implicated in the catalytic activity and investigated in this works are colored, while the residues depicted in grey have already been shown to be essential in Mn^{2+} (purple sphere) and UDP (black sticks) coordination. (B) Sequence alignment of human LH1, LH2 and LH3 GT domains, highlighting similarities and differences in the aminoacid residues within the active site. Residues shown in Figure 1A are indicated with a triangle. Colored boxes indicate the PolyAsp motif (brown) and the Glycoloop (cyan).

Figure 2: Evaluation of the effect of LH3 GT domain mutations in the GT site on glycosyltransferase activities. Evaluation of the GalT activity (A) and GlcT activity (B) of LH3 mutants compared to the wild-type. The plotted data are baseline-corrected, where the baseline was the background control.

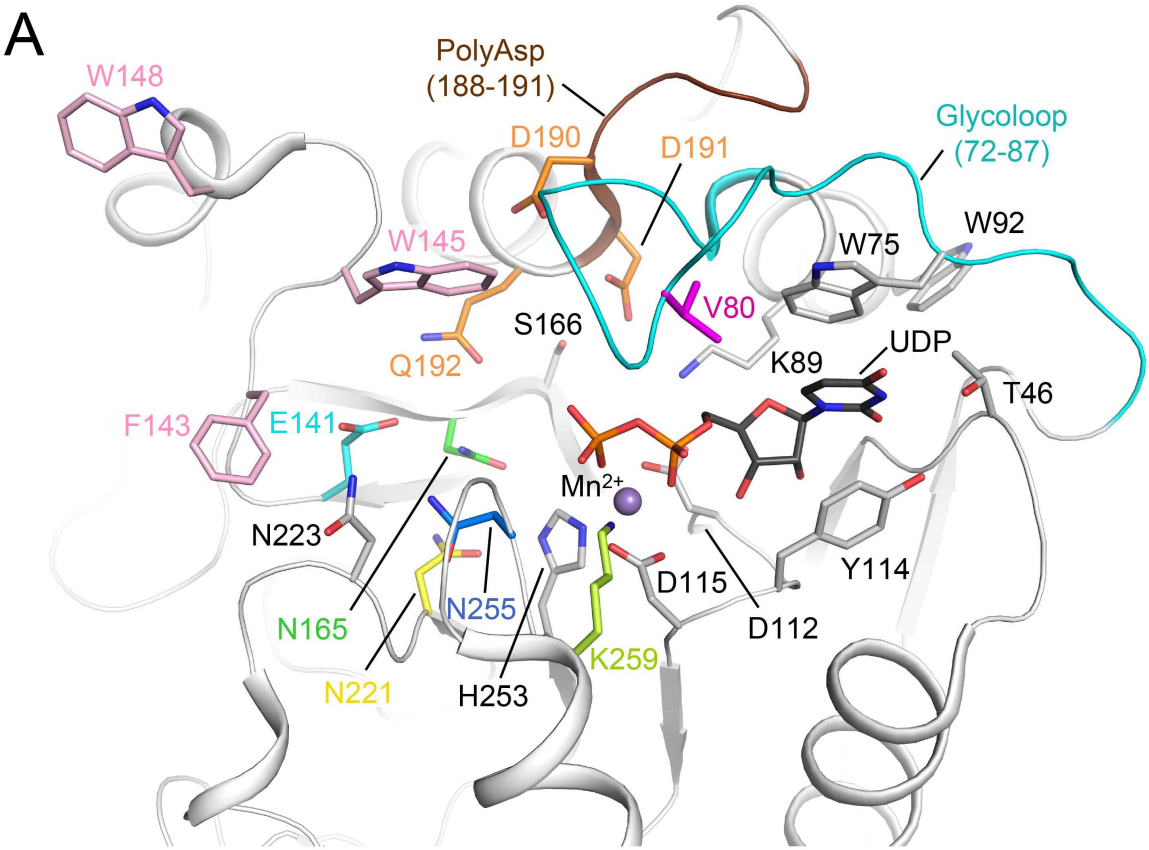
Figure 3. Structural characterization of the LH3 Val80Lys mutant. (A) Crystal structure of the LH3 Val80Lys mutant in complex with UDP-glucose and Mn^{2+}. Electron density is visible for the mutated lysine and the UDP donor substrate (green mesh, $2F_o$-F_c omit electron density map, contoured at 1.3 σ). Catalytic residues shaping the enzyme cavity are shown as sticks, Mn^{2+} is shown as purple sphere. Consistent with what observed in the crystal structure of wild-type LH3, the glucose moiety of the donor substrate is not visible in the experimental electron density. (B) Superposition of wild-type and Val80Lys LH3 crystal structures in substrate-free (cyan and yellow, respectively) with UDP-glucose bound (marine and orange,

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

respectively) states. Notably, the conformations adopted by the side chain of Trp145 upon ligand binding are consistent in the wild-type and in the mutant enzyme. As the glycoloop is flexible in substrate-free structures, the side chains of Val/Lys80 are visible only in the in UDP-sugar bound structures.

Figure 4. Characterization of UDP-sugar analogs. (A) Thermal stability of LH3 wild-type (solid green) using differential scanning fluorimetry (DSF) in presence of various Mn^{2+} and several UDP-sugars. A prominent stabilization effect is achieved in presence of the biological donor substrates UDP-galactose (solid blue), UDP-Glucose (solid purple) and free UDP (solid black). A milder stabilization effect is also obtained with UDP-xylose (red dash) and UDP-glucuronic acid (green dash). (B) Evaluation of GalT and GlcT enzymatic activities of LH3 in the presence of increasing concentrations of UDP-GlcA or UDP-Xyl. (C) Crystal structure of LH3 wild-type in complex with Mn^{2+} and UDP-glucuronic acid shows clear electron density for UDP ($2F_o$-F_c omit electron density maps, green mesh, contour level 1.2 σ). The glucuronic acid (shown in yellow) can be modelled even if with partial electron density. (D) Crystal structure of the LH3 Val80Lys mutant in complex with Mn^{2+} and UDP-glucuronic acid. Whereas the UDP backbone can be modelled in the electron density (black sticks) ($2F_o$-F_c omit electron density maps, green mesh, contour level 1.2 σ), in this case no electron density is present for the glucuronic acid (shown in yellow). In addition, the portion of the glycoloop containing the mutated lysine is flexible from residue 79 to 83 (shown as cyan spheres). (E) Crystal structure of LH3 wild-type in complex with Mn^{2+} and UDP-xylose. Similarly to UDP-GlcA, UDP shows clear electron density ($2F_o$-F_c omit electron density maps, green mesh, contour level 1.2 σ), whereas partial density is shown for the xylose moiety (shown in pink).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

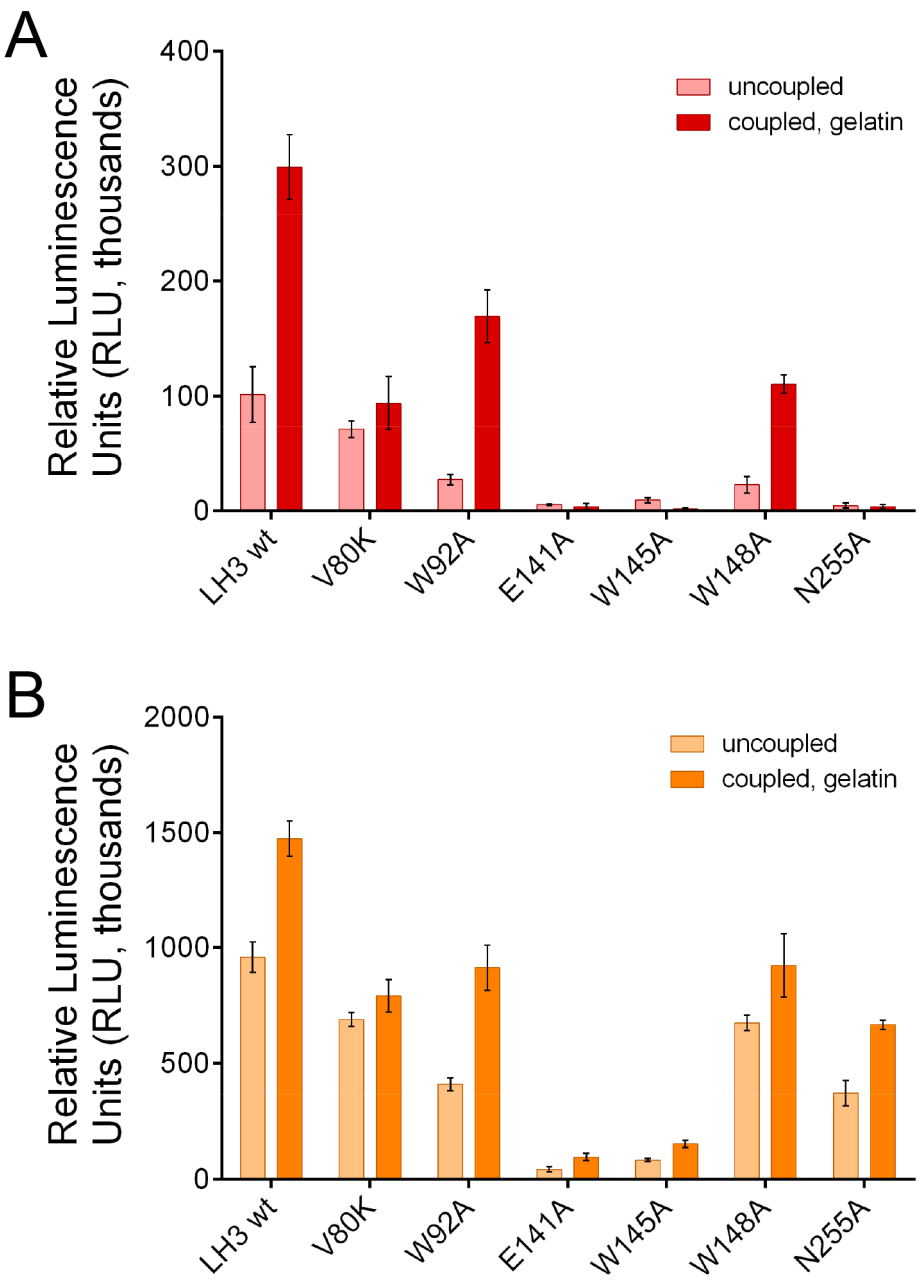


B

```
                                                                                Glycoloop
                 33          43          53          63          73          83
Human LH3        PVNPEKLLVI  TVATAETEGY  LRFLRSAEFF  NYTVRTLGLG  EEWRGGDVAR  TVGGGQKVRW
Human LH1        AKPEDNLLVL  TVATKETEGF  RRFKRSAQFF  NYKIQALGLG  EDWN-VEKGT  SAGGGQKVRL
Human LH2        SIPTDKLLVI  TVATKESDGF  HRFMQSAKYF  NYTVKVLGQG  EEWRGGDGIN  SIGGGQKVRL
consensus        .   ::***:  **** *::*:   ** .**::*  **.:..** *  *:*.  .     : *******

                 93         103         113         123         133         143
Human LH3        LKKEMEKYAD  REDMIIMFVD  SYDVILAGSP  TELLKKFVQS  GSRLLFSAES  FCWPEWGLAE
Human LH1        LKKALEKHAD  KEDLVILFAD  SYDVLFASGP  RELLKKFRQA  RSQVVFSAEE  LIYPDRRLET
Human LH2        MKEVMEHYAD  QDDLVVMFTE  CFDVIFAGGP  EEVLKKFQKA  NHKVVFAADG  ILWPDKRLAD
consensus        :*: :*::**  .:*::::*.:  .:**::*..*  *:**** ::     .::*:*:   : :*:. *

                                                                PolyAsp
                153         163         173         183         193         203
Human LH3        QYPEVGTGKR  FLNSGGFIGF  ATTIHQIVRQ  WKYKDDDDDQ  LFYTRLYLDP  GLREKLSLNL
Human LH1        KYPVVSDGKR  FLGSGGFIGY  APNLSKLVAE  WEGQDSDSDQ  LFYTKIFLDP  EKREQINITL
Human LH2        KYPVVHIGKR  YLNSGGFIGY  APYVNRIVQQ  WNLQDNDDDQ  LFYTKVYIDP  LKREAINITL
consensus        :** *  ***  :*.******:  *. : .:* :  *: :*.*.**  ****.:::**    ** :.:.*

                213         223         233         243         253         263
Human LH3        DHKSRIFQNL  NGALDEVVLK  FDRNRVRIRN  VAYDTLPIVV  HGNGPTKLQL  NYLGNYVPNG
Human LH1        DHRCRIFQNL  DGALDEVVLK  FEMGHVRARN  LAYDTLPVLI  HGNGPTKLQL  NYLGNYIPRF
Human LH2        DHKCKIFQTL  NGAVDEVVLK  FENGKARAKN  TFYETLPVAI  NGNGPTKILL  NYFGNYVPNS
consensus        **...***.*  :**:******  *: ...* .*    *:***: :  :******: *  **:***:*.

                273         283
Human LH3        WTPEGGCGFC  NQDRRTLPGG
Human LH1        WTFETGCTVC  DEGLRSLKGI
Human LH2        WTQDNGCTLC  EFDTVDLSAV
consensus        ** : ** .*  : .   * .
```

Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



615

616 **Figure 2**

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

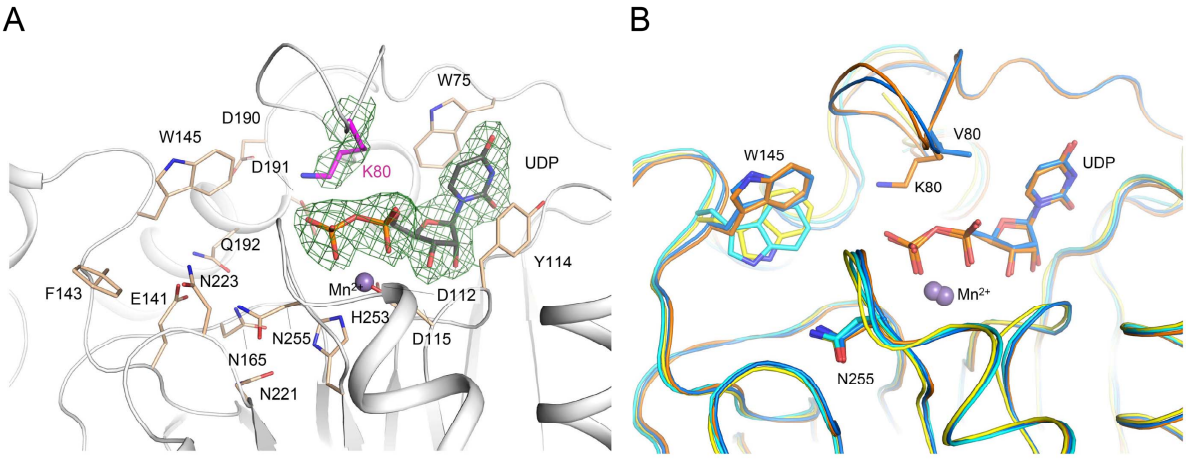


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

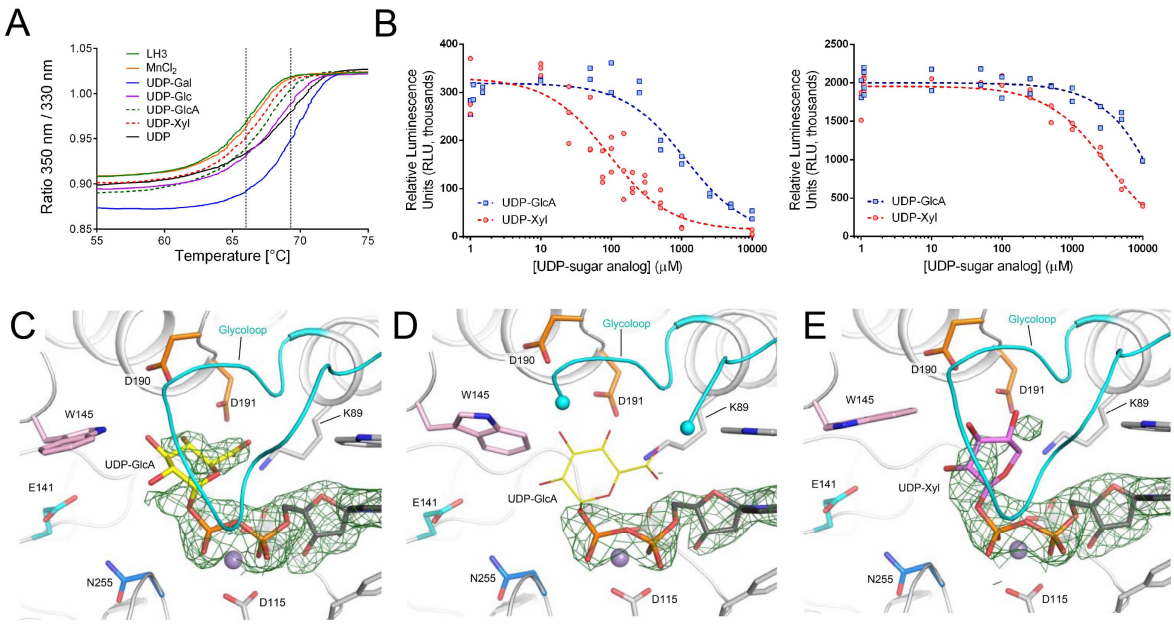


Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Myllyharju J & Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. (Translated from eng) *Trends Genet* 20(1):33-43 (in eng).

2. Luther KB, *et al.* (2011) Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. (Translated from eng) *J Biol Chem* 286(51):43701-43709 (in eng).

3. Cummings RD (2009) The repertoire of glycan determinants in the human glycome. (Translated from eng) *Mol Biosyst* 5(10):1087-1104 (in eng).

4. Spiro RG (1969) Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. (Translated from eng) *J Biol Chem* 244(4):602-612 (in eng).

5. Spiro RG (1967) The structure of the disaccharide unit of the renal glomerular basement membrane. (Translated from eng) *J Biol Chem* 242(20):4813-4823 (in eng).

6. Spiro MJ & Spiro RG (1971) Studies on the biosynthesis of the hydroxylsine-linked disaccharide unit of basement membranes and collagens. II. Kidney galactosyltransferase. (Translated from eng) *J Biol Chem* 246(16):4910-4918 (in eng).

7. Hamazaki H & Hamazaki MH (2016) Catalytic site of human protein-glucosylgalactosylhydroxylysine glucosidase: Three crucial carboxyl residues were determined by cloning and site-directed mutagenesis. (Translated from eng) *Biochem Biophys Res Commun* 469(3):357-362 (in eng).

8. Sternberg M & Spiro RG (1980) Studies on the catabolism of the hydroxylysine-linked disaccharide units of basement membranes and collagens: isolation and characterization of a new rat-kidney alpha-glucosidase of high specificity. (Translated from eng) *Ren Physiol* 3(1-6):1-3 (in eng).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

9. Sternberg M, Grochulski A, Peyroux J, Hirbec G, & Poirier J (1982) Studies on the alpha-glucosidase specific for collagen disaccharide units: variations associated with capillary basement membrane thickening in kidney and brain of diabetic and aged rats. (Translated from eng) *Coll Relat Res* 2(6):495-506 (in eng).

10. Bornstein P & Sage H (1980) Structurally distinct collagen types. (Translated from eng) *Annu Rev Biochem* 49:957-1003 (in eng).

11. Terajima M, *et al.* (2014) Glycosylation and cross-linking in bone type I collagen. (Translated from eng) *J Biol Chem* 289(33):22636-22647 (in eng).

12. Moro L, Romanello M, Favia A, Lamanna MP, & Lozupone E (2000) Posttranslational modifications of bone collagen type I are related to the function of rat femoral regions. (Translated from eng) *Calcif Tissue Int* 66(2):151-156 (in eng).

13. Schofield JD, Freeman IL, & Jackson DS (1971) The isolation, and amino acid and carbohydrate composition, of polymeric collagens prepared from various human tissues. (Translated from eng) *Biochem J* 124(3):467-473 (in eng).

14. Toole BP, Kang AH, Trelstad RL, & Gross J (1972) Collagen heterogeneity within different growth regions of long bones of rachitic and non-rachitic chicks. (Translated from eng) *Biochem J* 127(4):715-720 (in eng).

15. Rautavuoma K, *et al.* (2004) Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. (Translated from eng) *Proc Natl Acad Sci U S A* 101(39):14120-14125 (in eng).

16. Sipila L, *et al.* (2007) Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. (Translated from eng) *J Biol Chem* 282(46):33381-33388 (in eng).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Lehmann HW, *et al.* (1995) Composition and posttranslational modification of individual collagen chains from osteosarcomas and osteofibrous dysplasias. (Translated from eng) *J Cancer Res Clin Oncol* 121(7):413-418 (in eng).

18. Salo AM, *et al.* (2008) A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. (Translated from eng) *Am J Hum Genet* 83(4):495-503 (in eng).

19. Tenni R, Valli M, Rossi A, & Cetta G (1993) Possible role of overglycosylation in the type I collagen triple helical domain in the molecular pathogenesis of osteogenesis imperfecta. (Translated from eng) *Am J Med Genet* 45(2):252-256 (in eng).

20. Hennet T (2019) Collagen glycosylation. (Translated from eng) *Curr Opin Struct Biol* 56:131-138 (in eng).

21. Wang C, *et al.* (2002) The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. (Translated from eng) *Matrix Biol* 21(7):559-566 (in eng).

22. Heikkinen J, *et al.* (2000) Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. (Translated from eng) *J Biol Chem* 275(46):36158-36163 (in eng).

23. Savolainen ER, Kero M, Pihlajaniemi T, & Kivirikko KI (1981) Deficiency of galactosylhydroxylysyl glucosyltransferase, an enzyme of collagen synthesis, in a family with dominant epidermolysis bullosa simplex. (Translated from eng) *N Engl J Med* 304(4):197-204 (in eng).

24. Ewans LJ, *et al.* (2019) Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. (Translated from eng) *J Med Genet* (in eng).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. Ruotsalainen H, *et al.* (2006) Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. (Translated from eng) *J Cell Sci* 119(Pt 4):625-635 (in eng).

26. Schegg B, Hulsmeier AJ, Rutschmann C, Maag C, & Hennet T (2009) Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. (Translated from eng) *Mol Cell Biol* 29(4):943-952 (in eng).

27. Perrin-Tricaud C, Rutschmann C, & Hennet T (2011) Identification of domains and amino acids essential to the collagen galactosyltransferase activity of GLT25D1. (Translated from eng) *PLoS One* 6(12):e29390 (in eng).

28. Sricholpech M, *et al.* (2011) Lysyl hydroxylase 3 glucosylates galactosylhydroxylysine residues in type I collagen in osteoblast culture. (Translated from eng) *J Biol Chem* 286(11):8846-8856 (in eng).

29. Yamauchi M & Sricholpech M (2012) Lysine post-translational modifications of collagen. (Translated from eng) *Essays Biochem* 52:113-133 (in eng).

30. Baumann S & Hennet T (2016) Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase. (Translated from eng) *J Biol Chem* 291(35):18514-18524 (in eng).

31. Scietti L, *et al.* (2018) Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. (Translated from eng) *Nat Commun* 9(1):3163 (in eng).

32. Wang C, *et al.* (2002) Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). (Translated from eng) *J Biol Chem* 277(21):18568-18573 (in eng).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

33. Persson K, *et al.* (2001) Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. (Translated from eng) *Nat Struct Biol* 8(2):166-175 (in eng).

34. Flint J, *et al.* (2005) Structural dissection and high-throughput screening of mannosylglycerate synthase. (Translated from eng) *Nat Struct Mol Biol* 12(7):608-614 (in eng).

35. Charnock SJ & Davies GJ (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. (Translated from eng) *Biochemistry* 38(20):6380-6385 (in eng).

36. Pedersen LC, *et al.* (2000) Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. (Translated from eng) *J Biol Chem* 275(44):34580-34585 (in eng).

37. Lairson LL, Henrissat B, Davies GJ, & Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. (Translated from eng) *Annu Rev Biochem* 77:521-555 (in eng).

38. Ardevol A, Iglesias-Fernandez J, Rojas-Cervellera V, & Rovira C (2016) The reaction mechanism of retaining glycosyltransferases. (Translated from eng) *Biochem Soc Trans* 44(1):51-60 (in eng).

39. Gloster TM (2014) Advances in understanding glycosyltransferases from a structural perspective. (Translated from eng) *Curr Opin Struct Biol* 28:131-141 (in eng).

40. Albesa-Jove D, Sainz-Polo MA, Marina A, & Guerin ME (2017) Structural Snapshots of alpha-1,3-Galactosyltransferase with Native Substrates: Insight into the Catalytic Mechanism of Retaining Glycosyltransferases. (Translated from eng) *Angew Chem Int Ed Engl* 56(47):14853-14857 (in eng).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

41. Coutinho PM, Deleury E, Davies GJ, & Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. (Translated from eng) *J Mol Biol* 328(2):307-317 (in eng).

42. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. (Translated from eng) *Nucleic Acids Res* 42(Database issue):D490-495 (in eng).

43. Gomez H, Lluch JM, & Masgrau L (2012) Essential role of glutamate 317 in galactosyl transfer by alpha3GalT: a computational study. (Translated from eng) *Carbohydr Res* 356:204-208 (in eng).

44. Patenaude SI, *et al.* (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis. (Translated from eng) *Nat Struct Biol* 9(9):685-690 (in eng).

45. Kivirikko KI & Myllyla R (1979) Collagen glycosyltransferases. (Translated from eng) *Int Rev Connect Tissue Res* 8:23-72 (in eng).

46. Myllyla R, Anttinen H, Risteli L, & Kivirikko KI (1977) Isolation of collagen glucosyltransferase as a homogeneous protein from chick embryos. (Translated from eng) *Biochim Biophys Acta* 480(1):113-121 (in eng).

47. Scietti L, Campioni M, & Forneris F (2019) SiMPLOD, a structure-integrated database of collagen lysyl hydroxylase (LH/PLOD) enzyme variants. (Translated from eng) *J Bone Miner Res* (in eng).

48. Lairson LL, *et al.* (2004) Intermediate trapping on a mutant retaining alpha-galactosyltransferase identifies an unexpected aspartate residue. (Translated from eng) *J Biol Chem* 279(27):28339-28344 (in eng).

49. Raghunath M, Bruckner P, & Steinmann B (1994) Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. (Translated from eng) *J Mol Biol* 236(3):940-949 (in eng).

bioRxiv preprint doi: https://doi.org/10.1101/841486; this version posted November 15, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

50. Kabsch W (2010) Xds. (Translated from eng) *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125-132 (in eng).

51. Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? (Translated from eng) *Acta Crystallogr D Biol Crystallogr* 69(Pt 7):1204-1214 (in eng).

52. Tickle IJ, *et al.* (2018) STARANISO (Global Phasing Ltd. , Cambridge, United Kingdom).

53. McCoy AJ, *et al.* (2007) Phaser crystallographic software. (Translated from Eng) *J Appl Crystallogr* 40(Pt 4):658-674 (in Eng).

54. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D* 66(4):486-501.

55. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. (Translated from eng) *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221 (in eng).

56. Chen VB, *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. (Translated from eng) *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12-21 (in eng).