nature communications

Article

https://doi.org/10.1038/s41467-025-57768-9

The structural basis for the human procollagen lysine hydroxylation and dual-glycosylation

Received: 5 June 2024

Accepted: 3 March 2025

Junjiang Peng[1,2,5], Wenguo Li[1,2,5], Deqiang Yao[3,5], Ying Xia[2], Qian Wang[2], Yan Cai[2], Shaobai Li[2], Mi Cao[2], Yafeng Shen[2], Peixiang Ma[1,4], Rijing Liao[2], Jie Zhao[1,4], An Qin[1,4] & Yu Cao[1,2]

The proper assembly and maturation of collagens necessitate the orchestrated hydroxylation and glycosylation of multiple lysyl residues in procollagen chains. Dysfunctions in this multistep modification process can lead to severe collagen-associated diseases. To elucidate the coordination of lysyl processing activities, we determine the cryo-EM structures of the enzyme complex formed by LH3/PLOD3 and GLT25D1/ColGalT1, designated as the KOGG complex. Our structural analysis reveals a tetrameric complex comprising dimeric LH3/PLOD3s and GLT25D1/ColGalT1s, assembled with interactions involving the N-terminal loop of GLT25D1/ColGalT1 bridging another GLT25D1/ColGalT1 and LH3/PLOD3. We further elucidate the spatial configuration of the hydroxylase, galactosyltransferase, and glucosyltransferase sites within the KOGG complex, along with the key residues involved in substrate binding at these enzymatic sites. Intriguingly, we identify a high-order oligomeric pattern characterized by the formation of a fiber-like KOGG polymer assembled through the repetitive incorporation of KOGG tetramers as the biological unit.

Post-translational modifications (PTMs) are fundamental mechanisms for regulating the structure and function of proteins within the cell. Among the myriad PTMs, glycosylation stands out as one of the most intricate and heterogeneous, profoundly influencing protein stability, intracellular trafficking, and molecular recognition[1,2]. The cellular landscape of protein glycosylation encompasses N-linked glycosylation on asparagine residues and O-linked glycosylation on residues featuring hydroxyl sidechains, including serine, threonine, and tyrosine. Additionally, lysine residues can undergo O-linked glycosylation, following pre-hydroxylation into hydroxylysine (Hyl). Lysyl O-glycosylation (Lys-O-glyco) takes place within the endoplasmic reticulum (ER) and holds a pivotal role in the maturation of both secreted and membrane-bound proteins[3–5]. In the context of collagen, proper Lys-O-glycosylation is integral to the formation of mature collagen triple helices and cross-linking processes[6,7]. Furthermore, collagen-like proteins such as mannan-binding lectin and adiponectin, as well as other proteins like fibrinogen-like protein 1 (FGL1) and cellular communication network factor 1 (CCN1), also exhibit glucose-galactose-hydroxylysine modifications, participating in the regulation of various biological processes, including protein secretion, stability, and hyperpolymer formation[8–11].

---

[1]Department of Orthopaedics, Shanghai Key Laboratory of Orthopaedic Implant, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China. [2]Shanghai Institute of Precision Medicine, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, 115 Jinzun Road, Shanghai 200125, China. [3]Institute of Aging & Tissue Regeneration, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China. [4]Shanghai Frontiers Science Center of Degeneration and Regeneration in Skeletal System, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China. [5]These authors contributed equally: Junjiang Peng, Wenguo Li, Deqiang Yao. e-mail: dr_qinan@163.com; yu.cao@shsmu.edu.cn

The enzymatic orchestration required for Lys-O-glycosylation involves multiple sequential steps. The initial step necessitates the hydroxylation of lysine residues. In the context of collagen, the enzymes responsible for lysine hydroxylation are members of the procollagen-lysine-2-oxoglutarate-5-dioxygenase (PLOD) family, which encompasses genes *PLOD1-3*, encoding multifunctional procollagen lysine hydroxylase and glycosyltransferase LH1-3 proteins, respectively[12–14]. LH2, for instance, specializes in hydroxylating lysine residues within the telopeptide regions of procollagen. The resulting hydroxylated lysine is subsequently oxidized by lysine oxidase, generating an aldehyde group that plays a crucial role in intermolecular cross-linking within collagen[15–17]. In contrast, LH1 and LH3 predominantly catalyze lysyl hydroxylation within the helical region of collagen[18,19]. The subsequent stage in collagen Lys-O-glycosylation involves the enzymatic attachment of a galactosyl group to the hydroxyl group of hydroxylysine, catalyzed by glycosyltransferase 25 family member 1/procollagen galactosyltransferase 1 (GLT25D1/ColGalT1)[20,21]. Following this, LH3 facilitates the linkage of glucose to the galactosyl group, ultimately forming the characteristic glucose-galactose-hydroxylysine motif[22,23]. While LH3 has been traditionally considered the primary glucosyltransferase in the PLOD family, recent stutides revealed that LH1 and LH2 also exhibit glucosyltransferase activity in vitro[24,25], albeit at lower levels compared to LH3[22,26]. This intricate enzymatic cascade, LH1/2/3-ColGalT1-LH1/2/3, is central to the structural integrity and functional diversity of Lys-O-glycosylated proteins, particularly in the context of collagen and collagen-like molecules.

LH3/PLOD3 is a multifunctional enzyme that serves as a pivotal player in the Lysyl O-glycosylation (Lys-O-glyco) pathway by catalyzing both the initial lysyl hydroxylation and the final glucosyl transfer reactions. Notably, LH3/PLOD3 demonstrates robust glucosyltransferase activity. Studies in mice embryos with homozygous *PLOD3* knockout have shown a remarkable reduction of approximately 85% in glucosyltransferase activity compared to wild-type embryos[26]. LH3 works together with ColGalT1 to comprise the minimal functional unit performing the Lys-O-glycosylation of collagen, and the dysfunction of either enzyme can result in collagen defects, ultimately leading to severe disorders in connective, muscular, and cerebrovascular tissues. Mutations in the *COLGALT1* gene are associated with small cerebral vessel disease in humans[27,28], while *PLOD3* mutations are closely related to the occurrence of connective tissue diseases[29,30]. This underscores the critical roles of both GLT25D1/ColGalT1 and LH3/PLOD3 in maintaining the structural and functional integrity of collagen and related connective tissues.

Previous crystallographic studies have unveiled the structural architecture of the LH3 protein, revealing a tripartite domain organization. These domains consist of the C-terminal lysine hydroxylase domain, the N-terminal glycosyltransferase domain, and an intermediary accessory domain[31,32]. In contrast, GLT25D1/ColGalT1 has been predicted to comprise a two-domain structure, with its C-terminal domain taking on the primary responsibility for the galactosyl transfer reaction[33]. While individual enzymatic functions of the proteins involved in collagen glycosylation have been elucidated through biochemical and cell biological studies, the precise mechanisms underlying the coordinated enzymatic cascade remain incompletely understood. Similarly, the intricate details of the catalytic process involved in glycosyl transfer await a more comprehensive investigation.

In this study, we show that LH3 and ColGalT1 form a procollagen lysine hydroxylation and dual-glycosylation complex (KOGG complex for short) and present the cryo-electron microscopic (cryo-EM) structures of KOGG complex. The structures reveal a stoichiometry of 2:2 between the two enzymes in the LH3-ColGalT1 quaternary complex and show the spatial configuration of catalytic domains of the Lys-O-glyco reactive cascade. Additionally, through further EM analysis of the KOGG complex, we observe the existence of an enzyme matrix formed by interactions between quaternary complexes. These findings suggest a coupling mechanism for lysyl hydroxylation-galactosylation-glucosylation reactions in collagen processing.

Results

The human LH3/PLOD3 and GLT25D1/ColGalT1 form functional enzymatic complex

Recombinant human LH3 and ColGalT1 were successfully overexpressed via transient co-transfection of modified pcDNA3.4 vectors in HEK293F cells. Affinity chromatography targeting the twin strep tag at the N-terminus of ColGalT1 facilitated the purification of a stable protein complex, as confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Supplementary Fig. 1a). Subsequent purification through size-exclusion chromatography (SEC) yielded a mono-dispersed elution peak, indicating a 1:1 stoichiometry of LH3 and ColGalT1 as evaluated by SDS-PAGE (Supplementary Fig. 1b and c). The enzymatic activities of the LH3-ColGalT1 complex, including lysine hydroxylase (LH), galactosyltransferase (GalT), and glucosyltransferase (GlcT), were evaluated using pooled SEC peak fractions. The 2-oxoglutarate (2-OG)-dependent oxidation, indicative of LH activity, was detected using a synthetic peptide with sequence LPGTAGLPGMKGHRGFSGLDG to mimic the collagen substrate region (Fig. 1a). This sequence is derived from the human collagen alpha-1(I) chain (Uniprot ID P02452, residues 255-275), and K265 corresponds to α1 Lys-87 in the triple-helical region of type I collagen and represents a major substrate site for the hydroxylation and glycosylation by LH3[34,35]. Moreover, both GalT and GlcT activities were observed in an LH activity-dependent manner, affirming the prerequisite hydroxylation of lysine for subsequent glycosylation (Fig. 1a). Furthermore, GlcT activity was markedly enhanced in reaction solutions supplemented with both UDP-glucose and UDP-galactose compared to solutions containing only UDP-glucose as the glycosyl-donor, indicating a sequential hydroxylation-galactosylation-glucosylation process (Fig. 1a). Consequently, we designated the LH3-ColGalT1 protein complex as the KOGG complex (Lys-O-galactosyl-glucosyl).

Electron microscopy data were acquired for all protein complexes using a Titan Krios transmission electron microscope (FEI) operated at 300 kV. The data underwent processing utilizing RELION3 and cryoSPARC[36,37]. 2-D classification of particles prepared from the KOGG complex supplemented with 100 μM UDP-galactose (KOGG/UDP-Gal) revealed characteristic features of the quaternary complex, consisting of two copies each of LH3 and ColGalT1 (Supplementary Fig. 2). Subsequent 3-D classification and refinement yielded an electron microscopy density map with an overall resolution of 3.4 Å and a local resolution ranging from 2.5 Å to 6.5 Å at the core region of the complex. A similar data collection and processing protocol was applied to the KOGG complex supplemented with 100 μM UDP-glucose (KOGG/UDP-Glc) and the non-supplemented form (KOGG apo) (Supplementary Figs. 3 and 4). This resulted in electron microscopy density maps with overall resolutions of 3.75 Å and 3.93 Å, respectively, and local resolutions ranging from 3.0 Å–8.2 Å at the core regions of the KOGG/UDP-Glc and KOGG apo complexes (Supplementary Figs. 3 and 4, supplementary Tables 1 and 2).

The overview of the EM structure KOGG complex

Overall, the KOGG complexes, whether supplemented with UDP-galactose, UDP-glucose, or in apo state, exhibit a similar quaternary structure with overall RMSD (root mean square deviations) around 0.31-0.36 Å in structural comparison. For EM analysis, the KOGG/UDP-Gal complex was utilized as the representative structure for its highest resolution. Sequence analysis revealed that LH3 comprises three structural domains: the GlcT domain at the N-terminal, the LH domain at the C-terminal, and an accessory domain (AC) situated between the

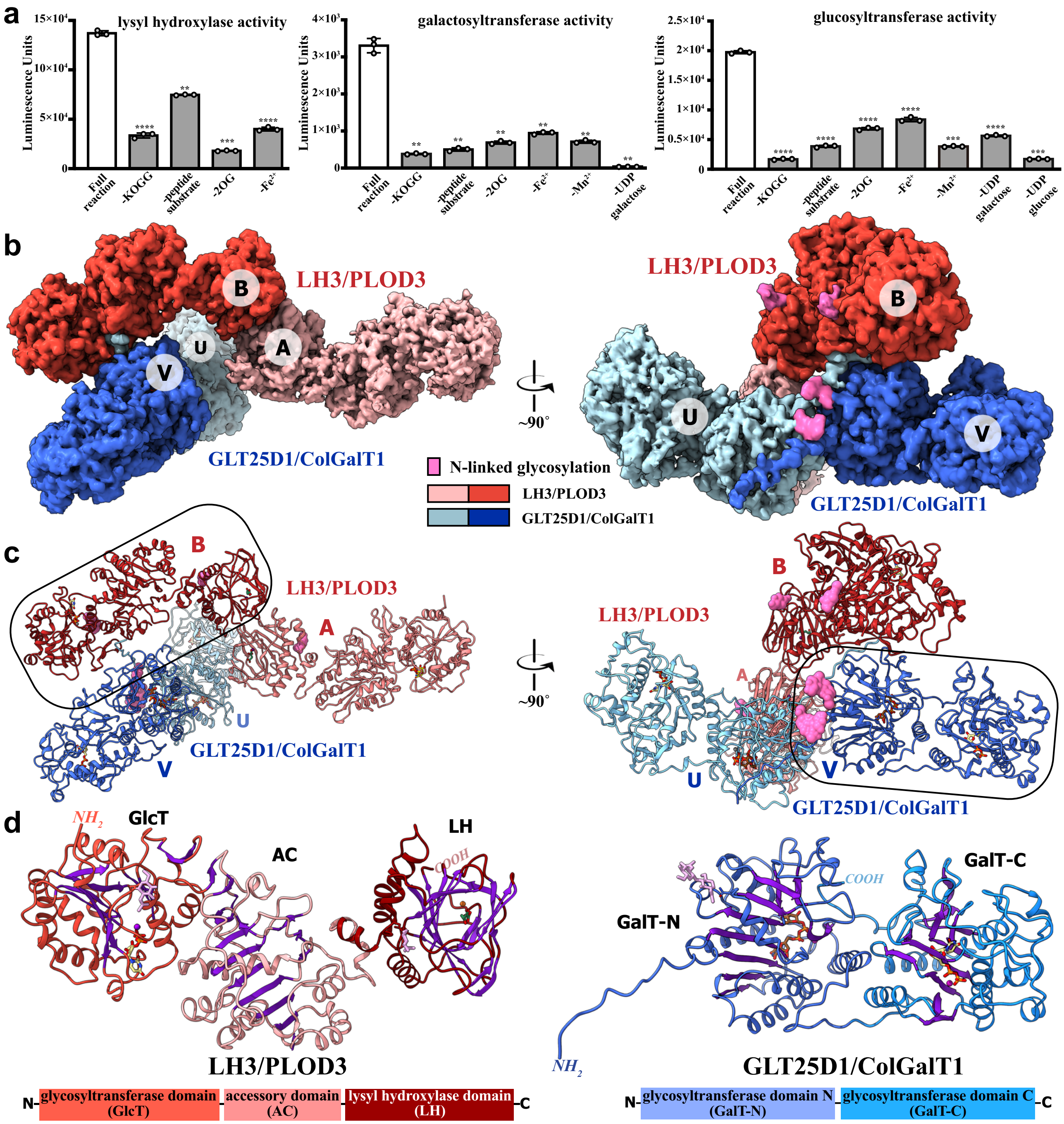

Fig. 1 | The overall architecture of the human KOGG complex. a Multi-functions of Purified LH3-ColGalT1 Complex. **Left**: 2-OG-dependent oxygenase activities of the purified KOGG complex. Oxygenase activities were minimal upon the removal of 2-OG from the reaction mixtures and substantially reduced in the absence of the KOGG complex sample, Fe^{2+}, or substrate peptide. **Middle**: Galactosyltransferase activities of the purified KOGG complex. The galactosyltransferase activities were minimal upon the removal of UDP-galactose from the reaction mixtures and substantially reduced in the absence of the KOGG complex sample, Mn^{2+}, or substrate peptide, as well as the necessary reaction components of 2-OG-dependent oxygenase activities, i.e., 2-OG and Fe^{2+}. **Right**: Glucosyltransferase activities of the purified KOGG complex. The glucosyltransferase activities were minimal upon the removal of UDP-glucose from the reaction mixtures and also substantially reduced in the absence of the KOGG complex sample, Mn^{2+}, or substrate peptide, as well as the necessary reaction components of 2-OG-dependent oxygenase activities, i.e. 2-OG and Fe^{2+}, and the necessary reaction components of galactosyltransferase activities, UDP-galactose. ($n = 3$ biologically independent samples). Data are presented as mean values +/− SD. Source data are provided as a Source Data file. Data analysis was performed using one-way ANOVA with Dunnett's multiple comparisons test. **, P-value < 0.01; ***, P-value < 0.001; ****, P-value < 0.0001. **b** Cryo-EM map of the heterotetrametic complex of human LH3 and ColGalT1. The cryo-EM density map of the human KOGG/UDP-Gal complex was viewed from two angles. The cryo-EM density map is colored in light red/red for $LH3^{A/B}$, cyan/blue for $ColGalT1^{U/V}$ and magenta for glycosyl modifications. **c** Molecular model of the KOGG complex. KOGG/UDP-Gal complex was depicted as a cartoon model, colored by protomers according to the corresponding cryo-EM maps shown above. **d** Structural analysis on LH3 (left) and ColGalT1 (right). The structure of $LH3^{B}$ and $ColGalT1^{V}$, indicated by the frames in the complex model above, are depicted as a cartoon model, colored by domains to the schemes shown below.

LH and GlcT domains[23,31], whereas ColGalT1 possesses two GalT domains[33], GalT-N and GalT-C, at the N- and C-termini, respectively (Supplementary Figs. 5 and 6). In the cryo-EM structure of the KOGG/UDP-Gal complex, both LH3 and ColGalT1 form homodimers. These two dimers then associate to form a T-shaped heterotetramer, where the long arm (LH3 dimer) intersects the short arm (ColGalT1 dimer) primarily at a contacting interface between the LH domain of LH3^{B} and the GalT-N domain of ColGalT1^{V} (Fig. 1b and c). Three N-linked glycosylation sites were identified in the cryo-EM map: Two of them are located in LH3 (N63 and N548), which aligns well with the observation by the crystallographic studies[31]; the 3rd N-glycosylation site was found in GLT25D1/ColGalT1, N184, located near the dimeric interface (Fig. 1b and c).

Both LH3 and ColGalT1 dimers exhibit a linear spatial configuration of structural domains. For the LH3 dimer, the arrangement is GlcT^{A}/AC^{A}/LH^{A}/LH^{B}/AC^{B}/GlcT^{B}, while for the ColGalT1 dimer, it is GalT-C^{U}/GalT-N^{U}/GalT-N^{V}/GalT-C^{V} (Fig. 1c and d). Interestingly, structural comparisons reveal a high similarity between GalT-N and AC (overall RMSD ~ 3.958 Å) and a moderate similarity between GalT-C and GlcT (overall RMSD ~ 14.468 Å), despite relatively low sequence identity between these domain pairs (23.83% for GalT-N vs. AC; 22.58% for GalT-C vs. GlcT) (Supplementary Fig. 7a, b). Further analysis using structural alignment of the core β-strands and helices of the GalT-C and GlcT domains provided a more reliable comparison, yielding a local RMSD of 5.385 Å (Supplementary Fig. 7c), suggesting a swapped N-to-C domain arrangement in GLT25D1/ColGalT1 compared to LH3/PLOD3 (Supplementary Fig. 7a).

The interfaces mediating the KOGG complex formation

The KOGG complex is a heterotetramer that is formed by the dimerization between the LH3 dimer and ColGalT1 dimer. Within the KOGG complex, the long LH3 dimer arm exhibits an overall structure similar to previously reported crystal structures (Supplementary Fig. 8), and the RMSDs between the cryo-EM and crystal structures range approximately from 1–3 Å (2.041 Å for LH3^{EM} vs. PDB ID 6TE3; 1.995 Å for LH3^{EM} vs PDB ID 6FXT)[31,32]. In addition to the hydrophobic interactions centered by the sidechain of L715 surrounded by F639 and Y642 from neighboring protomer, as established in the structural studies by Scietti et al.[31]., the cryo-EM structure reveals electrostatic interactions between R695 and D565 as a crucial mediating force for dimerization (Fig. 2a, b). The ColGalT1 dimer forms through contact between two GalT-N domains, where electrostatic interacting pairs (E46-K185, R53-D160, and E82-R248), as well as hydrogen bonds linking Q50 with R243/R248 and the glycosyl on N184 with the backbone carbonyl of S47, stabilize the ColGalT1 dimeric interface (Fig. 2c).

Critical to KOGG complex formation, the GalT-N domains not only mediate ColGalT1 dimerization but also facilitate ColGalT1 assembly onto LH3. Illustrated in Fig. 2a, d, the LH3-ColGalT1 interface entails a complex network involving GalT-N^{V}, GlcT^{B}, and the N-terminal loop from ColGalT1^{U} (N-loop^{U}), where GalT-N^{V} and GlcT^{B} form large, non-contacting surfaces, with a minimal slit width of approximately 7-8 Å between them. The N-terminal loop of ColGalT1^{U} extends from the neighboring ColGalT1 protomer, deeply inserting into the slit and engaging in electrostatic interactions with GalT-N^{V} (R42^{U}-E277^{V} and E46^{U}-K185^{V}) and LH3^{B} (E40^{U}-R452^{B}, E41^{U}-R241^{B}, and R42^{U}-E446^{B}). Additionally, N-loop^{U} (Y37, F38, and P39) binding into a hydrophobic groove, flanked by GlcT^{B} and AC^{B} (L226, V229, V230, L231, and L265) further stabilizes the interface (Fig. 2d, e). The co-immunoprecipitation assay revealed that mutations in the N-loop of ColGalT1, including single alanine substitutions at residues 37-42, 44, and 46, disrupted the formation of the KOGG complex while leaving the expression level of ColGalT1 largely unaffected (Fig. 2g). Similarly, alanine substitutions at interacting residues on LH3, such as 229-231, 241, 264, 265, 446, and 452, also resulted in a decrease or complete loss of complex formation (Fig. 2f).

The glycosyltransferase sites in GLT25D1/ColGalT1

The KOGG complex is a multi-functional enzyme assembly composed of five structural domains: GlcT, AC, and LH domains from LH3/PLOD3, and GalT-N and GalT-C domains from GLT25D1/ColGalT1 (Fig. 3a). It executes three enzymatic functions: hydroxylase, galactosyltransferase, and glucosyltransferase. In both KOGG/UDP-Gal and KOGG/UDP-Glc structures, large blobs of non-proteinous densities were identified within ColGalT1 protomers (Supplementary Fig. 9). The morphology and continuity of the non-proteinous density in the GalT-N domain corresponded well with UDP-Gal and UDP-Glc supplemented in KOGG/UDP-Gal and KOGG/UDP-Glc samples, respectively, thereby enabling us to model the respective substrates (Supplementary Fig. 9a, b). Conversely, the non-proteinous map in the GalT-C of KOGG/UDP-Gal exhibited a smaller size, and the density corresponding to the galactosyl group appeared fragmented (Supplementary Fig. 9b), suggesting an uncoupled enzymatic activity for the GalT-C domain in the cryo-EM sample preparation. Consequently, we modeled UDP, the hydrolysis product of UDP-Gal, in the GalT-C of KOGG/UDP-Gal. Intriguingly, almost no additional density was observed in the GalT-C of the KOGG complex supplemented with UDP-Glc, corroborating the established catalytic preference for galactosyl transfer by ColGalT1[32].

Many glycosyltransferases feature a metal ion-coordinating motif characterized by two aspartates separated by one or two residues and flanked by hydrophobic residues on both sides (DXD motif)[38,39]. In GLT25D1/ColGalT1, a previous study identified three DXD motifs based on sequence analysis, including residues 166-168 in the GalT-N domain, and residues 461-463 and 585-587 in the GalT-C domain. However, the third DXD motif (585-587) was found not to impact glycosyltransferase activity[33]. In the cryo-EM structure of GalT-N, the Mn^{2+} ion coordinates with the sidechains of D166 and D168, as well as the diphosphate moiety of UDP-Gal or UDP-Glc (Fig. 3b and Supplementary Fig. 9). The uridine moiety was positioned deeply into the binding pocket and stabilized by hydrogen bonds between the hydroxyl groups on ribosyl with the backbone carbonyl groups of L59 and A60, as well as the carbonyl on the uracil ring with the sidechain of Y126. The galactosyl group, on the other side of the diphosphate moiety, interacted with GalT-N via hydrogen bonds between its hydroxyl groups and the sidechains from R147, S236, D264, and D265 (Fig. 3b). Similarly, in KOGG/UDP-Glc, the UDP-Glc molecule was stabilized by Mn^{2+}, the DXD motif, and surrounding residues of GalT-N (Supplementary Fig. 9a). Mutations in the DXD motif and interacting residues of GalT-N within the KOGG complex, such as D166A, D168A, Y126A, R139A, and R147A, resulted in impaired galactosyltransferase activities in functional assays, corroborating the substrate-binding basis revealed by the structures (Fig. 3d).

Remarkably, structural analysis of the GalT-C domain of KOGG/UDP-Gal revealed that the proposed DXD motif, i.e., residues 461-463, is located on the outer surface of ColGalT1, far from the enzymatic site. Instead, a non-canonical EXD motif (E435-X-D437) was found within the inner cavity and coordinated with the Mn^{2+} ion (Fig. 3c). The diphosphate moiety of UDP interacted with the Mn^{2+} ion and R354, while the uridine moiety was stabilized by hydrogen bonds with the sidechain of S415 and the backbone amine groups of G377 and T571 (Fig. 3c). Mutations of the interacting residues identified within the GalT-C domain, including R354A, E435A, D437A, and T571A, abolished galactosyltransferase activity in functional assays, supporting the substrate-binding basis revealed by the structures (Fig. 3d). Distinguishing the specific role of GalT-N and GalT-C in the galactosyl transfer reaction could be challenging, as variations in the residues on substrate-binding pockets in either domain might disrupt ColGalT1's function. A closer examination of the local structures around the substrate-interacting residues in GalT-N and GalT-C reveals a hydrophobic environment around the β-face of the phosphate-sugar bond in GalT-N (Supplementary Fig. 9c), with no side-chain carboxylate available to serve as the base catalyst required for an inverting

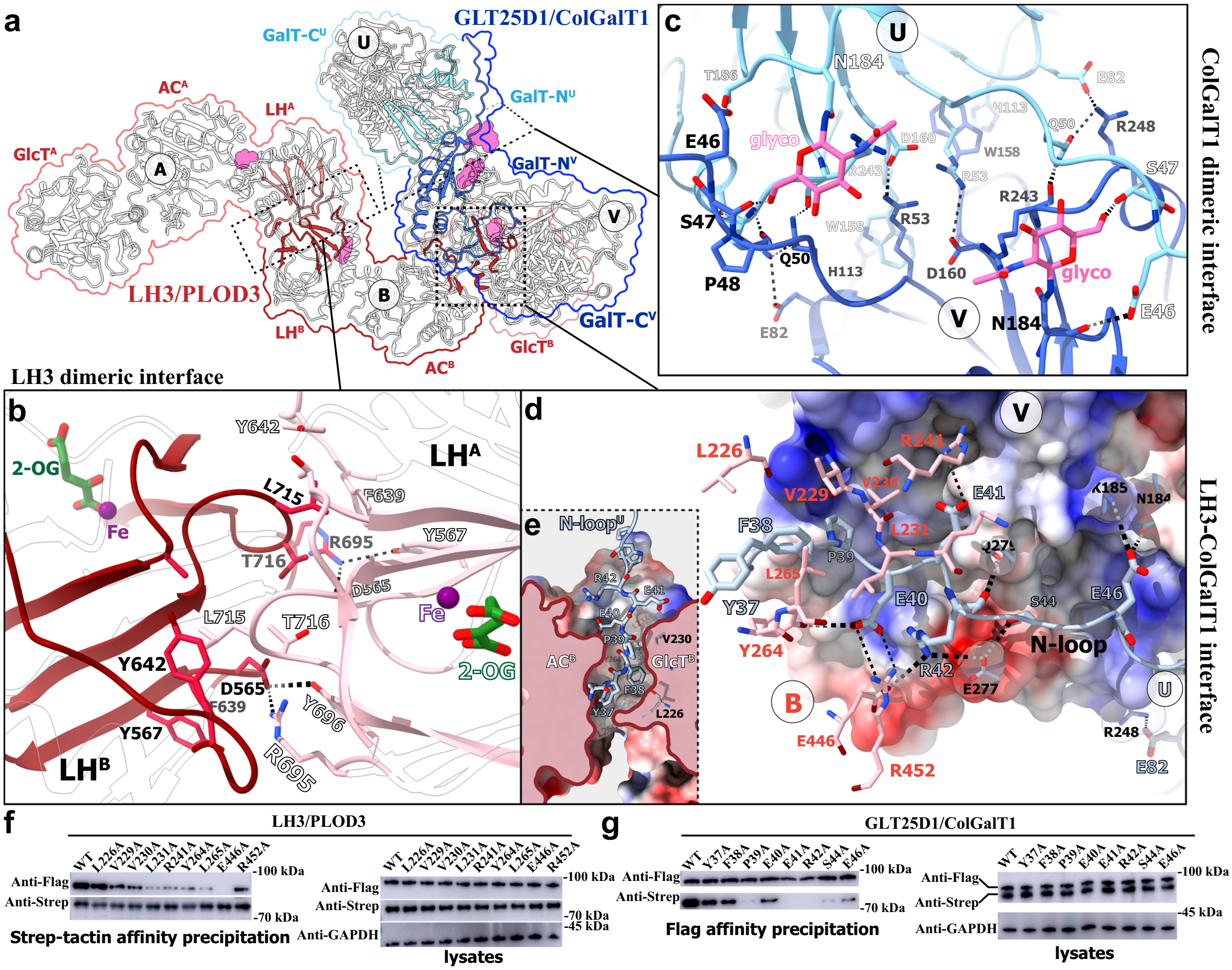

Fig. 2 | Molecular architecture of KOGG quaternary complex. a Overview of the intermolecular contacts within the KOGG complex. The protomers are depicted as a cartoon model, with secondary structures involved in complex formation highlighted in pink/red for LH3 and in blue/cyan for ColGalT1. Interfaces mediating complex assembly indicated by dotted frames. The N-linked glycosylations were depicted as surface models in magenta. **b** Enlarged views of the LH3 dimeric interface. LH3 is shown as a cartoon model, and the contacting residues are displayed as a stick model, colored according to the elements. **c** Enlarged views of the ColGalT1 dimeric interface. ColGalT1 is shown as a cartoon model, and the contacting residues are displayed as a stick model, colored according to the elements. The N-linked glycan moieties were shown as stick models, colored according to the elements (C-N-O: magenta-blue-red). **d** Enlarged view of the LH3-ColGalT1 interface. The N-loop of ColGalT1^U is shown as a cartoon model colored in light blue, while the GalT-N of ColGalT1^V is shown as a calculated solvent-accessible electrostatic surface potential map. The contacting residues from the N-loop^U, GalT-N^V, and GlcT^B are displayed as a stick model, colored according to the elements (light blue-red-blue for N-loop^U and GalT-N^V; pink-red-blue for LH3^B). **e** The interactions between N-loop of ColGalT1 and LH3. The protomer B of LH3 is shown as a calculated solvent-accessible electrostatic surface potential map and sliced to show the slit between its AC and GlcT domain. The N-loop of ColGalT1^U was shown as stick model colored by elements. **f**, **g** The interactions between LH3 and ColGalT1. The 293 T cells were co-transfected with LH3/PLOD3 wild-type or variants with flag-tag and GLT25D1/ColGalT1 wild-type or variants with Strep-tag. Shown were the representative immunoblot results from three independent experiments of the pulldown of LH3 and ColGalT1 by the Strep-tagged ColGalT1 using strep-tactin affinity resin (**f**) and the Flag-tagged LH3 using anti-Flag affinity resin (**g**), as well as the immunoblot results of lysates.

glycosyltransferase reaction[40]. The only carboxylate-containing residue near the β-phosphate in GalT-N, D265, is positioned on the distal side of the sugar anomeric carbon, making it unlikely to participate directly in the glycosylation reaction. In contrast, in the GalT-C domain, two carboxylates--D522 and E572--are located near the β-phosphate of UDP (Supplementary Fig. 9c). Structural comparisons with other functional GT-A enzymes, such as Bacillus subtilis SpsA (PDB ID: 1QGQ) and rabbit N-acetylglucosaminyltransferase GNT-I (PDB ID: 1FOA), reveal that D522 in ColGalT1 is structurally related to the catalytic residues D191 in SpsA and D291 in GNT-I (Supplementary Fig. 9c), positioning it as a likely candidate for the base catalyst in the inverting mechanism typical of Clan I GT-A enzymes[40,41]. This observation suggested that while GalT-N is essential for the functional integrity of ColGalT1, it might not play a direct catalytic role but rather contributes to maintaining the enzyme's structural stability.

The glycosyltransferase and lysyl hydroxylase sites in LH3

Similar structural analysis was conducted on the LH3 portion of the KOGG complex, revealing large, continuous, non-proteinous densities within the inner cavities of the LH and GlcT domains (Supplementary Fig. 10), corresponding to the metal ions and substrate/product molecules involved in lysyl hydroxylation (Fe and 2-OG) and glucosyl transfer reactions (Mn^{2+} and UDP) (Fig. 4a, b). Mutations in key residues within the LH domain, including Y656A, H667A, D669A, and H719A, disrupted all three enzymatic functions of the KOGG complex in functional assays (Fig. 4c, d). This supports the pivotal role of lysyl

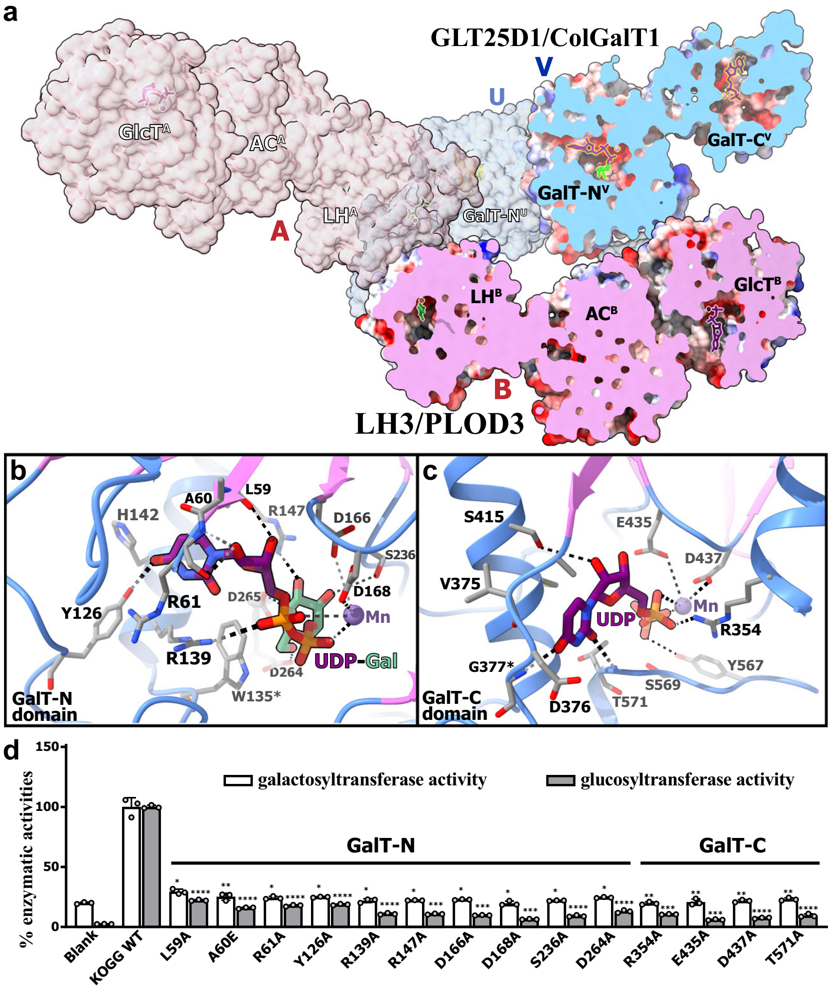

Fig. 3 | Catalytic sites in the KOGG complex. a Arrangement of the multiple catalytic sites in the KOGG complex. The KOGG/UDP-Gal complex was shown as surface model, with the protomer LH3^B and ColGalT1^V sliced to reveal the inner cavities in different domains. The metal ions were shown as spheres, and the UDP, UDP-Gal, and 2-OG were shown as stick models. **b** Enzymatic pocket in the GalT-N domain. The UDP-Gal bound in the pocket was shown as stick model colored by elements (carbon atoms in violet for UDP moiety and green for galactosyl). The Mn^{2+} ion was shown as a purple sphere. The interacting residues were shown as stick model colored by elements. **c** The enzymatic pocket in the GalT-C domain. The UDP bound in the pocket was shown as stick model colored by elements. The Mn^{2+} ion was shown as a purple sphere. The interacting residues were shown as stick model colored by elements. **d** The galactosyltransferase and glucosyltransferase activities of the purified KOGG complex with mutations at the catalytic sites of GalT-N and GalT-C ($n = 3$ biologically independent samples). Data are presented as mean values +/− SD. Source data are provided as a Source Data file. Data analysis was performed using one-way ANOVA with Dunnett's multiple comparisons test. *, P-value < 0.05; **, P-value < 0.01; ***, P-value < 0.001; ****, P-value < 0.0001.

hydroxylation in initiating the hydroxylation-galactosylation-glucosylation cascade of procollagens.

Particularly noteworthy is the identification of a long blob of density near the C1 carboxylate group of 2-OG and Fe ion in the LH domain of KOGG supplemented with either UDP-Glc or UDP-Gal (supplementary Fig. 10a and b), a position characteristic of the hydroxyl-acceptor substrate in 2-OG-dependent oxygenases[42,43], suggesting the presence of potential substrate peptides. The previous studies described a flexible capping loop (residues 590-610) near the catalytic center of the LH domain playing a gating role in response to Fe concentration[31,32]. Our analysis revealed that two potentially flexible segments -- the loop$^{588\text{-}612}$ (capping loop) and the C-terminal segment of LH3 (residues 729-738) -- were well-resolved with clear and continuous density (supplementary Fig. 10c and d). While this observation suggests that the unresolved density may not correspond to the capping loop, structural superposition between our cryo-EM structure and the previously published crystal structures shows a significant displacement of the capping loop (supplementary Fig. 10c). In the crystal structure, the second Fe ion prevents the capping loop from moving, maintaining a closed conformation that limits space around the 2-OG, thus decreasing accessibility to the substrate peptide. In contrast, in our cryo-EM structure, the capping loop moves outward in the absence of the second Fe ion, thereby opening up space in the reaction center and potentially allowing the substrate to enter. Given the current cryo-EM data, it remains challenging to definitively determine whether the unresolved density corresponds to the alternative conformation of the capping loop or to the substrate procollagen, which could have a low occupancy.

The non-proteinous densities in the GlcT domains of both KOGG/UDP-Gal and KOGG/UDP-Glc are well-fitted with UDP instead of UDP-Gal or UDP-Glc, respectively. These two UDP-only conformations for GlcT domains align well with observations from crystallography (Fig. 4a and supplementary Fig. 8c), implying that LH3 has at least hydrolytic capacity for UDP-Gal and UDP-Glc, if not a non-selective glycosyltransferase capacity[31,32]. The DXXD motif in the GlcT domain is located near the Fe ion, coordinating with D112 and D115 in the motif, as well as H253 and the diphosphate moiety of the bound UDP. Mutations in the interacting residues within the GlcT domain, such as

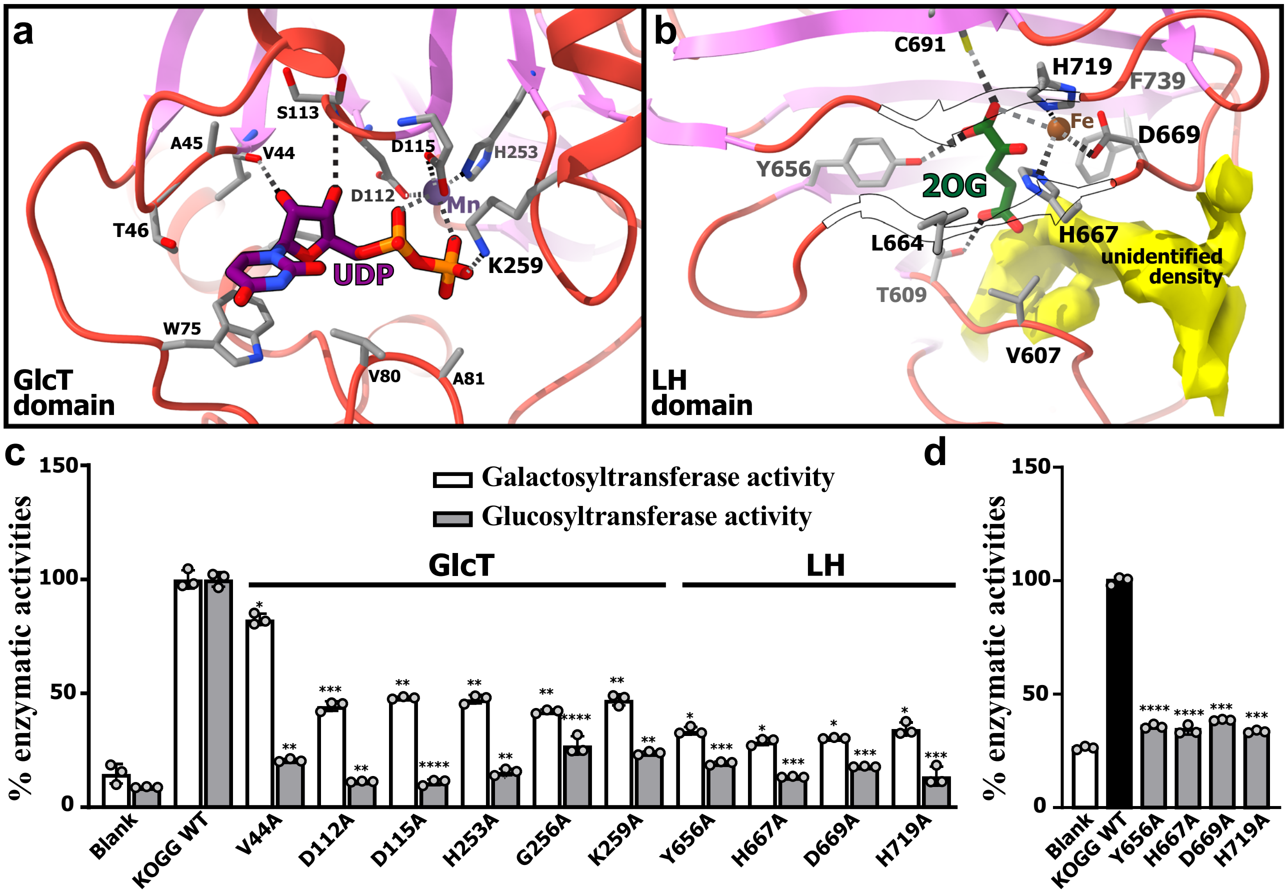

Fig. 4 | Catalytic sites in LH3. a Enzymatic pocket in the GlcT domain. The UDP bound in the pocket was shown as stick model colored by elements. The Mn^{2+} ion was shown as a purple sphere. The interacting residues were shown as stick model colored by elements. **b** The enzymatic pocket in the LH domain. The 2-OG bound in the pocket was shown as stick model colored by elements. The Fe ion was shown as a brown sphere. The interacting residues were shown as stick model colored by elements. The non-proteinous density in the cryo-EM map near the catalytic site was shown and highlighted in yellow. **c** The galactosyltransferase and glucosyltransferase activities of the purified KOGG complex with mutations at the catalytic sites of GlcT and LH domain of LH3 ($n = 3$ biologically independent samples). Data are presented as mean values +/− SD. Source data are provided as a Source Data file. Data analysis was performed using one-way ANOVA with Dunnett's multiple comparisons test. *, P-value < 0.05; **, P-value < 0.01; ***, P-value < 0.001; ****, P-value < 0.0001. **d** The lysyl hydroxylase activities of the purified KOGG complex with mutations at the catalytic site in the LH domain of LH3 ($n = 3$ biologically independent samples). Data are presented as mean values +/− SD. Source data are provided as a Source Data file. Data analysis was performed using one-way ANOVA with Dunnett's multiple comparisons test. ***, P-value < 0.001; ****, P-value < 0.0001.

V44A, D112A, D115A, and H253A, strongly impaired glucosyltransferase activity in functional assays. These findings support the substrate-binding interactions revealed by the structural data (Fig. 4c). Loose and fragmented densities could be found near the β-phosphate group of UDP in GlcT (supplementary Fig. 10a, b). However, due to the low resolution and the fragmented nature of these densities, reliably assigning their identity remains challenging. To ensure data precision, we have refrained from modeling the glycosyl portion of the substrates and instead fit only the UDP molecule into the main body of the non-proteinaceous densities.

The polymerization of KOGG quaternary complex

In 2D classification of cryoEM particles of KOGG, either supplemented with glycosyl-UDP and galactosyl-UDP, or in the apo state, image clusters displaying an elongated and continuous electron scattering signal shape were identified, suggesting the potential existence of KOGG in a higher oligomeric state. To rule out the possibility that these features were artifacts introduced during cryo-EM sample preparation, we performed a crosslinking assay with the purified LH3-ColGalT1 complex prior to cryo-EM preparation. The SDS-PAGEs showed that upon incubation with the crosslinking agent $BS(PEG)_5$, the migration rates of both LH3 and ColGalT1 shifted, with appearance of bands corresponding to tetrameric LH3-ColGalT1 complex (300-400 kD), as well as several higher molecular weight bands, exceeding 400 kD (the maximum size available for molecular weight standards) indicating the presence of even higher oligomeric states (Supplementary Fig. 11a). Additionally, native gel electrophoresis using cell lysates from untreated MC3T3-E1, SaoS-2, HEK293T, and Expi293F cell lines also suggested the existence of higher-order oligomers for the LH3-ColGalT1 complex (Supplementary Fig. 11a).

Upon data processing of the particles without the local refinement focusing on the core region in the highest resolution, a Coulomb potential map displaying a longer continuous signal than that after local refinement was generated (supplementary Figs. 2a, 3a, and 4a). Although the resolution in regions other than the core complex, i.e., protomers A, B, U, and V, is suboptimal in KOGG (supplementary Fig. 2d), the continuity allows for structural modeling of the backbones of one copy of the LH3 dimer (protomers C and D) and one copy of the ColGalT1 dimer (protomers W and X) on two sides of the KOGG quaternary complex, thereby generating KOGG octameric complexes composed of two LH3 dimers and two ColGalT1 dimers (Supplementary Fig. 11c and d). While only octameric complexes were modeled for KOGG apo and KOGG/UDP-Glc, we designated the octameric KOGG/UDP-Gal complex as the KOGG

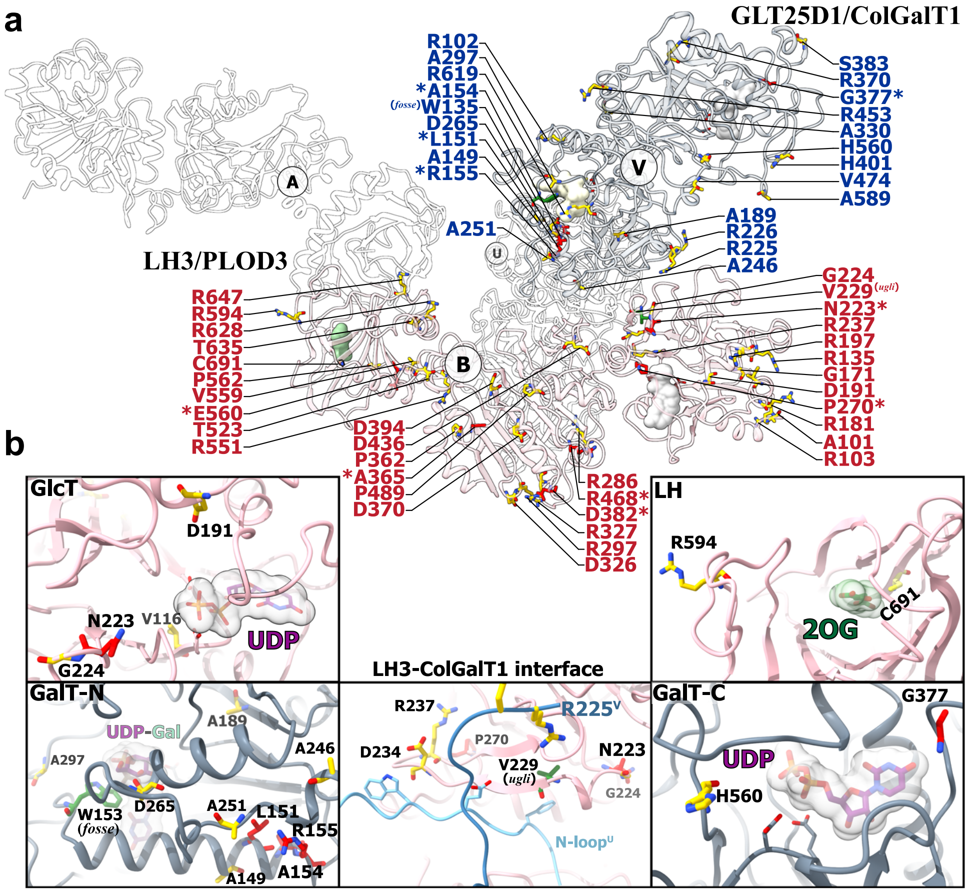

Fig. 5 | Structural pathological analysis of the KOGG complex. a The pathogenic mutations mapping on the KOGG complex. The KOGG/UDP-Gal was depicted as a cartoon model. On protomers LH3^B and ColGalT1^V, the disease-associated mutation sites were highlighted as stick models colored in red for those reported in human clinic cases, in yellow for mutation sites in the ClinVar database with potential pathogenicity, and in green for two mutations identified in mice (plod3^{ugli} and colgalt1^{fosse}). The substrates bound were shown as surface model. The protomers LH3^A and ColGalT1^U were set to transparent to facilitate observation. **b** The enlarged views on enzymatic sites and LH3-ColGalT1 interface of KOGG complex.

elongating complex to distinguish it from the tetrameric KOGG/UDP-Gal (Supplementary Table 1). Although the remaining parts of the map were left unmodeled in KOGG elongating, the overall shape and orientation suggest a linear polymerization for the KOGG complex with the 2 + 2 tetramer as a functional repeat unit (Supplementary Movie 1).

In the linear polymer of the KOGG complex, the 2 + 2 units were assembled with LH3-ColGalT1 interfaces. While a 2 + 2 unit is formed by the intra-unit interactions between protomers B-V, a new 2 + 2 unit would join by the inter-unit interactions between its protomer A and protomer U of the original 2 + 2 unit, facilitating the elongation of KOGG in a fiber-like manner (Supplementary Fig. 11e). In this fiber-like KOGG polymer, the length of each KOGG 2 + 2 unit is approximately 85 Å along the elongation axis, with approximately a 30° angle of torsion between neighboring units (Supplementary Fig. 11e).

The biological significance of KOGG's linear polymerization remains unclear, and further efforts to visualize this polymerization in a more native environment are needed to confirm its role as a natural state. We speculate that this long fiber-like arrangement might endow KOGG with the capacity to catalyze hydroxylation/glycosylation simultaneously on multiple lysyl sites on a single polypeptide chain, such as COL4A1[44]. Although further studies are warranted to elucidate the functional implications of KOGG polymerization, this high catalyzing efficacy in posttranslational modification could be important in the proper folding and assembly of procollagen chains, which appear to be unstable and prone to aggregation in the unmodified form[45,46].

Discussion

In this study, we elucidated the structure of the procollagen processing enzyme complex composed of LH3/PLOD3 and GLT25D1/ColGalT1, shedding light on diseases linked to collagen hydroxylation and glycosylation dysfunction, such as connective tissue disorders. Numerous mutations in *PLOD3* have been documented as pathogenic variants in clinical analyses or have potential pathogenicity, according to the ClinVar database [https://www.ncbi.nlm.nih.gov/clinvar]. These mutations are distributed across all three domains (Fig. 5a). For instance, the N223S mutation, associated with BCARD (bone fragility with contractures, arterial rupture, and deafness)[30], was situated in the catalytic site of GlcT, along with other potentially pathogenic mutations like V116M, D191N, and G224, indicating their potential impact on enzymatic activities (Fig. 5b). In addition to above mutations of consensus between the cryo-EM and crystal structures, a defect-related mutation site, V229, was identified at the LH3-ColGalT1 interface. This residue, corresponding to V232 in mouse Plod3, is situated in the hydrophobic groove between the GlcT and AC domains and mediates interactions with the N-loop of ColGalT1 (Fig. 2d and e). Substitution of Val with Glu results in a defective phenotype known as "*ugli*" in mice, which is characterized by craniofacial, ocular, and digital developmental defects[47].

On the other hand, a notable "hotspot" was identified in GalT-N, where several mutations were found (Fig. 5a), including L151R and A154P, linked to cerebral small vessel diseases[28]. Another disease-related mutation, G377R, was identified in the catalytic site of GalT-C, where it may disrupt the hydrogen bond between G377 and the carbonyl group on the uracil ring of the UDP molecule (Fig. 3c). These mutations in the GT catalytic sites could impair collagen modification by disrupting the galactosyltransferase activities of the KOGG complex. Similar to the ugli phenotype, another collagen glycosylation-related developmental defect in mice, known as the "fosse" phenotype, results from a mutation (W130R) in mouse LH3, corresponding to W135 in human LH3[48]. This mutation is located in the catalytic site of GalT-N, near the UDP-Gal molecule (Fig. 3b). As suggested by the biochemical assays conducted here and in previous studies, although GalT-N may not directly catalyze the reaction, its binding to UDP-sugar is essential for stabilizing ColGalT1. The W130R mutation likely disrupts its substrate binding, thereby impairing ColGalT1's enzymatic function.

Within the KOGG complex, the enzymatic functions of the LH, GlcT, and GalT-C domains are well-established, playing crucial roles in procollagen modification, including lysyl hydroxylation, galactosylation, and glucosylation. However, the biological functions of the AC and GalT-N domains remain unclear. Our structural analysis revealed a high structural similarity between the AC domain of LH3 and the GalT-N domain of ColGalT1 (overall RMSD ~ 3.958 Å) despite their low sequence identity (23.83%). While the inner cavity of the AC domain appeared void of distinguishable map density for small molecules (supplementary Fig. 7b), large, continuous map densities were identified with shape features corresponding to UDP-Gal or UDP-Glc (supplementary Fig. 9). Comparison between the AC and ColGalT1 domains highlighted similarities in beta-strands and most helices, including the DXD motif and glycosyl-contacting residues. However, structural differences, such as the longer helix 4 in ColGalT1 compared to its counterpart in LH3, helix 15, result in the critical residues (Y126, W135, and R139 of ColGalT1) in mediating the UDP moiety being absent in AC domain of LH3, suggesting potential differences in enzymatic activity between the two domains and the AC domain as a "silent" transferase with an incomplete glycosyl-UDP binding pocket.

Fibrillar collagens, like type I collagen, have fewer lysine modification sites compared to basement membrane collagens. LH1 hydroxylates lysines, and the LH3-ColGalT1 complex adds glycosylation, both essential for proper triple helix formation. Studies show *Plod3* mutant mice have disorganized collagen fibers[46], and GLT25D1/ColGalT1 disruption leads to type I collagen accumulation and growth arrest[21]. Basement membrane collagens, like type IV collagen, have many lysine modification sites. *Plod3* knockout mice show type IV collagen accumulation and fragmented basement membranes[46]. Mutations in *COLGALT1* impair enzymatic activity and cause cerebral small vessel disease[28]. This correlation between the pathological effects of LH3/PLOD3 and GLT25D1/ColGalT1, along with their spatial arrangement in the KOGG complex, suggests that the post-translational modifications carried out by the LH3-ColGalT1 complex are crucial for the proper localization and functional assembly of type IV collagen in the basement membrane.

Our cryo-EM structure of the KOGG quaternary complex and a potential polymerizing form provided insights into the organization of the LH3-ColGalT1 complex and the spatial arrangement of multi-enzymatic sites involved in the hydroxylation-galactosylation-glucosylation cascade of procollagens, as well as the details of substrate/product binding. However, several questions remain unanswered. One such question pertains to the recognition of substrate peptide chains and the forces driving them through the sequential reactions within the KOGG complex. Given that the three major catalytic sites are spaced by considerable distances, the complex must effectively capture target lysyl residues along with their flanking sequences in the nascent peptide and facilitate the long-distance transversing of the substrate region among the catalytic sites. To address this complex process, further structural studies involving large-scale complexes between the KOGG machinery and substrate peptides in different catalytic states are warranted to fully elucidate this reaction cascade.

Methods

Protein expression and purification

The cDNAs of human LH3/PLOD3(UniProt ID O60568) and GLT25D1/ColGalT1 (UniProt ID Q8NBJ5) were cloned from the CCSB-Broad Lentiviral Expression Library. PLOD3 was cloned into a modified pcDNA3.4 vector with a Tobacco Etch Virus (TEV) protease cleavage site, followed by a 3×Flag tag and an 8×histidine tag at the C-terminal of LH3/PLOD3. The cDNA of GLT25D1/ColGalT1 encoding the amino acid 30-622 was cloned into a modified pcDNA3.4 vector with an N-terminal haemagglutinin (HA) signaling peptide, followed by a twin strep tag and an HRV 3 C protease cleavage site at the N-terminal of ColGalT1. The QuikChange II Site-directed mutagenesis method was utilized to introduce point mutations in the constructs. All the DNA sequences of wild-type and mutants were confirmed by Sanger sequencing.

For the recombinant expression of LH3/PLOD3 and GLT25D1/ColGalT1, Expi293F cells (Thermo Fisher Scientific, A14527) were cultured in Union-293 medium (Union-Biotech, UP0050) at 37 °C with 5% CO_2 in an incubator shaker. When the cell density reached about 2×10^6 cells/mL with over 95% viability, cells were transiently co-transfected with plasmids of LH3/PLOD3 and GLT25D1/ColGalT1 (wild-type or the mutants) using polyethyleneimine (PEI, Polysciences, 24765-1). The transfected Expi293F cells were harvested after 72 h by centrifugation and resuspended in lysis buffer containing 150 mM NaCl, 20 mM HEPES pH7.4, and 10% glycerol, supplemented with 4 mM $MgCl_2$, 0.2 mg mL^{-1} Dnase I, and 1× protease inhibitor (MCE, HY-K0010).

The purification of the KOGG complex began with the sonication of the cell suspension on ice, followed by centrifugation at 46,000 g and 4 °C for 45 minutes. The supernatant obtained was then subjected to affinity chromatography using Strep-Tactin resin (IBA, 2-1201-025). The protein complex was eluted with lysis buffer supplemented with 50 mM D-biotin (Aladdin Biochemical, B105433). The eluate was concentrated and further purified by size-exclusion chromatography using a Superose 6 Increase 10/300 GL column (Cytiva, 29091596), pre-equilibrated with a buffer containing 150 mM NaCl and 20 mM HEPES at pH 7.4. Fractions corresponding to the peak with an elution volume of about 12.5 mL were pooled and concentrated to approximately 15 mg/mL, ready for cryo-EM sample preparation.

For cryo-EM sample preparation, the KOGG complex was supplemented with 0.05% (w/v) CHAPS (Anatrace, 75621-03-3) as a protective agent. For the preparation of KOGG complexes bound to UDP-galactose (KOGG/UDP-Gal) and UDP-glucose (KOGG/UDP-Glc), the respective ligands were added to the KOGG complex to a final concentration of 100 µM. The mixtures were incubated for 30 minutes, after which 0.05% (w/v) CHAPS was added.

Cryo-EM sample preparation and data collection

Cryo-EM sample preparation was carried out using a Vitrobot Mark IV (FEI) at 8 °C and 100% humidity. A 3 µL aliquot of KOGG samples was applied to glow-discharged holey carbon grids (Quantifoil R1.2/1.3 Au, 300 mesh), followed by a 20-second incubation. The grid was blotted with filter paper for 1 second, quickly immersed in liquid ethane, and cooled with liquid nitrogen. Images were captured using a 300 kV cryo-EM (FEI, Titan Krios) with the specimen maintained at liquid nitrogen temperature. High-resolution images were collected automatically with EPU software (FEI) on a K3 Summit direct electron detector (Gatan) in super-resolution counting mode. The GIF Quantum energy filter (Gatan) was used in zero-energy-loss mode with a 15 eV slit width. Data collection was typically carried out at a nominal magnification of 81,000 (corresponding to a physical pixel size of 1.1 Å), with a defocus range between −1.5 and −2.6 µm. A fluence rate of 15.7 electrons/$Å^2$/s was used, with a total exposure time of 3.85 seconds and a total fluence of 50 electrons/$Å^2$, fractionated into 32 frames.

Cryo-EM data processing

For the KOGG/UDP-Gal complex, a total of 3700 cryo-EM images were collected, and motion correction was performed on the dose-fractioned image stacks using MotionCor2 with dose-weighting[49,50]. The contrast-transfer function (CTF) parameters of each image were determined using Gctf[51], and automatic particle-picking was conducted with Gautomatch-v0.56 [https://www2.mrc-lmb.cam.ac.uk/download/gctf_gautomatch_cu10-1-tar-gz/]. Image processing was carried out using RELION-3.1[52] and cryoSPARC[36]. The data processing procedure is summarized in Supplementary Fig. 2a. Particles were initially extracted with 4× binning (4.4 Å/pixel), followed by two rounds of 2D classification to remove junk particles. The remaining particles were re-extracted without binning (1.1 Å/pixel) and subjected to heterogeneous refinement using four initial references generated by cryoSPARC. Particles corresponding to the best class were further processed with non-uniform refinement and 3D auto-refinement with C1 symmetry. Two rounds of 3D auto-refinement, CTF refinement, and Bayesian polishing were performed using RELION-3.1. A total of 427,054 good particles were then subjected to 3D auto-refinement, resulting in a high-resolution map with an overall resolution of 3.58 Å. For subsequent local refinement, cryoSPARC was used to generate a mask for the KOGG tetramer, yielding a final resolution of 3.40 Å after the local refinement.

For the KOGG/UDP-Glc dataset, a total of 3331 cryo-EM movie stacks were collected. Automatic particle-picking was performed with Gautomatch after motion correction and CTF estimation in RELION. The particles were initially extracted using 4× binning, followed by two rounds of 2D classifications to eliminate junk particles. The remaining particles were re-extracted without binning and subjected to heterogeneous refinement in cryoSPARC. Particles corresponding to the best class were further processed with several rounds of heterogeneous refinement and non-uniform refinement. After 3D auto-refinement, CTF refinement, and Bayesian polishing with RELION, the polished particles were subjected to cryoSPARC. Subsequently, non-uniform refinement with C1 symmetry was performed, resulting in a high-resolution map with an overall resolution of 3.75 Å (Supplementary Fig. 3a).

For the KOGG apo dataset, a total of 6351 cryo-EM movie stacks were collected. Automatic particle-picking was performed with Gautomatch after motion correction and CTF estimation in RELION. Similarly, particles were first extracted using 4× binning, followed by two rounds of 2D classifications to remove junk particles. The remaining particles were re-extracted without binning and subjected to heterogeneous refinement in cryoSPARC. Particles corresponding to the best class were further processed with several rounds of heterogeneous refinement and non-uniform refinement. After 3D auto-refinement, CTF refinement, and Bayesian polishing with RELION, the polished particles were subjected to cryoSPARC. This was followed by non-uniform refinement with C1 symmetry, resulting in a high-resolution map with an overall resolution of 3.93 Å (Supplementary Fig. 4a).

Model building and refinement

The AlphaFold2-predicted structural model of human ColGalT1 (Supplementary Fig. 13) [https://alphafold.ebi.ac.uk/entry/Q8NBJ5] and the crystal structure of human LH3 (PDB ID: 6FXK, [https://doi.org/10.2210/pdb6FXK/pdb]) were used as references for initial model building in Phenix[53]. These initial models were docked into the electron density map using Chimera[54], followed by iterative manual adjustments in COOT[55] and real space refinement using Phenix. During refinement, the UDP, UDP-galactose, UDP-glucose, Mn^{2+}, Fe^{2+}, and 2-oxoglutarate (2-OG) molecules were refined and fit into the cryo-EM density map using Real-space refinement and LigandFit programs within the Phenix software package, respectively.

Biochemical assessment of the hydroxylation activity of the KOGG complex

Reaction mixtures (5 µL total) containing 0.2 mg/mL wild-type or mutant KOGG samples were supplemented with 0.5 mM peptide substrate (LPGTAGLPGMKGHRGFSGLDG, synthesized by GL Biochem), 100 µM 2-OG, and 50 µM $FeCl_2$, as previously described protocol[56]. The mixture was incubated for 3 h at 37 °C. After incubation, the reactions were stopped by heating at 95 °C for 2 minutes and transferred to white 384-well plates (Corning, 3572). Subsequently, 5 µL of Succinate-Glo reagent I from the Succinate-GloTM JmjC Demethylase/Hydroxylase Assay kit (Promega, V7990) was added to each well and incubated at room temperature for 1 h. Afterward, 10 µL of Succinate-Glo reagent II was added, and the plates were incubated for 10 minutes. Luminescence detection was performed using a Nivo plate reader (PerkinElmer) according to the manufacturer's instructions. Data were analyzed and plotted using GraphPad Prism 9 software. All experiments were performed in triplicate, and control experiments were carried out by selectively omitting KOGG samples. For experiments involving mutant complexes, the luminescence units of the mutant complexes were normalized using the luminescence units of the wild-type complex. Data analysis and plotting were performed using GraphPad Prism 9 software.

Biochemical assessment of glycosylation activities

Reaction mixtures (total volume of 4.5 µL) containing 0.2 mg/mL wild-type or mutant KOGG samples were sequentially added with 0.5 mM peptide substrate, 100 µM 2-OG, and 50 µM $FeCl_2$, mixed thoroughly, and incubated for 30 min at 37 °C. To measure the galactosyltransferase activity, 50 µM $MnCl_2$ and 50 µM UDP-galactose were then added to the reaction mixture to the final volume of 5 µL and incubated at 37 °C for 2.5 h. For glucosyltransferase activity, 50 µM $MnCl_2$, 50 µM UDP-galactose, and 50 µM UDP-glucose were supplemented. Reactions were ended by heating at 95 °C for 2 min and transfer into white 384-well plates, then 5 µL of the UDP detection reagent from UDP-Glo™ Glycosyltransferase Assay kit (Promega, V6961) were added and let incubate at room temperature for 1 h. The plates were then transferred to a Nivo plate reader configured according to the manufacturer's instructions for luminescence detection. All experiments were repeated three times in parallel. Control experiments were performed by selectively removing KOGG samples under the same conditions. For detection of glucosyltransferase activity of mutant complex, the detected luminescence units subtracted the luminescence units analyzed for galactosyltransferase activity to give the final luminescence units. When detecting the enzyme activity of the mutant complex, the luminescence units of the mutant complex were normalized using the luminescence units of the wild-type complex. GraphPad Prism 9 software was used for data analysis and plotting.

Cell culture and immunoblotting

HEK293T (NCACC, SCSP-502), MC3T3-E1 (ATCC, CRL-2593), and SaoS-2 (FuHeng biology, FH0094) cells were cultured in Dulbecco's modified Eagle's medium (DMEM, HyClone, SH30243) with 5% fetal bovine serum (Wisent Corporation, 085-150) at 37 °C with 5% CO_2. For LH3 and ColGalT1 interaction studies, HEK293T cells were co-transfected with 5 µg of plasmids per 6-cm dish using PEI according to the manufacturer's instructions. The expression vectors for wild-type LH3/PLOD3 or mutants were co-transfected with wild-type GLT25D1/ColGalT1 or mutants at a 1:1 mass ratio. After 48 h of culturing, cells were harvested, washed with Dulbecco's phosphate-buffered saline (DPBS, Meilunbio, MA0010), and resuspended in lysis buffer with 1.5% (w/v) DDM, 4 mM $MgCl_2$, 0.2 mg/mL DNase I, and a 1× protease inhibitor cocktail. The lysates were incubated at 4 °C for 3 h and centrifuged at 10,000 g for 10 minutes at 4 °C. For immunoprecipitation assays, cleared supernatants were subjected to affinity chromatography using Strep-Tactin resin or Anti-DYKDDDDK (Flag) G1 Affinity Resin (GenScript, L00432-25). The beads were washed three times with lysis buffer and resuspended in a SDS-loading buffer of electrophoresis. The samples were then subjected to SDS-PAGE, transferred to PVDF membranes, and immunoblotted. Immunodetection was performed using the following primary antibodies: anti-DDDDK-Tag antibody (ABclonal, AE005, 1:5000) for LH3, anti-Strep II-Tag antibody (ABclonal, AE066, 1:5000) for ColGalT1, anti-GAPDH antibody (ABclonal, A19056, 1:10000), anti-COL1A1 antibody (ABclonal, A16891, 1:5000), and anti-ColGalT1 antibody (Proteintech, 16768-1-AP, 1:2000). The secondary antibodies used were anti-mouse IgG, HRP-linked antibody (Cell Signaling, 7076S, 1:10000) and anti-rabbit IgG, HRP-linked antibody (Transgen, HS101-01, 1:10000). All antibodies were diluted in 3% (w/v) BSA in TBST. Chemiluminescence imaging was performed using the GE Amersham Imager 600 RGB system.

Protein crosslinking assay

KOGG samples were diluted to 2 mg/mL, and 5 μL of the protein solution was mixed with BS(PEG)$_5$ to final concentrations of 0.32 mM and 0.64 mM, with a sample without crosslinker serving as the control. The crosslinking reaction was carried out on ice for 2 h, followed by the addition of Tris-HCl buffer (pH 7.5) to a final concentration of 50 mM to terminate the reaction. The samples were mixed with SDS-loading buffer, heated at 95 °C for 10 minutes, and analyzed by SDS-PAGE and immunoblotting.

Native-PAGE Analysis Coupled with Immunoblotting

Expi293F, HEK293T, MC3T3-E1, and SaoS-2 cells were harvested, and cell cultures were solubilized using DDM for 2 h at 4 °C. The lysates were centrifuged at 10,000 × g for 10 minutes at 4 °C, and the supernatants containing solubilized proteins were collected. The lysates were mixed with 2 × Native loading buffer and loaded onto a native gel along with a protein marker (Sangon Biotech, C560010). Electrophoresis was performed for 2 h at 120 V at 4 °C. After electrophoresis, proteins were transferred to a membrane, and immunoblotting was conducted as described.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The coordinates have been deposited in the PDB with the accession codes 8ZGH (KOGG apo complex, [)ZGE (KOGG/UDP-Gal complex, []. The cryo-EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) with accession codes EMDB-60079 (KOGG apo complex, [https://www.ebi.ac.uk/emdb/search/EMDB-60079]), EMDB-60078 ((KOGG/UDP-Glc complex, [https://www.ebi.ac.uk/emdb/search/EMDB-60078]), EMDB-60076 (KOGG/UDP-Gal complex, [https://www.ebi.ac.uk/emdb/search/EMDB-60076]), EMDB-60075 (KOGG elongated complex, [https://www.ebi.ac.uk/emdb/search/EMDB-60075]). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Unless otherwise stated, all data supporting the results of this study can be found in the article, supplementary, and source data files. Source data are provided with this paper.

References

1. Schjoldager, K. T., Narimatsu, Y., Joshi, H. J. & Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **21**, 729–749 (2020).
2. Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867 (2006).
3. Banushi, B. et al. Regulation of post-Golgi LH3 trafficking is essential for collagen homeostasis. *Nat. Commun.* **7**, 12111 (2016).
4. Liefhebber, J. M. P., Punt, S., Spaan, W. J. M. & van Leeuwen, H. C. The human collagen beta(1-O)galactosyltransferase, GLT25D1, is a soluble endoplasmic reticulum localized protein. *Bmc. Cell Biol.* **11**, 33 (2010).
5. Richards, A. A. et al. Adiponectin multimerization is dependent on conserved lysines in the collagenous domain: evidence for regulation of multimerization by alterations in posttranslational modifications. *Mol. Endocrinol.* **20**, 1673–1687 (2006).
6. Terajima, M. et al. Glycosylation and cross-linking in bone type I collagen. *J. Biol. Chem.* **289**, 22636–22647 (2014).
7. Takeyari, S. et al. 4-Phenylbutyric acid enhances the mineralization of osteogenesis imperfecta iPSC-derived osteoblasts. *J. Biol. Chem.* **296**, 100027 (2020).
8. Risteli, M. et al. Lysyl hydroxylase 3 modifies lysine residues to facilitate oligomerization of mannan-binding lectin. *Plos One* **9**, e113498 (2014).
9. Wang, Y., Xu, A. M., Knight, C., Xu, L. Y. & Cooper, G. J. S. Hydroxylation and glycosylation of the four conserved lysine residues in the collagenous domain of adiponectin - Potential role in the modulation of its insulin-sensitizing activity. *J. Biol. Chem.* **277**, 19521–19529 (2002).
10. Mori, K., Suzuki, T., Miura, K., Dohmae, N. & Simizu, S. Involvement of LH3 and GLT25D1 for glucosyl-galactosyl- hydroxylation on non-collagen-like domain of FGL1. *Biochemical Biophysical Res. Commun.* **560**, 93–98 (2021).
11. Ishizawa, Y. et al. Identification and characterization of collagen-like glycosylation and hydroxylation of CCN1. *Glycobiology* **29**, 696–704 (2019).
12. Takaluoma, K. et al. Tissue-specific changes in the hydroxylysine content and cross-links of collagens and alterations in fibril morphology in lysyl hydroxylase 1 knock-out mice. *J. Biol. Chem.* **282**, 6588–6596 (2007).
13. Saito, T. et al. Decrease of lysyl hydroxylase 2 activity causes abnormal collagen molecular phenotypes, defective mineralization and compromised mechanical properties of bone. *Bone* **154**, 116242 (2022).
14. Risteli, M., Niemitalo, O., Lankinen, H., Juffer, A. H. & Myllyla, R. Characterization of collagenous peptides bound to lysyl hydroxylase isoforms. *J. Biol. Chem.* **279**, 37535–37543 (2004).
15. Takaluoma, K., Lantto, J. & Myllyharju, J. Lysyl hydroxylase 2 is a specific telopeptide hydroxylase, while all three isoenzymes hydroxylate collagenous sequences. *Matrix Biol.* **26**, 396–403 (2007).
16. Herchenhan, A. et al. Lysyl oxidase activity is required for ordered collagen fibrillogenesis by tendon cells. *J. Biol. Chem.* **290**, 16440–16450 (2015).
17. Gjaltema, R. A., van der Stoel, M. M., Boersema, M. & Bank, R. A. Disentangling mechanisms involved in collagen pyridinoline cross-linking: the immunophilin FKBP65 is critical for dimerization of lysyl hydroxylase 2. *Proc. Natl Acad. Sci. USA* **113**, 7142–7147 (2016).
18. Eyre, D., Shao, P., Weis, M. A. & Steinmann, B. The kyphoscoliotic type of Ehlers-Danlos syndrome (type VI): differential effects on the hydroxylation of lysine in collagens I and II revealed by analysis of cross-linked telopeptides from urine. *Mol. Genet. Metab.* **76**, 211–216 (2002).
19. Ishikawa, Y. et al. Lysyl hydroxylase 3-mediated post-translational modifications are required for proper biosynthesis of collagen α1α1α2(IV). *J. Biol. Chem.* **298**, 102713 (2022).
20. Terajima, M. et al. Role of glycosyltransferase 25 domain 1 in type I collagen glycosylation and molecular phenotypes. *Biochemistry* **58**, 5040–5051 (2019).
21. Baumann, S. & Hennet, T. Collagen accumulation in osteosarcoma cells lacking GLT25D1 collagen galactosyltransferase. *J. Biol. Chem.* **291**, 18514–18524 (2016).

22. Heikkinen, J. et al. Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. *J. Biol. Chem.* **275**, 36158–36163 (2000).
23. Wang, C. G. et al. Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). *J. Biol. Chem.* **277**, 18568–18573 (2002).
24. Guo, H. F. et al. A collagen glucosyltransferase drives lung adenocarcinoma progression in mice. *Commun. Biol.* **4**, 482 (2021).
25. Koenig, S. N. et al. New mechanistic insights to PLOD1-mediated human vascular disease. *Transl. Res* **239**, 1–17 (2022).
26. Rautavuoma, K. et al. Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. *Proc. Natl Acad. Sci. USA* **101**, 14120–14125 (2004).
27. Teunissen, M. W. A. et al. Biallelic variants in the COLGALT1 gene causes severe congenital porencephaly a case report. *Neurol. Genet.* **7**, e564 (2021).
28. Miyatake, S. et al. Biallelic COLGALT1 variants are associated with cerebral small vessel disease. *Ann. Neurol.* **84**, 843–853 (2018).
29. Ewans, L. J. et al. Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. *J. Med. Genet.* **56**, 629–638 (2019).
30. Salo, A. M. et al. A connective tissue disorder caused by mutations of the Lysyl Hydroxylase 3 gene. *Am. J. Hum. Genet.* **83**, 495–503 (2008).
31. Scietti, L. et al. Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3 (vol 9, 3163, 2018). *Nat. Commun.* **9**, 3912 (2018).
32. Mattoteia, D. et al. Identification of regulatory molecular "Hot Spots" for LH/PLOD collagen glycosyltransferase activity. *Int. J. Mol. Sci.* **24**, 11213 (2023).
33. Perrin-Tricaud, C., Rutschmann, C. & Hennet, T. Identification of domains and amino acids essential to the collagen galactosyltransferase activity of GLT25D1. *PloS one* **6**, e29390 (2011).
34. Terajima, M. et al. Cyclophilin B control of lysine post-translational modifications of skin type I collagen. *PLoS Genet* **15**, e1008196 (2019).
35. Sricholpech, M. et al. Lysyl hydroxylase 3-mediated glucosylation in type I collagen: molecular loci and biological significance. *J. Biol. Chem.* **287**, 22998–23009 (2012).
36. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
37. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
38. Wiggins, C. A. & Munro, S. Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc. Natl Acad. Sci. USA* **95**, 7945–7950 (1998).
39. Luther, K. B., Schindelin, H. & Haltiwanger, R. S. Structural and mechanistic insights into lunatic fringe from a kinetic analysis of enzyme mutants. *J. Biol. Chem.* **284**, 3294–3305 (2009).
40. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev. Biochem* **77**, 521–555 (2008).
41. Ardevol, A., Iglesias-Fernandez, J., Rojas-Cervellera, V. & Rovira, C. The reaction mechanism of retaining glycosyltransferases. *Biochem Soc. Trans.* **44**, 51–60 (2016).
42. Markolovic, S. et al. Structure-function relationships of human JmjC oxygenases-demethylases versus hydroxylases. *Curr. Opin. Struct. Biol.* **41**, 62–72 (2016).
43. Copeland, R. A. et al. An iron(IV)-oxo intermediate initiating l-arginine oxidation but not ethylene production by the 2-oxoglutarate-dependent oxygenase, ethylene-forming enzyme. *J. Am. Chem. Soc.* **143**, 2293–2303 (2021).
44. Basak, T. et al. Comprehensive characterization of glycosylation and hydroxylation of basement membrane collagen IV by high-resolution mass spectrometry. *J. Proteome Res* **15**, 245–258 (2016).
45. Sipila, L. et al. Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. *J. Biol. Chem.* **282**, 33381–33388 (2007).
46. Ruotsalainen, H. et al. Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. *J. Cell Sci.* **119**, 625–635 (2006).
47. Geister, K. A., Timms, A. E. & Beier, D. R. Optimizing genomic methods for mapping and identification of candidate variants in ENU mutagenesis screens using inbred mice. *G3 Bethesda* **8**, 401–409 (2018).
48. Geister, K. A. et al. Loss of function of Colgant1 disrupts collagen post-translationa modification and causes musculoskeletal defects. *Disease Models & Mechanisms* **12**, dmm037176 (2019).
49. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6. *Elife* **4**, e06980 (2015).
50. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
51. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
52. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, https://doi.org/10.7554/eLife.42166 (2018).
53. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D. Struct. Biol.* **75**, 861–877 (2019).
54. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput Chem.* **25**, 1605–1612 (2004).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D. Biol. Crystallogr* **66**, 486–501 (2010).
56. Li, W. et al. The structural basis for the collagen processing by human P3H1/CRTAP/PPIB ternary complex. *Nat. Commun.* **15**, 7844 (2024).

Acknowledgements

The authors thank Drs. Ming Lei and Lijun Wang for scientific discussion. This work was supported by the National Natural Science Foundation of China (82272519 and 82072468, Y.C.; 82372430, A.Q.; 82130073, J.Z.; 22074092, R.L.), Shanghai Municipal Committee of Science and Technology (20S11902000, Y.C.), Shanghai Key Laboratory of Orthopedic Implants (KFKT202207, A.Q.). This work was also supported by the Shanghai Frontiers Science Center of Degeneration and Regeneration in Skeletal System (J.Z. and A.Q.) and the Innovative Research Team of High-level Local Universities (SHSMU-ZLCX20211700, Y.C.) from the Shanghai Municipal Education Commission. W.L. receives the post-doctoral fund from the Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine. We thank the staff members of the Electron Microimaging Center, Bioimaging Facility, and Proteomics Platform at Shanghai Institute of Precision Medicine for providing technical support and assistance in data collection.

Author contributions

Y.C. and A.Q. conceived the study. P.M., Y.C., J.P., W.L., A.Q., J.Z. and R.L. designed the experiments. J.P., W.L., Y.Cai. and Y.X. performed the functional assay. J.P., W.L., M.C., Q.W., Y.S., S.L. and Y.C. performed structural biology experiments. Y.C. and D.Y. built and refined structural models. Y.C., J.P., W.L. and D.Y., wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-57768-9.

Correspondence and requests for materials should be addressed to An Qin or Yu Cao.

Peer review information *Nature Communications* thanks James Blaza and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025