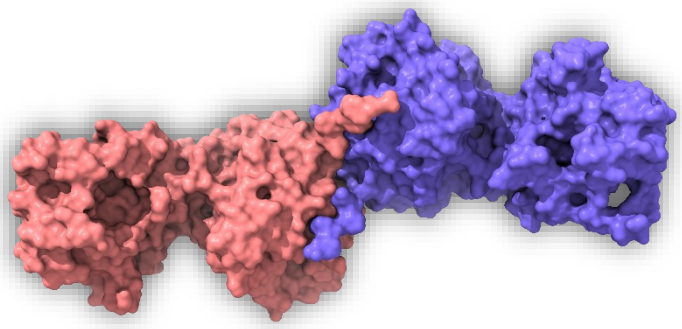
UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie 'Lazzaro Spallanzani'

Collagen Lysine Modifying Enzymes: An Integrative Structural Biology Approach



Matteo De Marco

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
Ciclo XXXVIII – A.A. 2022-2025.

Dedicato a mia madre Antonina Prometti

Summary

Chapter 1 AIM OF THE WORK

In the following chapter are discussed the structural and biochemical features of the main enzymes involved in collagen lysine post-translational modification (Fig.1). A detailed investigation related to the lysine glycosylation is presented in chapter 3 and 5, the works are comparative structural and functional analysis of the glycosyltransferase domains of both PLODS as well as GLT25D1 enzymes. Integrative approaches are used to provide precise mechanistic details of the enzymatic reactions underneath lysine collagen glycosylation. Chapter 3 describes the work which have been done to find promising strategies to make druggable the catalytic centre of the LH domain of PLOD enzymes. Metal ion chelators as 2,2'-bypyridine (BPY) can perturb the arrangement of a C-terminal segment that acts as a cap for the lysine hydroxylase activity. Finally, recent advancement in LH1 protein structure determination is introduced in chapter 6, through SPA cryo-EM.

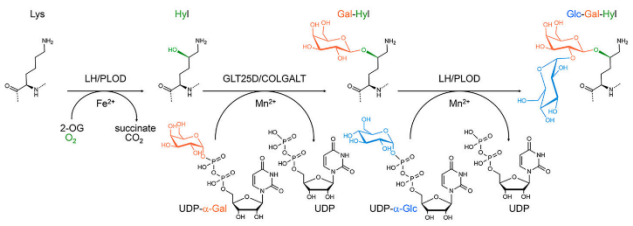


Figure 1. Schematic representation of collagen lysine residues modifications. Lysine residues are firstly hydroxylated by PLOD enzymes at their ε carbon generating the 5-hydroxylisine. The 5-hydroxylisine is then galactosilated by the procollagen galactosyltransferase GLT25D1 leading to the product Gal-(β1, O)-5-hydroxylisine. Finally, PLOD3 transfers a glucose moiety on the Gal-(β1, O)-5-hydroxylisine generating the final product Glc-(α1, 2)-Gal-(β1, O)-5-hydroxylysine.

Chapter 2 INTRODUCTION

THE EXTRACELLULAR MATRIX

The extracellular matrix is a heterogeneous milieu of proteins, glycoproteins and growth factors, which shape the tissues of multicellular organisms, giving them different biochemical properties [1–3]. The mechanical properties of each tissue are provided by a unique combination and interplay of specific proteins, which expression is programmed and regulated by external physical and chemical stimuli as well by the internal ones generated by the organisms themselves [4, 5]. Different proteins are found within the extracellular matrix, between the most common and ubiquitarians there are collagens, laminins, elastin, fibronectin, tenascins, MMP and glycoproteins such as decorin, perlecan and agrin [6]. Together the huge proteinaceous content, within the extracellular matrix are found reservoirs of growth factors; example of these signaling molecules is the vascular endothelial growth factor (VEGF), the fibroblast growth factor (FGF) and the transforming growth factor (TGF-β) [7–17]. Hence, the ECM is a biomaterial which has intrinsic structural roles, to support and compartmentalised tissue and organs, as well as signaling functions.

Collagen is the major constituent of the extracellular matrix, with 29 known representative family. All the collagens can be divided into two macro groups: the fibrillar collagens which assemble into the extracellular space as bundles of fibers (e.g. collagen type I, II, III) [18] and non-fibrillar collagens which form networks in the basement membranes (e.g. collagen type IV, VI) [19–21], fibril associated collagens with interrupted triple helices (FACIT, e.g. collagen type IX,XII) [22–30], transmembrane collagens (e.g. collagen type XIII and XXIII) [31–33] and multiplexin (e.g. collagen type XV) [34, 35]. Collagen molecules are right-handed triple helix formed by the intertwining of three precursor α-chains; each collagen type can have unique α chains as well different isoform of them, thus there are homotrimeric collagens as well heterotrimeric collagens. The nomenclature of collagens wants the collagens' family enumerated by romans number, followed by the specific chain composition of each collagen polypeptide: e.g. collagen type I $[\alpha1]_2[\alpha2]$.[36, 37]

Collagen biology is a huge biological topic which still requires extensive studies to decipher all the complexity underneath it. Different disciplines have adopted different approaches to elucidate collagen biosynthesis, modifications as well as biomechanical properties, which makes collagen research a very active research field. This has an impact not only in the context of basic academia research, but it is also of relevance for the biomedical field as well as for the nutraceutical and material industry. Collagen can be used for the most variegated usage; 1) it can be used to generate engineered tissues for regenerative medicine purposes [38–43] , 2) it is a sustainable material which can be used as biomaterial for packaging, in substitution to plastic [44], 3) or it can also be used to make food supplements due to the health benefits of collagenic peptides or cosmetics [45–51]. However, big challenges have still to be overcame to improve the sustainability and to reduce the carbon fingerprint of the

collagen's extraction methodology. Recent developments in collagen extraction were able to improve procedures to extract collagens from waste material such as fish scale, but further optimizations are still required to make these processes mor efficient.

The aim of this work is to describe fine details and mechanism of collagen modifications which it undergoes through its biosynthesis. In chapter 4 and 5 will be presented the multifunctional enzyme lysine hydroxylase LH3 (PLOD3), in chapter 5 will be discussed the recently characterized procollagen galactosyltransferase GLT25D1, while the chapter 6 will intrude the recent unpublished data related to the structure of the lysine hydroxylase isoform 1, LH1, obtained by single particle electron microscopy (SPA cryo-EM). The work will be introduced by a brief description about collagen synthesis, with particular attention to its secretion and the role of lysine and proline hydroxylation.

THROUGH THE COLLAGEN'S WALK OF LIFE

Collagen biosynthesis is a fascinating phenomenon which still has not be fully understood in his whole complexity. All the events leading to the final collagen fibrils are fine regulated by the cell biochemistry, by cell-cell interaction and by the abiotic mechanical stresses. Collagen, as fibrillar or non-fibrillar, can be considered as a biopolymer which requires initial basic molecular units for its final assembly[36, 52]. These initial units for collagen are tropocollagen molecules, firstly described by the pioneering work of G.C. Ramachandran in 1988 [53]. Tropocollagen is a right-handed triple helix, which is the result of three polyproline II left-handed helix coiling (also known as α-chains). Each collagen type owns its own specific set of precursors α-chains that can associate as homo- or hetero-trimer, following a complex synthesis path. Notwithstanding decades of studies on collagen biosynthesis and architecture, the fine temporal and spatial resolution of the series of events which occur within the cell, providing the biopolymer's precursors, and the subsequent assembly of collagen fibrils or network in the extracellular space are still to be fully clarified. Is to be clarified that most of the processes known about collagen biosynthesis are the results of studies focussing on fibrillar collagen, as type I, II and III the major constituents of connective tissues and blood vessels as well arteries [18, 54, 55]. For what concern non-fibrillar collagen the representative type which has been extensively studied is network forming type IV collagen, main collagen in basement membranes [56–60]. Although the two major class of collagens share common biosynthetic paths different characteristics distinguish them to each other, reflecting different developmental, organism specific and tissue specific requirements. Decades of studies on fibrillar collagen have been essential to lay the theoretical foundations for collagen biosynthesis models, that even if not highly explicative and precise for each collagen type, are of critical importance to understand 1) the intracellular synthesis and modification of tropocollagen molecules 2) the subsequent secretion of the product in the extracellular space 3) and the fibrillogenesis or the network nucleation events in the extracellular space.

TROPOCOLLAGEN SYNTHESIS

The basic units of collagens are polyproline II helix (PPII) which are commonly named α-chains [61]. For each collagen type there exist specific genes encoding different or single α-chains. Seven types of fibrillar collagens have been identified so far, each of which has specific genes encoding for different α chains (Table.1). PPII helices are known to be left-handed helix with 3 residues/turn with 3.1 Å rise/residue and dihedrals angles φ=-75° and ψ=145°. However, collagen's PPIIs show slightly different parameters, with 3.33 residues/turn, 2.9 Å rise/residues and similar values for dihedrals angles[62–72]. Different types of polyproline helices exist (PPI, PPII, PPIII, PPIV) [73], but to generate a mature and stable collagen triple helix is essential that the α-chains keep the PPII conformation. Different elements contribute to PPII formation and stability, of critical relevance are 1) the stereochemical characteristics of proline [62, 70], 2) the nature of amino acid flanking proline [66] and 3) the entropic contribution of the different amino acid's side chains to the overall helix structure [63, 65, 71, 74–76]. In 2009, the experimental work of Chiang and collaborators [77] revealed that the electron withdrawing groups at the 4R of proline increase the transition state barrier between PPII→PPI conformations, so stabilizing the PPII helix. The opposite phenomena, the decrease of the transition state barrier between PPII→PPI, stabilize the PPI conformation. This has been demonstrated by substituting, in collagen-like peptides, electron withdrawing groups at the 4R and 4S of proline (Hydroxyl-, fluoro-, methoxy-groups) and exploring the stereolectronic effects on transition barrier by time-dependent circular dichroism. Electron withdrawing groups on C_{γ} are not the only element that favour specific polyproline helix. In 1997 De Tar and Luthra defined two fundamental conformational states of pyrrolidine ring: the up-puckered configuration and the down-puckered configuration [78]. The proline configuration found in PPII helices is the down-puckered, in other words this means that the C_{γ} and the carbonyl group (C=O) lay on the same plane of the atoms C_{α}, N and C_{δ} of the prolyl ring. In the up-puckered configuration instead the C_{γ} and the carbonyl moiety lay on different planes. Thus, while PPI helices have a mixed puckering, proline residues in PPII helix are all in the down-puckered configuration. Proline residues are essential for PPII helix stability as well for the collagen triple helix proper folding. What is defined as tropocollagen is the result of the intertwining of three α-chains along the main axis with a right-handed twist, which generates a major helix pitch of 108°. The peculiarity of tropocollagen (in most collagen's families, and in all fibrillar collagens) is the repetition of the tripeptide unit Gly-X-Y, where X and Y can be any amino acid even if there is a frequency of the 20% for imino acids [64, 79]. The three chain are supercoiled together mainly due to the contribution of NH··CO hydrogen bonds. In this superstructure glycine residues point toward the internal part of the helix minimizing the interaction with the surrounding solvent molecules, while X and Y residues are arranged such that side chains are exposed to solvent [76] (fig x). The collagen's triple helix geometry unable interaction between three X (or three Y) residues, because of the large angular separation that is 103°. Interactions between X residues and the neighbouring chain's Y residues are instead favoured, such interaction can be direct through the free C=O moieties or mediated by water molecules [64, 79]. However, as mentioned beforehand prolines are driving forces in tropocollagen stability and assembly as well as they are essential for PPII stability. They can be found both in the X and Y position, determining thus different kind of modifications. The tripeptide repetition Gly-Pro-Hyp is one of the most important found in

Type	Genes	Expression
I	COL1A1, COL1A2	Bone, skin, tendon
II	COL2A1	Cartilage
III	COL3A1	Skin, vessels
IV	COL4A1–COL4A6	Basement membranes
V	COL5A1–COL5A3	Skin, bone, cornea
VI	COL6A1–COL6A6	Muscle ECM
VII	COL7A1	Dermal-epidermal junction
VIII	COL8A1, COL8A2	Cornea, vasculature
IX	COL9A1–COL9A3	Cartilage
X	COL10A1	Hypertrophic cartilage
XI	COL11A1–COL11A3	Cartilage
XII	COL12A1	Tendon, skin
XIII	COL13A1	Epithelia
XIV	COL14A1	Tendon, skin
XV	COL15A1	Muscle, microvessels
XVI	COL16A1	Smooth muscle, skin
XVII	COL17A1	Epidermis hemidesmosomes
XVIII	COL18A1	Endothelium
XIX	COL19A1	Brain, muscle
XX	COL20A1	Embryonic tissues
XXI	COL21A1	Heart, muscle
XXII	COL22A1	Myotendinous junction
XXIII	COL23A1	Skin, lung, kidney
XXIV	COL24A1	Bone, eye
XXV	COL25A1	Brain, neurons
XXVI	COL26A1	Testis
XXVII	COL27A1	Cartilage, bone
XXVIII	COL28A1	PNS, aorta

Table 1: The 28 collagen families and their tissue expression.

collagen, thanks to the hydroxyproline residues which play a key structural and functional role. In tropocollagen stability. The (2S,4R)-4-hydroxyproline (Hyp) is the post translation modification with the higher frequency in human, discovered in gelatin by Emil Fischer [80]. Its abundance is estimated to be near the 4% amongst animal protein and 38 % within collagen [81]. Hydroxylation is a modification that can deeply alter the nature of modified molecule, due to the highly electronegatively nature of oxygen [82]. Thus, through-bond inductive effect, the pK_a value of nitrogen is lowered from 10.8, in Pro, to 9.68, in Hyp. This results in a minor amidic resonance within the prolyl peptide bond, pushing it toward a more pyramidal configuration and increasing the rate of *cis-trans* bond isomerization [83, 84]. Thus, in other words, proline hydroxylation enables the *gauche effect* which is the tendency of a molecule to adopt the conformation that has the maximum number of adjacent polar bonds with a *gauche* (that is ±60°) dihedral angle. The *gauche* effect endows Hyp to adopt a C^{γ}-*exo* conformation, whereas Pro has a slightly preference for the C^{γ}-*endo* pucker [85, 86]. Proline hydroxylation has a direct impact also in the *trans:cis* ratio of the prolyl peptide bonds. Usually, folded proteins have both proline in *trans* and *cis* configuration, instead in collagen the *trans:cis* ratio is high with all the proline in the *trans* configuration. This high content of proline in trans configuration is achieved with 1) the activity of *trans-cis* prolyl isomerase, 2) electronegative substituent in the 4R position, 3) the C^{γ}-*exo* ring pucker, enforced by the gauche effect, which enables a strong $n \rightarrow \pi^*$ interaction between the oxygen of the prolyl peptide bond (O_{i-1}) [85, 86]. The nature of the proline ring puckering determines changes in the main-chain torsion angles φ and ψ, leading to the proper folding of the right-handed collagen triple-helix. Tropocollagen molecules are at the N- and C-terminal have

non-collagenous domains, which in fibrillar collagens are processed to enhance fibrillogenesis, while in type IV collagen the dimerization of the NC1 domains (C-terminal) helps the tail-to-tail association through the tetramerization of the 7S domain at the N-terminal thus generating networks [56]. In conclusion, tropocollagen is the basic unit of higher order collagen's structures, generated by the intertwining of three PPII helix into a supercoil right-handed triple helix. In these molecules the tripeptide repetition Gly-Pro-Hyp is the most representative 'collagen motif' which provide rigidity and stability to the overall helix, especially thanks to the nature of proline and its PTMs (Fig.2). Several studies have focused their attention on the contribution of charged residues in X and Y position, lateral and axial salt bridges can be generated between charged amino acids like Lys-Asp, Lys-Glu, Arg-Asp and Arg-Glu [87–92]. However, they don not seems to have a strong impact on triple helix stability, where instead the main driving force are glycine and proline which endeavour the intertwining of the PPII helix in the final triple helix. Thus, must of the solvent exposed side chains have a role in collagen's cross-linking, in surface-surface interaction between laterally staggered tropocollagen molecules or crucial protein-protein interactions. The last hypothesis can also explain the observed events of micro-unfolding, and the experimental fluctuation of the collagen's triple helix symmetry parameters. Usually, collagen triple helix symmetry can be described as a 10/3 or a 7/2 symmetry, however intermediated values between these two extremes have been described.

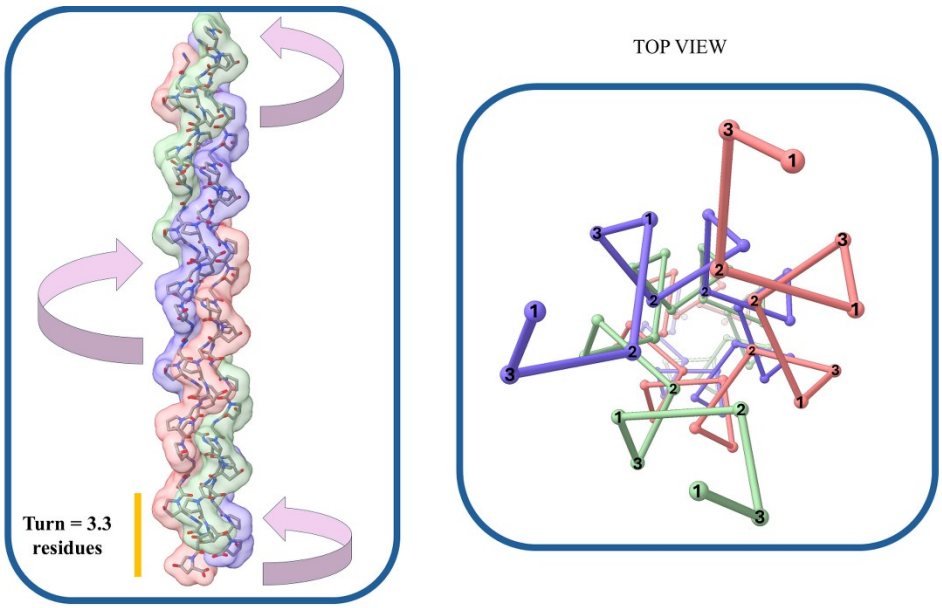


Figure 2. Crystal structure of the collagen-like peptide $(Gly\text{-}Pro\text{-}Hyp)_9$ [93] PDB code 3B0S. Collagen triple helix is characterised by a right handedness with 3.3 residues per turn. The unique collagenic right-handed triple helix is possible thanks to the high abundance of proline in collagen's sequences. From the top view is possible to understand that the glycine residues of the collagen residues are always buried within the collagen triple helix (2), while hydroxyproline (1) and proline (3) are arranged in a way that unable the interaction between three X (or three Y) residues (Gly-X-Y repetition). This is because the large angular separation (108°) around the triple helix. The allowed interactions are the one involving an X residue and a Y residue of neighbouring chains, or whit the available backbone C=O groups, directly or mediated by water molecules.

The mature tropocollagen molecules is approximately a 300 kDa, 300 nm in length and 1.5 nm in diameter macromolecules. This big moiety is then secreted outside the cell, and an obvious question arises: how is tropocollagen delivered into the extracellular compartment when known secretory vesicles are usually smaller?

TROPOCOLLAGEN SECRETION

As any other newly synthesised protein, tropocollagen molecules must be delivered to the proper functional compartment. Proteins are produced within the ER, transferred in the Golgi where they undergo post-translational modifications and then they are sorted within intracellular compartment or in the extracellular matrix [94]. These exchange of material between different compartments are much more complex respect what seems in a first look and require the coordination of different biophysical and biochemical events, such as membrane reorganization and the fine tuning of macromolecular complex's activity. In eukaryotes organisms, newly translated proteins accumulate in a specific compartment of the ER which is called ER exit site (ERES) which is linked to a neighbouring complex of membranes that is the ER-Golgi intermediate compartment (ERGIC). Based on current knowledge, the canonical secretory system which enable the exchange of protein between the ER and the Golgi is mediated by coat protein complex II and I (COPII-I), where the former one is responsible for the anterograde transport of proteins toward the ERGIC compartment and the last one is responsible for the retrograde transport of proteins back to the ER [95]. These complexes are not just clamps which helps the budding of ERES membrane, they are instead modular element with both structural role in vesicles formation and integrity, as well docking element for cargo uptake [96]. Secretion from the ERES starts with the assembly of COPII, the first step of this long series of events is the nucleotide exchange on the small GTPase secretion-associated RAS-related protein 1 (Sar1) [97], which reaction is catalysed by the guanine nucleotide-exchange factor Sec12. Sar-GTP is so partially inserted in the outer side of the ER membrane, thus recruiting Sec23-Sec24 complex and assembling the inner layer of the vesicle's coat. The fully mature coat is assembled by the ternary Sar1-Sec23-Sec24 complexes, which recruit the Sec13-Sec31 heterotetramers which are the components of the outer layer [94]. Coated vesicles are dynamic elements that undergo changes in the average size and diameter, leading to vesicles with an average size of 60-100 nm. Sar1 is the key element in assembly and disassembly of the coating, thus Sec23 can act as a GTPase-activating protein (GAP), process accelerated

by Sec31 binding, leading to Sar1's GTP hydrolysis and thus its detachment from the membrane and subsequent uncoating. Ras1 is a peculiar G protein of the Ras superfamily because it requires the presence of membrane to promote the GTP binding. When the nucleotide binds to Sar1 different conformational changes occur leading to important rearrangement. Thus, a hydrophobic pocket of Sar1, usually occupied by the N-terminal amphipathic α-helix, is disrupted and this allow the insertion of the N-terminal helix into the membrane. This features of Sar1 make its activity sufficient to induce the tubulation of membranes, also when other proteins involved in COPII formation are absent [98].

Tropocollagen molecules have a considerable size which exceed the average diameter of canonical COPII vesicles. The size mismatch between tropocollagen molecules and cargo vesicles lead to the speculation of different secretions model, until a genetic screen in *Drosophila* S2 cells lead to the discovery of Transmembrane Protein Transport and Golgi Organization 1 (TANGO1), encoded by the gene mia3 [99]. Thus, different research activity started to repot the involvement of TANGO1 into big macromolecules secretion as for tropocollagen or apolipoproteins [97, 100–102]. Have been identified different homologs of TANGO1 named as TANGO like protein or TALI. The most important TALI member is cTAGE5 that act in synergy with TANGO1 and further has a ubiquitous expression [103]. However, while TANGO1 is present in invertebrate as well in vertebrate, cTAGE5 is present only in vertebrate [104, 105]. TANGO1 is an ER resident protein localized at the ERES sites. It is a transmembrane protein with distinct domain in the ER lumen region and pin the cytoplasmatic region. The ER luminal part is a coiled-coil structures in which the N-terminal is an SH3-like domain, responsible of the binding with the protein Hsp47 [, 107]. The cytoplasmatic part is composed of two coiled-coil structures (CC1 and CC2) with a C-terminal proline-rich domain (PRD). The CC1 can recruit ERGIC-53-containing membranes through the TEER domain (Tether for ERGIC and the ER), CC2 interacts with cTAGE5, and PDR interacts with the COPII proteins Sec23 and Sec 16. Furthermore, has been described a shorter version of TANGO1 which is composed of 785 amino acids instead of the 1907 present in TANGO1 [108]. This short version lacks all the amino acid of the ER luminal part; thus, it is unable to bind cargos while all the cytoplasmatic interactions are preserved. The protein cTAGE5 share most of the feature of TANGO1, it is an 807 amino acids long protein with the same cytoplasmic domain as TANGO1 and a shorter ER luminal region; the CC1 domain interacts with Sec12, the CC2 domain with Sec22 and TANGO1 while the PDR domain interacts, like TANGO1, with SEC23. The interaction with the components of COPII vesicular system led to an interesting model that could explain the secretion of large macromolecules. In this model TANGO1 limits the formation of the COPII vesicles and thus allowing the formation of tubular structures [109]. This can be explained by the interaction of TANGO1 and cTAGE5 with Sec23 which compete with Sec23-Sec31 interaction. Sec23 without the interaction with Sec31 can't activate the GTP hydrolysis activity of Sar1, thus limiting the recruitment of outer layer of the COPII coat [110]. A recent model conceptualises collagen export from the ER as a mechanochemical ratchet, generated by the entropic contraction of the intrinsically disordered luminal region of TANGO1 which surrounds the ERES [111]. In more details, the partially extension of the disordered region of TANGO1 increases the probability of the binding with the trimeric tropocollagen molecules which are far from the ERES. TANGO1 is a transmembrane protein, its attachment to the membrane reduces the conformational entropy of the ensemble thus generating an entropic force that pulls the SH3 domain toward the ER membrane, lastly

pushing tropocollagen from the ER to the ERGIC. The interaction between tropocollagen and TANGO1 is not direct, HSP47 [112] is the docker element that enable the binding between the two molecules within the ER's lumen. Has been postulates that molecular gradients regulate the polarization of these series of events, in particularly pH and HSP47 concentration. Is important to highlight that the affinity of HSP47 for the SH3 domain of TANGO1 is higher than the one toward tropocollagen (260 nM vs. 2.25 uM), this means that when HSP47 is released in the ERGIC compartment it quickly rebinds to TANGO1 in the ER. This model proposing a TANGO1-induced forces for the initial formation of ER/ERGIC tunnel is very promising, but it still requires extensive experimental validation [111].

A lot of questions remain to be addressed; more experimental data are required to understand the biochemistry and the biophysical properties of cargo systems for proteins secretion. The whole system of tropocollagen secretion is even more complex if are considered all the post-translation modifications which modifies or decorate it. Processing of tropocollagen molecules involved different steps: 1) The cleavage of the signal peptide at the N-terminal within the ER, which allow the entrance of tropocollagen into the secretory pathway [113]. 2) The formation of disulfide bonds by protein disulfide isomerase (PDI) family, which is necessary to the formation and rearrangement of disulfide bonds ant the N- and C-terminal telopeptides. This ensure the trimerization of three α-chains into the collagenous triple helix whit the correct register [114]. 3) The hydroxylation of proline residues by the enzymes procollagen-proline,2-oxoglutarate-4-dioxygenases (P4H) and by the prolyl 3-hydroxylase (P3H). As previously described prolines modifications are essential for the stability of PPII helix and intertwining of tropocollagen helix [115, 116]. 4) The hydroxylation of lysine residues, which is a peculiar PTM's found in collagen and collagen-like proteins. This modification is carried out by the class of enzymes known as procollagen-lysine ,2-oxoglutarate 5-dioxygenases (LH), also named PLODs as the gene which encode them. Hydroxylation of lysine is crucial for the formation of cross-links, which enable the lateral association of tropocollagen molecules into the extracellular space leading to the mature collagen molecules. [115, 117, 118] 5) The glycosylation of lysine residues by the procollagen galactosyltransferase 1 (GLT25D1) and by LH3. The former enzyme has a galactosyltransferase activity on 5-hydroxylisines, whereas LH3 transfers a glucose moiety on the β-(1, O)-galactosyl-5-hydroxylisine thus generating the final product α-(1,2)-glucosyl-β-(1, O)-galactosyl-5-hydroxylisine [119, 120]. 6) The oxidative deamination of lysine and hydroxylysine to generate cross-links by the lysyl oxidase (LOX) enzymes [121–123]. 7) Minor modifications such as phosphorylation and sulfidation, which function is still under investigation [124–127].

In the last years, several studies have been made on proline and lysine modifications as well as on PDI's function and telopeptide's processing, which details are going to be presented in the next chapters.

PROLINE AND LYSINE MODIFICATION

In 1967 the research activity of Kivirikko and Prockop revealed that collagen's lysine and proline residues were enzymatically hydroxylated throughout the biosynthesis path [115].

The enzymes which perform these modifications are the procollagen lysine hydroxylase 1-3 (LH1-3) and the proline 4-hydroxylase and proline 3-hydroxylase (P4H and P3H). They are member of a large family of non-heme containing Fe^{2+} and 2-oxoglutarate dependent dioxygenases (2-OGDO enzymes) which is a family distributed through the three of life, and different evolutionary conserved subfamily can be found in prokaryotes, plants as well as metazoan. Regulation of O_2 and Fe^{2+} concentrations by the activity of 2-OGDO enzymes could have been an important sensor mechanism developed by multicellular organisms of the Cambrian period, when O_2 level started to increase in the atmosphere [128]. Thus, 2-OGDO enzymes are cellular sensor 1) for the energy metabolism, sensing the availability of 2-OG, 2) for the oxygen level, promoting responses for hypoxia and 3) for the metabolism and redox properties of Fe^{2+}, hence sensing oxidative stresses. Thus, 2-OGDO can be found in different biological context as epigenetic and non-epigenetic regulators. The Ten-Eleven Translocations (TET1-3) for example is a crucial enzyme which controls the DNA demethylation through hydroxylation of 5-methylcytosine [129]; several lysine demethylases (KDM2-7) are 2-OGDO enzymes with Jumonji C domain [130]; prolyl hydroxylase domain-containing proteins 1-3 (PHD1-3), which regulate hypoxic response through the hydroxylation of two proline residues in HIF-1α, belong to the super family of 2-OGDO[131].

The common features of this wide family of enzymes are in the mechanism of the catalytic reaction. The structural signature of 2-OGDO enzymes is a double-stranded β-helix structure (DSBH) [132], also known as jelly-roll fold, which comprises eighth antiparallel β-strands at the core surrounded by helix or flexible loops. The catalytic reaction of 2-OGDO enzymes starts with the binding of Fe^{2+} and 2-oxoglutarate into highly evolutionary conserved binding site of jelly-roll fold. Thus, O_2 makes a complex with Fe^{2+} which enhance the oxidative decarboxylation of 2-OG to succinate and CO_2, so oxidizing Fe^{2+} to the ferryl intermediate ($Fe^{3+/4+}$). The generated high-valent iron oxidant can thus hydroxylate the substrate molecules, generating the final reaction product. Before a new cycle can start, the ferryl iron must be reduced through ascorbate and/or glutathione. The O_2 consumed has a dual role in the catalytic reaction, one oxygen atom is employed in the oxidative decarboxylation of 2-OG where the other oxygen takes part in the hydroxylation of the substrate. The decarboxylation of 2-OG leads to the formation of succinate which can compete for the binding in the active site with 2-OG, and like fumarate can inhibit the activity of these enzymes *in vitro* and *in vivo* [128].

PROCOLLAGEN PROLINE HYDROXYLASE

The procollagen-proline, 2-oxoglutarate-4-dioxygenase (P4H) has been identified in different organisms spanning from algae to metazoan [133–136]. In all the taxa in which is found it performs the same enzymatic activity, namely the hydroxylation of proline residues[137]. However, different taxa have different P4H with different structural and biochemical properties. Tetramers of P4H are found only in vertebrate, whereas in invertebrate like *C.elegans* only P4H αβ dimers are physiological and in plants P4H is found as monomer of α subunits [135].

Vertebrates have three P4H isoforms, where P4H type I Is the most abundant in most of the tissues. All the isoforms are heterotetramers of α and β subunits, type I P4H is a $[\alpha(I)_2]\ \beta_2$ while type II P4H is a $[\alpha(II)_2]\ \beta_2$. The α(I) and α(II) subunits are 517 and 514 amino acids long proteins, with signal peptide of 17 or 21 residues. The two isoforms share a sequence identity of 67%, with the highest identity at the C-terminal region which is involved in substrate recognition. The α subunits consists of four domains: the N-terminal domain, which engages in dimerization, the peptide-substrate-binding (PSB) domain, a linker region and the catalytic domain at the C-terminal (CAT). The catalytic sites are quite similar with similar Km values for Fe^{2+} and 2-OG, while differences have been described for Km value of different peptide substrate and for Ki value for the competitive inhibitor poly (L-proline) and the metal ions Zn^{2+}, Co^{2+}, Cd^{2+} and Ni^{2+} [138, 139]. The two α subunits differ in the cysteine content, the α(I) subunit has 5 cysteine residues which make interchain disulfide bonds (the second with the third and the fourth with the fifth residue). Instead, the α(II) subunit contains an additional cysteine residue which is located between the fourth and the fifth residue. The interchain disulfide bonds are essential to ensure stability to the native state of the α subunits, however they do not have a direct role in tetramer assembly and stability [140, 141]. The β subunit of P4H is identical to the protein disulfide isomerase (PDI), it consists of four domain structure known as a, b, b' and a' sharing the thioredoxin fold. As forehead mentioned, PDI subunit is not present in all P4H. Its role is not still completely understood, it is essential to provide solubility to the α subunits and to drive them to the proper catalytically competent conformation [142, 143]. However, are not known the mechanistic details of how PDI is able to assemble with the α subunits and its involvement in interchain disulfide bridges. PDI subunits are synthesized in large excess in respect of the α subunits, but they are not the essential elements which drive the tetramer assembly; however, they have the KDEL endoplasmic reticulum retention signal at the C-terminal, hence ensuring the proper subcellular compartmentalization of P4H. As others 2-OGDO enzymes, P4H activity can be described as a two-step reaction dependent by the presence of Fe^{2+}, 2-OG and ascorbate: in a first instance there is the generation of the hydroxylating species, which are then used for 4-hydroxyproline formation. The evolutionary conserved -His-X-Asp-…-His- motif is the essential binding site for the Fe^{2+} metal ion. The 2-OG cofactor binds in between the major and minor β-sheets of the main jelly-roll domain, the coordination involved both positively charged amino acid around the binding pocket as well by the enzyme-bound Fe^{2+} which interacts with the C1-C2 moiety. While the 2-OG is consumed stoichiometrically, the consumption and use of ascorbate is not clear. What is clear about P4H's activity is that it cannot perform its enzymatic reaction on free proline, the minimal unit required for proper enzymatic activity is the tripeptide unit -X-Pro-Gly [144]. The Km values of P4H for different peptide substrates is inversely proportional to their length, it has a higher affinity for longer peptides reflecting its incapacity to process isolated proline. The mechanism of P4H activity is a series of time events which involved the rearrangement of the substrate as well the enzyme-substrate conformations [145]. The X-Pro segment of the peptide substrate may have to assume a poly (L-proline) II conformation to bind to the PSB domain, while the Pro-Gly segment may have to adopt a β-turn for the proline residues to enable the proper enzymatic conversion into the catalytic site. The hydroxylation of collagen's proline is an event which occur before the triple helix has been folded, different studies have shown the inability of P4H to get access to the buried proline residues in the mature collagen's triple helix [146]. However, collagen is not the only protein whit -X-Y-Gly- motifs and there is evidence of non-collagenous protein with a triple helical

collagen-like domain. Examples of such proteins are the subcomponent C1q of complement, a C1q-like factor, several humoral lectins, adiponectin, the tail structure of acetylcholinesterase, three macrophage receptors, elastin, ectodysplasin, gliomedin, elastic fiber-associated glycoproteins and a src-homologous-and-collagen protein [147, 148]. Hence, the presence of different P4H isoforms with different substrate specificity can be related to their subcellular/extracellular localization to regulate different biological responses; further studies are still needed to elucidate the whole scenario.

The 4R,2S-hydroxyproline is not the only modification that can be found on collagen's proline residues. The less frequent modification of proline into 3S,2S-hydroxyproline (3Hyp) is present in fibrillar as well in non-fibrillar collagen [149]. The enzyme able to perform this enzymatic reaction is the prolyl 3-hydroxylase, in complex with the cartilage-associated protein (CRTAP) and Cyclophilin B (PPIB) [150] in a 1:1:1 ratio [151]. P3H and CTRAP belong to a unique family of gene products which includes P3H family, with the isozymes P3H1 (LEPRE1), P3H2 (LEPRE2), P3H3 (LEPRE3), as well as two potential coenzymes CRTAP (LEPREL3) and Sc65 (LEPREL4). The role of P3H complex is not clear yet, but there are assumptions about its role as macromolecular chaperon for collagen molecular assembly. The role of the hydroxylase activity of P3H toward collagen residues is still debated, 3-Hyp provides a slight increase in the thermal stability of triple helix compared to the one provided by 4-Hyp. However, there is a high concentration of 3-Hyp at the N-terminal region of type IV collagen, also known as 7S domain[152]. This domain is made up by four antiparallel triple helices that are tethered together by hydrophobic interactions, ionic interactions and disulfide bonds. Is not clear why this domain is highly populated in 3-Hyp residues but is a putative starting point to understand the role of this uncommon modification.

PROCOLLAGEN LYSINE HYDROXYLASE

Collagen's cross-linking is a fundamental enzymatic event which enable the proper maturation of collagen fibrils in the extracellular space, once are secreted into the extracellular space. The enzymes responsible for lysine hydroxylation are member of the macrofamily of 2-oxoglutarate dependent dioxygenases; in vertebrate are known three isoforms, LH1, LH2 and LH3 encoded respectively by the gene PLOD1, PLOD2 and PLOD3 [153]. The function of each isoforms has a specificity toward different consensus motif on the collagen's collagenic region or in its telopeptide region. The common consensus region is the tripeptide repetition of -X-Lys-Gly-, while in the telopeptide region the consensus motif is the -X-Lys-Ala- repetition. Hydroxylation is the main enzymatic function of LH enzymes, although slightly differences are present between the different isoforms. The enzymes LH1 and LH3 have mainly a hydroxylase activity toward the collagenic core of tropocollagen molecules [120, 154], while the isoform LH2 (which has two splicing variants; $LH2_a$ and $LH2_b$ with an extra 63bp exon 13A) has a specificity toward the collagen's telopeptide region [155, 156]. Several research studies demonstrated that LH enzymes are multifunctional enzyme, meaning that together the lysine hydroxylase activity they have also a glucosyltransferase activity. The only isoform of which is present a full structural and biochemical characterization is the enzyme LH3, which can perform a retaining glucosyltransferase activity toward the Gal-(β1, O)-5-hydroxylisine leading to the final product Glc-(α1,2)-Gal-(β1, O)-5-hydroxylisine [120, 157]. The galactosylation of the 5-

hydroxylation residue is performed by the procollagen galactosyltransferase GLT25D1 [119, 158], which will be described in more detail in the next chapter. Glycosyltransferase activity has been detected also in the other LH enzymes; however, structural and biochemical date are still lacunose to understand their efficiency in respect to the one of LH3. The degree of collagen lysine hydroxylation is different for each collagen type and for each tissue; the range is between 5-70 Hyl in a 1000 amino acids long molecule. This modification is crucial to generate cross-links between the collagen molecules in the extracellular space, thus leading to the final mature collagen fibrils. In fibrillar collagens the telopeptide domains are deaminated by lysine oxidases (LOX) [159, 160], generating the reactive aldehydes allysine or hydroxyallysine[161]. The reactive species at the telopeptide regions can than interacts with lysine or hydroxylysine at the collagenic domain of the collagen molecules. When two telopeptidyl hydroxyallysine interact with an hydroxylysine residue within the triple helix the hydroxylysylpyridinoline (HP) cross-link is formed. Instead, the lysylpyridinoline (LP) cross-link is the interaction between two telopeptidyl hydroxyallysine and an hydroxylysine residues within the collagen triple helix. The most abundant cross-link in the connective tissues such as bone, cartilage and skin is the HP cross-link, while it is completely absent in skin and cornea [162, 163]. The role of collagen hydroxylysine in collagen cross-links generation has been extensively studied [117, 163–169] ; however, the role of glycosylation of collagen hydroxylysine remains yet to be completely understood. In a sequential series of events, galactose is linked to the 5-hydroxylisine residue through a β-glycosidic bond while glucose is enzymatically linked to the galactose moieties via a retaining α-glycosidic bond. For several years LH3 has been considered the enzyme responsible for both the galactosyl- and glucosyltransferases activity; however, in 2018 the high resolution LH3 crystal structure has been obtained making clearer the entire lysine modification pathway [120].

LH3 is a dimer, each monomer has three distinct domains: the N-terminal domain is enzymatically active in glycosyltransferase activity, it is an inverting glycosyltransferase characterized by a Rossmann-fold with an uncommon and peculiar D-x-x-D motif which coordinate the UDP-Glucose donor substrate and the catalytic Mn^{2+} metal ion. Between the N-terminal GGT domain and the C-terminal lysine hydroxylase domain (LH) there is an accessory domain (AC) which has a structural organization like canonical inverting glycosyltransferase; however, it lacks key motif essential to coordinate the substrate and the cofactors for the enzymatic activity. The C-terminal domain shares a DSBH fold, as other members of the 2-OGDO family, which host the catalytic Fe^{2+} and the 2-OG; however, an additional iron has been identified in the proximity of the gating loop at the catalytic pocket [170]. The functional role of this additional metal ion will be better described in chapter 4. LH enzymes have been identified as ER resident protein, except for LH3 which has been identified also into the extracellular space [157]. Enzymatic assay on LH3 established clearly that the only suitable acceptor substrate is the Gal-(β1, O)-5-hydroxylisine, *in-vitro* LH3 cannot transfer the galactose moieties on the 5-hydroxylysine. This agrees with the discovery, by the group of Hennet and colleagues [171], of an inverting glycosyltransferase (GLT25D1) which transfer via an inverting β-glycosidic bond a galactose moiety on the 5-hydroxylisine. Thus, lysine glycosylation requires the coordinated activity of two distinct procollagen glycosyltransferases [172], and what is of particular interest is the high similarity in the quaternary structure of these enzymes. However, unpublished data (that will be described in chapter 7), revealed that LH1 has a different quaternary structure compared to the others LH isoforms, it is a pseudo-symmetric dimer of dimers, with the interface of

oligomerization between the different LH domains. The different superstructure of these enzymes is probably related to the specificity toward different substrates and/or for the recognition and interaction with other proteins partners; however, the scenario in not clear yet and deeper investigation are required.

PROCOLLAGEN GALACTOSYLATION

As mentioned in the previous chapter, through the lysine modification pathway glycosylation are carried out by distinct glycosyltransferase; distinct enzymatic mechanisms require distinct structural folds which for glycosyltransferase have been systematically annotated into the CAZY database (http:/www.cazy.org). Glycosyltransferase can be classified into two main superfamily: the GT-A glycosyltransferases and the GT-B glycosyltransferase [173]. Each family has than an internal sub-systematic division into glycosyltransferase and non-glycosyltransferase, and further into inverting and retaining glycosyltransferase. Thus, belonging to one of the two superfamily is not a sufficient element to discriminate between inverting and retaining glycosyltransferase. Structurally, enzymes of the GT-A superfamily have a fold which is a reminiscence of two abutting Rossmann-fold, which is common in nucleotide-binding proteins. This fold can be described as the proximity of two β/α/β domains which lead to the formation of an internal extended β-sheet. Variation of the core structure can be found between GT-A glycosyltransferases, such as the presence of additional α-helices or β-sheet which have mostly a relevant role in multimerization [174]. A common motif found in GT-A glycosyltransferases is the tripeptide unit D-x-D, which coordinate the enzymatic bivalent metal ion, usually Mn^{2+} or Mg^{2+}, and the ribose of the UDP-sugar. This peculiar signature is found in the β-hairpin or in the small β-sheet which is near the residues involved in the catalysis. One of the strands that characterized this structure is always located after the fourth strand of the Rossmann core (fig), while the second one is at the C-terminal of the strand 7. Even if very common in GT-A glycosyltransferases, the D-x-D motif is found also in protein that are not inverting glycosyltransferases; hence, as for the overall structural organization this motif is not a sufficient criterion to precisely identify GT-A glycosyltransferases [175–177]. The GT-B glycosyltransferases, as the GT-A glycosyltransferase, have the main structural core which is a Rossmann-like fold. However, the difference is in the proximity of the two β/α/β domain which are less tight in GT-B glycosyltransferase. The two domains are facing each other defining a central cleft in which usually is located the substrate binding site. In GT-B glycosyltransferases the departure of the leaving group is mediated in a metal ion independent fashion, thus the D-x-D motif is not a common motif of GT-B glycosyltransferases [173, 178]. Recently, 8 family of glycosyltransferase have been grouped into the superfamily of the GT-C glycosyltransferases [179]. These enzymes are large hydrophobic protein which are usually localized into the ER or in association with the plasma membrane. No experimental structural data are available yet; however, some predictions have been made based on sequence analysis. GT-C glycosyltransferases have multiple predicted transmembrane domains; interestingly, the N-terminal extra cytoplasmic loop is the most conserved element between the protein which belong to this superfamily. Notably, this loop contained different motif, like the ExD, DxE, DDx or DEx, that are

variation of the canonical GT-A D-x-D signature. Notwithstanding the close similarity between these motifs, up to now there are no evidence of common evolutionary origin.

The procollagen glycosyltransferase found by the Hennet's group has been predicted, based on sequence analysis and functional characterization, to be an inverting glycosyltransferase belonging to the GT-A family. Furthermore, the recent high resolution crystal structure of the human full-length GLT25D1 validate the initial guest and adds additional, remarkable, observations about its structural and functional peculiarities [119, 171]. Three isoforms have been identified for the human procollagen galactosyl transferase encoded by the *colgalt* genes: the GLT25D1 which is expressed in almost all tissue, the isoform GLT25D2 [171, 180] which has been identified mainly in the nervous system, finally the GLT25D3, also known with the name CERCAM [171], which is expressed in secretory tissues (e.g salivary glands, placenta and pancreas) as well in the nervous system. However, while GLT25D1 and GLT25D2 are enzymatically active, the isoform CERCAM is inactive. The procollagen galactosyltransferases are ER resident protein, all the three isoforms have a C-terminal ER-retention motif RDEL, and there are no data suggesting extracellular localization as in the case of PLOD3. As mentioned before, the only isoform fully characterized structurally and biochemically is GLT25D1. It is a glycosyltransferase belonging to the GT-A superfamily, and to the inverting-clan I family. GLT25D1 is an elongated head-to-head homodimer; each monomer is composed of two globular domains connected by a well rigid loop of 21 amino acids. The N-terminal domain (residues 51-320), defined as GT1 domain, is directly involved in the formation of the dimerization interface; it has the canonical topology of GT-A glycosyltransferase and in the proximity of the catalytic site is present the D-x-D signature. The GT1 domain has a very high structural similarity with the accessory domain (AC) of PLOD3 (R.M.S.D $\approx$ 1 Å) while it has a poor similarity with the LH3 GGT domain (R.M.S.D $\approx$ 3 Å) [119, 120]. The C-terminal domain, the GT2 domain, has some peculiar features which make it vastly different from the GT1 domain, as well from the AC and the GGT domains of PLOD3. It is remarkable the presence of an uncommon EDD motif which coordinate the bivalent metal ion Mn^{2+}, and an overall less evolutionary conserved amino acids composition at the core Rossman-like domain. Both the domains can bind the substrate and the cofactors for the enzymatic reaction; however, only the GT2 domain has been demonstrated to be solely responsible for the enzymatic activity of GLT25D1, while the GT1 domain is the interface of dimerization and helps to coordinate the protein-protein interaction with PLOD3 [118]. The reaction mechanism of GLT25D1, like other glycosyltransferases is a direct displacement S_N2-like reaction. The side chain of a residues in proximity of the catalytic sites acts as a base catalyst which deprotonate the nucleophile of the acceptor substrate, thus helping the displacement of the activated phosphate leaving group. The detailed description of GLT25D1 will be described in chapter 6.

Notwithstanding the recent achievement in procollagen galactosyltransferase understanding, the precise role of lysine glycosylation is still unclear. Different collagen types have a different degree of glycosylation; this is remarkably evident when non-fibrillar type IV collagen and fibrillar collagen are compared [56]. In fibrillar collagen usually the axial distribution of glycosylated lysine is regular, following the D-period of collagen fibrils. In collagen type I, the (α1) chain has a total of 38 hydroxylysine of which only 18 are than glycosylated. The frequency of glycosylation is even lower in the collagen type I (α2) chain, where on 30 hydroxylysine only 7 are glycosylated. The frequency of glycosylation is highly

organism and tissue dependent, however the major glycosylation load on collagen type IV is systematically seen. Indeed, the collagen type IV has 21 conserved sites which undergo lysine glycosylation and on the (α1) chain 35 hydroxylysine on 39 are further glycosylated. Functionally, in fibrillar collagen lysine glycosylation regulate and lead the fibrils assembly, the cross-links formation, the fibrils diameter and the mineralization process. However, in type IV collagen the role of glycosylation is not so clear, the main hypothesis is that they play a crucial role into collagen secretion [181], assembly and formation of the basement membrane [182–185].

*I*NTO THE EXTRACELLULAR SPACE: COLLAGEN FIBRILLOGENESIS

Once the tropocollagen molecules have been assembled, they serve as building blocks for the formation of extracellular ultrastructure as collagen fibrils [186] or intricated collagen networks [56]. The most known collagen ultrastructure are the fibrils building the connective tissues. In these tissues, the macromolecules which decorate the extracellular space must hold to extensive mechanical stresses generated by the external environment as well the forces generated by the organism itself (e.g. cell migration, blood vessel shear, interstitial fluid pressure). The high variability of physical forces that an organism must manage is mirrored in the presence of extracellular matrices with heterogeneous mechanical features. Hard tissues, for example, can reach a stiffness in the order of the GPa, while soft tissues have a stiffness ranging from 0.1 kPa to 100 kPa [187, 188]. Between the different molecules which determine the mechano-chemical features of a tissue, collages play a vital role. In loading-bearing tissues the fibrillar collagen is the main constituent that guarantees mechanical integrity. The ultrastructure generated by fibrillar collagens must be seen as heterotypic polymers, made of three collagen types: 1) a ‘major’ fibrillar component, as collagen type I or collagen type II, 2) a ‘minor’ fibrillar collagen, as collagen type V in cornea or collagen type XI in cartilage, 3) bridging collagens as FACIT or type IX, XII, XIV collagens [189]. Collagen type I and II are the essential component to start collagen fibrillogenesis, where collagen type V and XI have a nucleating role for the fibril’s formation, and further they might be the collagen which links the ensemble to the cell surface through collagen-binding integrins [190–192]. The tertiary collagenic components help the lateral association between different collagen fibrils and other component of the ECM such as proteoglycans, which with their abundance can enable the sliding between different collagen fibres in respect to each other [193, 194]. Connective tissues have a mechanotransducive role, collagen fibers have been described as viscoelastic polymers with a viscous module (time-dependent) and an elastic module (time independent) which can carry out this task. It has been proposed that the elastic behaviour of collagen molecules is mainly determined by the stretching of cross-linked collagen molecules, while the viscous behaviour is involved in the sliding of different fibrils in the deformation of tissues [188, 195, 196]. However, has been seen that the viscoelastic behaviour of collagen is non-linear at large strain. After tissue unloading residual strain has been observed to remain applied on collagen molecules, thus suggesting a plastic behaviour of collagen molecules at large forces [187, 188, 197]. However, there are few experimental data available to describe the plastic behaviour of collagen, what has been speculated is a plastic answer in relation to collagen concentration and plasticity as a stretch-dependent formation of weak cross-links.

One of the most important obstacles which makes complicated the understanding of collagen mechanotransduction is the lack of precise models which can provide precise spatial and temporal information describing the formation of collagen higher order structures. There are different models that try to model fibrillogenesis, which can be mainly divided into cell-dependent models and cell-independent models (collagen self-assembly models). In 1996 K.Kadler and colleagues proposed a self-assembly model to describe collagen fibers synthesis, where early unipolar and bipolar fibrils can self-assemble thanks to intrinsic chemical properties [186]. Briefly, the procollagen molecules are 300 nm long molecules flanked by a trimeric globular C-propeptide domain and a trimeric N-propeptide domain (. Once tropocollagen molecules are secreted into the extracellular space, collagen N-proteinase (PNP) and C-proteinase (PCP, also known as Bone Morphogenetic Proteins, BMP) remove the propeptide domains; thus, leading to the head-to-head, or head-to-tail, fusion of tropocollagen molecules. The removal of the C-terminal propetide is necessary to start the collagen self-assembly, while the N-terminal propetide could be more like a regulatory element to modulate the shape of fibrils. The last step of fibrils stabilization is achieved by the later association that can arise trough lysine or hydroxylysine cross-links. Altogether the collagen self-assembly theory, different research groups started to gather data about a prominent role of the cell into collagen fibrils formation [198]. In this model the cells are not only the hub responsible for the synthesis of tropocollagen molecules, but they can also actively take part in the fibrils' assembly of tropocollagen molecules in the extracellular space. Interaction between the cell and the collagen molecules in the extracellular space are mediated by important proteins. 1) The small leucin rich proteoglycans (SLRP) as decorin, fibromodulin and lumican, can prevent the fibril-fibril surface interaction; thus, regulating collagen fibrils branching [194]. 2) Integrins are between the main mediators for cell mechanotransduction, they are transmembrane heterodimers made by the association of an α and β subunit. In mammals have been identified 18 α subunits and 8 β subunits, which can originate 24 different types of heterodimers. The most important integrins in collagen extracellular reorganization are the $\alpha_5\beta_1$ and the $\alpha_2\beta_1$ dimers. While the $\alpha_5\beta_1$ binds fibronectin and so indirectly bridge collagen to the cell through fibronectin-collagen interaction, $\alpha_2\beta_1$ binds directly collagen and plays a direct role in collagen fibril assembly [199–201]. 3) The minors type V and type XI collagens are the elements which bridge the interaction between the major collagens and integrins, thus they start the collagen nucleation at the cell-surface. Importantly, these collagens retain the N-propeptide which sterically regulate the lateral growth of collagen fibrils as well proteoglycans interactions [190–192]. This model has been implemented considering the contribution of cells movement in collagen fibrillogenesis and remodelling.

Collagen gels for the Anisotropic Biphasic Theory (ATB) can be seen as continuous biphasic materials, behaving like Maxwell-like fluid which compaction and mechanical properties are determined by fibroblasts' movement [202–204]. The factors which drive collagen compaction are the stress generated by the cells' movement, the cells orientation and concentration within different tissues. At the beginning of the 2000s, have been conceptualized models which were able to integrate cell-dependent and cell-independent way for collagen fibers formation and organization. There are different key concepts to fully understand the current models which describe collagen fibrillogenesis, between them are worthy of mention the fibripositors, the diffusion-limited aggregation (DLA), the flow induced crystallization/flow induced aggregation (FIC/FIA) and the fine-tuned collagen

PTMs [205–209]. The fibripositors are actin-driven membranes invagination, which have been described in the early 2000s. These invaginations are much way bigger in respect to the canonical secretory vesicles, here tropocollagen molecules are accumulated [208]. In this compartment collagen molecules, due to their rigid, mesogenic triple-helix structure, have a lyotropic liquid crystal behaviour; when dissolved in specific solvent (like water or acetic acid) at specific concentrations (for collagen lower than 30 mg/ml) tropocollagen molecules can exhibit liquid crystalline state, as he cholesteric (chiral nematic) phase. Hence, tropocollagens are concentrated at the 'necking' region of fibripositors, and when extensional flow is applied, the collagen isotropic solution can shift toward more ordered structures that are aligned along the flow direction. Hence, when the cholesteric phase is reached collagen molecules are aligned in the same direction, thus enhancing the mechanical and optical properties of collagen [210–212]. Is to highlight that pH, electrostatic interaction and the ionic strength of the medium, are additional elements which influence collagen fibrillogenesis. The secretion of tropocollagen molecule into the extracellular space seems to be coupled by a re-uptake of the excess of non-utilized monomers by fibroblasts [213–216]. So collagen fibrillogenesis is more likely to be a complex phenomenon regulated on different levels: 1) at the microscale by the intrinsic chemical properties of tropocollagen molecules, with its diversified spectrum of PTMs which can undergoes [168, 217, 218], 2) at the mesoscale by the physical forces generated by cells migrations as well the one generated by the interplay of cells and the ECM's molecules [199, 200], 3) at the macroscale by the environmental forces acting on an organism, which lastly lead to the huge diversification of connective tissues in different taxa [193, 219–222].

Different models have been conceptualized to described collagen fibrillogenesis and its role in ECM remodelling. However, stays a huge gap of knowledge about the non-fibrillar molecules which shape basement membrane.

BASEMENT MEMBRANE AND NON-FIBRILLAR COLLAGEN

Basement membranes are sheet-like structure which are found beneath all the epithelial cell layers, they also wrap around epithelial and endothelial tubes. The main function of basement membranes is to keep compartmentalized the different tissues that are found in multicellular organism [60, 223, 224]. Thus, is not surprising that basement membranes are found in different organs and tissues. Between the most known basement membrane present in humans there are 1) the glomerular basement membrane (GBM) [225], which is a component in the kidney filtration barrier, 2) the lens basement membrane (LBM) which is located around the eye lens supporting the lens epithelium, 3) the Reichert's basement membrane which surround the embryo through its development [226, 227]. Basement membranes, in addition to their role in tissues compartmentalization, have a fundamental regulatory role in cell polarity and migration and it is a hub for different signalling stimuli [228]. Basement membranes, differently to other connective tissues shows peculiar and unique biomechanical features. The main constituents of basement membranes are collagen type IV, laminin, nidogens, heparan sulphate proteoglycans (HSPGs) as perlecan and agrin [229–231]. The fundamental elements to start the formation of basement membranes sheets are laminins and collagen type IV [56]. When investigated through *in-vitro* experimental setup, laminins and collagen type IV are able, independently, to self-assemble into sheetlike structures. However, have been then proved that *in-vivo* the primary deposition of laminins

is essential for the following polymerization of the collagen type IV sheets. The thickness of basement membranes varies in respect to the tissue and in relation to the age of the organism, for example the basement membrane of the capillary is in the range of 100-200 nm while the one found in eye (cornea and lens) can reach the diameter of 10-20 μm [232, 233].

Collagen type IV is the collagenic constituent of basement membranes, and differently to fibrillar collagens, it has unique features which prevent its assembly into fibrillar superstructure while favouring its network-like superstructural organization [234]. As for fibrillar collagens, the tropocollagen molecules of type IV collagen are triple helix whit the characteristic collagen motif Gly-X-Y. However, within the collagenic core the tripeptide repetitions of type IV collagens are more often interrupted, which confers a much higher degree of flexibility to these molecules. Further, the N-terminal and C-terminal of non-fibrillar collagen are different in respect of the one found in the fibrillar collagens. The N-terminal domain is known as 7S domain, while the C-terminal domain is defined as the non-collagenous domain (NC1). The collagen type IV network is assembled in the extracellular space through the first non-covalent interaction between 4 7S domains and 2 NC1 domains. Once the initial network has been assembled in the extracellular space it further undergoes other 7S-7S and NC1-NC1 cross-links, which will determine its biomechanical properties [235]. The cross-links of collagen type IV are mediated by the lysine oxidase isoform 2 (LOX2), and the reaction mechanism is the same of the one for fibrillar collagens [236]. The presence of different collagen type IV α-chains, in addition to the cross-links, makes this type of collagen very heterogeneous in its mechanical properties. In vertebrates have been identified three different protomers for the type IV collagen: the $[\alpha1]_2[\alpha2]$, the [α3] [α4] [α5] and the $[\alpha5]_2[\alpha6]$ heterotrimer. The three protomers have a differential tissue expression, the $[\alpha1]_2[\alpha2]$ is the most ubiquitous while the [α3] [α4] [α5] shows a higher compactness, due to a higher number of interchain disulfide bonds, and is found in specific vertebrates' tissues such as the GBM. As aforementioned, the assembly of collagen type IV involves the tetramerization of the 7S domain, which are cross-linked at the cysteine residues by interchain disulfide bonds and at the lysine residues (mediated by LOX2) [237], and the tetramerization of the globular NC1 domains [238]. A very specific cross-link which is highly conserved in animal collagen IV is the sulfonation between an exposed lysine reside in the NC1 domain and a methionine (catalysed in the extracellular space by the enzyme peroxidasin) [239]. Compared to the fibrillar collagens, type IV collagen is much softer, thus it is suited to resist to extensional forces instead of the compressional and tensional ruling tissues such as tendons. Hence, the elastic Young's module of non-fibrillar collagen is much lower respect the one of tendons.

The full mechanical properties of basement membranes are not merely determined by the abundance and quality of type IV collagen [240]. The hydration state of the basement membranes is tightly regulated by the HSPG glycoproteins Perlecan and Agrin [241, 242]. The negative charge of these molecules is the key determinant which makes them like hydrogels; materials which can be highly hydrated, thus increasing the resistance of the basement membranes to the compressional forces [243]. Another glycoprotein which plays a crucial role into basement membranes integrity is nidogen, that is a monomeric glycoprotein which main role is to link the two distinct layers of laminins and collagen type IV within the basement membranes [224, 244]. Lastly, laminins are heterotrimeric protein which elemental units are the three chains α, β and γ. These proteins are synthesized and

assembled within the cell, the C-terminal of each chains form a coiled-coil rod (long arm), which terminal extremity is the globular domain of the α chain [224]. The N-terminal regions are well separated in an umbrella-like fashion. The C-terminal globular domain engages in the interaction with the cell membrane receptors, while the N-terminal domains engage in laminins' polymerization, which result in the well-known triskelion-like ensemble. Laminin polymerization is strictly dependent on calcium concentration; hence it can be started only in the extracellular space. The resulting flat laminin layers is so able to interact with cell membrane proteins, such as integrins, as well as with glycoproteins as syndecans and HSPG [60]. Is to be highlighted that mutations or different expression levels of laminins are less detrimental, for the basement membrane integrity, compared to the ones which involve collagen type IV or perlecan.

To conclude, is possible to define the basement membranes as materials which behave like non-linear, strain-stiffening, poro-visco-elastic material. These properties are highly variable and can strongly change from tissues to tissue in answer to different mechanical stimuli.

ECM RELATED DISEASES

The extracellular matrix is a very heterogeneous environment, with several part which interplay synergistically to give specific biomechanical properties to the different tissues and organs which build the multicellular organisms. When the homeostasis of the ECM is broken different pathological condition can arise, with different degrees of severity. As already mentioned, collagen is one of the most abundant proteins within the ECM and different pathologies, defined as collagenopathies, are associated to collagen's mutations, non-physiological post-translational modifications or altered expression [245, 246]. Collagen has also an active role in cancer progression and metastatic events [247]. Fibrotic tissues are indeed characteristic marker of cancer development in specific tissues; hence promoting cancer's cell development and migration, as well a solid protection from the immune system (Tab.2) [248–250].

Table.2. Interactive table of ECM related diseases (https://bdee9f94.diseases-cm7.pages.dev/)

Different pathologies which involve collagen have been described, worthy of mention are osteogenesis imperfecta (OI) [251], Ehlers-Danlos syndrome (EDS) [252], Alport syndrome [253] and the type II collagenopathies (see table 2). Osteogenesis imperfecta is a hereditable

disease which progression leads to bone fragility or skeletal, skin and soft tissues abnormalities. The onset of the OI's symptoms shows a spectrum of severity from mild to severe; OI type II is perinatal lethal; OI type III shows severe symptoms while OI type I shows mild symptoms. The most common mutations which lead to OI are the ones of collagen I α-chains, which account for almost the 90% of the cases of osteogenesis imperfecta [254–257]. However, have been identified and described different mutations of the enzymes involved in collagen synthesis, modification and secretion. Examples of such enzyme include the CRTAP, P3H1, CYPB complex, OASIS, FKBP65, HSP47, LH2, BMP1 and proteins involved in collagen I mineralization such as PEDF, SPARC as well as the signalling protein WNT [258–261]. The severity of osteogenesis imperfecta is much more related to the quality of collagen rather than its altered expression. Hence, mutations associated with collagen modifying enzymes, as well mutations which drastically change the collagen architecture, are associated with severe onset or perinatal death [252]. Another collagen related disease which manifests a broad spectrum of symptoms is the Ehlers-Danlos syndrome (EDS). The main feature which characterised EDS is a non-physiological collagen type I fibrillogenesis, thus leading to a disorganized population of collagen fibers with different width and length. EDS can be classified in different types, however common feature of the disease are the skeletal abnormalities, the hyperextensibility of the skin, the joint hypermobility as well the spontaneous vascular rapture [262]. Mutations on different collagen types as well as mutations on enzymes involved in collagen's modification defined different EDS types: 1) mutations of collagen type V α1 and α2 chains define the classic type EDS [263]; 2) mutations of the collagen type III α1 chain led to the vascular type EDS [264–266]; 3) mutations of lysyl hydroxylase (PLOD/LH) are the cause of kyphoscoliotic type EDS [267]; 4) mutations of collagen I α1 and α2 chains are responsible of arthrochalasia EDS [268]; 5) mutations of the procollagen I N-terminal proteinase ADAMTS-2 define dermatosparactic type EDS [269]. Much more rare mutations that can result in EDS involve enzymes which are involved in GAG biosynthesis, as the β3galactosyltransferase-II, the DS epimerase and the dermatan 4-O-sulfotransferase. Mutations of type IV collagen determine pathology of the basement membranes, such as the Alport syndrome. This syndrome is determined by mutations of the α3, α4 and α5 chains of collagen type IV, which result into an altered expression, or secretion, of collagen type IV thus causing nephritic syndrome and deafness. Has been also identified a frameshift mutation in exon 49 of the collagen type IV α1 chain, which produce a truncated version, at the C-terminal, of the chain [253]. All the collagenopathies involving collagen type II are defined as type II collagenopathies, the spectrum of the symptoms' severity is extremely broad form mild to severe conditions. The most severe phenotypes are the one of a-chondrogenesis type I, hypo-chondrogenesis and the platyspondyly lethal skeletal dysplasia, Torrance type. These diseases led to severe skeletal abnormalities, due to unproper collagen deposition through the development, in most cases type II collagen is replaced by type I or type III collagen which alters the biomechanical properties of bone tissues: leading to premature death [270]. Severe conditions can be also observed in Ulrich congenital muscular dystrophy, cause by collagen type VI mutations and in dystrophic epidermolysis bullosa which is the results of mutated collagen type VII. [271, 272]

Not only mutations of collagens are the cause of severe diseases, but indeed mutations on collagen also modifying enzymes are also actively involved in the development of fibrotic tissues and cancer progression. The lysine hydroxylase isoform 1, PLOD1 [273], has a

central role for the bone mineral density (BMD), PLOD2 plays a major role in collagen cross-links at the telopeptides region, while PLOD3 has an important role into the biosynthesis of collagen type IV and VI. PLODs expression is highly regulated at the transcriptional level, with different pathways and cytokines identified as regulators. The hypoxia-inducible factor (HIF-1α) can induce PLOD2 expression and thus inducing the hypoxia-induced Epithelial-Mesenchymal Transition (EMT) in glioma cells and breast cancer. The regulation of PLOD2 is achieved by the miR-26a-5p and by the miR-26b-5 as well by TGF-β signaling. E2Fs and FOXA1 regulate PLOD2 in cancer progression, furthermore PLOD2 expression is a potential marker in bladder cancer and renal cell carcinoma [274–276]. The HIF-1 factor can also promote the expression of PLOD1, which has been observed in breast cancer cells [277, 278]. However, the role of PLOD1 [279] and PLOD3 in cancer progression and fibrillogenesis is less understood compared to the role of PLOD2 [280–283]. The overexpression of LOX enzymes causes an alteration of the collagen cross-links and, as for PLOD enzymes, this results in tissues with altered mechanical properties which enhance the cancer cells progression and so the metastatic events. The role of the glucosyltransferase GLT25D is instead still partially understood, different clinical and experimental evidence suggest its direct role into hepatocellular carcinoma (HCC), as well as in the cerebral small vessel disease (CSVD). Its overexpression is associated to poor prognosis in fibrotic cancers, suggesting its active role in promoting nonphysiological collagen fibrillogenesis [284–287]. Compared to PLOD and LOX enzymes, less is known in relation of P4H involvement in ECM diseases or cancer progression. It has been found that in head and neck squamous cell carcinoma (HNSC), P4Hs are highly overexpressed which are associated with advanced tumor stage and poor patient prognosis. Furthermore, P4H expression is linked to the infiltration of immune cells, thus they can modify the response of cancer microenvironment to immunotherapy [288, 289].

To conclude, the deep investigation of the collagen modifying enzymes is of crucial importance to develop new therapeutic approaches to limit collagen fibrillogenesis, and thus cancer progression and metastasis; as well to better understand the processes underneath the complexity of ECM biology.

Chapter 3 IDENTIFICATION OF REGULATORY MOLECULAR "HOT SPOTS" FOR LH/PLOD COLLAGEN GLYCOSYLTRANSFERASE ACTIVITY

Daiana Mattoteiaa,*, Antonella Chiapparinoa,*,1, Marco Fumagalli a,*, Matteo De Marcoa,#, Francesca De Giorgia,2,#, Lisa Negroa, Alberta Pinnolaa, Silvia Faravellia, Tony Rosciolib,c,d, Luigi Sciettia,3, Federico Forneris.

Contribution: Recombinant protein production, enzymatic assay development.

ABSTRACT

Hydroxylysine glycosylation is a critical post-translational modification for the maturation and homeostasis of both fibrillar and non-fibrillar collagens. In humans, collagen lysyl hydroxylase 3 (LH3/PLOD3) and the collagen galactosyltransferase GLT25D1 are known to be involved in lysine glycosylation, but the exact division of labour between the two enzymes is unclear. Historically, LH3/PLOD3 was suggested to catalyse two glycosyltransferase reactions at an active site: an inverting β-1, O-galactosylation of hydroxylysine (Gal-T) and a retaining α-1,2-glucosylation of galactosyl-hydroxylysine (Glc-T).

In this study, we used indirect luminescence assays, direct mass-spectrometric readouts, and structural studies to show that LH3/PLOD3 has Glc-T activity only, and GLT25D1 only mediates Gal-T activity. Structure-guided mutagenesis showed that Glc-T catalysis required first-shell residues in the glycosyltransferase active site, but also longer-range residues in the same GT domain. We also determined the structures of human LH3/PLOD3 with different UDP-sugar analogues to investigate their interaction profiles and used structural information to develop LH3/PLOD3 glycosyltransferase inhibitors. Collectively, this data provides new tools to directly interrogate collagen hydroxylysine PTMs and provides a better appreciation for the geometric, electrostatic, and interaction networks that enable LH3/PLOD3 glycosyltransferase function, advancing mechanistic understanding of the hydroxylysine glycosylation pathway and informing on how to modulate these enzymes for biomedical and biotechnological uses.

INTRODUCTION

Collagens represent the most diverse protein family in humans and are highly conserved across metazoan evolution, from sponges to mammals [1–2]. The oligomeric assembly and functions of collagens are mediated primarily by post-translational modifications, in particular hydroxylation of proline and lysine and the highly specific glycosylation of hydroxylysine [3]. The Hyl-linked disaccharide displays a conserved glucosyl(α-1,2)-galactosyl [4] and monosaccharidic galactosyl-[5–6].

The extent of glycosylation appears to depend on collagen type [7–9], tissue microanatomical context [10–12], stage of development [13], and disease state [14–16]. Although thoroughly studied the precise mechanisms controlling collagen glycosylation and their roles in collagen homeostasis remain less well defined. The specific identity and stereochemistry of the Glc(α-1,2)-Gal [17]. Because both donor sugars have α-glycosides, the first step must be catalyzed by a galactosyltransferase [18]. Human isoenzyme 3 (LH3/PLOD3) has been shown to have in vitro Gal-T activity [19], but vivo studies have shown that reduced levels of LH3/PLOD3 and/or pathogenic GT-domain variants selectively impaired Glc-T activity [16, 20–21]. Examples include p.Asn223Ser, which generates an additional site of glycosylation within the GT domain and results in osteogenesis imperfecta [16], and p.Pro270Leu, which has recently been associated with a Stickler-like syndrome with vascular complications in association with features of Ehlers–Danlos syndrome and epidermolysis bullosa [21]. Mouse studies have also shown that LH3/PLOD3 Glc-T activity alone is necessary for collagen IV biosynthesis and basement-membrane formation during embryonic development [13, 22], and likely co-existed with additional collagen galactosyltransferases.

Two O-galactosyltransferases, GLT25D1 and GLT25D2, have been characterized [23–24]. Interesting to note, GLT25D1 and LH3/PLOD3 have been shown to utilize collagen substrates in a cooperative process [23, 25–26]. In osteosarcoma cell lines, producing more than ample fibrillar collagens, deletion of both GLT25D1 and GLT25D2 resulted in growth arrest, owing to loss of glycosylation, which demonstrates that the Gal-T activity of LH3/PLOD3, is of lesser importance than its Glc-T activity [27].

These overall findings are supportive of a specialized and conserved O-glycosylation machinery where separable proteins/protein complexes are directing both the Gal-T and Glc-T steps in vivo. That lastly returns us to broad questions about the spatial and temporal regulation of this pathway. Current structural insights are still primarily limited to 3D models of human LH3/PLOD3 bound to UDP-sugar donors [28] and a handful of mutagenesis studies examining Mn^{2+}-dependent glycosyltransferase features [19, 29].

Using structure-guided mutagenesis with standard indirect assays, and a novel mass-spectrometry-based direct assay that tracks the conversion of Lys to Glc-Gal-Hyl, we delineated a wide cooperative landscape of residues, within the LH3/PLOD3 GT domain, required for glycosyltransferase activity. Given the remarkable conservation of this

landscape across human LH/PLOD isoforms with reported Glc-T activity for LH1/PLOD1 [30] and LH2/PLOD2 [18], have exhibited the distribution of these catalytic features throughout the family, providing an extensive molecular overview of Glc-T activity. Our data also confirmed that GLT25D1 alone has Gal-T activity on Hyl, creating the substrate for the next LH/PLOD-mediated Glc-T step that produces Glc-Gal-Hyl. Finally, we have identified and characterized UDP-sugar analogs that inhibit LH3/PLOD3 glycosyltransferase activity, allowing for the development of the first targeted inhibitors of collagen Glc-T.

MATERIALS AND METHODS

CHEMICALS

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

MOLECULAR CLONING AND SITE-DIRECTED MUTAGENESIS

Coding sequences for LH3/PLOD3 (UniProt Q60568; Source BioScience), LH1/PLOD1 (UniProt Q02809; Source BioScience), and GLT25D1 (UniProt Q8NBJ5; codon-optimized, GeneWiz), lacking their N-terminal signal peptides, were amplified with oligonucleotides bearing in-frame 5′-BamHI and 3′-NotI sites (Suppl. Table 3) and inserted into pCR8, which served as the template for subsequent mutagenesis. LH/PLOD variants were generated using the Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific), with whole plasmid amplification employing phosphorylated primers; in each case, the forward primer introduced the desired substitution (Suppl. Table 3). Linear mutagenized plasmids were phosphorylated with T4 polynucleotide kinase (Life Technologies) prior to ligation. All constructs were verified by Sanger sequencing before subcloning into expression vectors. Mature wild-type and mutant LH3/PLOD3 and LH1/PLOD1 sequences were cloned into the pUPE.106.08 mammalian expression vector (U-protein Express BV) in frame with an N-terminal 6×His tag followed by a TEV protease site. GLT25D1 was subcloned into a modified pET28b-SUMO vector (Novagen), yielding an N-terminal 8×His-SUMO fusion for recombinant expression in E. coli.

LH/PLOD RECOMBINANT EXPRESSION AND PROTEIN PURIFICATION

To generate protein samples, suspension-adapted HEK293-F cells (Life Technologies, UK) at a density of 1×10^6 cells mL^{-1} were transfected with 1 µg plasmid DNA and 3 µg linear polyethyleneimine (PEI; Polysciences). Six days post-transfection, cell cultures were harvested by centrifugation at 1,000 × g for 15 min and the clarified supernatant was passed

through a 0.2 mm syringe filter, adjusted to pH 8.0, and processed by affinity purification as described above (Scietti and coworkers, 2018). Proteins from the supernatant were captured by their 6×His tags using a HisTrap Excel column (GE Healthcare) and further purified by size-exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare) in 25 mM, HEPES/NaOH; 200 mM, NaCl; pH 8.0, to yield a homogeneous preparation. Fractions containing the target protein were pooled and concentrated to 1 mg mL^{-1}.

GLT25D1 RECOMBINANT EXPRESSION AND PROTEIN PURIFICATION

The pET28b-SUMO-GLT25D1 construct was transformed into E. coli BL21 codon plus (DE3)-RP Plus (Agilent). A single colony was inoculated into 100 mL LB with kanamycin (0.1 mg mL^{-1}; 1:1000 v/v) and grown overnight at 37 °C, 220 rpm. The preculture was then used to inoculate 6 L of autoinducing ZYP5052 medium [31] for large-scale expression. Cultures were maintained for 4 h at 37 °C, then transitioned to 20 °C and grown for a total of 24 h prior to harvesting (4,000 × g, 20 min). The pellet was resuspended/homogenized in 100 mL buffer A (25 mM HEPES/NaOH, 0.5 M NaCl, pH 8) at a 1:5 (w/v) ratio and lysed by sonication (16 cycles, 9 s on/6 s off). Cell debris was removed by centrifugation (50,000 × g, 40 min, 4 °C), and the supernatant was filtered through a 1 µm Minisart GF syringe filter (Sartorius).

The clarified lysate was applied to a 5 mL HisTrap Excel column (GE Healthcare) pre-equilibrated in buffer A and eluted stepwise to 250 mM imidazole with an NGC chromatography system (Bio-Rad). Fractions containing His–SUMO–GLT25D1 were pooled, desalted on a 5 mL HiTrap Desalting column (GE Healthcare), and supplemented with His-tagged SUMO protease to a concentration of 3 µg mL^{-1} (1:300 v/v) and incubated for 2 h at room temperature. The mixture was reloaded onto the 5 mL HisTrap Excel column; cleaved GLT25D1 was in the flow-through. The sample was concentrated to 5 mg mL^{-1} using a 30 kDa Amicon Ultra-15 concentrator (Merck) and further polished by size-exclusion chromatography on a Superdex 75 Increase 10/300 GL column (GE Healthcare) equilibrated in SEC buffer (25 mM HEPES/NaOH, 200 mM NaCl, pH 8). GLT25D1 peak fractions, verified by SDS-PAGE, were pooled, concentrated to 5 mg mL^{-1}, flash-frozen in liquid N_2, and stored at −80 °C.

DIRECT MASS SPECTROMETRY ACTIVITY ASSAYS

For direct LH assays, the reaction mixtures were composed of 5 µM LH/PLOD (either wild type or evaluated variants), 50 µM $FeCl_2$, 100 µM 2-oxoglutarate, 500 µM ascorbate, and 1 mM peptide substrate. In addition to the first mixture, reaction conditions for the Gal-T were comprised of 5 µM GLT25D1, 50 µM Mn^{2+}, and 100 µM UDP-Gal; for Glc-T assays, an additional 100 µM UDP-Glc was included into the reaction mixture. Control reactions utilized the same mixture compositions, except the control reactions did not include LH/PLOD or GLT25D1, as showed in the text and figure legends. All the reactions were run for 3 h at 37 °C. For analysis, aliquots (10 µL) were made by diluting into 38 µL of Milli-Q

water, and acidifying with 2 µL of formic acid, to a final volume of 50 µL before being analysed on a UHPLC system (AB Sciex) either on a high-resolution QTOF mass spectrometer (AB Sciex X500B) with Turbo V ion source and TwinSprayer ESI probe, using analysis software controlled by SCIEX OS 2.1.

To make use of RP-HPLC for peptide separation, a Hypersil Gold C18 column (150 × 2.1 mm, 3 µm, 175 Å; Thermo Fisher Scientific), was used to run a linear gradient, for 15 min, from 2% to 50% B (solvent A: 0.1% aqueous FA; solvent B: acetonitrile with 0.1% FA) at a flow rate of 0.2 mL min^{-1}. The mass spectrum was taken in positive mode. The instrument parameters were as follows: ion spray voltage 4,500 V; declustering potential 100V; curtain gas 30 psi; ion source gas 1, 40 psi; ion source gas 2, 45 psi, source temperature was 350 °C; and the collision energy was set to 10 V. The data processing was performed using SCIEX OS 2.1.

INDIRECT LUMINESCENCE-BASED ACTIVITY ASSAYS

LH and Glc-T activities were measured using luminescence-based assays on a GloMax Discovery reader (Promega, USA) following Scietti et al. [32]. For Glc-T competitive inhibition experiments, we introduced minor modifications: reactions were started by adding 1 µL of a mixture containing 250 µM $MnCl_2$, UDP-glucose, and increasing concentrations of either UDP-GlcA or UDP-Xyl to the enzyme and gelatin substrate. All measurements were performed in triplicate, with controls conducted under identical conditions by omitting LH/PLOD enzymes. Data were processed and graphed in GraphPad Prism 7 (GraphPad Software, USA).

DIFFERENTIAL SCANNING FLUORIMETRY (DSF)

For differential scanning fluorimetry (DSF) studies, using a Tycho NT.6 (NanoTemper Technologies, Germany), we used wild-type and mutant LH3/PLOD3 enzymes. Enzyme stock solutions (1 mg mL^{-1}) were prepared in 25 mM HEPES ,500 mM NaCl, pH 8. For ligand binding experiments, LH3/PLOD3 was pre-incubated for 30 minutes at 23 °C in 50 µM $MnCl_2$ and with 5 mM UDP or UDP-sugar donor substrates (or analogues). The data was processed and plotted using GraphPad Prism 7 (GraphPad Software, USA).

CRYSTALLIZATION, DATA COLLECTION, STRUCTURE DETERMINATION AND REFINEMENT

The co-crystallization of wild-type and mutant LH3/PLOD3 was done using the hanging-drop vapor-diffusion method from Scietti et al., [32] where 0.5 µL of the enzyme (3.5 mg mL^{-1}) was mixed with 0.5 µL of the reservoir, which contained 600 mM sodium formate,

12% poly-glutamic acid (PGA-LM; Molecular Dimensions), 100 mM HEPES/NaOH pH 7.8, the metal salts, 500 µM $FeCl_2$ and 500 µM $MnCl_2$, and the UDP-sugar analog indicated (UDP, UDP-glucose, UDP-glucuronic acid, UDP-xylose). Crystals were cryoprotected in the mother liquor supplemented with 20% ethylene glycol, harvested by MicroMounts Loops (MiTeGen), flash cooled and stored in liquid nitrogen for data collection.

X-ray diffraction was performed at beamlines from the ESRF (Grenoble, France) and the SLS (Villigen, Switzerland). Images were indexed and integrated via XDS [33] and scaled via AIMLESS [34]; collection statistics are indicated in Supplementary Table 1. Due to significant anisotropy, datasets generated were then processed with STARANISO [35] prior to solving and refining the structures. The structures were solved by molecular replacement in PHASER [36] using the LH3/PLOD3 model bound to Fe^{2+}, 2-OG and Mn^{2+} [28] as the search model, then refined in iterative manual rebuilding in COOT [37] and automated refinement in phenix.refine [38]. The quality of the models was determined using MolProbity [39] and final refinement statistics listed in Supplementary Table 1.

RESULTS

A DIRECT MS-BASED ASSAY TO EVALUATE LYS-TO-GLC-GAL-HYL CONVERSION

Prior work has consistently shown that the N-terminal GT domain is the sole catalytic glycosyltransferase module in LH/PLOD enzymes [17–18, 28, 30]. This leaves unresolved the debated possibility that two chemically opposite activities, an inverting Gal-T and a retaining Glc-T, could coexist within a single active site. The identification of the collagen galactosyltransferases GLT25D1/2 further questioned the specificity of the initially reported LH3/PLOD3 Gal-T reaction [28]. To address this directly, we recombinantly produced human LH3/PLOD3 and GLT25D1 and refined an MS-based assay previously used for Lys-to-Hyl conversion [28] to detect Gal-Hyl and Glc-Gal-Hyl on synthetic collagen peptides.

First, LH activity was confirmed in the presence of Fe^{2+}, 2-OG, ascorbate, and the $(GIK)_4$ peptide, yielding the expected +16 Da mass shift consistent with hydroxylation of a single Lys (Figure 1A; Suppl. Fig. 1). The same mixture containing GLT25D1 but lacking LH3/PLOD3 produced no peptide modification (Suppl. Fig. 2). To test the putative inverting Gal-T of LH3/PLOD3, we added Mn^{2+} and UDP-Gal; no new peaks were observed (Figure 1B). Repeating the assay with GLT25D1 present revealed an additional peak consistent with galactosylation of a single Hyl (Figure 1C), demonstrating that galactosylation is mediated by GLT25D1, not LH3/PLOD3. Finally, inclusion of UDP-Glc generated a fourth peak corresponding to the nude peptide bearing one Glc-Gal-Hyl residue (Figure 1D).

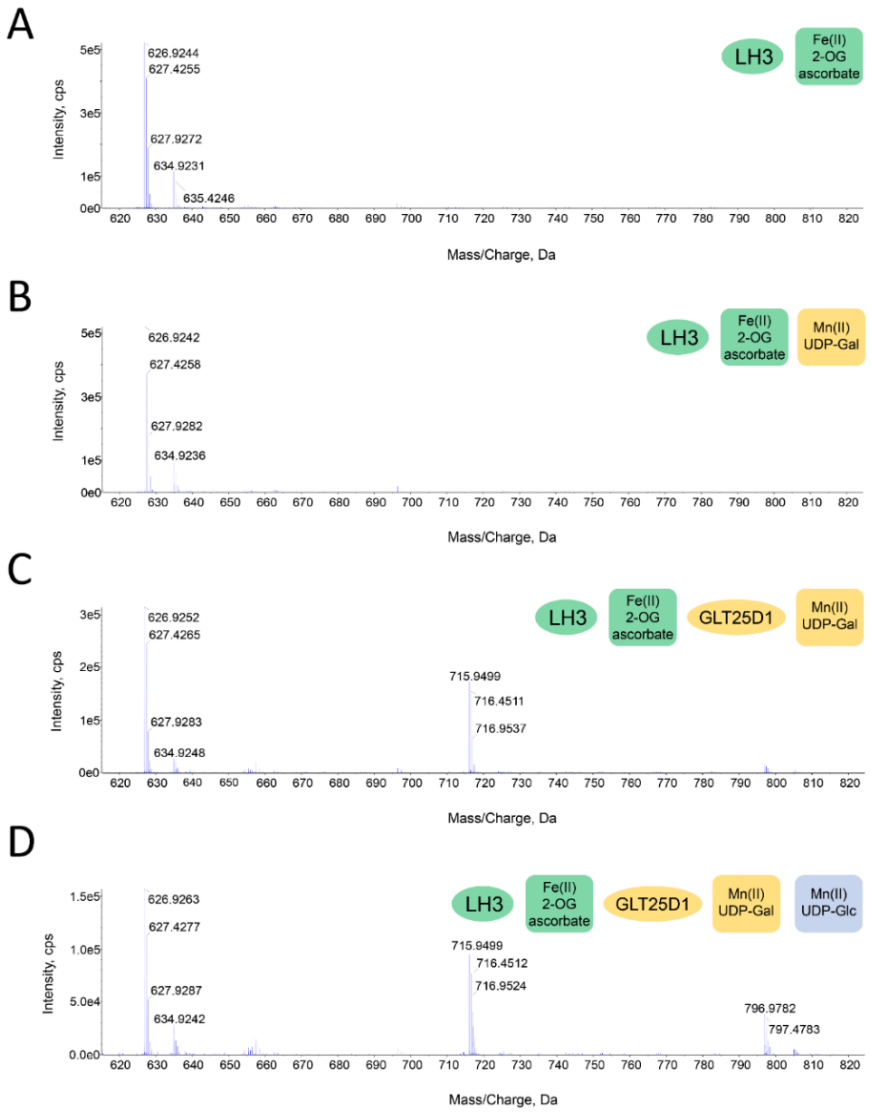


Figure 1: A direct assay to probe Lys-to-Glc-Gal-Hyl conversion. The MS spectra derive from incubations of a synthetic GIKGIKGIKGIK peptide (MW 1,254 Da) with the indicated enzymes and cofactors (see figure legends). All signals correspond to doubly charged ions and therefore appear at m/z values approximating one-half of the neutral mass. (A) With LH3/PLOD3 plus LH cofactors (2-OG and Fe^{2+}), a peak at m/z ≈ 635 is observed, consistent with +16 Da (single Lys hydroxylation). (B) Adding Gal-T cofactors (UDP-Gal and Mn^{2+}) to the mixture in (A) produces no additional peaks. (C) Inclusion of GLT25D1 under the conditions in (B) yields new ions at m/z ≈ 716, indicative of Gal-Hyl. (D) Further addition of UDP-Glc to the mixture in (C) produces ions at m/z ≈ 797, corresponding to Glc–Gal–Hyl.

THE AMINO ACID RESIDUES SHAPING THE UDP BINDING SITE ARE ESSENTIAL FOR LH3/PLOD3 GLC-T ACTIVITY

The N-terminal GT domain of LH3/PLOD3 shows a Mn^{2+} dependent GT-A type fold, which contains a UDP-donor binding pocket extending into the catalytic cavity [28], while the Glc-T activity of the two related enzymes LH1/PLOD1 and LH2/PLOD2 are only somewhat defined to date. Previous work demonstrated that LH1/PLOD1 Glc-T activity using indirect luminescence assays was ~10-fold lower than LH3/PLOD3 activity [30], and other independent work implicated LH2/PLOD2 Glc-T activity in lung adenocarcinoma progression [18]. To identify catalytic determinants for LH3/PLOD3, we used our prior homology models of LH1/PLOD1 and LH2/PLOD2 [40] to perform a side-by-side comparison across all human LH/PLOD isoform GT domains (Suppl. Fig. 3). Superimposition of the LH1 and LH2 models was particularly useful in this regard because of their close match to the LH3/PLOD3 crystal structure. Thus, side chain comparisons could be performed at the residue level (Suppl. Fig. 3B).

To identify unique characteristics in the UDP-binding pocket, we compared LH3/PLOD3 with lower Glc-T activity isoforms, LH1/PLOD1 and LH2/PLOD2 [18, 30]. Most residues that coordinate Mn^{2+} and bind the UDP are conserved among paralogs, however there was a notable difference: Val80 in LH3/PLOD3 corresponds to Lys68 in LH1/PLOD1 and Gly80 in LH2/PLOD2 (Figure 2A, Suppl. Fig. 3). The discrepancy in the donor binding cavity suggested a possible determinant of functional variation among the LH/PLOD GT domains. In LH3/PLOD3, Val80 is in the middle of a flexible “glycoloop” (Gly72–Gly87) that is disordered in the apoenzyme but becomes ordered in the presence of UDP ([28]) near the UDP-sugar ribose. We reasoned that we could induce an impairment of donor binding by replacing Val with a bulky, positively charged Lys (LH1/PLOD1) or removing side chain bulk altogether (Gly; LH2/PLOD2). Accordingly, we created the LH3/PLOD3 Val80Lys mutant. This mutant was correctly folded as seen in analytical gel filtration and DSF, it retained catalytically competent lysyl hydroxylase activity like the wild type (Suppl. Fig. 4) but lacked detectable Glc-T activity in our MS-based assays (Figure 2B, Table 1A).

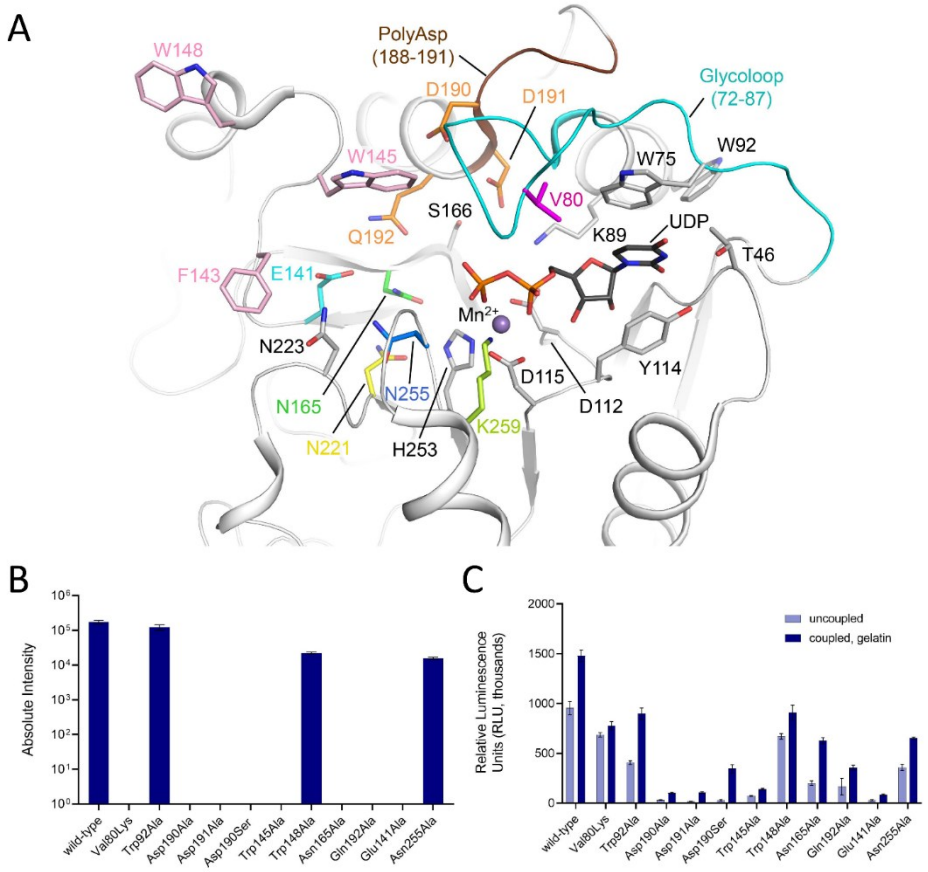


Figure 2: Structural and functional features of the LH3/PLOD3 glycosyltransferase (GT) domain. (A) Schematic of the LH3/PLOD3 GT domain (PDB 6FXR) showing catalytic-site residues as sticks. Motifs included in the interaction with UDP-sugar donors are the Poly-Asp motif (brown) and the glycoloop (cyan). Residues evaluated for catalytic role are colored; the gray residues are previously established ligands to Mn^{2+} (purple sphere) and UDP (black sticks).(B) Summary of LH3/PLOD3 mutant Glc-T activities compared to wild type, using a direct MS assay.(C) Summary of Glc-T activities for the same mutants compared to wild type, using a previously described indirect luminescence assay. Each plot shows activity measured without acceptor ("uncoupled," light blue), as well as in the presence of gelatin as an acceptor; measurements are background-corrected. In the B and C plots error bars show standard deviations averaged from independent experiments (N > 3).

Table 1A: results of activity assays (performed using either MS or luminescence) for all LH3/PLOD3 mutants described in this work, compared to wild-type LH3/PLOD3

Mutation	Localization	Folding state/ LH activity	Glc-T activity (MS) (%)	Glc-T activity (Luminescence) (%)	
				uncoupled	coupled, gelatin
wild-type	N/A	Yes	100	100	100
Val80Lys	glycoloop	Yes	N.D.	72 ± 2	53 ± 3
Trp92Ala	UDP-binding cavity	Yes	71 ± 14	43 ± 2	60 ± 4
Asp190Ala	poly-Asp helix	Yes	N.D.	3 ± 0.27	7 ± 0.5
Asp191Ala	poly-Asp helix	Yes	N.D.	2 ± 0.3	7 ± 0.6
Asp190Ser	poly-Asp helix	Yes	N.D.	2 ± 0.4	17 ± 2
Trp145Ala	acceptor substrate cavity and gates	Yes	N.D.	8 ± 0.5	9 ± 0.6
Trp148Ala	acceptor substrate cavity and gates	Yes	13 ± 2	70 ± 3	61 ± 5
Asn165Ala	region proximate UDP-sugar	Yes	N.D.	21 ± 3	42 ± 2
Gln192Ala	region proximate UDP-sugar	Yes	N.D.	17 ± 9	24 ± 2
Glu141Ala	region proximate UDP-sugar	Yes	N.D.	3 ± 0.6	6 ± 0.6
Asn255Ala	region proximate UDP-sugar	Yes	9 ± 0.5	38 ± 3	44 ± 0.7
Pro270Leu	interface of AC and GT domains	No	N.D.	N.D.	N.D.

Since LH3/PLOD3 activates UDP-sugar donors and releases UDP even in the absence of a collagen acceptor [28], we compared the Val80Lys mutant under acceptor-free ("uncoupled") and acceptor-present ("coupled") conditions through luminescence assays. Results indicate that most of the apparent Glc-T signal for the mutant is due to uncoupled turnover, consistent with the direct MS assay (Figure 2C, Table 1A). Thus, it is likely that Val80 facilitates the productive positioning of the donor in the enzyme to laterally transfer the glycan to the acceptor and does not only simply stabilize the UDP moiety in the active site.

To clarify Val80's role in LH3/PLOD3 Glc-T activity, we determined a 3.0 Å crystal structure of the Val80Lys mutant with Mn^{2+} and a 2.3 Å structure of the same mutant bound to both Mn^{2+} and UDP-Glc (Figure 3A; Suppl. Table 1). Both models superimpose closely with wild-type LH3/PLOD3 across all domains (Suppl. Fig. 5A). In the Mn^{2+}-only structure, the glycoloop containing residue 80 remained unresolved due to flexibility, mirroring the wild type (Figure 3C). By contrast, in the UDP-donor bound complex, the Lys80 side chain was unambiguously defined in the electron density. Despite added steric bulk, Lys80 adopts a conformation compatible with concurrent occupancy of UDP-Glc in the active site; however, as in the wild type, the sugar moiety is disordered and not visible (Figure 3A). Overall, these findings indicate that the Val80Lys substitution perturbs LH3/PLOD3 glycosyltransferase function only partially.

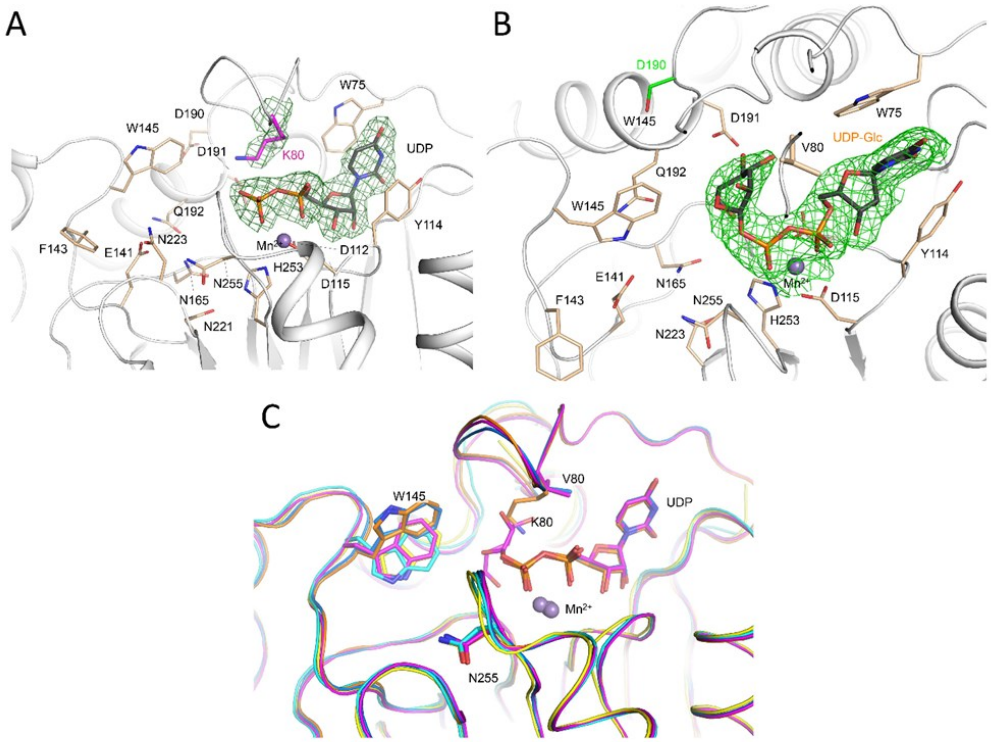


Figure 3: Structural characterization of LH3/PLOD3 mutants. (A) Crystal structure of the LH3/PLOD3 Val80Lys variant bound to UDP-glucose and Mn^{2+}. Electron density is resolved for the substituted lysine and for the UDP portion of the donor (green mesh; 2Fo–Fc omit map contoured at 1.3 σ). Catalytic-site residues are shown as sticks; Mn^{2+} appears as a purple sphere. As in the wild type, the glucose moiety is not visible in the experimental density. (B) Crystal structure of the LH3/PLOD3 Asp190Ser variant with UDP-glucose and Mn^{2+}. Electron density is evident for the substituted serine and for the entire donor, including the sugar (green mesh; 2Fo–Fc omit map at 1.3 σ). Colouring and representations match panel (A). (C) Superposition of available apo and UDP-glucose–bound structures for wild type, Val80Lys, and Asp190Ser: apo (wild type, cyan; Val80Lys, yellow) and donor-bound (wild type, marine; Val80Lys, orange; Asp190Ser, magenta). Notably, ligand binding induces similar Trp145 side chain conformations in wild type and mutants. Owing to glycoloop flexibility in apo states, Val/Lys80 side chains are resolved only in UDP-sugar bound structures.

THE LH3/PLOD3 GLC-T ACTIVITY IS AFFECTED BY LONG RANGE REARRANGEMENT OF TRP92 AND TRP75

The glycoloop is a structural element unique to LH/PLOD GT domains. It includes Trp75, whose aromatic side chain π-stacks with the uridine of the donor, and, together with Tyr114 from the hallmark DxxD motif [28], creates an aromatic "sandwich" around the donor (Figure 2A). Both residues are essential for LH3/PLOD3 Glc-T activity [28]. Upon UDP-donor binding, the glycoloop adopts an ordered conformation without major rearrangements of neighboring residues, aside from subtle shifts in the distant, nonconserved Trp92 (Figure 2A; Suppl. Fig. 3), whose bulky side chain rotates toward Trp75. Motivated by this, we generated the Trp92Ala variant, which retained proper folding and LH activity (Suppl. Fig. 4). In contrast, the mutant exhibited undetectable Glc-T activity relative to wild-type LH3/PLOD3 (Figure 2B–C; Table 1A), whereas the specific Trp92Ala substitution itself only modestly reduced Glc-T (Figure 2B–C; Table 1A). These observations suggest that LH3/PLOD3-specific long-range interactions within the GT domain help stabilize productive glycoloop conformations in donor-bound states.

A POLY-ASP SEQUENCE NEAR THE DONOR SUGAR BINDING SITE IS ESSENTIAL FOR GLC-T ACTIVITY IN LH1/PLOD1 AND IN LH3/PLOD3

We also investigated the poly-Asp stretch (Asp188–Asp191; Fig. 2A), present in UDP-sugar bound structures, which interacts with the glycoloop, and is partially conserved among isoforms of LH/PLOD (Suppl. Fig. 3). Previous research showed that substitutions at Asp190 and Asp191 can prevent LH3/PLOD3 glycosyltransferase activity [29] and aligns with our crystal structures indicating these side chains project into the Glc-T catalytic cavity. We made individual alanine variants for Asp190 and Asp191, both of which could be properly folded and had general LH/PLOD activity (Suppl. Fig. 4). However, both mutants destroyed Glc-T activity blocking donor UDP-sugar activation and transfer as confirmed by MS and luminescence-based assays. Taken together, these data suggest that the poly-Asp motif, specifically Asp190 and Asp191, is important for the positioning and recognition of donor and/or acceptor substrates.

Recent work reported that the LH1/PLOD1 Ser178Arg substitution, aligned with LH3/PLOD3 Asp190 within the poly-Asp segment, causes heritable thoracic aortic disease [30]. In LH3/PLOD3, replacing Asp190 with Arg severely compromises enzyme function by drastically reducing Glc-T activity. To probe residue-specific differences between LH1/PLOD1 and LH3/PLOD3, we engineered the LH3/PLOD3 Asp190Ser variant to mimic the LH1/PLOD1 identity at this position (Suppl. Fig. 4). Both MS and luminescence assays showed that this single substitution abolishes LH3/PLOD3 Glc-T activity (Figure 2B–C). We also determined the 2.30 Å crystal structure of full-length human LH3/PLOD3 Asp190Ser bound to Mn^{2+} and UDP-Glc, which is overall nearly indistinguishable from the wild type (Suppl. Fig. 5A). Notably, electron density at the GT active site clearly resolves the glucose moiety of the UDP-Glc donor (Figure 3B). This observation: absent in wild-type

LH3/PLOD3 [28], underscores the Poly-Asp loop's importance for Glc-T catalysis. To test the residue's broader relevance, we examined the human LH1/PLOD1 ortholog by introducing Ser178Asp. Remarkably, this single substitution substantially rescued LH1/PLOD1 Glc-T activity, detectable by both direct MS-based and indirect luminescence assays (Figure 4); by contrast, wild-type LH1/PLOD1 Glc-T is too low to observe by MS (Figure 4). Collectively, these results provide direct in vitro evidence for LH1/PLOD1 Glc-T activity and further establish Asp190 (LH3/PLOD3 numbering) as a key determinant in donor activation prior to transfer to the collagen acceptor.

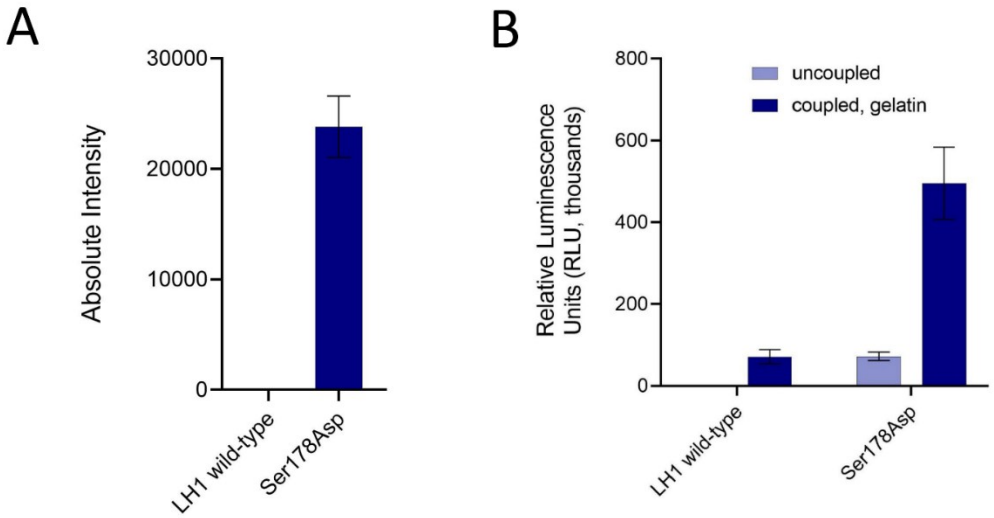


Figure 4: Ser178 in LH1/PLOD1, corresponding to Asp190 in LH3/PLOD3, is a key residue for Glc-T activity for both enzyme isoforms. (A) Direct MS assays comparing Glc–Gal–Hyl signal intensities for wild-type LH1/PLOD1 and the Ser178Asp variant. (B) Indirect, luminescence-based assessment of Glc-T activity for LH1 wild type and Ser178Asp, reporting both coupled and uncoupled reactions (as shown in Figure 2C). In both panels, error bars indicate standard deviations from independent (N > 3) experiments.

TWO GATING TRP RESIDUES MODULATE GLC-T ACTIVITY BY AFFECTING ACCEPTOR SUBSTRATE BINDING

Having characterized residues that stabilize the UDP moiety of donor substrates, we turned to a second set of positions in the GT active site, situated opposite the putative location of the flexible sugar rings (Figure 2A). Many LH3/PLOD3 residues shaping this sector of the

catalytic cavity align with catalytic determinants described for other GT-A glycosyltransferases (Suppl. Fig. 6) [41]. Notably, Trp145, located in an LH3/PLOD3-specific loop within the GT domain, has been implicated as a modulator of Glc-T activity, adopting distinct rotamers in apo versus donor-bound states that alter the contour and steric profile of the catalytic pocket [28]. Strikingly similar conformational shifts are evident when comparing substrate-free and donor-bound structures of the Val80Lys and Asp190Ser LH3/PLOD3 variants (Figure 3C). Inhibition of Glc-T activity because of a Trp145 to Ala replacement (Figure 2B–C) with no other properties impaired (Suppl. Fig. 4), strengthens the proposed "gatekeeper" function of Trp145 within the GT active site, and enables the donor sugar to correctly orient for transfer. Structure-based alignment of different GTs (including distantly related enzymes) reveals that several related enzymes utilized aromatic side chains that stem from different structural elements in the active site in the same manner as the LH3/PLOD3 Trp145 example. Examples include: Tyr186 in Neisseria meningitidis LgtC, Trp314 in the N-acetyllactosaminide α-1,3-galactosyltransferase GGTA1, Trp300 in the ABO blood-group transferase, and Trp243/Phe245 in the glucuronyltransferases B3GAT3 and B3GAT1 (Suppl. Table 2; Suppl. Fig. 6). This is notable given that the 142–163 loop is a structural element unique to LH/PLOD enzymes [28]. The observation underscores the remarkable versatility of glycosyltransferases, whose structural plasticity enables closely related catalytic mechanisms to act on a wide array of specific donor and acceptor substrates. Previously, comparisons of apo and substrate-bound LH3/PLOD3 suggested a concerted mechanism in which conformational changes of a no conserved surface aromatic residue (Trp148; Fig. 2A) act in tandem with Trp145 [28]. To test this, we generated the Trp148Ala mutant. This variant displayed wild-type folding and LH activity (Suppl. Fig. 4) but showed markedly reduced Glc-T activity in the presence of both donor and acceptor substrates (Fig. 2B–C, Table 1A). Although the effect was less severe than for Trp145 mutation, the data support a synergistic model linking long range substrate recognition at the enzyme surface with active site rearrangements upon substrate binding.

ADDITIONAL RESIDUES FACING BOTH DONOR AND ACCEPTOR SUBSTRATES AFFECT THE LH3/PLOD3 GLC-T ACTIVITY

In LH3/PLOD3 complexes with UDP-Glc and Mn^{2+}, electron density adjacent to the UDP pyrophosphate was weak and only partly consistent with the donor sugar, suggesting multiple conformations coexisting within the binding pocket [28]. To pinpoint additional catalytic determinants, we surveyed the neighbouring region of the active site for residues, particularly those bearing carboxylate or amide side chains, which could serve as candidate catalytic nucleophiles, potentially forming a covalent glycosyl–enzyme intermediate prior to transfer to the acceptor [42]. In GT-6 retaining glycosyltransferases, a conserved Glu often serves as the catalytic nucleophile [43–44]. Inspection of LH3/PLOD3 structures revealed Gln192, Asn165, and Glu141 projecting into the pocket that accommodates the donor sugar (Fig. 2A). We generated Ala substitutions at each site and, in all cases, obtained properly folded, functional LH enzymes (Suppl. Fig. 4). Assessment of Glc-T activity showed that Asn165Ala and Gln192Ala retained the ability to activate UDP donors and transfer sugar, albeit with reduced efficiency, whereas Glu141Ala completely abolished Glc-T function

(Fig. 2B–C, Table 1A). These data suggest Glu141 is necessary for catalysis, and implicate a surrounding negatively charged pocket (Asp190, Asp191, Gln192, and Asn165) as a larger cooperative network assisting with the glycosyltransferase function. Asn255, which is closest to the UDP phosphate-sugar linkage near Glu141 within the Glc-T site, is completely conserved between the LH / PLOD isoforms (Suppl. Fig. 3), but there is no equivalent residue in any of the structurally characterized GT-A glycosyltransferases. In the crystal structures for all six LH3/PLOD3 structures, the Asn255 side chain pointed away from the donor substrate (Fig. 2A), raising the possibility that its amide might play a role in catalysis via acceptor recognition. However, the properly folded Asn255Ala (Suppl. Fig. 4) variant retained Glc-T activity with only a small reduction (Fig. 2B–C, Table 1A), disputing a strictly catalytic and mandatory role.

Pathogenic mutations in the LH3/PLOD3 GT domain affect protein folding.

A recently reported pathogenic LH3/PLOD3 variant, p.Pro270Leu, maps to the interface between the AC and GT domains [21]. Although this residue lies in a loop that helps shape the GT cavity, its position makes a direct catalytic role unlikely. To assess its impact on enzymatic function, we attempted recombinant expression of the Pro270Leu mutant; however, protein yields were nearly undetectable, precluding in vitro assays. Given the otherwise robust reproducibility of expressing numerous LH3/PLOD3 point mutants, these observations suggest that Pro270Leu primarily compromises overall enzyme stability rather than catalytic activity, leading to severely reduced protein expression in vitro, and likely in vivo as well.

Molecular structures of LH3/PLOD3 in complex with UDP-sugar analogs provide insights glycan moieties processing in the catalytic cavity

A common challenge in characterizing glycosyltransferases is the pronounced flexibility of the donor sugar within the active site, a limitation that is accentuated for enzymes, such as LH3/PLOD3, that can turn over UDP-sugars in the absence of acceptor substrates. Considering our prior [28] and current co-crystallization data, we asked whether the reaction product, free UDP, remains bound in the LH3/PLOD3 GT domain with efficiency comparable to physiological donors after catalysis. Differential scanning fluorimetry revealed a larger thermal stabilization for free UDP ($\Delta T \approx 3.5$ °C) than for UDP-sugar donors ($\Delta T \approx 2–2.5$ °C) (Figure 5A). These findings suggest that free UDP can bind LH3/PLOD3, potentially with higher affinity than UDP–sugar donors, implying that Glc-T catalysis may be susceptible to product inhibition. Unexpectedly, this increased thermal stabilization did not translate into detectable product trapping in wild-type LH3/PLOD3 crystal structures: regardless of the UDP concentration used for co-crystallization or soaking, no electron density for free UDP was observed, and the models were indistinguishable from the apo enzyme (Suppl. Fig. 7). Notably, Kivirikko and colleagues previously reported that UDP-

glucuronic acid (UDP-GlcA) can act as a competitive inhibitor of collagen glycosyltransferases [45], and UDP-GlcA has been employed to isolate LH3/PLOD3 from chicken embryo preparations [19].
Follow-up biochemical data were lacking, so we evaluated the effect of UDP-GlcA on LH3/PLOD3 by DSF and luminescence-based Glc-T assays. DSF indicated weak binding, a modest thermal shift of ~1–1.5 °C (Fig. 5A), providing less stabilization than either UDP-sugar donors or free UDP. Enzymatic measurements confirmed competitive inhibition (Fig. 5B) with IC_{50} values in the millimolar range (Table 1B). We also obtained a 2.2 Å crystal structure of wild-type LH3/PLOD3 bound to Mn^{2+} and UDP-GlcA (Suppl. Table 1), showing that UDP-GlcA effectively occupies the donor-substrate site (Fig. 5C). Additional electron density for the glucuronic acid moiety was present but could not be fit by a single conformation; nonetheless, it clearly adopts a bent geometry, with the sugar buried deep in the catalytic pocket near Lys89, Asp190, and Asp191, and positioned away from catalytically critical residues (Trp145, Asn255, Glu141), thereby leaving space in the cavity to accommodate acceptor substrates.

Given the alternative conformations inferred for GlcA from electron-density analysis and its proximity to LH3/PLOD3 Val80, we tested whether the Val80Lys substitution affects inhibitor binding. We therefore determined the 2.7 Å structure of the Val80Lys mutant bound to UDP-GlcA (Suppl. Table 1) and observed a partial displacement of the glycoloop; its normally well-defined density in UDP-sugar–bound wild-type structures was not evident (Fig. 5D). Concurrently, electron density for the glucuronic acid moiety was not improved and in fact was poorer than in the wild type, indicating that the sugar's intrinsic flexibility is not governed by specific glycoloop conformations but rather by the absence of stabilizing protein–ligand contacts that would lock the ring into a single geometry. Because the glucuronic acid lacked a defined pose in the crystal, we examined another UDP-sugar analog lacking the carboxylate, UDP-xylose (UDP-Xyl). Like UDP-GlcA, UDP-Xyl acted as a weak Glc-T inhibitor (Fig. 5A–B) with IC_{50} values in the high micromolar range (Table 1B). The 2.0 Å structure of LH3/PLOD3 with Mn^{2+} and UDP-Xyl showed the inhibitor in the catalytic site, but with weak density for the sugar, consistent with multiple xylose conformations, as seen for UDP-GlcA (Fig. 5E). Collectively, these findings indicate that the LH3/PLOD3 GT pocket accommodates diverse UDP-sugar ligands and that inhibitory potency correlates with reduced ligand flexibility (i.e., increased stabilization) within the cavity. Accordingly, UDP-sugar analogs that form stronger interactions with residues adjacent to the glycan portion of UDP-GlcA/UDP-Xyl may serve as more potent inhibitors of LH/PLOD glycosyltransferase activity.

Table 1B: evaluation of the competitive binding of UDP-GlcA and UDP-Xyl in wild-type LH3/PLOD3 in presence of UDP-Gal or UDP-Glc and acceptor substrates.

Inhibitor	UDP-Gal (IC_{50}, □M)	UDP-Glc (IC_{50}, □M)
wild-type + UDP-GlcA	1130 ± 370	> 10000
wild-type + UDP-Xyl	91 ± 23	3170 ± 211

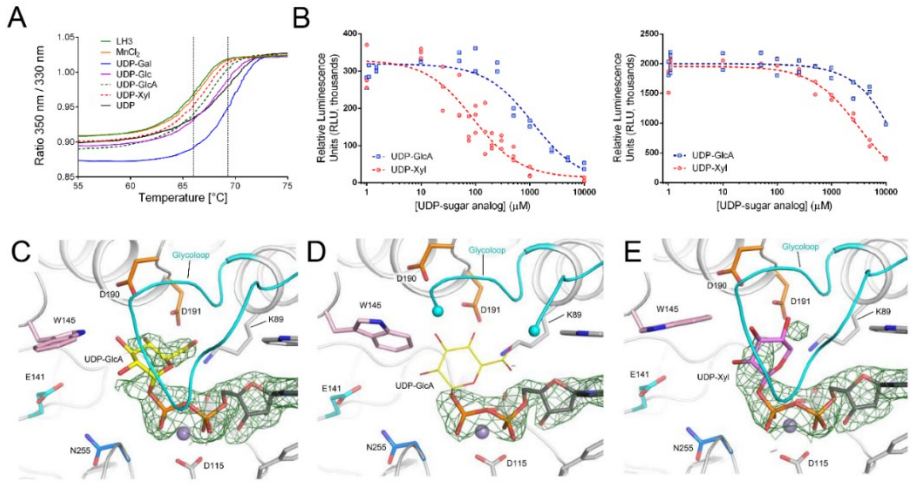


Figure 5: Characterization of UDP-sugar analogues. (A) Differential scanning fluorimetry (DSF) of LH3 wild type reveals substantial thermal stabilization by Mn^{2+} in addition to physiological donors, UDP-galactose (solid blue), and UDP-glucose (solid purple), and free UDP (solid black), where UDP-xylose (red dashed) and UDP-glucuronic acid (green dashed) only demonstrate modest stabilization (solid green trace, LH3 alone). (B) Luminescence-based competition assays of increasing UDP-GlcA or UDP-Xyl binding to wild-type LH3/PLOD3 with UDP-Gal (left) and UDP-Glc (right), with gelatin as an acceptor. (C) Crystal structure of wild-type LH3 with Mn^{2+} and UDP-GlcA shows significant electron density for UDP (2Fo-Fc omit map, green mesh, contoured at 1.2 σ). The glucuronic acid moiety (yellow) can be modeled, although it only occupies partial density. (D) Crystal structure of the LH3 Val80Lys mutant with Mn^{2+} and UDP-GlcA has density for the UDP backbone (black sticks; 2Fo–Fc omit map at 1.2 σ) whereas there is no density for the glucuronic acid (yellow). The glycoloop segment comprised of residues 79–83 is disordered (cyan spheres). (E) Crystal structure of wild-type LH3 with Mn^{2+} and UDP-xylose shows significant density for UDP (2Fo-Fc omit map at 1.2 σ) as well as only partial density for the xylose moiety (pink).

DISCUSSION

Glycosyltransferases are incredibly versatile and capable of catalyzing broad ranges of substrates, however when examined closely they share similar framework and mechanism features that allow for comparative analysis even in the absence of structural similarities and sequence conservation. LH3/PLOD3 has long been recognized as multifunctional and catalysing several steps in the Lys→Glc–Gal–Hyl posttranslational pathway [1], while recent examples suggest a similar multifunctionality for the paralogs LH1/PLOD1 and LH2/PLOD2 [18, 30]. Our in vitro data demonstrates the partitioning of function by combining direct MS based assays with indirectly read-out glucosyltransferase products. Collagen galactosyltransferases like GLT25D1 carry out only hydroxylysine glycosylation, whereas LH/PLOD enzymes will only glucosylate a galactosyl-hydroxylysine. Thus, the LH/PLOD enzymes are retaining-type glucosyltransferases.

Moreover, the direct MS assay enabled real-time tracking of the first fully synthetic Glc–Gal–Hyl collagen peptide generated in vitro via a one-pot total synthesis.

To interrogate the GT domains of LH/PLOD enzymes, we initiated structure guided mutagenesis by targeting the sole non conserved residue immediately surrounding the UDP-donor site, Val80 in LH3/PLOD3 (aligned with Lys68 in LH1/PLOD1 and Gly80 in LH2/PLOD2). This position lies within the glycoloop and abuts the ribose of the UDP-sugar donor(s). Biochemically, the Val80Lys substitution compromised LH3/PLOD3 catalysis. Structurally, the LH3/PLOD3 Val80Lys mutant displayed a disordered glycoloop, offering a mechanistic rationale for the loss of Glc-T activity. These findings suggest that Val80 supports productive positioning of the donor, particularly its glycan moiety, yet also indicate that sequence divergence at this single site is insufficient to account for the diminished Glc-T activity of LH1/PLOD1 and LH2/PLOD2. Comparative analysis of apo versus UDP-Glc bound LH3/PLOD3 revealed three second-shell residues that adopt altered side chain conformations, implicating them in catalysis. Among these, the non-conserved Trp92 in LH3/PLOD3 (Leu80 in LH1; Leu92 in LH2a/b) orients its aromatic ring to stabilize the glycoloop and thereby facilitate reaction activity. Substitution of Trp92 did not completely abolish LH3/PLOD3 Glc-T activity, suggesting that diminished Glc-T function in LH1/PLOD1 and LH2/PLOD2 likely reflects this and additional subtle sequence differences acting in concert. We next examined a second, catalytically critical region of the GT domain, the poly-Asp helix [29]. Asp190 in LH3/PLOD3, located at the catalytic-site entrance, proved essential: the Asp190Ser variant (mirroring LH1/PLOD1) markedly reduced Glc-T activity and, notably, yielded the first structure with resolvable electron density for the donor sugar. Conversely, the reciprocal change in LH1/PLOD1 (Ser178Asp) substantially enhanced Glc-T activity relative to wild type, identifying this position as a key determinant for donor-glucose activation prior to transfer to the acceptor substrate.

In the GT active site, LH3/PLOD3 has a non-conserved loop that defines the catalytic cavity with two large aromatic residues, Trp145 and Trp148. These residues could be working together, based on a comparison of the apoprotein and donor complex structures [44], for

example, the α-1,3-galactosyltransferase GGTA1, where a conserved glutamate resides on the β-face of the donor sugar [46–47]. The corresponding position in LH3/PLOD3 is Gln192. Although Gln192 lies adjacent to the poly-Asp helix and is spatially removed from the donor-binding site, making a direct catalytic role unlikely, our mutagenesis indicates that its side chain is nonetheless important for LH3/PLOD3 glycosyltransferase efficiency. Nearby, we identified two additional residues, Glu141 and Asn165, whose side chains point directly toward the donor glycan (Fig. 2A), implicating them in donor activation and/or transfer of the sugar to the acceptor. Whereas the Asn165Ala substitution reduced glycosyltransferase activity by roughly twofold (Fig. 2B–C, Table 1A), the Glu141 carboxylate proved indispensable: replacing Glu141 with alanine rendered LH3/PLOD3 completely inactive. In LH3/PLOD3, Glu141 occupies a position analogous 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 (Suppl. Fig. 6; Suppl. Table 2). Because loss of this carboxylate abolishes activity, we propose that the residue corresponding to Glu141 constitutes a catalytic "hot spot" in glycosyltransferases that present a carboxylate or amide at this position. Our study updates the biochemical framework for the Lys→Glc-Gal-Hyl conversion using robust, orthogonal in vitro assays and new three-dimensional structures of LH3/PLOD3 and variants, including complexes with UDP-sugar analogs that act as mild inhibitors (Fig. 5). Despite substantial flexibility of the bound glycans, these structures furnish actionable insight for structure-based discovery of LH3/PLOD3 Glc-T inhibitors; agents that could inspire therapeutic strategies for conditions marked by excessive collagen glycosylation, such as osteogenesis imperfecta. Together with a comprehensive mutational scan of the GT active site, our results delineate the intricate network of shapes, electrostatics, and interactions that enable LH3/PLOD3 Glc-T catalysis.

REFERENCES

1. Myllyharju, J., and Kivirikko, K.I. (2004). Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20, 33-43.

2. Luther, K.B., Hulsmeier, A.J., Schegg, B., Deuber, S.A., Raoult, D., and Hennet, T. (2011). Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J Biol Chem 286, 43701-43709.

3. Cummings, R.D. (2009). The repertoire of glycan determinants in the human glycome. Mol Biosyst 5, 1087-1104.

4. Spiro, R.G. (1967). The structure of the disaccharide unit of the renal glomerular basement membrane. J Biol Chem 242, 4813-4823.

5. Spiro, M.J., and Spiro, R.G. (1971). Studies on the biosynthesis of the hydroxylsine-linked disaccharide unit of basement membranes and collagens. II. Kidney galactosyltransferase. J Biol Chem 246, 4910-4918.

6. Sternberg, M., and Spiro, R.G. (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. Ren Physiol 3, 1-3.

7. Sternberg, M., Grochulski, A., Peyroux, J., Hirbec, G., and 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. Coll Relat Res 2, 495-506.

8. Hamazaki, H., and Hamazaki, M.H. (2016). Catalytic site of human protein-glucosylgalactosylhydroxylysine glucosidase: Three crucial carboxyl residues were determined by cloning and site-directed mutagenesis. Biochem Biophys Res Commun 469, 357-362.

9. Spiro, R.G. (1969). Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. J Biol Chem 244, 602-612.

10. Bornstein, P., and Sage, H. (1980). Structurally distinct collagen types. Annu Rev Biochem 49, 957-1003.

11. Terajima, M., Perdivara, I., Sricholpech, M., Deguchi, Y., Pleshko, N., Tomer, K.B., and Yamauchi, M. (2014). Glycosylation and cross-linking in bone type I collagen. J Biol Chem 289, 22636-22647.

12. Schofield, J.D., Freeman, I.L., and Jackson, D.S. (1971). The isolation, and amino acid and carbohydrate composition, of polymeric collagens prepared from various human tissues. Biochem J 124, 467-473.

13. Toole, B.P., Kang, A.H., Trelstad, R.L., and Gross, J. (1972). Collagen heterogeneity within different growth regions of long bones of rachitic and non-rachitic chicks. Biochem J 127, 715-720.

14. Moro, L., Romanello, M., Favia, A., Lamanna, M.P., and Lozupone, E. (2000). Posttranslational modifications of bone collagen type I are related to the function of rat femoral regions. Calcif Tissue Int 66, 151-156.

15. Rautavuoma, K., Takaluoma, K., Sormunen, R., Myllyharju, J., Kivirikko, K.I., and Soininen, R. (2004). Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. Proc Natl Acad Sci U S A 101, 14120-14125.

16. Tenni, R., Valli, M., Rossi, A., and Cetta, G. (1993). Possible role of overglycosylation in the type I collagen triple helical domain in the molecular pathogenesis of osteogenesis imperfecta. Am J Med Genet 45, 252-256.

17. Lehmann, H.W., Wolf, E., Roser, K., Bodo, M., Delling, G., and Muller, P.K. (1995). Composition and posttranslational modification of individual collagen chains from osteosarcomas and osteofibrous dysplasias. J Cancer Res Clin Oncol 121, 413-418.

18. Salo, A.M., Cox, H., Farndon, P., Moss, C., Grindulis, H., Risteli, M., Robins, S.P., and Myllyla, R. (2008). A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. Am J Hum Genet 83, 495-503.

19. Hennet, T. (2019). Collagen glycosylation. Curr Opin Struct Biol 56, 131-138.

20. De Giorgi, F., Fumagalli, M., Scietti, L., and Forneris, F. (2021). Collagen hydroxylysine glycosylation: non-conventional substrates for atypical glycosyltransferase enzymes. Biochemical Society Transactions accepted for publication.

21. Wang, C., Luosujarvi, H., Heikkinen, J., Risteli, M., Uitto, L., and Myllyla, R. (2002a). The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. Matrix Biol 21, 559-566.

22. Guo, H.F., Bota-Rabassedas, N., Terajima, M., Leticia Rodriguez, B., Gibbons, D.L., Chen, Y., Banerjee, P., Tsai, C.L., Tan, X., Liu, X., Yu, J., Tokmina-Roszyk, M., Stawikowska, R., Fields, G.B., Miller, M.D., Wang, X., Lee, J., Dalby, K.N., Creighton, C.J., Phillips, G.N., Jr., Tainer, J.A., Yamauchi, M., and Kurie, J.M. (2021). A collagen glucosyltransferase drives lung adenocarcinoma progression in mice. Commun Biol 4, 482.

23. Savolainen, E.R., Kero, M., Pihlajaniemi, T., and Kivirikko, K.I. (1981). Deficiency of galactosylhydroxylysyl glucosyltransferase, an enzyme of collagen synthesis, in a family with dominant epidermolysis bullosa simplex. N Engl J Med 304, 197-204.

24. Ewans, L.J., Colley, A., Gaston-Massuet, C., Gualtieri, A., Cowley, M.J., Mccabe, M.J., Anand, D., Lachke, S.A., Scietti, L., Forneris, F., Zhu, Y., Ying, K., Walsh, C., Kirk, E.P., Miller, D., Giunta, C., Sillence, D., Dinger, M., Buckley, M., and Roscioli, T. (2019).

Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. J Med Genet.

25. Ruotsalainen, H., Sipila, L., Vapola, M., Sormunen, R., Salo, A.M., Uitto, L., Mercer, D.K., Robins, S.P., Risteli, M., Aszodi, A., Fassler, R., and Myllyla, R. (2006). Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. J Cell Sci 119, 625-635.

26. Schegg, B., Hulsmeier, A.J., Rutschmann, C., Maag, C., and Hennet, T. (2009). Core glycosylation of collagen is initiated by two beta(1-O) galactosyltransferases. Mol Cell Biol 29, 943-952.

27. Perrin-Tricaud, C., Rutschmann, C., and Hennet, T. (2011). Identification of domains and amino acids essential to the collagen galactosyltransferase activity of GLT25D1. PLoS One 6, e29390.

28. Sricholpech, M., Perdivara, I., Nagaoka, H., Yokoyama, M., Tomer, K.B., and Yamauchi, M. (2011). Lysyl hydroxylase 3 glucosylates galactosylhydroxylysine residues in type I collagen in osteoblast culture. J Biol Chem 286, 8846-8856.

29. Yamauchi, M., and Sricholpech, M. (2012). Lysine post-translational modifications of collagen. Essays Biochem 52, 113-133.

30. Baumann, S., and Hennet, T. (2016). Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase. J Biol Chem 291, 18514-18524.

31. Scietti, L., Chiapparino, A., De Giorgi, F., Fumagalli, M., Khoriauli, L., Nergadze, S., Basu, S., Olieric, V., Cucca, L., Banushi, B., Profumo, A., Giulotto, E., Gissen, P., and Forneris, F. (2018). Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. Nat Commun 9, 3163.

32. Wang, C., Risteli, M., Heikkinen, J., Hussa, A.K., Uitto, L., and Myllyla, R. (2002b). 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.

33. Koenig, S.N., Cavus, O., Williams, J., Bernier, M., Tonniges, J., Sucharski, H., Dew, T., Akel, M., Baker, P., Madiai, F., De Giorgi, F., Scietti, L., Faravelli, S., Forneris, F., Mohler, P.J., and Bradley, E.A. (2022). New mechanistic insights to PLOD1-mediated human vascular disease. Transl Res 239, 1-17.

34. Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41, 207-234.

35. Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr 66, 125-132.

36. Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69, 1204-1214.

37. Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2018). "STARANISO". (Cambridge, United Kingdom: Global Phasing Ltd.).

38. Mccoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr 40, 658-674.

39. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica Section D 66, 486-501.

40. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., Mccoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213-221.

41. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12-21.

42. Scietti, L., Campioni, M., and Forneris, F. (2019). SiMPLOD, a structure-integrated database of collagen lysyl hydroxylase (LH/PLOD) enzyme variants. J Bone Miner Res.

43. Lairson, L.L., Henrissat, B., Davies, G.J., and Withers, S.G. (2008). Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem 77, 521-555.

44. Gloster, T.M. (2014). Advances in understanding glycosyltransferases from a structural perspective. Curr Opin Struct Biol 28, 131-141.

45. Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. J Mol Biol 328, 307-317.

46. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42, D490-495.

47. Kivirikko, K.I., and Myllyla, R. (1979). Collagen glycosyltransferases. Int Rev Connect Tissue Res 8, 23-72.

48. Patenaude, S.I., Seto, N.O., Borisova, S.N., Szpacenko, A., Marcus, S.L., Palcic, M.M., and Evans, S.V. (2002). The structural basis for specificity in human ABO(H) blood group biosynthesis. Nat Struct Biol 9, 685-690.

49. Gomez, H., Lluch, J.M., and Masgrau, L. (2012). Essential role of glutamate 317 in galactosyl transfer by alpha3GalT: a computational study. Carbohydr Res 356, 204-208.

50. Lairson, L.L., Chiu, C.P., Ly, H.D., He, S., Wakarchuk, W.W., Strynadka, N.C., and Withers, S.G. (2004). Intermediate trapping on a mutant retaining alpha-galactosyltransferase identifies an unexpected aspartate residue. J Biol Chem 279, 28339-28344.

Chapter 4 A Fe^{2+}-dependent self-inhibited state influences the druggability of human collagen lysyl hydroxylase (LH/PLOD) enzymes.

Luigi Scietti1,#,*,a, Elisabetta Moroni2,#, Daiana Mattoteia1,#, Marco Fumagalli1, Matteo De Marco1, Lisa Negro1, Antonella Chiapparino1, Stefano A. Serapian3, Francesca De Giorgi1, Silvia Faravelli1, Giorgio Colombo3, Federico Forneris1,*

Contribution to the project: Recombinant protein production and enzymatic assay development.

ABSTRACT

Multifunctional human lysyl hydroxylases (LH/PLODs) are crucial enzymes that catalyze the hydroxyation, and ultimately glycosylation, of lysine on collagen. The requisite modifications make it possible for collagen to mature and ultimately polymerize into higher order structures in the extracellular matrix (ECM). Tumors are characterized by LH/PLOD over-expression that leads to characteristic collagen post-translational modifications recently associated with higher aggressive metastatic behavior in many solid tumors, highlighting their potential therapeutic targeting. The literature surrounding LH/PLOD has rapidly multiplied with recent structural data and high-throughput assays used to screen interesting compounds of potential drugs. Using a combination of biochemical assays and in silico methods we are working to deconvolute the bifunctional role of Fe^{2+} as both a co-factor and autoinhibitrory agent of lysyl hydroxylase activity while beginning to examine the widely used Fe^{2+} chelating agent 2,2′-bipyridyl. In a somewhat unanticipated result, we show for low concentrations of 2,2′-bipyridyl that LH activity is increased, presumably because it counteracts the inhibition from high concentrations of Fe^{2+}. Collectively, these results provide insight into the delicate balance between Fe^{2+}-dependent catalysis and self-inhibition by Fe^{2+}, also point out some notable differences between LH/PLODs and other Fe^{2+}/2-oxoglutarate dioxygenases, and suggest that a structure based strategy of developing inhibitors is potentially not feasible. These data provide clarity around the druggability of the LH/PLOD catalytic site as well as a starting point to future screening and medicinal chemistry projects.

INTRODUCTION

Collagen's supramolecular assembly in the extracellular matrix (ECM) is guided by several biosynthesized post translational modifications (PTMs). Of these, lysine (Lys) hydroxylation is critical for fibrillogenesis and accordingly establishing the ECM's physicochemical characteristics [1]. In humans, hydroxylation is the function of the lysyl hydroxylase (LH/PLOD) family: LH1/PLOD1, LH2/PLOD2, and LH3/PLOD3, which are specified by the procollagen-lysine, 2-oxoglutarate 5-dioxygenase (PLOD) gene [2]. These Fe^{2+}/2-oxoglutarate/ascorbate/O_2-dependent enzymes hydroxylate collagen Lys at the C5 position to make 5-hydroxylysine (Fig.1) [3]. An imbalance in the LCC: HLCC cross-link ratio promotes deposition of elevated and highly organized collagen fibers-pathognomonic for advanced tissue fibrosis and a well-established hallmark of cancer overall [4]. Tumor cells will use these aligned "collagen highways" to migrate towards vasculature and metastasize [5–7]. Hypoxia-dependent and independent overexpression and mislocalization of LH enzymes contributes to greater metastatic capability in many solid tumors and these enzymes are viewed as poor prognostic indicators [3–4, 8–9].Early work reported an hypoxia induced PLOD2 overexpression as a prognostic biomarker across various malignancies which included hepatocellular carcinoma [10–11], lung and colon cancers [11], glioma [12–13], oral squamous cell and endometrial carcinoma [14–15], and bone and breast metastases [6, 16–17]. Notably, inhibition of the LH2/PLOD2 isoform in kidney cancer cell lines via tumor suppressive miRNA inhibited cell migration, and invasion [18], indicating a role for LH

enzymes in promoting tumor progression. As such, it was subsequently studied, as well, that the LH1/PLOD1 and LH3/PLOD3 isoforms were also recognized as biomarker in several solid tumors. There have been reported increases in PLOD1 expression in gastrointestinal carcinoma [19], osteosarcoma [20], glioma [21–22], and bladder cancer [23]. LH3/PLOD3 has been shown be upregulated in glioma [24–25], gastric cancer [26] and colorectal cancer [27–28], it has further been suggested that it acts as a modifier to metastatic spread across tumor types [29].

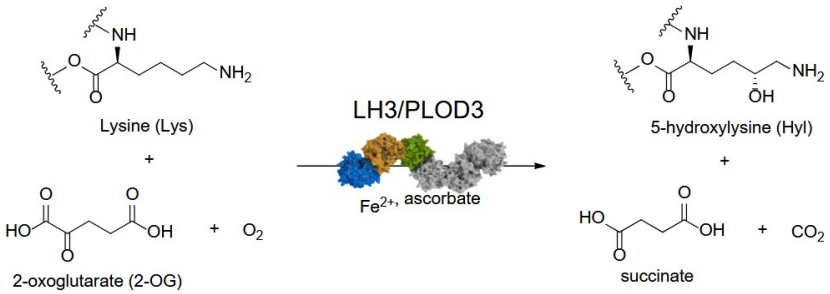


Figure 1: Reaction scheme for collagen lysine hydroxylation by LH3/PLOD3. The reaction is Fe^{2+}/2-oxoglutarate–dependent: the enzyme uses Fe^{2+} as a catalytic cofactor, O_2 and 2-oxoglutarate (2-OG) as co-substrates, and requires ascorbate to maintain the metal in the ferrous state during turnover. During lysine hydroxylation, 2-OG is converted to succinate with concomitant release of CO_2.

Similar findings for PLOD2, knockdown of PLOD3 similarly inhibits malignant behavior in renal cell carcinoma [30]. More broadly, dysregulation across the LH/PLOD family correlates with metastatic progression in many solid tumors, including hepatocellular and renal carcinoma [31–32], gliomas [33], gastric [34], and ovarian cancers [35–36]. For example, the first crystal structure of full-length LH3/PLOD3, a multifunctional enzyme that, unlike the other isoforms, catalyzes both collagen lysine hydroxylation and glycosylation, has established the configuration of LH3/PLOD3 catalytic pockets and assists in future in silico drug-design efforts by predicting drug interactions with the crystal structure [2, 37]. It has been shown, in vitro, that LH3/PLOD3 can perform sequential galactosylation and glucosylation of Hyl to produce α-[3–4]-glucosyl–β- [38]. All the LH/PLOD enzymes are classified within the Fe^{2+}/2-oxoglutarate dependent dioxygenase superfamily, a large and diverse superfamily of biocatalysts that promote oxidative transformations such as epimerization, demethylation, and hydroxylation, which all exhibit a conserved double-stranded β-helix [36]. In human LH3/PLOD3, the catalytic Fe^{2+} is coordinated by His667, Asp669, and His719 within the active site. (Fig. 2A–B). In a viral LH/PLOD homolog, the catalytic Fe^{2+} is indispensable: chelation from the active site abolishes proper folding, dimerization, and enzymatic function [39]. Foundational studies [40–42] further underscored the enzymes' metal-binding promiscuity and its effects on hydroxylase activity. Structural analyses have also revealed a putative second, non-catalytic

Fe^{2+} uniquely positioned within the LH domain; an arrangement not seen in other Fe^{2+}/2-OG dioxygenases. In human LH3/PLOD3, this metal is coordinated by His595, Asp597, Asp611, and His613, and its interactions stabilize a "capping loop" (Gly590–Glu610) that occludes the catalytic entrance and positions Arg599 directly opposite the 2-OG co-substrate, effectively mimicking the collagen Lys substrate ([36]; Fig. 2C–D). In the absence of this Fe^{2+}, the region is highly flexible [36, 39], although crystallographic stabilization effects cannot be fully excluded. The coexistence of canonical Fe^{2+}/2-OG dioxygenase features with LH/PLOD-specific elements makes this family particularly attractive for structure-based inhibitor design. Here, we integrate molecular dynamics with structure-guided mutagenesis of LH3/PLOD3, coupled to biochemical assays, to define how the capping loop regulates active-site access. These insights lay a foundation for LH/PLOD-directed drug discovery aimed at limiting cancer metastasis.

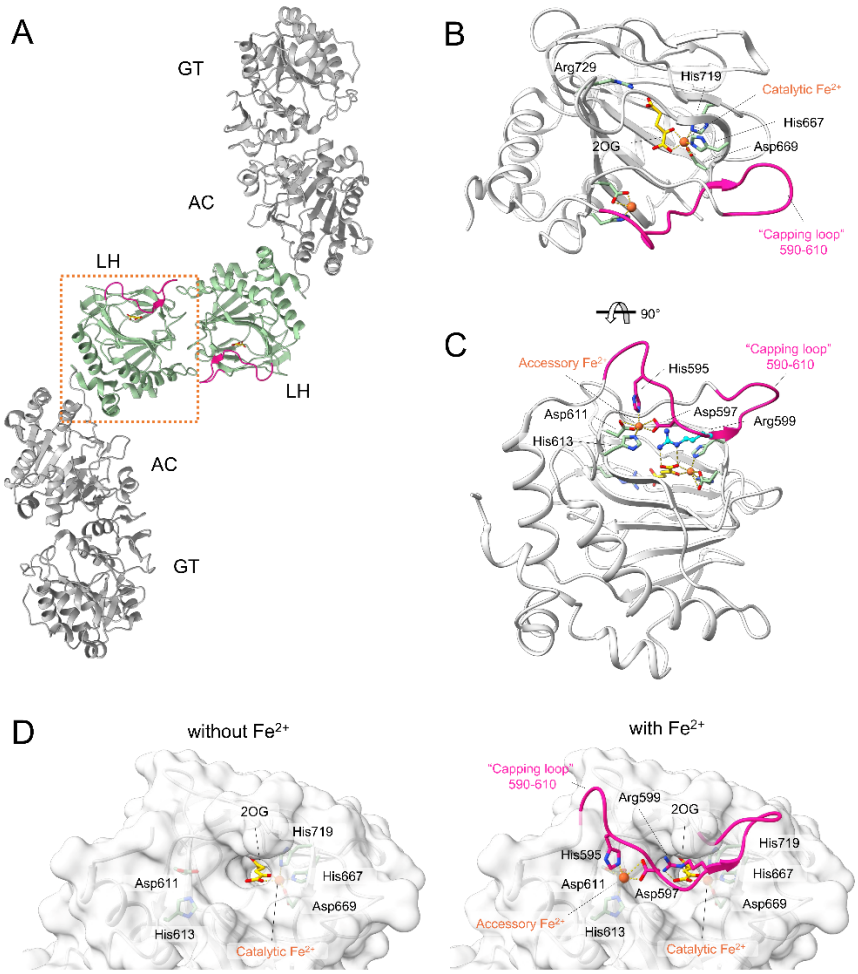


Figure 2: Structural features of the LH domain and catalytic site in LH3/PLOD3. (A) Cartoon of the dimeric, multidomain assembly of LH3/PLOD3 and labeled domains (GT, AC, LH). The LH domain (green) is at the dimer interface (in box). (B) Close-up of the LH catalytic site. The Fe^{2+} cofactor and 2-oxoglutarate (2-OG) shown with surrounding side chains illustrated as sticks. The capping loop (magenta) indicates the substrate entry path into the LH catalytic site. (C) Non-catalytic Fe^{2+} site that locks the capping loop in a self-inhibited conformation. Coordinating residues are shown as sticks; Arg599 (blue) adopts a pose's similar to the incoming collagen lysine. (D) Surface views of active site accessibility without (left) and with (right) the capping loop engaged. Catalytic residues and residues that promote the self-inhibitory state are colored as shown in B, and C.

MATERIALS AND METHODS

Chemicals: All chemicals were purchased from Sigma-Aldrich (Merck) unless specified otherwise.

MOLECULAR CLONING AND SITE-DIRECTED MUTAGENESIS

The wild-type human PLOD3 coding sequence (GenBank BC011674.2) was sourced from Source Bioscience. Oligonucleotides bearing in-frame 5′-BamHI and 3′-NotI sites were designed to subclone the ORF—lacking its N-terminal signal peptide—into pCR8, which also served as the template for subsequent manipulations. Single–amino acid substitutions were introduced using Phusion Site-Directed Mutagenesis (Invitrogen) with primers listed in Table 1. Linear, mutagenized plasmids were phosphorylated with T4 polynucleotide kinase (Invitrogen) and ligated using T4 DNA ligase (Invitrogen). All constructs were verified by Sanger sequencing prior to transfer into the pUPE.106.08 expression vector. This vector (courtesy of U-protein Express, BV) provides an N-terminal cystatin signal peptide, an N-terminal 6×His tag, and a TEV protease recognition site upstream of the in-frame BamHI–NotI cloning cassette, followed by an in-frame stop codon.

Table 1. List of oligonucleotides used to generate LH3/PLOD3 mutants.

Oligonucleotide name	**Oligonucleotide sequence**
D597A-Fw	CTTCAAGGCTGGCTGGAGGCTAC
D597A-Rv	CCTCATGCCGGCCGCCTGAC
D611A-Fw	CCATCCACATGAAGCAGGTGGGG
D611A-Rv	CCACGGTGGGCACATTCTCGTAG

EXPRESSION OR RECOMBINANT PLOD3/LH3 USING TRANSIENTLY TRANSFECTED HEK293F CELLS

Recombinant, epitope-tagged LH3/PLOD3 was expressed in HEK293F suspension cultures (Invitrogen) maintained and transfected as described in [43]. Cells were neither authenticated nor screened for mycoplasma. Transfections were performed at 1×10^6 cells mL^{-1} using polyethyleneimine (PEI; Polysciences) at 3 µg PEI per 1 µg pUPE.106.08-LH3 plasmid DNA per mL of culture. Four hours post-transfection, cultures were supplemented with 0.6% Primatone RL. Conditioned medium containing secreted LH3/PLOD3 was harvested 6 days later by centrifugation at 1,000 × g for 15 min.

PURIFICATION OF RECOMBINANT LH3/PLOD3 ENZYMES

The conditioned medium from HEK293F cultures expressing LH3/PLOD3 was clarified using a 0.8 mm syringe filter (Sartorius). The pH and ionic strength were adjusted with 5X buffer stock to 25 mM HEPES/NaOH, 500 mM NaCl, and 30 mM imidazole at pH 8.0 for recombinant LH3/PLOD3 purification on Äkta systems (GE Healthcare), following the same procedures as Scietti et al., [44]: the clarified filtrate was loaded on a 20 mL bed volume His-Prep FF column and eluted with 250 mM imidazole, then desalted on a 5 mL HiPrep Desalting FF column equilibrated at 25 mM HEPES/NaOH, 500 mM NaCl, pH 8.0. The N-terminal His tag was digested overnight at 4 °C with His-tagged TEV protease, and TEV and the cleaved tag were captured on a 5 mL HisTrap FF column. The enzyme was then concentrated to 5 mg/mL using 30 kDa MWCO Vivaspin Turbo filters (Sartorius) and subsequently purified by size-exclusion chromatography using Superdex 200 (10/300 GL for preparative runs or 5/150 GL for analytical runs) equilibrated with 25 mM HEPES/NaOH, 200 mM NaCl, pH 8.0. Fractions containing LH3/PLOD3 (confirmed by SDS-PAGE) were pooled, concentrated, and stored at −80 °C.

LH ASSAYS USING LC-MS AND ANALYSIS OF 2,2’-BIPYRIDYL EFFECTS ON ENZYMATIC ACTIVITY

Synthetic collagen peptides (China Peptides) corresponding to the sequences ARGIKGIRGFS and GIKGIKGIKGIK ([36]) were used as the substrates. Reactions contained 5 µM LH3/PLOD3, 50 µM $FeCl_2$, 100 µM 2-oxoglutarate, 500 µM ascorbate, 1 mM peptide, and 0–500 µM 2,2′-bipyridine and were incubated for 3 h at 37 °C. For LC–MS analysis, 10 µL of each reaction was diluted with 38 µL Milli-Q water and acidified with 2 µL formic acid for a total volume of 50 µL. Samples were analyzed on a UHPLC–HRMS/MS (AB Sciex) platform made up of an ExionLC AD system (column oven 40 °C; autosampler 10 °C, binary pump) and an X500B QTOF with Turbo V ion source and Twin Sprayer ESI, controlled via SCIEX OS 2.1. Peptides were separated by RP-HPLC on a Hypersil Gold C18 column (150 × 2.1 mm, 3 µm, 175 Å; Thermo Fisher Scientific) using a linear gradient from 2% to 50% solvent B over 15 minutes at a flow rate of 0.2 mL min^{-1} (solvent A: 0.1% formic acid in water; solvent B: acetonitrile with 0.1% formic acid). Spectra were collected in positive ion mode under constant conditions: ion spray voltage 4,500 V; declustering potential 100 V; curtain gas 30 psi; Ion source gas 1 (40 psi) and gas 2 (45 psi); source temperature 350 °C; collision energy 10 V. Data processing was completed in SCIEX OS 2.1. Statistical comparisons between uncoupled and coupled assay readouts were performed using paired Student’s t-tests in Prism 7 (GraphPad).

LUMINESCENCE-BASED LH ASSAYS

As in [36], reaction mixtures (5 µL total) were prepared by sequential addition of LH3/PLOD3 (0.2 mg mL^{-1}), either peptide substrate (0 - 1 mM) or gelatine (4 mg mL^{-1}; pre-solubilized at 95 °C for 10 min), ascorbate (500 µM), 2-oxoglutarate (100 µM), and $FeCl_2$ (0 - 200 µM). Reactions were incubated for 1 h at 37 °C and stopped by heating to 95 °C for

2 min before transferring to Proxiplate white 384-well plates (PerkinElmer). Succinate production was measured according to the Promega Succinate-Glo assay: 5 µL of Reagent I followed by 1 h incubation at 25 °C, then 10 µL of Reagent II and 10 min incubation at 25 °C; luminescence was read on a GloMax Discovery plate reader. All assays were conducted in triplicate. Controls were performed under the same conditions without LH3/PLOD3, 2-OG, or peptide substrate. Data was analyzed and graphed in Prism 7 (GraphPad), and paired Student's t-tests were used to compare uncoupled vs. coupled conditions.

DIFFERENTIAL SCANNING FLUORIMETRY (DSF) ASSAYS

Differential scanning fluorimetry (DSF) was performed with the wild-type LH3/PLOD3 using a Tycho NT.6 (NanoTemper). The samples (1 mg mL^{-1}) were prepared in 25 mM HEPES/NaOH, 200 mM NaCl, pH 8.0. The samples were mixed with increasing concentrations of $FeCl_2$ and 2,2′-bipyridyl. The data were then analysed and plotted using GraphPad Prism 7.

MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamics [36] edited in COOT [45]. Two systems were made: one with both Fe^{2+} sites (LH3Fe2) and one with just the catalytic Fe^{2+} site (LH3Fe1). The protonation states at physiological pH were assigned using PROPKA 3.1 (Søndergaard et al., 2011), which left a single disulfide (Cys563–Cys698); His546, His586, His643, His681, and His717 were protonated on Nε2; and His595, His613, His667, His711, and His719 were protonated on Nδ1 (which covered the histidines in both Fe^{2+} sites). The crystallographic waters were taken from 6FXR. Hydrogens and terminal caps (NH_3^+ at the N terminus and COO^- at the C terminus) were added with tleap (AmberTools 19). All simulations were performed using AMBER 18 [46] with pmemd. cuda for equilibration and production, and sander otherwise. Each system (LH3Fe2 and LH3Fe1) consisted of three independent replicas (different random seeds). A cutoff at 8.0 Å for Lennard-Jones and short-range Coulomb interactions was used while long-range electrostatics beyond the cutoff were handled with particle mesh Ewald [47]. Each production trajectory was 1 µs per replica using the NpT ensemble (2-fs timestep) at 300 K with a Langevin thermostat (collision frequency 1 ps^{-1}) and 1 atm using a Berendsen barostat [48]. The preproduction consisted of minimization (10 steps steepest descent followed by 290 steps conjugate gradient), heating (20 ps, NpT, 25→300 K; harmonic restraints on Cα atoms with k = 5.0 kcal mol^{-1} $Å^{-1}$; collision frequency 0.75 ps^{-1}, 2 fs timestep), and equilibration (1.0 ns, NpT, 300 K; collision frequency 1 ps^{-1}, 2 fs timestep). The trajectory analyses were performed with CPPTRAJ (AmberTools 19) and in-house developed scripts.

Distance Fluctuation: To investigate the role of the additional Fe^{2+} on the internal dynamics of the LH domain in LH3 we used our previously described distance-fluctuation [49]. For each MD trajectory of the two systems, we generated the DF matrix for the merged meta-

trajectory, in which the (i,j) entry reports the DF value for residue pair i–j. DF is defined as the variance of the time-dependent Cα–Cα distance

$$DF_{ij} = \langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle$$

where the angle brackets denote time averages over the trajectory. DF is invariant to overall translation and rotation and does not rely on choosing a single reference structure as for covariance analyses. The DF matrix represents the intrinsic flexibility of a protein: pairs of residues located in the same quasi-rigid unit would exhibit low DF (indicates they were highly coordinated), while the residue pairs located in different dynamic regions would have high DF.

PARAMETRIZATION AND CHARGE DERIVATION OF COFACTOR 2-OG AND FE2+ BINDING SITES

Lennard–Jones and intramolecular bonded terms for the catalytic 2-oxoglutarate (2-OG) were assigned with GAFF [50] using AmberTools antechamber and parmchk2, after adding hydrogens with reduce [46]. Partial charges and the metal–ligand (2-OG–Fe^{2+}) bonded terms were handled as detailed below. Both Fe^{2+} coordination sites were parameterized with MCPB.py [51] in combination with DFT calculations in Gaussian09 [52]. Metal–ligand bonded parameters for all coordinating residues and for 2-OG were derived by MCPB.py via the Seminario method [53], using a Hessian obtained at the B3LYP [54–55]/6-31G [51] consisted of: (i) the Fe^{2+} ion for the target site; (ii) Cβ-truncated coordinating residue side chains with the Cα–Cβ bond replaced with a shortened Cβ–H; and (iii) for the catalytic site, the full 2-OG cofactor. During geometry optimization, the Cβ atoms of all truncated residues and the carboxylate oxygens distant from the Fe^{2+} of 2-OG were constrained, and the Cβ atoms of the truncated residues and the oxygen atoms of the carboxylate of 2-OG were omitted in subsequent frequency analyses. The atomic point charges for the Fe^{2+}, all coordinating residues, and the 2-OG (the complete structure including the fragment) were generated with MCPB.py through RESP fitting [56] to electrostatic potentials (ESP) obtained at a higher-level DFT (density functional theory) on "large" site models [57] with the following DFT method: B3LYP [54–55]/6-31G(d) with def2-SV(P) for the Fe^{2+} [58]. The "large" models included Fe^{2+}, 2-OG, the full coordinating residues (with full backbone), and backbone fragments of Glu596 and His668 adjacent to His595/Asp597 and His667/Asp669, respectively; all peptide ends were capped with acetyl (N-term) and N-methyl (C-term) groups. For the Gaussian09 [52] ESP grid calculations, we used 10 concentric shells with 17 points per square Bohr; the Fe^{2+} radius was the default of MCPB.py: 1.409 Å [51]. In all DFT steps, and B3LYP/6-31G(d) testified lower energy than triplet or singlet (data not shown), we modeled the catalytic and noncatalytic sites Fe^{2+} centers in a quintet spin state. For all remaining protein residues, including the intra-residue terms for the site ligands, we described as ff14SB [59]; Na^{+} counterions were modeled with Joung–Cheatham-parameters [60] that are compatible with TIP3P water [61].

RESULTS

THE LH CATALYTIC SITE DOES NOT ACCOMMODATE COMPETITIVE INHIBITORS

Conserved catalytic-site residue networks across Fe^{2+}/2-oxoglutarate (2-OG)–dependent dioxygenases offer a rationale for structure-guided inhibitor design. Characterizing the environment surrounding the catalytic Fe^{2+} and 2-OG in these enzymes suggests that 2-OG mimetics could inhibit co-substrate binding through a competitive mechanism [57]. In human LH3/PLOD3, the catalytic Fe^{2+} is bound by His667, Asp669, and His719 as part of a DSBH fold, with additional binding to 2-OG. The site is then shielded from solvent by a flexible "loop" of 21 amino acids (the capping loop, residues 590–610; Fig. 2). To identify inhibitors of LH activity, we screened library of 2-OG analogs (Table 2) via nano-DSF and luminescence-based activity assays. None of the compounds showed detectable thermal stabilization or inhibited the conversion of 2-OG to succinate. Furthermore, an in-silico set of putative candidates that were predicted to bind the LH3/PLOD3 active site based on structural modeling also did not yield inhibition under these assay conditions.

Table 2. List of 2-OG analogs tested. For each compound, we list the chemical formula of its free-acid form. Across the panel, none showed detectable binding to LH3/PLOD3 (by DSF) or inhibition of 2-OG to succinate turnover under either uncoupled (no acceptor) or coupled (gelatin) conditions.

Compound	**Formula**
Formate	CH_2O_2
Oxalate	$C_2H_2O_4$
Malonate	$C_3H_4O_4$
Tartronate	$C_3H_4O_5$
Mesoxalate	$C_3H_2O_5$
Aminomalonate	$C_3H_5NO_4$
Fumarate	$C_4H_4O_4$
Oxalacetate	$C_4H_4O_5$
Malate	$C_4H_6O_5$
Aspartate	$C_4H_7NO_4$
Tartrate	$C_4H_6O_6$
Glutamate	$C_5H_9NO_4$
Glutarate	$C_5H_8O_4$
Acetonedicarboxylate	$C_5H_6O_5$
2-hydroxyglutarate	$C_5H_8O_5$
Adipate	$C_6H_{10}O_4$

A NON-CATALYTIC SECOND FE2+ BINDING SITE ON THE CAPPING LOOP MODULATES ACCESSIBILITY TO THE LH CATALYTIC SITE

Motivated by the striking resistance of the LH active site to inhibition by 2-OG analogues, we turned to features that distinguish this enzyme from other Fe^{2+}/2-OG dioxygenases; most notably the capping loop and its stabilized conformation under excess Fe^{2+}. This "interlocked" state likely impedes inhibitor access, making the loop's positioning a key determinant of pocket accessibility. Attempts to quantify Fe^{2+} occupancy at the two LH3 metal sites proved inconclusive, owing to multiple metal-binding sites distributed across the

LH and GT domains and the tight coordination of Fe^{2+} in the catalytic center [28-30,61]. We therefore combined in-silico analyses of loop dynamics with targeted mutagenesis in-vitro.

To assess how the non-catalytic Fe^{2+} influences the conformation of loop 590–610, we performed explicit-solvent all-atom MD on the C-terminal LH domain that mediates dimerization, comparing simulations with the second Fe^{2+} present (LH3Fe2) versus absent (LH3Fe1). For each system, three independent 1 μs replicas were run with identical protocols (Materials and Methods), differing only in Maxwell-distributed initial velocities. Visual inspection indicates both systems exhibit limited atomic fluctuations, consistent with a generally rigid domain that remains close to the crystallographic conformation, except for mobility in the gate loop and selected regions at the dimer interface. To probe how the second Fe^{2+} shapes C-terminal dynamics in LH3/PLOD3, we first mapped locally flexible regions by computing the local fluctuation (LF) metric, i.e., for each residue i, the mean variance of the Cα–Cα distances between i and its nearest neighbors ($i \pm 1$, $i \pm 2$). Comparing the two systems (LH3Fe1 vs. LH3Fe2) revealed that the gate loop (residues 593–611) in both monomers is markedly more flexible without the second Fe^{2+} (LH3Fe1) (Fig. 3A–B). A second LF peak at residues 640–646, contacts from one monomer that engage the partner's gate loop, is also more pronounced in LH3Fe1 (Fig. 3A–B). To quantify these differences, we analyzed RMSD distributions after optimal rigid-body superposition of backbone atoms from each MD snapshot onto the crystal structure, which features the gate loop in the closed conformation.

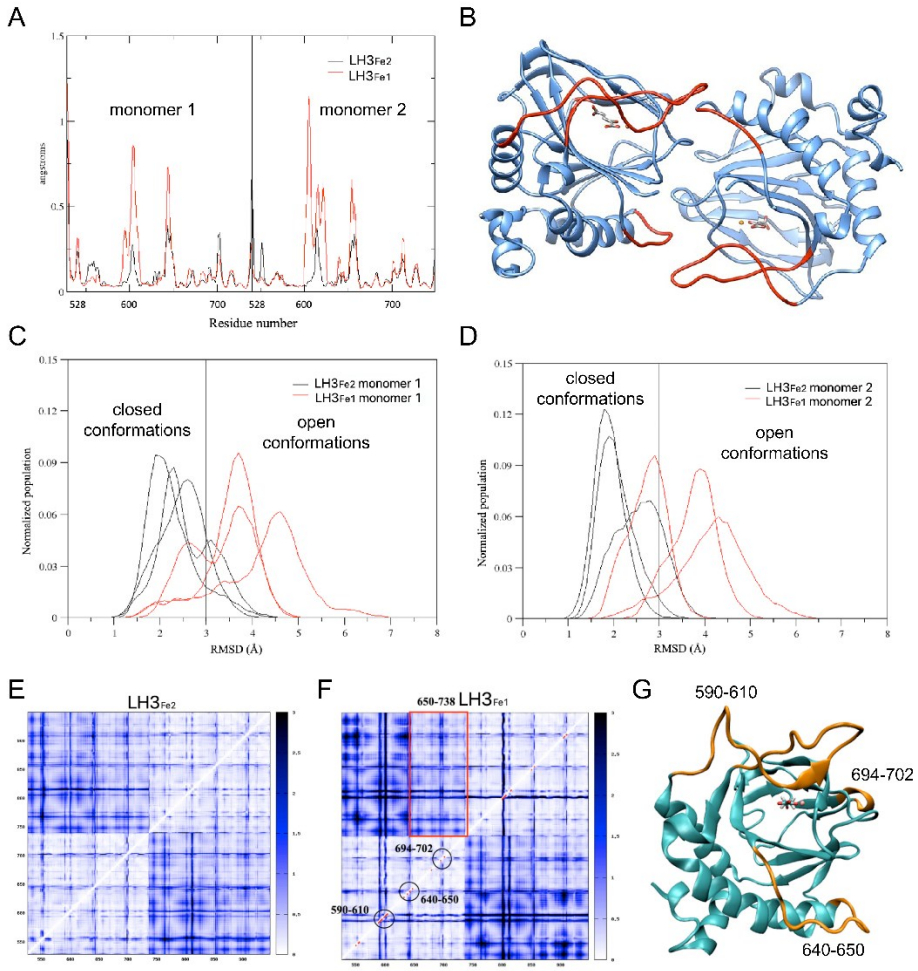


Figure 3: Computational analysis of the dynamic behavior of LH capping loop in modulating accessibility to the LH catalytic site. (A) Average mean local fluctuations (LF) per residue for the two LH domain dimers, LH3Fe1 (red) and LH3Fe2 (black); based off an average of three replica simulations. (B) Labelled LH domain dimer with LF hotspots from LH3Fe1 coloured with the gate loop residues (593 - 611) and residues 640 - 646 in red. (C-D) RMSD distributions of the gate loop (590-610) for each monomer, the LH3Fe2 (black) and LH3Fe1 (red) dimers presented here clearly demonstrate many more opening/variability without the help of a second Fe^{2+}. (E-F) Distance fluctuation (DF) maps for LH3Fe2 (E) and LH3Fe1 (F); white = coordinated (quasi-rigid), blue = less coordinated. There are red dots in F showing pairs of residues that were markedly different in DF versus LH3Fe1 (F-test). The red rectangle denotes the pair of residues 650 - 738 of one of the monomers (β-sheet core) that demonstrated coordinated movements with the partner monomer. (G) LH domain and residues that had significant motion difference between systems shown in orange colours; demonstrating the second-Fe^{2+}-dependent stabilization that occurred directed proximal to the gate loop and dimer interface.

We analyzed RMSD distributions for the gate loop (residues 590–610) in both monomers across individual trajectories (Fig. 3C–D). Based on visual inspection and overlay of representative MD snapshots, we set a 3 Å RMSD threshold to distinguish "closed" from "open" conformations. In LH3Fe2, the loop is predominantly stabilized in the closed state, whereas removal of the second Fe^{2+} shifts the LH3Fe1 population toward the open state, although both systems sample both conformations.

To see how the second Fe^{2+} modulates intradomain coordination, we calculated distance fluctuation (DF) maps on the meta-trajectories we created when concatenating the three replicates per condition. DF, the variance of the inter-residue Cα–Cα distance $d_{(ij)}$ over time, was calculated for all pairs of residues. The resulting matrices (Fig. 3E–F) report intrinsic flexibility and coordinated motion within the LH domains, with low DF values indicating coherently moving residue pairs. Contrasting LH3Fe1 and LH3Fe2 reveals coordination changes attributable to the presence or absence of the second Fe^{2+}.Overall, the two matrices are highly similar, exhibiting largely overlapping regions of low and high fluctuations. In both systems, intradomain motions show small atomic displacements, consistent with a generally rigid architecture; the largest DF values occur in poorly ordered segments, reflecting their inherent flexibility. Interdomain motions display greater fluctuations than intradomain motions. Notably, in LH3Fe1, residues 650–738 of one monomer (the β-sheet core) exhibit more coordinated movement with the opposing monomer than does the rest of the protein (Fig. 3F). Quantifying the differences in the DF maps provided by F-test statistics (with LH3Fe2 as the reference) indicated that many residues with red dots in LH3Fe1 (Fig. 3F) are significantly different from the residue in LH3Fe2. The most extensive differences localized in the gate loop region (590-610), and the residues 640-650 that link an α -helix (618-639) to a β -sheet core (and which make up part of the dimer interface), and further into residues 694-702 that show elevated DF in LH3Fe1 (Fig. 3F-G). To assess how the second Fe^{2+} modulates access to the catalytic pocket via capping-loop positioning, we computed the solvent-accessible surface area (SASA) of active-site residues (those within 6 Å of 2-OG and Fe^{2+}) over all MD frames using VMD. In both monomers, SASA distributions for LH3Fe1 are shifted to slightly higher values relative to LH3Fe2, indicating that absence of the second Fe^{2+} biases the loop toward more open conformations (Fig.4). Nonetheless, both systems sample open and closed states, consistent with the RMSD analysis. To resolve how the second Fe^{2+} tunes capping-loop dynamics, we quantified the time–persistence of (i) native contacts and (ii) newly formed interactions between loop 590–610 and the rest of the protein. Most native contacts persisted in both systems (Supp. Fig. 2A–B), but several were weakened in LH3Fe1 relative to LH3Fe2. Consistent with the loop's greater mobility, LH3Fe1 failed to establish any new stable contacts, whereas in LH3Fe2 the loop formed persistent interactions with residues 643–648 of the opposite monomer (Supp. Fig. 2C). In the crystal structure obtained under excess Fe^{2+}, Arg599 from the capping loop forms a salt bridge with 2-OG, mimicking the positioning of the collagen lysine substrate. Tracking Arg599–2-OG hydrogen/salt-bridge contacts across the MD trajectories showed high occupancy in both systems, with a slight increase in LH3Fe2 (Supp. Fig. 2D–F). Together, these analyses indicate that the second Fe^{2+} stabilizes additional inter-subunit contacts, maintains native interactions, and modestly increases Arg599–2-OG engagement; pushing the LH domain toward a more closed, interlocked and less mobile catalytic environment. To

understand how the second Fe^{2+} site regulates catalysis and access to the active site, we designed the LH3/PLOD3 variants Asp597Ala (a capping-loop ligand that is on the surface Fe^{2+}) and Asp611Ala (a counterpart on the Fe^{2+}-coordination platform; Fig. 2C–D). Both mutants expressed and purified cleanly, with yields, stability, and oligomeric state comparable to wild type (Suppl. Fig. 1). In the enzymatic assays both mutants had a minor increase in activity (~1.5×, with respect to wild type; Fig. 5A, right panel). Both Asp597Ala and Asp611Ala had some intriguing and significant increases (~3–5× wild type activity) when the reaction was uncoupled (there was no acceptor substrate, uncoupled turnover of 2-OG to succinate; Fig. 5A left panel).

We next examined Fe^{2+} dependence. High [Fe^{2+}] strongly dampened wild-type uncoupled activity, whereas both Asp597Ala and Asp611Ala were markedly less inhibited (Fig. 5B, left). In contrast, Fe^{2+} modulation of coupled activity was essentially unchanged by these substitutions (Fig. 5B, right). Together with our MD results, these data support a model in which binding of a non-catalytic surface Fe^{2+} stabilizes a closed, self-inhibited conformation of the capping loop that limits acceptor-independent 2-OG turnover, while leaving acceptor-driven catalysis largely intact.

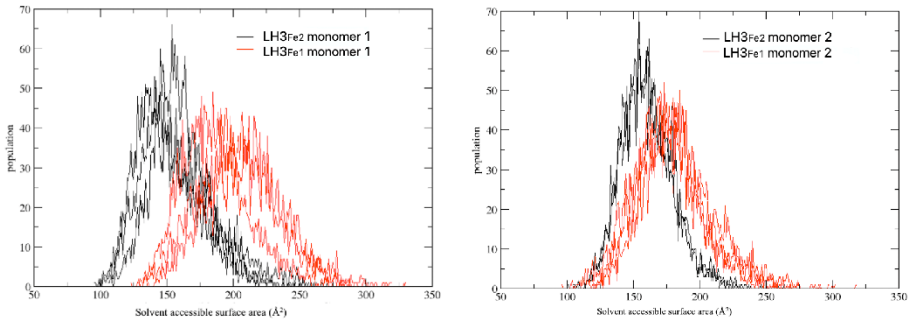


Figure 4: Distribution of the solvent accessible surface area calculated for the three replicate simulations. Simulations were analyzed independently for each system (LH3Fe1, red; LH3Fe2, black) and for each monomer, displayed in the left and right panels, respectively.

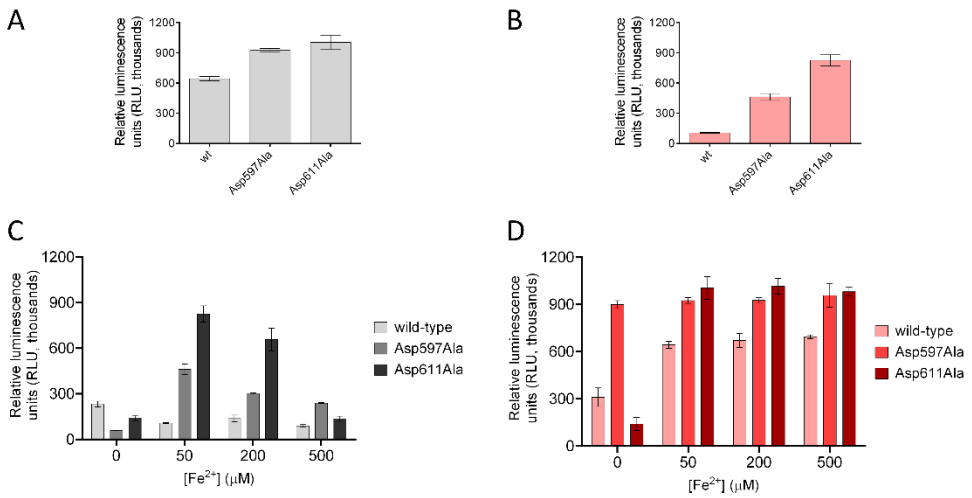


Figure 5: Comparison of enzymatic activity of LH3/PLOD3 wild-type and mutants affecting the non-catalytic second Fe2+ binding site. (A) Luminescence readout of succinate production in uncoupled assays (no acceptor; left) and coupled assays (gelatin acceptor; right) for wild type LH3 and mutants. Bars show mean ± SD (n = 3). Statistics: two-tailed Student's t-test versus wild type; ns, not significant;, $P < 0.05$;, $P < 0.01$;, $P < 0.001$; , $P < 0.0001$. (B) Uncoupled (left) and coupled (right) activities had increasing Fe^{2+} concentrations evaluated as in panel A (mean ± SD, n = 3; same significance levels).

FE2+ CHELATING AGENTS PRODUCE UNEXPECTED EFFECTS ON LH3/PLOD3 ENZYMATIC ACTIVITY

Driven by the unique proximity of two Fe2+ sites in LH/PLODs, with competing consequences on catalysis, we wondered whether small metal-chelating molecules could regulate substrate processing. Despite extensive screening yielding apparent hits, genuine inhibitors of lysyl-hydroxylase activity have yet to be identified. The only compound available is 2,2'-bipyridine (BPY), a preferential non-competitive broad-spectrum chelator composed of two nitrogenated pyridyl rings, assumed to sequester the catalytic Fe2+ in the LH domain [33, 37, 57, 70].To define the effects of BPY on the human LH3/PLOD3, we initially tried concentration response ligand-binding measurements with a luminescence assay, but BPY interfered with the Mg2+ dependent luciferase readout. So, we used a direct mass-spectrometry workflow on synthetic collagen peptides, where we detected

hydroxylysine formation by quantifying the ratio of the non-hydroxylated to hydroxylated Lys. The MS method was useful because it enabled straightforward, dose-dependent quantitation of BPY-mediated inhibition (Figs 6 A-B). Absence of BPY followed collagen peptide hydroxylation as prior literature [61] has documented. With BPY having a 3:1 Fe^{2+} stoichiometry for chelation and a 40 μM Fe_{20}-EC_{50} of at 20 μM Fe^{2+} [70] we expected that we would see a decline in LH activity that would infer inhibition by BPY at an initial ratio of [Fe^{2+}]:[BPY] of roughly 2:1. Rather than inhibition we found that BPY up to ~25 μM increased the Hyl/Lys ratio, i.e., increased catalysis, while we only observed inhibition at higher BPY (Fig. 6B-C). Since standard assays use 50 μM Fe^{2+} we postulated that low BPY was preferentially stripping Fe^{2+} from the non-catalytic capping-loop site (so that self-inhibition was being removed) while higher BPY was chelating the catalytic Fe^{2+} and inhibiting turnover. To examine this, we co-varied [Fe^{2+}] and [BPY] - LH3/PLOD3 activity always peaked at 2:1 [Fe^{2+}]: [BPY] (Fig. 6D; Supplementary Fig. 3) and peaks had narrow maximum and were restricted due to the opposing consequences of disinhibition by removing capping-loop Fe^{2+} and depletion of active-site Fe^{2+}. These findings caution against advancing BPY as a straightforward lead for LH/PLOD inhibition, as its concentration-dependent, bidirectional action can paradoxically activate the enzyme by unlocking the Fe^{2+}-dependent capping loop.

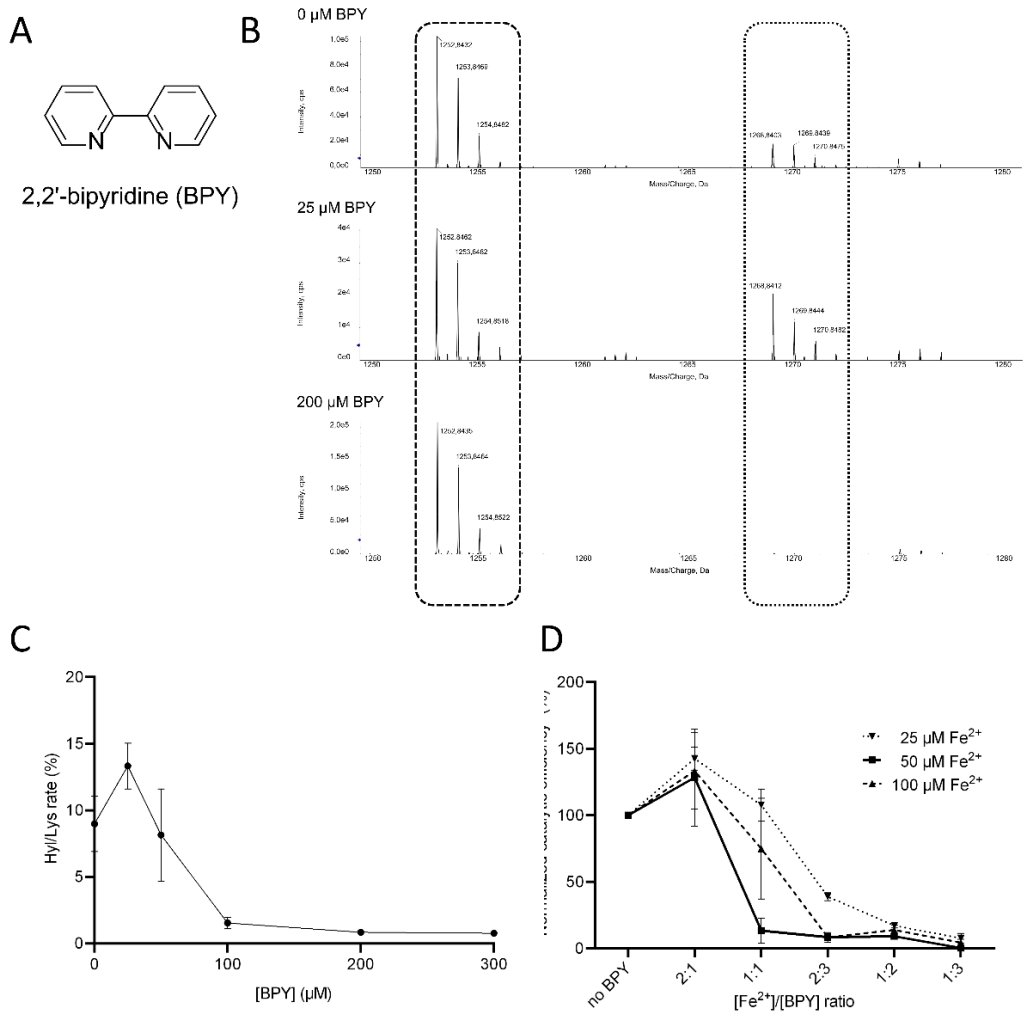


Figure 6: Evaluation of the effect of 2,2'-bipyridyl (BPY) on LH3/PLOD3 enzymatic activity. (A) Chemical structure of 2,2′-bipyridine (BPY). (B) Direct MS readout of Lys→Hyl conversion in synthetic peptides, illustrating unmodified Lys peaks (dashed box) and additional Hyl peaks upon reaction (dotted box). (C) LH3/PLOD3 activity versus BPY concentration: low BPY increases the Hyl/Lys signal (indicating activation) while high BPY produces the inhibition that we would expect, and which is consistent with chelation of catalytic Fe^{2+}. Points are mean ± SD (n = 3). Statistics: two-tailed Student's t-test comparing $+Fe^{2+}$ vs $-Fe^{2+}$ for each BPY concentration; ns, $P \geq 0.05$;, $P < 0.05$; , $P < 0.01$; , $P < 0.001$; , $P < 0.0001$. (D) Enzymatic activity as a function of the [BPY]/[Fe^{2+}] ratio, showing that the low-BPY activation depends on residual catalytic Fe^{2+} availability. Data are mean ± SD (n = 3). One-way ANOVA across different absolute Fe^{2+} concentrations at the same ratio shows no significant differences; additional pairwise statistics are reported in Supplementary Fig. 3.

DISCUSSION

Human LH/PLOD enzymes have emerged as compelling cancer targets: they drive collagen fibrillogenesis within the tumour microenvironment and correlate with metastatic risk. Given the link between metastasis and excessive Lys→Hyl conversion, selective LH/PLOD inhibitors are highly desirable. Although the full-length human LH3/PLOD3 structure now provides an atomic template [61], no inhibitors exist to date. Here, we conducted a small-molecule screen guided by structural and mechanistic insights from related Fe^{2+}/2-oxoglutarate dioxygenases and uncovered additional, system-specific hurdles, underscoring the need for extra care in designing truly selective LH/PLOD inhibitors. Even though there was substantial overlap between the LH3 / PLOD3 active site and other Fe^{2+} / 2-oxoglutarate (2-OG) dioxygenases (many of which are susceptible to 2-OG–mimetic inhibitors), with our 2-OG-competitive inhibitors and structure-guided libraries, we found no viable LH / PLOD hits. Molecular dynamics (MD) simulated and justified our speculation of the resistance: an excess of Fe^{2+} induces a capping-loop conformation that "self-locks" the catalytic pocket, obstructing ligands access. Previous structural studies demonstrated that the capping-loop of PLOD3 is organized in a unique fashion by an Fe^{2+} center complexed by His595, Asp611, and His613 on the surface of LH and Asp597 in the capping-loop. In support of this characterization, and remarkably, alanine substitution of either Asp611 or Asp597 induced large improvements in uncoupled 2-OG→succinate turnover, indicating increased accessibility of the pocket and providing evidence for a physiological, Fe^{2+}-dependent, autoinhibited state mediated by the capping loop.

Our findings show a very narrow Fe^{2+} concentration working range for LH/PLOD enzymes where both too little Fe^{2+} means loss of catalysis, as there is no co-factor for the active site; and too much Fe^{2+} means the LH domain capping loop is locked in an allosteric self-inhibited state. This behaviour is unique to LH/PLOD enzymes when placed beside the Fe^{2+}/2-oxoglutarate dioxygenases and explains the seemingly paradoxical effects of metal chelators on LH/PLOD activity. We observed that low concentrations of 2,2'-bipyridine (BPY), enhanced catalysis instead of inhibited activity. This is understandable, as BPY preferentially scraps the non-catalytic surface Fe^{2+}, which is what is preventing opening of the capping loop. The force of enhancement was proportional to the Fe^{2+}: BPY ratio, which demonstrated that processing by the activity chelator really changed the weighting factor between the interaction of releasing the interlocked loop and stripping of the catalytic Fe^{2+}.Therefore, our findings open our eyes to view substrate processing and to design inhibitors. The interesting stability of the resulting capping loop conformations with the allosteric position of the non-catalytic Fe^{2+} from your β-sheet interaction with Arg599 and substrate mimic conformation, proposes that productive turnover could be even more involved and the long range (and possibly across the LH domain surface) release of the second non-catalytic Fe^{2+}. Given the extended geometry of collagen chains, such allosteric triggering is plausible and merits investigation, including potential roles of metal ions bound to collagen during and after lysine modification. Finally, from a drug/discovery context, the implication of a metal-dependent autoinhibited state adds another level of complexity. It will most likely necessitate a sophisticated integrated approach where effective inhibition is required to first disengaging the interlocked loop from the capping position; and then subsequently blocking catalysis, and either competing with 2-OG or the side chain of the lysine which must be modified.

REFERENCES

1. Baek, J.H., Yun, H.S., Kwon, G.T., Lee, J., Kim, J.Y., Jo, Y., Cho, J.M., Lee, C.W., Song, J.Y., Ahn, J., Kim, J.S., Kim, E.H., and Hwang, S.G. (2019). PLOD3 suppression exerts an anti-tumor effect on human lung cancer cells by modulating the PKC-delta signaling pathway. Cell Death Dis 10, 156.

2. Bayly, C.I., Cieplak, P., Cornell, W., and Kollman, P.A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. The Journal of Physical Chemistry 97, 10269-10280.

3. Becke, A.D. (1993). Density-functional thermochemistry. III. The role of exact exchange. The Journal of Chemical Physics 98, 5648-5652.

4. Berendsen, H.J.C., Postma, J.P.M., Gunsteren, W.F.V., Dinola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics 81, 3684-3690.

5. Besler, B.H., Merz Jr., K.M., and Kollman, P.A. (1990). Atomic charges derived from semiempirical methods. Journal of Computational Chemistry 11, 431-439.

6. Blanco, M.A., Leroy, G., Khan, Z., Aleckovic, M., Zee, B.M., Garcia, B.A., and Kang, Y. (2012). Global secretome analysis identifies novel mediators of bone metastasis. Cell Res 22, 1339-1355.

7. Case, D.A., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham, T.E., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Ghoreishi, D., Gilson, M.K., Gohlke, H., Götz, A.W., Greene, D., Harris, R., Homeyer, N., Huang, Y., Izadi, S., Kovalenko, A., Kurtzman, T., Lee, T.S., Le Grand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D., Merz, K.M., Miao, Y., Monard, G., Nguyen, C., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D.R., Roitberg, A.E., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C.L., Smith, J., Salomon-Ferrer, R., Swails, J.M., Walker, R.C., Wang, J., Wei, H., Wolf, R.M., Wu, X., Xiao, L., York, D.M., and Kollman, P.A. (2018). "AMBER 2018". University of California, San Francisco).

8. Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. J Comput Chem 26, 1668-1688.

9. Chen, Y., Terajima, M., Yang, Y., Sun, L., Ahn, Y.H., Pankova, D., Puperi, D.S., Watanabe, T., Kim, M.P., Blackmon, S.H., Rodriguez, J., Liu, H., Behrens, C., Wistuba, Ii, Minelli, R., Scott, K.L., Sanchez-Adams, J., Guilak, F., Pati, D., Thilaganathan, N., Burns, A.R., Creighton, C.J., Martinez, E.D., Zal, T., Grande-Allen, K.J., Yamauchi, M., and Kurie, J.M. (2015). Lysyl hydroxylase 2 induces a collagen cross-link switch in tumor stroma. J Clin Invest 125, 1147-1162.

10. Chen, Y.L., Guo, H.F., Terajima, M., Banerjee, P., Liu, X., Yu, J., Momin, A.A., Katayama, H., Hanash, S.M., Burns, A.R., Fields, G.B., Yamauchi, M., and Kurie, J.M. (2016). Lysyl Hydroxylase 2 Is Secreted by Tumor Cells and Can Modify Collagen in the Extracellular Space. J Biol Chem 291, 25799-25808.

11. Clifton, I.J., Mcdonough, M.A., Ehrismann, D., Kershaw, N.J., Granatino, N., and Schofield, C.J. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. J Inorg Biochem 100, 644-669.

12. Costas, M., Mehn, M.P., Jensen, M.P., and Que, L., Jr. (2004). Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. Chem Rev 104, 939-986.

13. Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98, 10089-10092.

14. De Giorgi, F., Fumagalli, M., Scietti, L., and Forneris, F. (2021). Collagen hydroxylysine glycosylation: non-conventional substrates for atypical glycosyltransferase enzymes. Biochemical Society Transactions accepted for publication.

15. Deng, X., Pan, Y., Yang, M., Liu, Y., and Li, J. (2021). PLOD3 Is Associated with Immune Cell Infiltration and Genomic Instability in Colon Adenocarcinoma. Biomed Res Int 2021, 4714526.

16. Devkota, A.K., Veloria, J.R., Guo, H.F., Kurie, J.M., Cho, E.J., and Dalby, K.N. (2019). Development of a High-Throughput Lysyl Hydroxylase (LH) Assay and Identification of Small-Molecule Inhibitors against LH2. SLAS Discov 24, 484-491.

17. Du, H., Chen, Y., Hou, X., Huang, Y., Wei, X., Yu, X., Feng, S., Wu, Y., Zhan, M., Shi, X., Lin, S., Lu, L., Yuan, S., and Sun, L. (2017a). PLOD2 regulated by transcription factor FOXA1 promotes metastasis in NSCLC. Cell Death Dis 8, e3143.

18. Du, H., Pang, M., Hou, X., Yuan, S., and Sun, L. (2017b). PLOD2 in cancer research. Biomed Pharmacother 90, 670-676.

19. Eisinger-Mathason, T.S., Zhang, M., Qiu, Q., Skuli, N., Nakazawa, M.S., Karakasheva, T., Mucaj, V., Shay, J.E., Stangenberg, L., Sadri, N., Pure, E., Yoon, S.S., Kirsch, D.G., and Simon, M.C. (2013). Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. Cancer Discov 3, 1190-1205.

20. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica Section D 66, 486-501.

21. Ewans, L.J., Colley, A., Gaston-Massuet, C., Gualtieri, A., Cowley, M.J., Mccabe, M.J., Anand, D., Lachke, S.A., Scietti, L., Forneris, F., Zhu, Y., Ying, K., Walsh, C., Kirk, E.P., Miller, D., Giunta, C., Sillence, D., Dinger, M., Buckley, M., and Roscioli, T. (2019).

Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. J Med Genet 56, 629-638.

22. Faravelli, S., Campioni, M., Palamini, M., Canciani, A., Chiapparino, A., and Forneris, F. (2021). Optimized Recombinant Production of Secreted Proteins Using Human Embryonic Kidney (HEK293) Cells Grown in Suspension. Bio-protocol 11, e3998.

23. Flashman, E., and Schofield, C.J. (2007). The most versatile of all reactive intermediates? Nat Chem Biol 3, 86-87.

24. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, Ö., Foresman, J.B., Ortiz, J.V., Cioslowski, J., and Fox, D.J. (2009). "Gaussian 09 Revision A.2".).

25. Gilkes, D.M., Bajpai, S., Wong, C.C., Chaturvedi, P., Hubbi, M.E., Wirtz, D., and Semenza, G.L. (2013). Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. Mol Cancer Res 11, 456-466.

26. Gkretsi, V., and Stylianopoulos, T. (2018). Cell Adhesion and Matrix Stiffness: Coordinating Cancer Cell Invasion and Metastasis. Front Oncol 8, 145.

27. Gong, S., Duan, Y., Wu, C., Osterhoff, G., Schopow, N., and Kallendrusch, S. (2021). A Human Pan-Cancer System Analysis of Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenase 3 (PLOD3). Int J Mol Sci 22.

28. Guo, H.F., Bota-Rabassedas, N., Terajima, M., Leticia Rodriguez, B., Gibbons, D.L., Chen, Y., Banerjee, P., Tsai, C.L., Tan, X., Liu, X., Yu, J., Tokmina-Roszyk, M., Stawikowska, R., Fields, G.B., Miller, M.D., Wang, X., Lee, J., Dalby, K.N., Creighton, C.J., Phillips, G.N., Jr., Tainer, J.A., Yamauchi, M., and Kurie, J.M. (2021a). A collagen glucosyltransferase drives lung adenocarcinoma progression in mice. Commun Biol 4, 482.

29. Guo, H.F., Tsai, C.L., Terajima, M., Tan, X., Banerjee, P., Miller, M.D., Liu, X., Yu, J., Byemerwa, J., Alvarado, S., Kaoud, T.S., Dalby, K.N., Bota-Rabassedas, N., Chen, Y., Yamauchi, M., Tainer, J.A., Phillips, G.N., Jr., and Kurie, J.M. (2018). Pro-metastatic collagen lysyl hydroxylase dimer assemblies stabilized by Fe(2+)-binding. Nat Commun 9, 512.

30. Guo, T., Gu, C., Li, B., and Xu, C. (2021b). PLODs are overexpressed in ovarian cancer and are associated with gap junctions via connexin 43. Lab Invest 101, 564-569.

31. Hausinger, R.P. (2004). FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. Crit Rev Biochem Mol Biol 39, 21-68.

32. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph 14, 33-38, 27-38.

33. Ikeda, H., Ogata, I., and Fujiwara, K. (1994). Evidence that impaired intracellular collagen synthesis reduces proliferation in cultured rat hepatocytes. Biochem Biophys Res Commun 200, 1701-1707.

34. Jiang, H., Guo, W., Yuan, S., and Song, L. (2020). PLOD1 Is a Prognostic Biomarker and Mediator of Proliferation and Invasion in Osteosarcoma. Biomed Res Int 2020, 3418398.

35. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 79, 926-935.

36. Joung, I.S., and Cheatham, T.E. (2008). Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. The Journal of Physical Chemistry B 112, 9020-9041.

37. Jover, E., Silvente, A., Marin, F., Martinez-Gonzalez, J., Orriols, M., Martinez, C.M., Puche, C.M., Valdes, M., Rodriguez, C., and Hernandez-Romero, D. (2018). Inhibition of enzymes involved in collagen cross-linking reduces vascular smooth muscle cell calcification. FASEB J 32, 4459-4469.

38. Kivirikko, K.I., and Prockop, D.J. (1967). ENZYMATIC HYDROXYLATION OF PROLINE AND LYSINE IN PROTOCOLLAGEN[*]. Proceedings of the National Academy of Sciences 57, 782-789.

39. Kurozumi, A., Kato, M., Goto, Y., Matsushita, R., Nishikawa, R., Okato, A., Fukumoto, I., Ichikawa, T., and Seki, N. (2016). Regulation of the collagen cross-linking enzymes LOXL2 and PLOD2 by tumor-suppressive microRNA-26a/b in renal cell carcinoma. Int J Oncol 48, 1837-1846.

40. Lee, C., Yang, W., and Parr, R.G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B 37, 785-789.

41. Levental, K.R., Yu, H., Kass, L., Lakins, J.N., Egeblad, M., Erler, J.T., Fong, S.F., Csiszar, K., Giaccia, A., Weninger, W., Yamauchi, M., Gasser, D.L., and Weaver, V.M. (2009). Matrix crosslinking forces tumor progression by enhancing integrin signaling. Cell 139, 891-906.

42. Li, G., Wang, X., and Liu, G. (2021). PLOD2 Is a Potent Prognostic Marker and Associates with Immune Infiltration in Cervical Cancer. Biomed Res Int 2021, 5512340.

43. Li, P., and Merz, K.M., Jr. (2016). MCPB.py: A Python Based Metal Center Parameter Builder. J Chem Inf Model 56, 599-604.

44. Li, S.S., Lian, Y.F., Huang, Y.L., Huang, Y.H., and Xiao, J. (2020). Overexpressing PLOD family genes predict poor prognosis in gastric cancer. J Cancer 11, 121-131.

45. Loenarz, C., and Schofield, C.J. (2008). Expanding chemical biology of 2-oxoglutarate oxygenases. Nat Chem Biol 4, 152-156.

46. Loenarz, C., and Schofield, C.J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. Trends Biochem Sci 36, 7-18.

47. Loncharich, R.J., Brooks, B.R., and Pastor, R.W. (1992). Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. Biopolymers 32, 523-535.

48. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. Journal of Chemical Theory and Computation 11, 3696-3713.

49. Martinez, S., and Hausinger, R.P. (2015). Catalytic Mechanisms of Fe (II)- and 2-Oxoglutarate-dependent Oxygenases. J Biol Chem 290, 20702-20711.

50. Moroni, E., Agard, D.A., and Colombo, G. (2018). The Structural Asymmetry of Mitochondrial Hsp90 (Trap1) Determines Fine Tuning of Functional Dynamics. J Chem Theory Comput 14, 1033-1044.

51. Morra, G., Potestio, R., Micheletti, C., and Colombo, G. (2012). Corresponding functional dynamics across the Hsp90 Chaperone family: insights from a multiscale analysis of MD simulations. PLoS Comput Biol 8, e1002433.

52. Myllylä, R., Schubotz, L.M., Weser, U., and Kivirikko, K.I. (1979). Involvement of superoxide in the prolyl and lysyl hydroxylase reactions. Biochemical and Biophysical Research Communications 89, 98-102.

53. Noda, T., Yamamoto, H., Takemasa, I., Yamada, D., Uemura, M., Wada, H., Kobayashi, S., Marubashi, S., Eguchi, H., Tanemura, M., Umeshita, K., Doki, Y., Mori, M., and Nagano, H. (2012). PLOD2 induced under hypoxia is a novel prognostic factor for hepatocellular carcinoma after curative resection. Liver Int 32, 110-118.

54. Pankova, D., Chen, Y., Terajima, M., Schliekelman, M.J., Baird, B.N., Fahrenholtz, M., Sun, L., Gill, B.J., Vadakkan, T.J., Kim, M.P., Ahn, Y.H., Roybal, J.D., Liu, X., Parra

Cuentas, E.R., Rodriguez, J., Wistuba, Ii, Creighton, C.J., Gibbons, D.L., Hicks, J.M., Dickinson, M.E., West, J.L., Grande-Allen, K.J., Hanash, S.M., Yamauchi, M., and Kurie, J.M. (2016). Cancer-Associated Fibroblasts Induce a Collagen Cross-link Switch in Tumor Stroma. Mol Cancer Res 14, 287-295.

55. Provenzano, P.P., Eliceiri, K.W., Campbell, J.M., Inman, D.R., White, J.G., and Keely, P.J. (2006). Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med 4, 38.

56. Puistola, U., Turpeenniemi-Hujanen, T.M., Myllyla, R., and Kivirikko, K.I. (1980). Studies on the lysyl hydroxylase reaction. II. Inhibition kinetics and the reaction mechanism. Biochim Biophys Acta 611, 51-60.

57. Rose, N.R., Mcdonough, M.A., King, O.N., Kawamura, A., and Schofield, C.J. (2011). Inhibition of 2-oxoglutarate dependent oxygenases. Chem Soc Rev 40, 4364-4397.

58. Saito, T., Uzawa, K., Terajima, M., Shiiba, M., Amelio, A.L., Tanzawa, H., and Yamauchi, M. (2019). Aberrant Collagen Cross-linking in Human Oral Squamous Cell Carcinoma. J Dent Res 98, 517-525.

59. Salomon-Ferrer, R., Gotz, A.W., Poole, D., Le Grand, S., and Walker, R.C. (2013). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem Theory Comput 9, 3878-3888.

60. Sato, K., Parag-Sharma, K., Terajima, M., Musicant, A.M., Murphy, R.M., Ramsey, M.R., Hibi, H., Yamauchi, M., and Amelio, A.L. (2021). Lysyl hydroxylase 2-induced collagen cross-link switching promotes metastasis in head and neck squamous cell carcinomas. Neoplasia 23, 594-606.

61. Scietti, L., Chiapparino, A., De Giorgi, F., Fumagalli, M., Khoriauli, L., Nergadze, S., Basu, S., Olieric, V., Cucca, L., Banushi, B., Profumo, A., Giulotto, E., Gissen, P., and Forneris, F. (2018). Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. Nat Commun 9, 3163.

62. Scietti, L., and Forneris, F. (2020). "Full-Length Human Collagen Lysyl Hydroxylases," in Encyclopedia of Inorganic and Bioinorganic Chemistry, ed. R.A. Scott. Wiley), 1-12.

63. Seminario, J.M. (1996). Calculation of intramolecular force fields from second-derivative tensors. International Journal of Quantum Chemistry 60, 1271-1277.

64. Shi, J., Bao, M., Wang, W., Wu, X., Li, Y., Zhao, C., and Liu, W. (2021). Integrated Profiling Identifies PLOD3 as a Potential Prognostic and Immunotherapy Relevant Biomarker in Colorectal Cancer. Front Immunol 12, 722807.

65. Sondergaard, C.R., Olsson, M.H., Rostkowski, M., and Jensen, J.H. (2011). Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J Chem Theory Comput 7, 2284-2295.

66. Song, Y., Zheng, S., Wang, J., Long, H., Fang, L., Wang, G., Li, Z., Que, T., Liu, Y., Li, Y., Zhang, X., Fang, W., and Qi, S. (2017). Hypoxia-induced PLOD2 promotes proliferation, migration and invasion via PI3K/Akt signaling in glioma. Oncotarget.

67. Tian, L., Zhou, H., Wang, G., Wang, W.Y., Li, Y., and Xue, X. (2021). The relationship between PLOD1 expression level and glioma prognosis investigated using public databases. PeerJ 9, e11422.

68. Tsai, C.K., Huang, L.C., Tsai, W.C., Huang, S.M., Lee, J.T., and Hueng, D.Y. (2018). Overexpression of PLOD3 promotes tumor progression and poor prognosis in gliomas. Oncotarget 9, 15705-15720.

69. Van Der Slot, A.J., Zuurmond, A.M., Van Den Bogaerdt, A.J., Ulrich, M.M., Middelkoop, E., Boers, W., Karel Ronday, H., Degroot, J., Huizinga, T.W., and Bank, R.A. (2004). Increased formation of pyridinoline cross-links due to higher telopeptide lysyl hydroxylase levels is a general fibrotic phenomenon. Matrix Biol 23, 251-257.

70. Vasta, J.D., and Raines, R.T. (2015). Selective inhibition of prolyl 4-hydroxylases by bipyridinedicarboxylates. Bioorg Med Chem 23, 3081-3090.

71. Wan, J., Qin, J., Cao, Q., Hu, P., Zhong, C., and Tu, C. (2020). Hypoxia-induced PLOD2 regulates invasion and epithelial-mesenchymal transition in endometrial carcinoma cells. Genes Genomics 42, 317-324.

72. Wang, B., Xu, L., Ge, Y., Cai, X., Li, Q., Yu, Z., Wang, J., Wang, Y., Lu, C., Wang, D., Wang, Y., Chen, X., and Gu, Y. (2019). PLOD3 is Upregulated in Gastric Cancer and Correlated with Clinicopathologic Characteristics. Clin Lab 65.

73. Wang, D., Zhang, S., and Chen, F. (2018). High Expression of PLOD1 Drives Tumorigenesis and Affects Clinical Outcome in Gastrointestinal Carcinoma. Genet Test Mol Biomarkers 22, 366-373.

74. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general amber force field. J Comput Chem 25, 1157-1174.

75. Wang, Z., Shi, Y., Ying, C., Jiang, Y., and Hu, J. (2021). Hypoxia-induced PLOD1 overexpression contributes to the malignant phenotype of glioblastoma via NF-kappaB signaling. Oncogene 40, 1458-1475.

76. Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. Physical Chemistry Chemical Physics 7, 3297-3305.

77. Xie, D., Li, J., Wei, S., Qi, P., Ji, H., Su, J., Du, N., and Zhang, X. (2020). Knockdown of PLOD3 suppresses the malignant progression of renal cell carcinoma via reducing TWIST1 expression. Mol Cell Probes 53, 101608.

78. Xu, W.H., Xu, Y., Wang, J., Tian, X., Wu, J., Wan, F.N., Wang, H.K., Qu, Y.Y., Zhang, H.L., and Ye, D.W. (2019). Procollagen-lysine, 2-oxoglutarate 5-dioxygenases 1, 2, and 3 are potential prognostic indicators in patients with clear cell renal cell carcinoma. Aging (Albany NY) 11, 6503-6521.

79. Xu, Y., Zhang, L., Wei, Y., Zhang, X., Xu, R., Han, M., Huang, B., Chen, A., Li, W., Zhang, Q., Li, G., Wang, J., Zhao, P., and Li, X. (2017). Procollagen-lysine 2-oxoglutarate 5-dioxygenase 2 promotes hypoxia-induced glioma migration and invasion. Oncotarget 8, 23401-23413.

80. Yamada, Y., Kato, M., Arai, T., Sanada, H., Uchida, A., Misono, S., Sakamoto, S., Komiya, A., Ichikawa, T., and Seki, N. (2019). Aberrantly expressed PLOD1 promotes cancer aggressiveness in bladder cancer: a potential prognostic marker and therapeutic target. Mol Oncol 13, 1898-1912.

81. Yamauchi, M., and Sricholpech, M. (2012). Lysine post-translational modifications of collagen. Essays Biochem 52, 113-133.

82. Yang, B., Zhao, Y., Wang, L., Zhao, Y., Wei, L., Chen, D., and Chen, Z. (2020). Identification of PLOD Family Genes as Novel Prognostic Biomarkers for Hepatocellular Carcinoma. Front Oncol 10, 1695.

83. Zhao, Y., Zhang, X., and Yao, J. (2021). Comprehensive analysis of PLOD family members in low-grade gliomas using bioinformatics methods. PLoS One 16, e0246097.

Chapter 5 Molecular Structure and Enzymatic Mechanism of the Human Collagen Hydroxylysine Galactosyltransferase GLT25D1/COLGALT1

Matteo De Marco1, #, Sristi Raj Rai1,#, Luigi Scietti1,#,§, Daiana Mattoteia1, Stefano Liberi1, Elisabetta Moroni2, Alberta Pinnola3, Alice Vetrano4, Claudio Iacobucci4, Carlo Santambrogio5, Giorgio Colombo6, Federico Forneris1,7,*

Contribution: recombinant protein production, GLT25D1 crystallization, enzymatic assay development, biophysical characterization (NS-EM, SEC-SAXS).

ABSTRACT

During collagen biosynthesis, lysine residues undergo extensive post-translational modifications mediated by two distinct families of metal ion–dependent enzymes (LH/PLODs and GLT25D/COLGALT). These sequential modifications ultimately generate the highly conserved - [1–2]-glucosyl--(1, O)-galactosyl-5-hydroxylysine motif. Dysfunctions in these enzymes are associated with developmental disorders and extracellular matrix changes that promote the aggressiveness of solid tumours. In this work, we investigated human GLT25D1/COLGALT1 and found that it adopts an elongated head-to-head homodimeric structure. Each monomer is composed of two domains (GT1 and GT2), both unexpectedly capable of binding metal ion cofactors and UDP--galactose donor substrates, creating four potential catalytic sites per dimer. Our analysis pinpointed the GT2 domain as the true catalytic site, characterized by an unusual Glu-Asp-Asp motif essential for Mn2+ binding. This excluded a direct catalytic role for GT1, although its bound Ca2+ and UDP--galactose cofactors proved crucial for structural stability. While dimerization is not strictly required for GLT25D1/COLGALT1 activity, it provides an important molecular interface that facilitates multi-enzyme complex formation with partner multifunctional LH/PLOD lysyl hydroxylase–glycosyltransferases.

INTRODUCTION

Collagen is really conserved across animal life from the most robust invertebrates to higher vertebrates. Their amino acid sequence, post-translational modifications, and assembly of larger structures provide collagens with a range of physical and mechanical properties. This flexibility is precisely why collagens are the most dominant protein family in the animal kingdom and an important part of many tissues including bone, cartilage, tendons, and skin [1–3].

While collagens are being synthesized by the cell, they undergo extensive post-translational modifications that influence the ability for the protein to fold into a right-handed triple helix. The repeating unit is a Gly-Xaa-Yaa motif in which Xaa and Yaa are commonly proline and hydroxyproline, respectively [4–7]. Among these modifications, proline hydroxylation has been studied most and is considered essential for collagen's triple helix to wind correctly [8–10].

Lysine residues occur less frequently than proline in collagen polypeptides and are predominantly located at the Yaa position of the Gly–Xaa–Yaa motif [11]. The functions of lysine and its post-translational modifications are chiefly associated with the intermolecular cross-linking that takes place after collagen is secreted into the extracellular milieu. In collagens and collagen-like sequences, lysine is processed sequentially by dedicated biosynthetic enzymes to generate a simple yet highly conserved modification: α-[1–2]-glucosyl-β-(1,O)-galactosyl-5-hydroxylysine (Glc-Gal-Hyl) [12–16]. Formation of Glc-Gal-Hyl begins with hydroxylation of lysine to 5-hydroxylysine (Hyl), catalyzed by Fe2+ dependent collagen lysyl hydroxylases (LH; also termed procollagen, lysyl 2-oxoglutarate dioxygenases, PLOD). In humans, the LH/PLOD family comprises three isoforms: LH1/PLOD1, LH2/PLOD2, and LH3/PLOD3 [17–19].

The LH reaction is carried out within the C-terminal domain of the dimeric lysyl hydroxylase LH/PLOD enzymes. The resulting Hyl residues are subsequently β-galactosylated by Mn^{2+} dependent galactosyltransferases via an inverting mechanism that transfers the galactose moiety from uridine diphosphate-α-galactose (UDP-α-Gal) to the Hyl hydroxyl group, yielding β-(1,O)-galactosyl-5-hydroxylysine (Gal-Hyl; Gal-T activity). Gal-Hyl is then further modified by multifunctional LH/PLOD enzymes through a retaining reaction requiring Mn^{2+} and uridine diphosphate-α-glucose (UDP-α-Glc) (Glc-T activity), producing Glc-Gal-Hyl. Early models posited that a single multifunctional LH/PLOD enzyme mediated the entire conversion from lysine to Glc-Gal-Hyl [20–21]. The subsequent identification of a dedicated family of collagen galactosyltransferases (GLT25D/COLGALT) established their essential roles in collagen biosynthesis in vivo [22–27]. Additional biochemical analyses of human LH/PLOD and GLT25D/COLGALT enzymes ultimately refuted the single-enzyme hypothesis. The emerging view is that two enzyme systems, LH/PLOD and GLT25D/COLGALT, act in concert, alternating on the same

substrate and possibly interacting directly [28], to achieve the full Glc-Gal-Hyl post-translational modification of collagen (Fig. 1a).

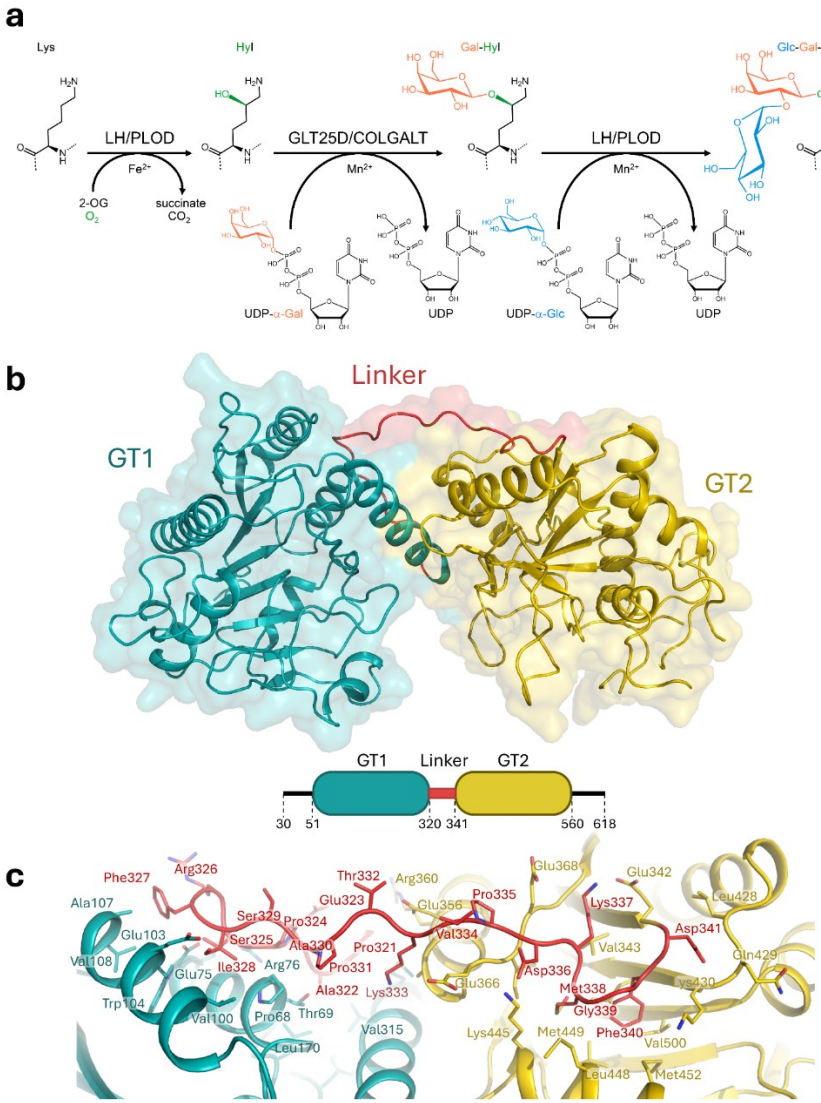


Figure 1. Biochemical and structural features of human GLT25D1/COLGALT1. (a) Schematic of the post-translational pathway converting collagen lysine to Glc-Gal-Hyl, driven by the alternating activities of LH/PLOD and GLT25D1/COLGALT1, with donor substrates and products highlighted. (b) Cartoon of the human GLT25D1/COLGALT1 crystal structure depicting the domain organization: GT1 and GT2 are linked by a rigid connector that engages both domains via hydrophobic anchor points detailed in panel (c).

While the multifunctional lysyl hydroxylase/glucosyltransferase (LH/PLOD) enzymes have been extensively studied at molecular and structural levels, the same cannot be said for the GLT25D/COLGALT family, which catalyzes the central galactosyltransferase reaction. The human GLT25D/COLGALT proteins are soluble, endoplasmic reticulum-resident factors, retained with a C-terminal Arg–Asp–Glu–Leu (RDEL) sequence, and co-localize with LH/PLOD enzymes [28–29]. Three human isoform genes have been identified: GLT25D1/COLGALT1, constitutively expressed across tissues; GLT25D2/COLGALT2, primarily found in the nervous system; and GLT25D3/COLGALT3, found in secretory organs (e.g., salivary glands, pancreas, placenta), and neural tissues. Of the three human isoform genes, only the first two are catalytically active, while isoform 3 (also known as CERCAM) is regarded as inactive in vitro and its modification role on collagen lysine may be unrelated.

Of importance the colgalt genes and their protein products (GLT25D/COLGALT enzymes) are associated with a variety of pathologies that include cerebral small vessel disease, liver disease, osteoarthritis, musculoskeletal malformations, and cancer [26, 30–32]. Liver fibrosis is the result of chronic extracellular matrix deposition with an acute endpoint of cirrhosis with hepatocellular carcinoma. In recent years, high-level plasma GLT25D1/COLGALT1 levels have been reported in patients with cirrhosis when compared with unaffected individuals, which has been speculated to result from a stimulatory effect on hepatic stellate cell activity [33].

Efforts to define the structure–function principles governing GLT25D/COLGALT activity on collagen have yielded valuable clues but also revealed multiple Asp-x-Asp (DxD) motifs along the polypeptide that could contribute to catalysis, underscoring the need for deeper structure-based analyses. In this study, we integrated computational and experimental structural biology with biochemical and biophysical approaches to resolve the quaternary organization of GLT25D1/COLGALT1. Using experimental phasing, we determined X-ray crystal structures of full-length human GLT25D1/COLGALT1 in both the apo state and bound to the donor substrate UDP-α-Gal. Complementary small-angle X-ray scattering, negative-stain electron microscopy, and mass photometry converged on an elongated architecture reminiscent of LH/PLOD enzymes, comprising a head-to-head homodimer in which each protomer contains two homologous Rossmann-fold domains capable of coordinating metal ions and engaging the donor substrate. Guided by the structural data, site-directed mutagenesis, native mass spectrometry, and molecular dynamics simulations pinpointed the C-terminal domain as the catalytic center, whereas the N-terminal domain retains UDP-α-Gal in an atypical Ca^{2+}-dependent manner that serves solely to stabilize folding. Cross-linking mass spectrometry demonstrated that the homodimer interface is essential for engaging partner LH/PLOD enzymes. Taken together, these findings illuminate the molecular architecture and functional roles of a key human enzyme in collagen biosynthesis.

RESULTS

GLT25D1/COLGALT1 IS A Mn^{2+} DEPENDENTGALACTOSYLTRANSFERASE.

We developed protocols for recombinant expression and purification of full-length human GLT25D1/COLGALT1 in E. coli, yielding enzyme preparations suitable for in vitro assessment of galactosyltransferase activity (Supplementary Fig. 1). Activity was quantified either by high-resolution mass spectrometric detection of Gal-T conversion on synthetic collagen peptides or by monitoring UDP formation resulting from consumption of the UDP-α-Gal donor in the presence of collagen hydroxylysines. Using luminescence-based detection of UDP as the product, we established a new biochemical assay for steady-state kinetic measurements (Supplementary Fig. 2). The kinetic parameters obtained from Michaelis–Menten analysis (Supplementary Table 5; Supplementary Fig. 2) were consistent with prior studies employing radiolabeled UDP-α-Gal donor substrates [34]. Using these assays, we first assessed metal cofactor preference and observed a pronounced requirement for Mn^{2+}: no activity was detected with other divalent cations, except for Mg^{2+}, which supported markedly lower activity (Supplementary Fig. 3a–b). Considering that the related multifunctional LH/PLOD enzymes can process multiple donor substrates and exhibit turnover even without an acceptor, we tested GLT25D1/COLGALT1 and found, in contrast to LH/PLOD, strict specificity for UDP-α-Gal and only limited donor processing in the absence of acceptor (i.e., uncoupled activity; Supplementary Fig. 3c).

GLT25D1/COLGALT1 FEATURES AN ELONGATED MULTI-DOMAIN ARCHITECTURE

The 2.8 Å crystal structure of human wildtype GLT25D1/COLGALT1 was solved by experimental phasing, employing single-wavelength anomalous dispersion (SAD) from an Hg^{2+} soaked crystal (Supplementary Tables 2–3). Electron density was well defined for the globular cores of the two monomers in the asymmetric unit (Supplementary Fig. 4), whereas flexible segments at the N terminus (residues 30–41) and C terminus (residues 572–618) lacked interpretable density and were therefore not modeled.

Each GLT25D1/COLGALT1 polypeptide comprises two globular domains, designated GT1 and GT2 (Fig. 1b), which share a glycosyltransferase-like fold but differ in overall topology (Supplementary Fig. 5a). In GLT25D1/COLGALT1, the domains interact through a combination of electrostatic and hydrophobic contacts centered on helix α7 (GT1) and helix α8 (GT2). A 25-residue linker connects the domains and shields otherwise solvent-exposed hydrophobic patches, primarily via two phenylalanines (Phe327 and Phe340) (Fig. 1c).

While both domains deviate from previously described Rossmann-fold topologies, the GT1 domain shows a closer resemblance to known glycosyltransferases, producing DALI Z-scores >15 against several bacterial and human enzymes despite low levels of sequence identity (Supplementary Fig. 5b) [35]. The closest experimentally determined structural

homolog is the central accessory (AC) domain of human LH3/PLOD3, which shares 24% sequence identity but has a similar overall structure (RMSD 1.2 Å; Supplementary Fig. 5c). Furthermore, we explicitly searched Protein BLAST databases with the GLT25D1/COLGALT GT1 sequence, and numerous hits were found across vertebrates, invertebrates, and even giant viruses (Supplementary Fig. 6). These viral hits were particularly noteworthy, as prior work showed that Acanthamoeba Polyphaga mimivirus bears post-translationally modified collagens exhibiting high homology to mammalian counterparts [36]. Nonetheless, BLAST searches performed with LH/PLOD homologs excluded returned chiefly the additional vertebrate isoforms GLT25D2/COLGALT2 and CERCAM, along with GLT25D/COLGALT orthologs in insects (Supplementary Fig. 6). Collectively, these observations emphasize the conservation of the GT-A fold within the GT1 domain, revealing closely aligned folding topologies despite minimal primary-sequence conservation (Supplementary Fig. 5b). The GLT25D1/COLGALT1 GT2 domain exhibits only remote structural similarity to characterized glycosyltransferases, and its superpositions are less definitive than those obtained for GT1 (Supplementary Fig. 5d). Similarly, when aligned with LH3/PLOD3, the GT2 domain shows limited resemblance to the N-terminal GT domain of LH3/PLOD3, with an RMSD of 4.7 Å and a DALI Z-score below 10 [35] (Supplementary Fig. 5e). BLAST analysis of the GT2 sequence retrieved more hits than for GT1: spanning vertebrates, invertebrates, giant viruses, and bacteria (Supplementary Fig. 7). The inclusion of bacterial and viral sequences in these alignments supports the possibility of ancestral gene fusion events that ultimately produced the two-domain architecture of GLT25D/COLGALT.

GLT25D1/COLGALT1 IS A DIMERIC ENZYME

The asymmetric unit accommodates two GLT25D1/COLGALT1 molecules arranged antiparallel, engaging through GT1-mediated contacts that wrap around an unstructured N-terminal loop spanning residues 42–51 (Fig. 2a). The dimer buries ~1,250 Å² of surface and is stabilized by 22 hydrogen bonds and 5 salt bridges, with 35 residues per protomer contributing to the interface. Mass photometry corroborated this assembly, revealing a predominant solution species of ~140 kDa, consistent with a dimer, and showing minimal dispersion (Fig. 2b). Concordant masses of 131 ± 10 kDa were obtained by SEC–MALS and SEC–SAXS (Supplementary Table 6; Supplementary Fig. 8a–b). The SEC–SAXS data yielded ab initio sphere models of an elongated complex that matched the crystallographic dimer's dimensions and envelope with ~95% correlation (Supplementary Fig. 8c). Orthogonal comparison of the unrefined crystal model against solution SAXS profiles gave an initial χ^2 of 7.62, improving to 1.89 after incorporating flexible N- and C-terminal segments (Fig. 2c). Additionally, single-particle negative-stain EM produced low-resolution 2D classes of elongated particles closely matching both the SEC–SAXS solution models and the crystallographic dimers (Fig. 2d).

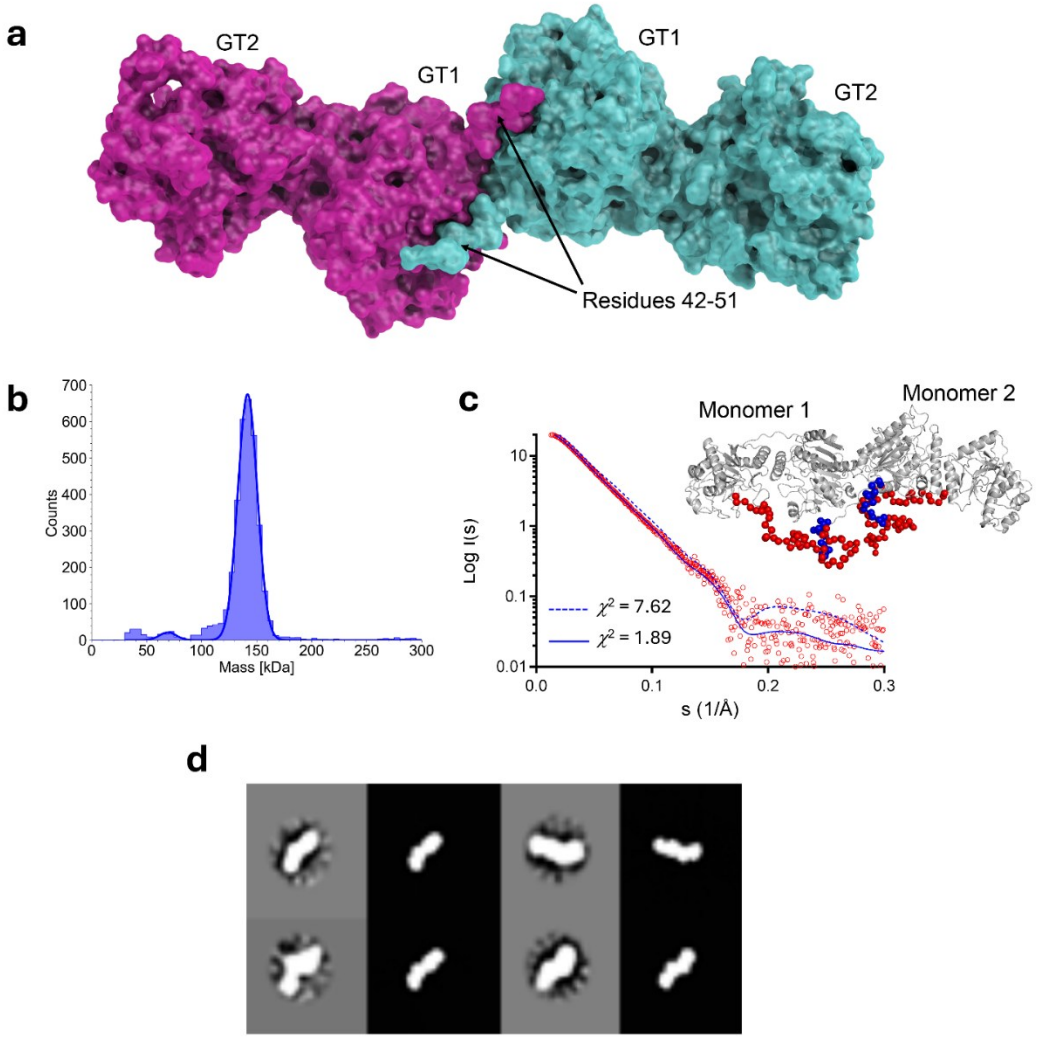


Figure 2. GLT25D1/COLGALT1 forms stable elongated dimers. (a) The structure of the full-length GLT25D1/COLGALT1 homodimer was determined, based on the orientation of the two protomers which associate through the N-terminal domains. (b) Mass photometry profile of recombinant wild-type GLT25D1/COLGALT1. (c) Based on a $\chi^2 = 7.62$ (dashed line), our comparison of the crystallographic dimer and our solution SAXS data (red circles) supports the structure; however, including modeled flexible N- and C-terminal segments for a dimer model (white cartoon) improved fit to the SAXS data ($\chi^2 = 1.89$, solid line; blue loops and spheres represent flexible N-and C-terminal segments). (d) Negative-stain EM single-particle 2D classes (gray) aligned with computed projections of the crystal structure (black) supporting that we are measuring the same aspect of the molecule.

PROBING THE ROLE OF GLT25D1/COLGALT1 DIMERIZATION

To assess the functional role of dimerization in GLT25D1/COLGALT1, we identified interface residues within the crystallographic dimer: Trp158, whose side chain nests within a pocket formed by Arg53, Ala85, His113, and Met157 of the opposite protomer, and Asp160, which participates in a hydrogen-bonding network with Arg53 from both subunits (Fig. 3a). Substitution of Trp158 with Arg disrupted the dimer, yielding a monomeric species in solution as determined by mass photometry and SEC–MALS (Fig. 3b, c). When tested for activity, via luminescence-based detection of free UDP in the presence of gelatin and by HRMS quantification of galactosyl-hydroxylysine on synthetic peptides, the Trp158Arg variant exhibited catalytic performance comparable to wild type (Fig. 3d; Supplementary Table 5), indicating that dimerization is not strictly required for enzymatic activity in vitro.

Prompted by this finding, we used cross-linking mass spectrometry to probe whether dimerization facilitates interactions with the multifunctional lysyl hydroxylase–glucosyltransferase LH3/PLOD3. A 1:1 mixture of wild-type and Trp158Arg GLT25D1/COLGALT1 was cross-linked with disuccinimidyl dibutyric urea (DSBU), digested with trypsin, and analyzed by nano-LC–MS/MS. In both datasets, we detected a robust interprotein cross-link between GLT25D1 Lys145 (GLT25D1Lys145) and LH3 Lys645 (LH3Lys645) (Supplementary Fig. 9). Notably, the abundance of this GLT25D1Lys145–LH3Lys645 cross-link was approximately sevenfold higher in complexes containing wild-type GLT25D1/COLGALT1 relative to the Trp158Arg mutant (Fig. 3e). This difference could not be attributed to variations in protein levels or lysine reactivity, as total protein abundances and the frequencies of GLT25D1Lys145 and LH3Lys645 mono-links were essentially unchanged across experiments (Fig. 3e). Together, these results indicate that GLT25D1/COLGALT1 dimerization is a critical determinant for assembly of higher-order complexes with LH3/PLOD3.

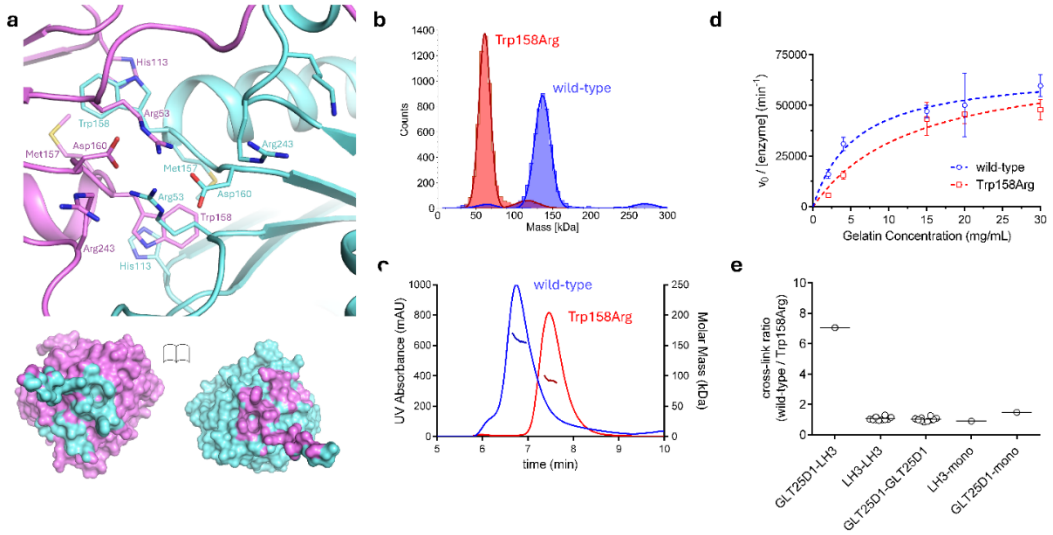


Figure 3. Probing GLT25D1/COLGALT1 dimeric architecture and its functional significance. (a) Close-up of the dimer interface with residues that mediate direct interprotomer contacts. Bottom panel shows an open-book view of the GLT25D1/COLGALT1 monomers, mapping the interaction footprint of each partner onto their respective surfaces. (b) Mass photometry of wild-type GLT25D1/COLGALT1 versus Trp158Arg variant designed to eliminate dimerization. (c) SEC–MALS elution profiles of wild type and Trp158Arg showing shifts in elution volume with corresponding shifts in molar mass. (d) Gal-T activities of wild-type versus Trp158Arg measured by luminescence. Michaelis–Menten plots show apparent turnover (v_0/[enzyme]) as a function of gelatin concentration; KM and kcat values of both enzymes are provided in Supplementary Table 5. (e) Relative abundances of GLT25D1/COLGALT1–LH3/PLOD3 cross-linked peptides quantified by MS/MS with either wild-type or Trp158Arg mutant.

BOTH GT1 AND GT2 DOMAINS BIND DONOR SUBSTRATES AND METAL IONS

Examination of the experimental electron density for GLT25D1/COLGALT1 revealed pronounced difference peaks in the GT1 domain of both protomers near the DxD motif (Asp166–Asp168) situated in the β4–α4 loop, consistent with UDP-α-Gal coordinated to a metal ion (Fig. 4a). Notably, the Asp166 carboxylate does not directly chelate the metal. The donor substrate is embedded within an extensive interaction network: the Tyr126 side chain hydrogen bonds to the C4 carbonyl of the uridine ring, which is further stabilized by a π–σ contact with Arg61 and a π–alkyl interaction with Val143, within a hydrophobic cavity formed by Leu59, Asp91, and Arg139. The ribose adopts a well-defined conformation stabilized by Trp135 and by hydrogen bonds involving the backbone carbonyl and amide

groups of Leu59, Arg61, and Ala167. Additional contacts include interactions of Arg147 and Asp265 with hydroxyls on the galactose moiety (Fig. 4a, b). Of note, Trp135 corresponds to Trp130 in murine GLT25D1/COLGALT1; mutation at this position (the “fosse” allele) has been linked to skeletal and muscular defects due to markedly reduced gene expression [22].

Upon co-crystallization with Mn^{2+} and UDP-α-Gal, electron density consistent with cofactors and donor substrate was also observed in the GT2 domain of one protomer in the asymmetric unit (Fig. 4c). Here, the metal ion is coordinated by a Glu–Asp–Asp motif arranged analogously to Mn^{2+}-dependent glycosyltransferases: the carboxylates of Glu435 and Asp437 directly chelate the metal, while Asp436 engages the ribose hydroxyls of the cofactor via hydrogen bonds (Fig. 4c,d). The uridine ring is constrained by Gly408 and Gly411 of helix α10 within a hydrophobic groove defined by Cys412 (same helix) and by Leu348 and Ala374 on the opposing side (Fig. 4c, d). No interpretable density was evident for the galactose, suggesting conformational heterogeneity of the sugar, likely reflecting the absence of an acceptor substrate within the broad cleft formed by Trp495, Leu497, Tyr567, and Pro558 (Fig. 4c, d).

Comparison of GT2 structures with and without donor bound revealed a conformational rearrangement of the otherwise flexible C terminus (residues 562–580), which wraps around the donor-binding cavity and positions the backbone carbonyls of Ser569 and Thr571 to form productive hydrogen bonds with both phosphates of the UDP moiety (Fig. 4e). The observation that cofactors and donor were visible in only one crystallographic protomer is attributable to distinct crystal packing environments that, in the second molecule, restrict the GT2 C-terminal loop from adopting the rearranged conformation due to lattice contacts (Supplementary Fig. 9).

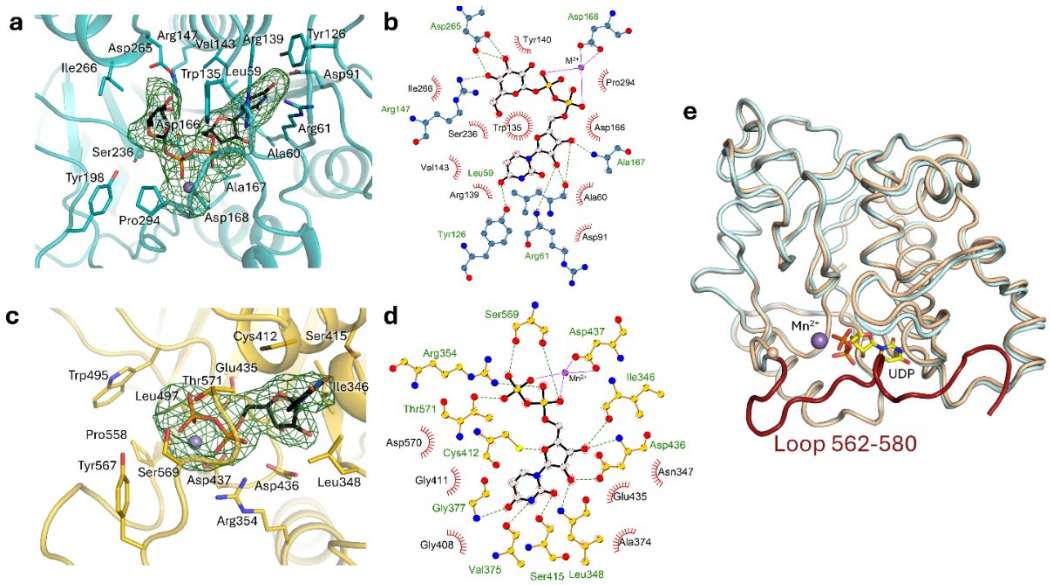


Figure 4. Structural insights into GLT25D1/COLGALT1 enzymatic activity and assembly. (a) The GT1 domain is constitutively occupied by a divalent cation (M^{2+}; purple sphere) and UDP-α-Gal, as evidenced by the Fo–Fc omit map contoured at 3.2σ. Residues coordinating the metal and engaging the donor are shown as sticks. (b) LIGPLOT+[71] schematic of the GT1 interaction network with donor and cofactors. (c) Upon co-crystallization with excess Mn^{2+} and UDP-α-Gal, Fo–Fc omit density (3.2σ) also reveals ligands in GT2; the galactose moiety is unresolved, so UDP is modeled as sticks. Coordinating and interacting residues are depicted as sticks. (d) LIGPLOT+[71] schematic of the GT2 interaction network with donor and cofactors. (e) Binding of Mn^{2+} and UDP-α-Gal induces conformational changes in GT2 by ordering the otherwise flexible C-terminal loop via direct contacts with cofactor and donor. Superposition of apo (light orange) and ligand-bound (light blue) GT2 highlights the 560–574 loop (brown), which is stabilized only upon ligand binding.

MUTATIONS IN THE GT1 DOMAIN IMPACT ON ENZYME FOLDING STABILITY

Guided by the structural data, we engineered GT1-domain substitutions predicted to disrupt metal-ion or UDP-α-Gal binding (Fig. 4a,b), including Asp166Ala (eliminating coordination to the bound metal), Asp265Ala (removing hydrogen bonds to the galactose), and Ile266Gln (hindering rigid positioning of the galactose within the GT1 cavity). None of these variants yielded recombinant protein suitable for biochemical analysis, in line with prior work on the Asp166Ala/Asp168Ala double mutant designed to abolish the GT1 DxD motif [34]. We also produced a Trp135Arg mutant corresponding to the murine fosse allele (Trp130Arg) [22]. This variant expressed at markedly lower levels than wild-type GLT25D1/COLGALT1,

consistent with the reduced expression reported for fosse, and SDS–PAGE profiles deviated substantially from wild type, indicating severe folding instability (Supplementary Fig. 1). By contrast, a control mutant, Ile267Gln (adjacent to Ile266 but not expected to perturb UDP-α-Gal binding or orientation), produced well-folded, catalytically active enzyme (Supplementary Fig. 1), mirroring the previously reported Pro292Asn substitution that was intended to render a GLT25D1–CERCAM chimera inactive yet retained activity [34].

THE GT2 DOMAIN IS RESPONSIBLE FOR CATALYTIC ACTIVITY

Our structural analyses indicate that, relative to GT1, the GT2 domain binds metal ions and donor substrates more transiently, provoking conformational rearrangements of the extended C terminus that may influence catalysis (Fig. 4). The presence of an atypical EDD motif, rather than the canonical DxD, for metal coordination in GT2, together with a UDP-α-Gal donor exhibiting a flexible sugar moiety (Fig. 4c, d), points toward the location of the catalytic site. Conservation mapping across vertebrate homologs shows that residues mediating metal-ion and UDP-α-Gal recognition are highly conserved in both GT1 and GT2 (Fig. 5a).

Guided by these observations, and by evidence that the N-terminal GT1 domain primarily serves structural roles, we engineered GT2 point mutants targeting metal-ion and donor binding, as well as residues potentially involved in acceptor engagement during galactose transfer. All variants were purified and displayed biophysical properties consistent with proper folding (Supplementary Fig. 1). Substitutions Glu435Ala and Asp437Ala within the EDD motif abolished activity in both luminescence and HRMS assays (Fig. 5b; Supplementary Fig. 3; Supplementary Table 5). A similar loss of function was observed for Cys412Ser, which disrupts productive positioning of the uridine–ribose within a hydrophobic pocket (Fig. 5b; Supplementary Fig. 3; Supplementary Table 5). Motivated by established roles for aromatic and charged residues at key positions in inverting glycosyltransferases [37–38], we generated Trp495Ala and Asp522Ala; both mutations eliminated GLT25D1/COLGALT1 Gal-T activity (Fig. 5b; Supplementary Fig. 3; Supplementary Table 5).

The GT2 C-terminal rearrangement also implicated Arg351, whose side chain reorients to hydrogen-bond with Asp570 upon cofactor and donor binding (Fig. 5c). Accordingly, we produced Arg351Ala and Asp570Ala variants. Although Arg351Ala yielded insufficiently robust preparations for biochemical testing, Asp570Ala was obtained successfully and proved completely inactive (Fig. 5b; Supplementary Table 5; Supplementary Fig. 1).

The structural data further rationalize previously examined positions: Asp336 at the C terminus of the GT1–GT2 linker forms a hydrogen bond with GT2 Lys445 that appears only modestly affected by substitution to serine; and Asp461 and Asp433, the latter engaged in a hydrogen-bonding network with Lys508 and His547 likely contribute critically to GT2 fold stability. Collectively, the structural and mutational evidence unambiguously identifies the GLT25D1/COLGALT1 catalytic site and delineates its unusual features.

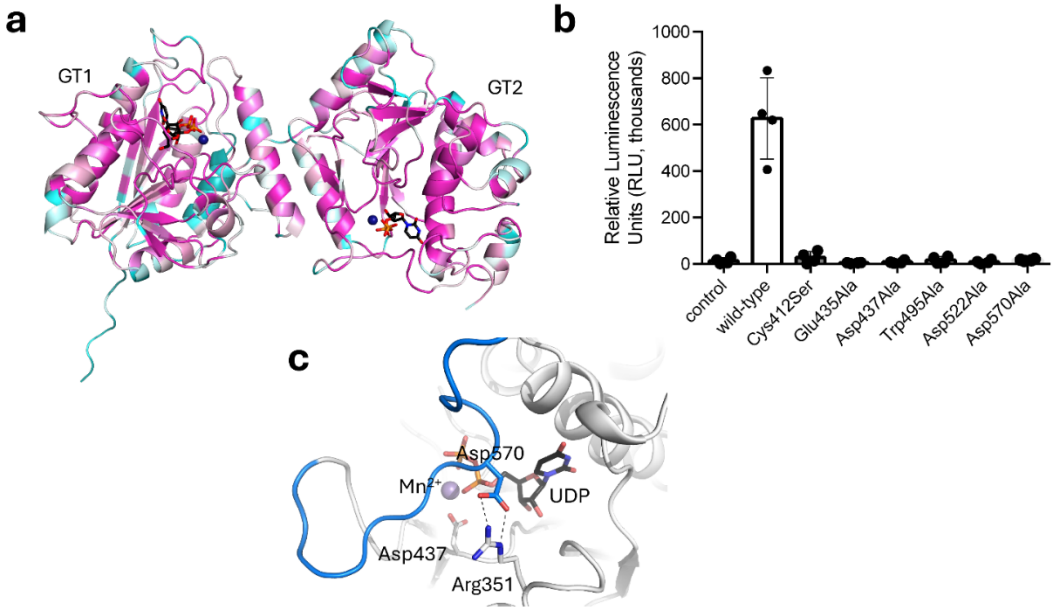


Figure 5. Structure-guided interpretation of GLT25D1/COLGALT1 functional features. (a) Amino acid conservation in GLT25D/COLGALT homologs, color-coded from dark blue (low conservation) to dark purple (high conservation); the sequence set and alignment are in Supplementary Figs.6–7. (b) Luminescence-based Gal-T activity for wild-type GLT25D1/COLGALT1 versus GT2-site mutants that disrupt cofactor/substrate binding; bars reflect means of three biological replicates with individual data points overlaid as dots; error bars represent standard deviation from three independent experiments. (c) In the cofactor-bound state, the GT2 active site is sculpted by a salt bridge between the flexible C-terminal loop (blue) Arg351 and Asp570. Mn^{2+} (purple sphere) is coordinated by Glu437 and UDP and side chains that interact are displayed as sticks. (d) Structural mapping of pathogenic variant residues; mutants Leu151Arg, Ala154Pro, and Gly377Arg from cerebral small vessel disease and Trp135 from the murine fosse are highlighted in red with arrows.

MOLECULAR SIGNIFICANCE OF GLT25D1/COLGALT1 PATHOGENIC MUTATIONS

The molecular structure and biochemical profiling of multiple GLT25D1/COLGALT1 variants allow us to relate enzyme alterations to disease-associated mutations. The murine fosse substitution Trp130Arg—linked to developmental abnormalities—maps to Trp135 in the human ortholog. In GLT25D1/COLGALT1, Trp135 shapes the GT1-domain UDP-α-Gal binding pocket by wedging its side chain between the uridine and galactose moieties of the cofactor (Fig. 4a). Experimentally, this substitution resulted in markedly reduced recombinant expression and pronounced instability in vitro, precluding further biochemical analysis (Supplementary Fig. 1a). Mapping cerebral small vessel disease variants [39–40] onto the GLT25D1/COLGALT1 structure provides a mechanistic rationale for their effects

(Fig. 5d). Leu151Arg and Ala154Pro map to the C terminus of helix α3 in the GT1 domain. Leu151 is buried in a hydrophobic pocket; substituting a bulky, charged Arg at this site is expected to destabilize GT1 substantially. Likewise, replacing Ala154 with Pro is likely to compromise the structural integrity of helix α3 in proximity to the homodimer interface. The Gly377Arg variant [39] affects a residue that forms an α-turn linking strand β14 to helix α9, adjacent to the uridine-binding site of the catalytic UDP-α-Gal donor. Introducing Arg at this position would be predicted to perturb GT2 conformational stability, leading to loss of function.

Mn^{2+} BINDS TRANSIENTLY IN THE GT2 DOMAIN DURING CATALYSIS

Alongside the dimeric assembly which could potentially have four aligned catalytic sites, we wondered whether the metal ions and donor substrates might have a structural role. To test this speculation, we purified recombinant GLT25D1/COLGALT1 with no supported cofactors and assessed all the various bound metals. Using differential scanning fluorimetry (DSF), we saw significant thermal stabilization upon the binding of Mn^{2+} only (Fig 6a), which aligns with biochemical data indicating Mn^{2+} can be a transiently bound cofactor that is catalytically relevant. The Glu435Ala and Asp437Ala versions show no stabilization by Mn^{2+} in their respective DSF profiles (Fig 6a) which is also aligned with structural data depicting transient metals bound within the GT2 domain (Fig 4).

ROLES OF Ca^{2+} AND UDP-A-GAL DONOR SUBSTRATE IN THE GT1 DOMAIN

In parallel with Mn^{2+} binding studies, we examined the effect of removing the tightly bound GT1 metal by incubating GLT25D1/COLGALT1 with varying concentrations of ethylenediaminetetraacetic acid (EDTA). Differential scanning fluorimetry (DSF) consistently showed destabilization, with thermal shifts of 3.0–3.5 °C relative to untreated samples (Fig. 6b). Comparable destabilization was observed when EDTA was applied to the Glu435Ala and Asp437Ala variants (Fig. 6b). Because these substitutions target only the transient metal site in GT2, we inferred that EDTA chiefly perturbs protein stability by chelating the GT1-bound metal. Despite this effect, DSF profiles of EDTA-treated samples remained indicative of partial folding, prompting tests of conformational stability. Mass photometry of EDTA-treated wild-type enzyme showed no substantive change in oligomeric state (Fig. 6c), whereas size-exclusion chromatography revealed an altered elution profile consistent with a modified hydrodynamic radius (Fig. 6d).

Unexpectedly, ICP–MS aimed at detecting Mn associated with GT1 found no evidence for this ion. We therefore probed the GT1-bound ligand(s) more directly. Native mass spectrometry of wild-type and Trp158Arg GLT25D1/COLGALT1 under controlled partial unfolding revealed coexisting free and ligand-bound monomers, implicating Ca^{2+} as the most probable GT1-bound metal (Fig. 6e); ICP–MS corroborated a 1:1 Ca^{2+}: GLT25D1/COLGALT1 stoichiometry. Consistent with this assignment, DSF in the presence

of the Ca^{2+}- selective chelator EGTA produced a destabilization profile essentially identical to that observed with EDTA (Fig. 6b).

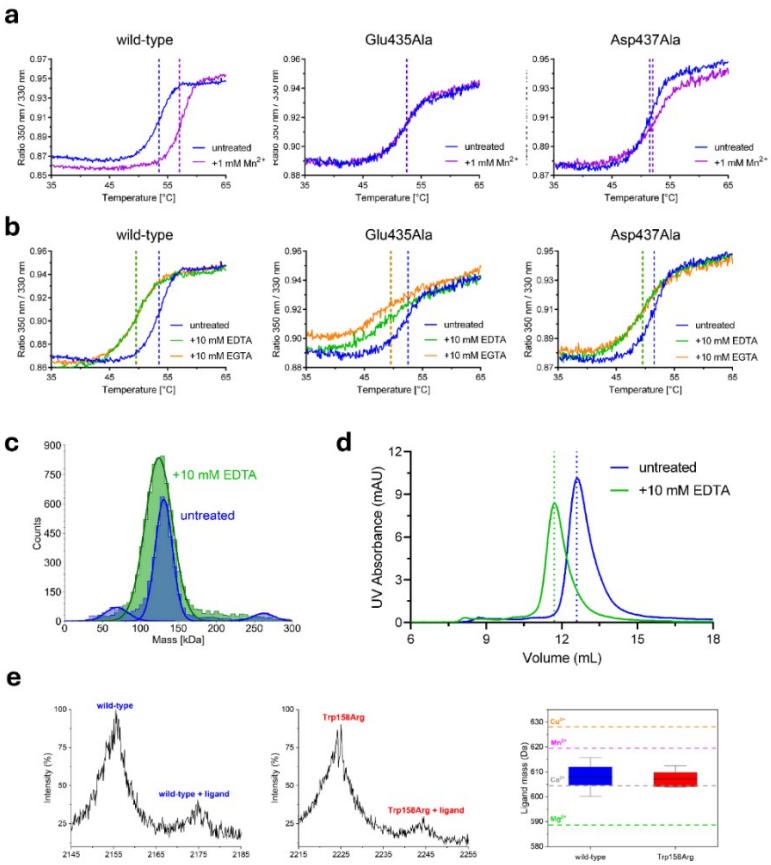


Figure 6. Probing the structural role of Mn2+ in GLT25D1/COLGALT1 GT1 domain. (a) The differential scanning fluorimetry (DSF) profiles of wild-type GLT25D1/COLGALT1 (left) show pronounced thermal stabilization following the addition of Mn^{2+} (purple) when compared to the untreated control (blue). Mutating GT2 metal-binding residues (e.g., Glu435, center; Asp437) abolish Mn^{2+} stabilization. (b) DSF profiles following metal chelation show relative instability when comparing EDTA (green) or EGTA (orange) to untreated (blue) samples; wild-type (left) and GT2 mutant proteins are similarly unstable. (c) Mass photometry of EDTA-treated samples shows higher dispersity. (d) Size-exclusion chromatography of EDTA-treated samples show different elution behavior. (e) Native mass spectrometry under partially denaturing conditions comes from representative spectra for wild-type (30^+ charge state, left) and Trp158Arg (31^+ charge state, center) with peaks for the apo protein and protein-ligand complexes annotated; the box plot (right) summarizes ligand-mass estimates across the full charge-state distribution, with reference lines for theoretical ligand masses of each divalent cation.

Collectively, these data establish a uniquely Ca^{2+}-bound GT1 domain and support structural roles for this metal cofactor and the UDP-α-Gal donor in GT1. The lack of a direct catalytic function for GT1 aligns with prior GLT25D1/COLGALT1–CERCAM chimera studies and is further supported by the absence of essential negatively charged residues near the donor's galactose moiety (Fig. 5a, b) that are required for activity in inverting GT-A glycosyltransferases.

MD SIMULATIONS SUPPORT DISTINCT ROLES FOR GT1 AND GT2 DOMAINS

To refine the structural picture of GLT25D1/COLGALT1, we performed comparative molecular dynamics on the isolated GT1 and GT2 domains with and without bound metal ions and donor substrates. No large-scale conformational changes accompanied ligand binding in either domain. Nevertheless, analyses of internal motions indicated that ligand binding markedly stabilizes GT1, whereas the effect on GT2 is more modest.

Principal component analysis (PCA) was conducted by merging bound and unbound trajectories for each domain and projecting snapshots onto the first two components (PC1, PC2), which capture most of the variance. Normalized projection histograms visualize the conformational space sampled (Supplementary Fig. 11a–d). For GT1, bound and unbound distributions along PC1 overlap, but the unbound state exhibits broader, less pronounced peaks, indicative of greater flexibility; the bound state shows sharper, taller peaks, consistent with restriction to a narrower ensemble (Supplementary Fig. 11a). Along PC2, the bound state likewise remains more confined than the multi-peaked unbound profile (Supplementary Fig. 11b). In GT2, unbound projections along PC1 are dominated by a single peak centered at zero, reflecting minimal motion and intrinsic stability, whereas the bound state becomes bimodal, with one peak overlapping the unbound distribution and a second, sharper peak, suggesting cofactor-induced access to an additional conformation (Supplementary Fig. 11c). Along PC2, the unbound state remains narrowly centered at zero, while the bound state shifts slightly toward negative values, indicating subtle adjustments upon binding (Supplementary Fig. 11d). Collectively, these results imply that ligand binding chiefly stabilizes GT1 by curbing its dynamics, while GT2 is already stable in the apo form and gains only a minor, additional conformational substate upon binding.

Hydrogen-bond persistence analyses corroborate this view. In GT1, the bound state shows a larger fraction of persistent hydrogen bonds (60–90% of the simulation) than the unbound state (<60%), consistent with increased rigidity (Supplementary Fig. 11e). GT2 exhibits only small differences between conditions, maintaining a similar, moderately dynamic H-bond network in both states (Supplementary Fig. 11f).

Distance fluctuation (DF) analysis further supports these findings. In apo GT1, the loop spanning residues 130–140, the C-terminal segment of helix α7, and the interdomain linker are highly mobile and poorly coordinated with the domain core (Supplementary Fig. 12a). Ligand binding enhances internal coordination and restricts motions in these regions. By

contrast, GT2 shows minimal DF changes between bound and unbound trajectories, with only slight variations in flexible segments (Supplementary Fig. 12b). Notably, in the absence of stabilizing cofactors, the GT1 130–140 loop, the α7 C terminus, and the GT1–GT2 linker remain particularly dynamic and weakly constrained (Supplementary Fig. 12c).

Overall, the MD results and accompanying analyses support a model in which cofactor binding substantially enhances GT1 structural stability while exerting a comparatively modest influence on GT2 dynamics.

DISCUSSION

Galactosylation of hydroxylysine in collagen is a pivotal step in the post-translational pathway that produces Glc-Gal-Hyl—the most prevalent O-linked glycosylation in the animal kingdom [12–13, 16]. Following decades of biochemical work on the enzymes and their functions, recent efforts have shifted toward a detailed molecular description of the reactions and participants, with particular focus on the multifunctional LH/PLOD lysyl hydroxylase–glucosyltransferases [17, 41]. Yet, the absence of experimental structural information for GLT25D/COLGALT galactosyltransferases has constrained mechanistic insight, despite significant prior studies [16, 28, 34]. In this study, we resolved the experimental structure of full-length human GLT25D1/COLGALT1 and combined biochemical, biophysical, and site-directed mutagenesis approaches to define its in vitro functionality.

Our data define an elongated, multidomain homodimer in which each protomer contains two Rossmann-like domains, GT1 and GT2, connected by a rigid linker, with dimer contacts encircling the N terminus of each monomer (Fig. 1b, c). Unexpectedly, both GT1 and GT2 are required for catalysis and each binds metal ions and the UDP-α-Gal donor (Fig. 4), with their ligand-binding regions exhibiting the highest sequence conservation (Supplementary Figs. 6–7). The N-terminal GT1 domain aligns closely with the inactive central accessory (AC) domain of LH/PLOD enzymes and shares hallmarks of inverting GT-A glycosyltransferases, whereas GT2 displays a distinctive topology lacking close counterparts among well-characterized glycosyltransferases.

Prior investigations of GLT25D1/COLGALT1 DxD motifs demonstrated that substitutions at the GT1 metal-coordinating residues Asp166 and Asp168 abolish enzymatic activity [34]. The same study reported that replacing the enzyme's entire N terminus with that of the inactive paralog CERCAM did not impair collagen glycosylation [34]. These findings align with our experimental observation of Ca^{2+} and UDP-α-Gal bound within the GT1 domain (Fig. 6b–e). Unlike the common Mn^{2+} or Mg^{2+} cofactors in GT-A Rossmann-fold glycosyltransferases [37–38], we found no evidence of Ca^{2+} occupying catalytic sites in other members of this family. Together with the absence of essential negatively charged residues at key positions required for inverting GT-A catalysis [37–38], the uniquely stable Ca^{2+}–UDP-α-Gal complex in GT1 argues against a direct catalytic role for this domain.

Consistently, MD simulations comparing ligand-free and ligand-bound GT1 further support a structural function for the bound metal and donor substrate, revealing increased flexibility around residues 130–140 and 310–330 in the absence of Ca^{2+} and UDP-α-Gal (Supplementary Figs. 11–12).

Notably, these segments include residues adjacent to the UDP-α-Gal binding pocket, such as Trp135, which corresponds to the loss-of-function murine fosse mutation [22], and the principal GT1 contact surface with the neighboring GT2 domain, where cerebral small vessel disease–linked variants map (Fig. 5d). Elevated flexibility within this region could substantially disrupt the globular packing of the multidomain enzyme. Consistent with this notion, EDTA/EGTA treatment produces folding instability and solution-phase conformational changes without appreciable alteration of the oligomeric state (Fig. 6b–d).

The GT2 domain displays strong sequence conservation around the site transiently occupied by Mn^{2+} and UDP-α-Gal during catalysis (Fig. 5a; Supplementary Fig. 7). Structural analysis was therefore crucial to localize the catalytic site of GLT25D1/COLGALT1. Prior efforts to identify catalytically essential residues were inconclusive, likely because GT2 employs an atypical EDD motif for Mn^{2+} coordination, rather than the canonical DxD, and because multiple DxD motifs are scattered throughout the polypeptide (Supplementary Figs. 6–7).

To delineate the GT2 catalytic network, we engineered point mutations around the Mn^{2+} cofactor and donor substrate observed in the crystal structure, revealing an extensive array of charged and hydrophobic side chains that are indispensable for UDP-α-Gal recognition, activation, and catalysis (Fig. 5b–c).

Overall, the apo and donor-bound GT2 active sites are highly similar (Fig. 4), aside from a rearrangement of the otherwise flexible C-terminal segment. This region is enriched in acidic residues whose direct participation in binding and processing UDP-α-Gal was confirmed by site-directed mutagenesis, in agreement with earlier in silico predictions [42].

Consistent across orthogonal solution studies and mutagenesis, the dimeric quaternary structure is not required for GLT25D1/COLGALT1 catalytic activity in vitro (Fig. 3). Nevertheless, cross-linking mass spectrometry shows that the dimer architecture furnishes a direct contact surface for GLT25D1/COLGALT1–LH3/PLOD3 interactions, providing experimental support for multimeric biosynthetic hetero-complexes that effect complete conversion of collagen Lys to Glc-Gal-Hyl [29, 31]. Notably, our biophysical data consistently indicate a minor dispersed fraction of the recombinant enzyme, hinting at the need for additional stabilizing homo- or heterotypic interactions of physiological relevance in vivo.

The unusually elongated GLT25D1/COLGALT1 dimer mirrors the tail-to-tail dimer of LH3/PLOD3, a geometry that may facilitate extensive crosstalk between the two enzymes acting sequentially on collagen lysines and raises the possibility of descent from a common ancestor. This notion is reinforced by similarities between their noncatalytic domains (Supplementary Fig. 5c): the GLT25D1/COLGALT1 GT1 domain, which binds Ca^{2+} and hosts UDP-α-Gal solely to maintain fold integrity, and the LH/PLOD AC domain, which lacks capacity to bind metals and/or co-substrates [41]. These hypotheses warrant further

investigation, particularly considering the atypical trafficking of multifunctional LH/PLOD enzymes through the secretory pathway in the absence of canonical ER-retention signals [43–46].

METHODS

Chemicals

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

DNA CONSTRUCTS

The human GLT25D1/COLGALT1 coding sequence lacking the N-terminal signal peptide and C-terminal RDEL ER-retention motif (UniProt Q8NBJ5; residues 30–618) was synthesized (Genewiz) and subcloned into pCR8 using in-frame 5′-BamHI and 3′-NotI sites. Variants were generated with the Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific), amplifying the entire plasmid with primers listed in Supplementary Table 1. Linearized, mutagenized plasmids were phosphorylated with T4 polynucleotide kinase (Invitrogen) prior to ligation. All constructs were verified by Sanger sequencing (Microsynth) and subsequently transferred into a modified pET28b-SUMO vector (Novagen). The expression plasmids encode an N-terminal 8×His tag followed by a SUMO tag upstream of the in-frame 5′-BamHI site and include an in-frame stop codon downstream of the 3′-NotI site.

PRODUCTION OF RECOMBINANT GLT25D1/COLGALT1

Chemically competent E. coli BL21 cells (Invitrogen) were heat-shock–transformed with the pET28b-SUMO-GLT25D1 construct. A single colony was inoculated into LB supplemented with kanamycin (100 µg mL^{-1}) and grown overnight at 37 °C in a shaking incubator (New Brunswick). The preculture was then diluted 1:50 into 1 L ZYP5052 autoinducing medium [47] in a 5 L Erlenmeyer flask, incubated at 37 °C for 3 h at 180 rpm, then shifted to 17 °C and grown overnight. Cells were harvested (5,000 g), resuspended at 1:10 (w/v) in lysis buffer (25 mM HEPES/NaOH, 500 mM NaCl, 10 µM leupeptin, 10 µM pepstatin, 0.3 mg mL^{-1} lysozyme, 500 µM $MnSO_4$, pH 8.0), and sonicated (16 cycles of 9 s on/6 s off). The lysate was clarified by centrifugation (60,000 g, 45 min, 4 °C; Avanti J26, Beckman Coulter), filtered (MiniSart GF 0.8 µm; Sartorius), loaded onto a 5 mL Ni Sepharose Excel column (Cytiva), and eluted stepwise with 25 mM HEPES/NaOH, 500 mM NaCl, 500 mM imidazole, pH 8.0. Fractions containing His-SUMO-GLT25D1/COLGALT1 (by SDS-PAGE) were pooled and dialyzed overnight against 2 L of 25 mM HEPES/NaOH, 500 mM NaCl, pH 8.0, while concurrently cleaving the N-terminal 8×His-SUMO tag with His-tagged SUMO protease (1 mg mL^{-1}; 1:300 v/v). The protease and tag were removed by

re-passing the sample over Ni Sepharose and collecting the flow-through.The protein was buffer-exchanged using 30 kDa MWCO Vivaspin Turbo filters (Sartorius) into 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0, applied to a HiScreen Capto Q column (Cytiva) pre-equilibrated in the same buffer, and eluted with a linear NaCl gradient (peak at ~250 mM NaCl). After concentration (30 kDa MWCO), the sample was polished by size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0. Final quality was verified by reducing and non-reducing SDS-PAGE. GLT25D1/COLGALT1-containing fractions were pooled, concentrated to 4 mg mL^{-1}, and stored at −80 °C. SDS-PAGE quality controls for all proteins used are shown in Supplementary Fig. 1.

PROTEIN CRYSTALLIZATION

GLT25D1/COLGALT1 spherulites were initially observed in several nanoliter sitting drops (0.1 µL protein at 4 mg mL^{-1} plus 0.1 µL reservoir). These drops were prepared by a Gryphon crystallization robot (Art Robbins) on SwissSci vapor-diffusion plates using commercially available screens. Microcrystals were observed using the Morpheus screen [48] (Molecular Dimensions), after a few days at 4 °C, that developed in an expected manner compared to the crystals from PEG precipitants reported in the literature, suggesting a preference for PEG precipitants at neutral pH between 6.0 and 7.0. Using 0.5 µL protein at the concentration of 3.5 mg mL^{-1} was mixed with 0.5 µL of the optimized reservoir which with 8% PEG 4000, 100 mM MES/NaOH, 20 % glycerol, and a pH of 6.5, to obtain diffracting quality crystals. Co-crystals were obtained under those same conditions, with the addition of 500 µM $MnCl_2$ and 1 mM UDP-α-Gal in the protein solution as catalysts. Once the crystals were grown, they were harvested using mounted Litholoops (Molecular Dimensions), which were flash-cooled and put in liquid nitrogen for data collection. To isolate heavy-atom derivatives, crystals (with requisite optimizations), were soaked overnight at a minimum of 5 hours at 4 °C in mother liquor containing 1 mM K_2HgBr_4, harvested, cryoprotected and flash-cooled in liquid nitrogen.

X-RAY DATA COLLECTION, STRUCTURE DETERMINATION AND REFINEMENT

Single-crystal X-ray datasets were recorded at 100 K from multiple beamlines at the ESRF (Grenoble, France) and the SLS (Villigen, Switzerland). The experimental phasing employed data collected at the mercury absorption inflection point on ESRF beamline ID23-1 while making use of an Eiger2 16M (Dectris) detector. Diffraction images were indexed and integrated with XDS [49] and scaled and merged in AIMLESS [50]; collection statistics are on Supplementary Table 2. The SAD phasing was performed with the SHELXC/D/E pipeline and HKL2MAP [51]. SHELXD identified 13 heavy-atom sites at 4.0 Å, and using SHELXE the hand was unambiguously established using density modification while scaling the working resolution to 2.80 Å (estimated solvent content, 62%). The density-modified map allowed the initial model building to progress with iterative ARP/wARP and BUCCANEER [52]. The completion of the model and its refinement involved cycles of

manual rebuilding in COOT and automated refinement with phenix.refine [53] and REFMAC5 [54]. Complexes of GLT25D1/COLGALT1 with donor substrates were solved by molecular replacement in PHASER, using the experimentally phased model as the search model. The validation of the model was done with MolProbity and Protein Data Bank validation tools [55]; final refinement statistics are provided in Supplementary Table 3. Structural images were created in PyMOL, including superpositions with the "super" command, and computed all-atom RMSD values.

MASS PHOTOMETRY

Mass photometry was performed on a Refeyn Two instrument (Refeyn) using 24 × 50 mm^2 glass coverslips. Purified human GLT25D1/COLGALT1 (wild type and variants) prepared as 200 nM stocks were diluted 1:10 into 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0, and recorded over a 4 μm × 11 μm field of view at 500 Hz. Data were acquired with AcquireMP and processed/visualized in DiscoverMP (Refeyn). Summary metrics are reported in Supplementary Table 4.

SIZE EXCLUSION CHROMATOGRAPHY COUPLED WITH MULTI-ANGLE LIGHT SCATTERING (SEC-MALS)

30 uL of recombinant GLT25D1/COLGALT1 (4 mg mL^{-1}) were injected onto a Protein KW-802.5 analytical SEC column (Shodex) and eluted in phosphate-buffered saline at 1 mL min^{-1} on a Shimadzu Prominence HPLC. A miniDAWN multi-angle light-scattering detector (Wyatt) was connected in-line with a RID-20A differential refractive index detector (Shimadzu) for total mass quantification, and an SPD-20A UV detector (Shimadzu) to monitor protein specific to the elution volume. Chromatograms were processed in ASTRA 7 (Wyatt) using an estimated dn/dc of 0.185 mL g^{-1}. System calibration was confirmed by injection of 10 μL of 2.5 mg mL^{-1} monomeric BSA.

DIFFERENTIAL SCANNING FLUORIMETRY (DSF)

Differential scanning fluorimetry (DSF) was conducted on recombinant GLT25D1/COLGALT1 (wild type and variants) at 1 mg mL^{-1} in 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0, using a Tycho NT.6 instrument (NanoTemper Technologies GmbH). Data were processed in GraphPad Prism 7 [56].

EVALUATION OF GAL-T ACTIVITY THROUGH LUMINESCENCE

Reaction mixtures were prepared by preincubating GLT25D1/COLGALT1 (final concentration 1 mg mL-1, 5 μL; 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0) for three hours at 37 °C with a variety of bovine skin gelatin concentrations that had previously been solubilized through thermal denaturation (95 °C, 10 min). Reactions were initiated with the

addition of UDP-α-sugar (100 µM) and either MnCl2 or MgCl2 (50 µM). At designated time points (1−120 min), reactions were stopped by heating to 95 °C for two minutes, transferred to a Proxiplate white 384-well plate (PerkinElmer), following which they were incubated for one hour at 25 °C with the addition of UDP-Glo luminescence reagent (5 µL; Promega). Luminescence was measured using a GloMax Discovery plate reader (Promega) according to the manufacturer's protocols.

Initial velocities were determined as the slopes of linear fits to time-course data using GraphPad Prism 7 [56]. Apparent turnover with errors were assessed for linearity using double-reciprocal plots (Supplementary Fig. 2c), and apparent kcat and Km were fitted directly to the Michaelis–Menten equation in Prism 7 with proper error propagation [57]. Each assay was performed three times and controls that omitted enzyme, donor, or acceptor were included using the same conditions. Results are collected in Supplementary Table 5.

EVALUATION OF GAL-T ACTIVITY THROUGH HR-LCMS

Recombinant human LH3/PLOD3 (5 µM; produced in-house as previously described) was incubated with GLT25D1/COLGALT1 (5 µM; wild type or variants) in the presence of 50 µM $FeCl_2$, 100 µM 2-oxoglutarate (2-OG), 500 µM sodium ascorbate, 50 µM $MnCl_2$, 100 µM UDP-α-sugar, and 1 mM peptide substrate (Ac-GIKGIKGIKGIK-COOH). Reactions proceeded for 3 h at 37 °C. Aliquots (5 µL) were diluted with 43 µL Milli-Q water, acidified with 2 µL formic acid to a final volume of 50 µL, and analyzed by full-scan LC–ESI-HRMS/MS on an Exion LC AD UHPLC (AB Sciex) coupled to an X500B system (AB Sciex). The column oven was maintained at 40 °C and the autosampler at 10 °C. Peptide separation employed reverse-phase HPLC on a Hypersil Gold C18 column (150 × 2.1 mm, 3 µm, 175 Å; Thermo Fisher Scientific) with a 15-min linear gradient from 2% to 50% solvent B (solvent A: 0.1% formic acid in water; solvent B: acetonitrile with 0.1% formic acid) at 0.2 mL min^{-1}. Mass spectra were acquired in positive mode under fixed settings: ion spray voltage 4,500 V; declustering potential 100 V; curtain gas 30 psi; ion source gas 1, 40 psi; ion source gas 2, 45 psi; source temperature 350 °C; collision energy 10 V. Data processing used SCIEX OS 2.1 (AB Sciex). Summary results are provided in Supplementary Table 5.

SIZE EXCLUSION CHROMATOGRAPHY COUPLED WITH SMALL-ANGLE X-RAY SCATTERING (SEC-SAXS)

Solution scattering was performed at the ESRF beamline BM29 utilizing a Pilatus 1M detector at 1 frame s^{-1}, and 2.87 m from the sample, providing an overall q-range of 0.01–4.00 nm^{-1}. SEC–SAXS measurements were executed with a Nexera HPLC system (Shimadzu) coupled online to the SAXS capillary. In these runs 50 µL of GLT25D1/COLGALT1 at 4 mg mL^{-1} were injected onto a Superdex 200 PC 3.2/300 Increase column (GE Healthcare) pre-equilibrated in 25 mM HEPES/NaOH, 200 mM NaCl, pH 8.0.

Frames corresponding to the elution peak of GLT25D1/COLGALT1 were picked, buffer-subtracted, and averaged using CHROMIXS.

SEC-SAXS MODELING AND DATA ANALYSIS

Radii of gyration (Rg), molar mass estimates, and distance distribution functions P(r) were computed in PRIMUS [58] as part of the ATSAS package [59]. Experimental SAXS curves were compared to models for which crystal structures were available, with CRYSOL [60]. All collection and analysis metadata can be found in Supplemental Table 6. Ab initio bead models were created from the SAXS data using GASBOR the P(r) functions were computed from the experimental profiles alongside the number of residues in the GLT25D1/COLGALT1 monomers, without constraints for internal symmetry. Twenty independent GASBOR reconstructions were performed and compared for similarity using SUPCOMB and DAMSEL (ATSAS) and the one with the least normalized spatial discrepancy was selected. Superpositions were performed with SUPALM by using the saspy ATSAS plugin for PyMOL [61]. Rigid body SAXS modeling was performed on a monomer extracted from the experimental crystal structure, subjected to loop modeling in CORAL [62], and the model-data agreement was assessed with CRYSOL [60].

SINGLE PARTICLE NEGATIVE STAINING ELECTRON MICROSCOPY

200 mesh copper grids (Ted Pella) were glow-discharged on a PELCO EasiGlow (pressure 40 mBar, 15 mA, 30 s). A 2 µL aliquot of GLT25D1/COLGALT1 at 1 mg mL^{-1} was applied to each grid and incubated for 60 s, excess liquid was blotted with Whatman paper, and a 25 µL drop of 2% (w/v) uranyl acetate was added with gentle agitation for 1 min. After blotting the stain, grids were air-dried overnight. A total of 350 micrographs were recorded on a JEM 1200EXIII transmission electron microscope (JEOL) equipped with a MegaView III CCD camera (Olympus). Image processing in RELION included auto picking and multiple rounds of 2D classification. Comparison of 2D classes from single-particle analysis with crystal-structure–derived GLT25D1/COLGALT1 models was performed using AlignProjections on the COSMIC2 web server [63].

GLT25D1/COLGALT1-LH3/PLOD3 CROSS-LINKING AND ENZYMATIC PROTEOLYSIS

GLT25D1/COLGALT1 (5 µM; wild type and Trp158Arg) was incubated with human LH3/PLOD3 (5 µM; total protein = 10 µM) in 50 mM HEPES/NaOH, 100 mM NaCl, pH 8.0. All samples were incubated with DSBU (0.5 mM) at RT for 1 h; DSBU was freshly prepared in neat DMSO before adding to the samples. Reactions were quenched with 2 µL of 1 M Tris/HCl. The cross-linked mixtures were processed using S-Trap microcolumns

(Protifi). Following loading, the samples were washed with 90:10 methanol/50 mM ammonium bicarbonate, and then digested with trypsin (Promega) at a 1:20 trypsin: protein for 1.5 h at 47 °C. The peptides were eluted with 50 mM ammonium bicarbonate, and subsequently two more elutions with 0.2% formic acid, and 0.2% formic acid with 50% acetonitrile. The fractions were pooled, dried down via vacuum centrifugation, and reconstituted in 5% acetonitrile, 0.1% trifluoroacetic acid for LC–MS/MS.

LIQUID CHROMATOGRAPHY COUPLED TO MASS SPECTROMETRY FOR CROSS-LINKING PEPTIDE ANALYSIS

Proteolyzed, cross-linked samples were analyzed by LC–MS/MS using an UltiMate 3000 RSLC nano-HPLC (Thermo Fisher Scientific) coupled with a timsTOF Pro mass spectrometer (Bruker Daltonics) with a CaptiveSpray source: peptides were first trapped in a C18 precolumn (Acclaim PepMap 100, 300 µm × 5 mm, 5 µm, 100 Å; Thermo Fisher Scientific) at 50 °C, then separated using a self-packed Picofrit nanospray emitter (New Objective; 360 µm o.d. × 75 µm i.d. × 400 mm, 15 µm tip), packed with C18 (3.0 µm, 120 Å; Dr. Maisch GmbH). After trapping, elution took place using a 90 min linear gradient from 3% to 50% (v/v) acetonitrile in water.

timsTOF Pro settings were: mobility-dependent collision energy ramping was set to 95 eV at an inverse reduced mobility ($1/k_0$) of 1.6 V·s/cm² and 23 eV at 0.73 V·s/cm², with linear S-5 interpolation between and constant values for anything else. TIMS scans were not concatenated. Target intensity per PASEF precursor was set to 20,000. The scan window was 0.6–1.6 V·s/cm², ramp accumulation time was 166 ms. 14 PASEF MS/MS scans were acquired for each 2.57 s cycle, up to seven precursors per mobilogram. Precursors were selected for fragmentation within m/z of 100–1700, 3+ to ≤8+ charges. Active exclusion was enabled for 0.4 min (mass width 0.015 Th; $1/k_0$ width 0.015 V·s/cm²).

XL-MS DATA ANALYSIS

Cross-linked peptides were identified with MeroX [64] (v2.0.1.7; https://stavrox.com). Searches assumed tryptic specificity (C-terminal to Lys/Arg) with up to four missed cleavages, peptide lengths of 3–30 residues, a fixed modification of Cys alkylation (iodoacetamide), and variable Met oxidation. The cross-linker was defined to react with Lys side chains and peptide N termini. Identification used RISEUP mode permitting up to three missing ions, with precursor and fragment tolerances of 10 ppm and 20 ppm, respectively, and a signal-to-noise threshold >2.0. PSM-level FDR was controlled at 1% with a minimum score of 100; to refine FDR estimates, a contaminant (CRAP) database was included.Proteomic data analysis employed MaxQuant [65] (v2.6.2.0) and Skyline [66] (v24.1.1.339). Bruker raw files were processed by DDA searches against the UniProt Human Reference Proteome (UP000005640; 20,406 canonical proteins), with contaminants filtered via the default MaxQuant list. Parameters were tryptic cleavage C-terminal to Lys/Arg, up to two missed cleavages, peptide lengths 6–25 amino acids; fixed Cys alkylation (iodoacetamide); variable Met oxidation and N-terminal acetylation. For LFQ, at least three

unique peptides per protein were required. MS1 and MS2 tolerances were set to 10 ppm and 20 ppm, respectively. Match Between Runs was enabled (retention time window 0.4 min; ion mobility window 0.05). FDR at both PSM and protein levels was controlled at 1%. MaxQuant and MeroX outputs were imported into Skyline for MS/MS spectrum review and MS1-level peptide quantification. Protein quantification used the ten most abundant validated peptides per protein. MS/MS spectra corresponding to the PLOD3–GLT25D1 interprotein cross-link and associated dead-end products were imported and quantified in Skyline following established workflows [67].

ICP-MS MEASUREMENTS

To quantify metal ions in recombinant GLT25D1/COLGALT1, 900 µL of protein (4 mg mL^{-1}; purified without added divalent cations) were digested with 300 µL of 65% ultrapure HNO_3 and 200 µL of 30% w/w H_2O_2, brought to 5 mL with ultrapure water, and analyzed by single-quadrupole ICP-MS (iCAP RQ, Thermo Fisher Scientific) equipped with a quartz cyclonic chamber (3 °C), MicroMist nebulizer (400 µL min^{-1}), quartz torch, Ni sampler and skimmer cones, and a QCell pressurized with He (3 V, KED mode). Manganese and calcium were quantified by external calibration using four daily prepared standards (0, 1, 5, 10 µg L^{-1}) in the same buffer, dilution, and acid conditions. Manganese was not detected, whereas calcium was present at ~1:1 stoichiometry.

NATIVE MASS SPECTROMETRY (NATIVE-MS)

The metal cofactors of GLT25D1/COLGALT1 were identified using native mass spectrometry (native- MS). Recombinant wild-type and Trp158Arg enzymes (20 µM) were analyzed under partially denaturing conditions (50 mM ammonium bicarbonate, 40% acetonitrile), which increased droplet desolvation and minimized the formation of adducts during electrospray. For both enzymes, aliquots were infused directly into an Orbitrap Fusion (Thermo Fisher Scientific) mass spectrometer at a flow rate of 200–300 nL/min using metal-coated borosilicate emitters (~1 µm i.d.). Instrument settings were resolving power 120,000 at m/z 200; IRM pressure 3 mTorr (intact-protein mode); spray voltage 1.1–1.2 kV; ion-transfer tube 275 °C; in-source fragmentation 75 V; AGC target 4×10^5; maximum injection time 100 ms. Final spectra were averaged from 60 s of acquisition duration. The ligand masses were determined as the mass difference between the signals of the bound and apo proteins over the entirety of the charge-state distribution (Supplementary Fig. 10).

MOLECULAR DYNAMICS SIMULATIONS

Molecular Dynamics (MD) simulations were performed with AMBER (v24) using the FF19SB force field. Equilibration and production runs were performed with pmemd.cuda (the GPU accelerated PMEMD engine), and equilibrations and restrained simulations used the standard sander. Both systems were set-up with the xleap program in AMBER, and the specific simulation protocols and conditions are described in detail in the Supplementary

Information. The simulation trajectory analyses were performed using cpptraj (AmberTools v24) and some user-defined scripts.

To assess the dominant motions of the GT1 and GT2 domains, a principal component analysis (PCA) was performed on the apo and holo trajectories from the two domains collectively. For the PCA of each trajectory, we calculated a matrix of backbone covariance, which was then diagonalized to yield principal components (eigenvectors) and eigenvalues. The eigenvalue indicated the percent of total variance described by that mode. The first two principal components, PC1 and PC2, described the most variance, so trajectory snapshots were projected onto PC1 and PC2, and occupancy histograms were created from the normalized distributions along the PC1 and PC2 axes. The persistence of hydrogen bonds calculated using the hbond command in cpptraj, with a geometric definition (i.e., maximum donor–acceptor distance of $\leq$3.5 Å and H–D–A angle of $\geq$135°), tracked the presence of a bond for each trajectory over time and normalized the frequency of each bond against the total number of hydrogen bonds to yield a percent occupancy, which was then binned into 10% increments to generate persistence histograms. To calculate distance fluctuation (DF) for each residue pair i, j, $Dij=\langle(dij-\langle dij\rangle)\rangle$, where dij indicates the instantaneous Cα–Cα distance, and angle brackets indicate time averages over the entire trajectory. This final matrix indicates the motion, and correlated flexibility of residue pairs.

REFERENCES

1. Bourhis, J.M. et al. Structural basis of fibrillar collagen trimerization and related genetic disorders. Nat Struct Mol Biol 19, 1031-6. 2012. doi: 10.1038/nsmb.2389

2. Fidler, A.L., Boudko, S.P., Rokas, A. & Hudson, B.G. The triple helix of collagens - an ancient protein structure that enabled animal multicellularity and tissue evolution. J Cell Sci 131, jcs203950. 2018. doi: 10.1242/jcs.203950

3. Myllyharju, J. & Kivirikko, K.I. Collagens and collagen-related diseases. Ann Med 33, 7-21. 2001. doi: 10.3109/07853890109002055

4. Ishikawa, Y. & Bachinger, H.P. A molecular ensemble in the rER for procollagen maturation. Biochim Biophys Acta 1833, 2479-91. 2013. doi: 10.1016/j.bbamcr.2013.04.008

5. Gelse, K., Poschl, E. & Aigner, T. Collagens--structure, function, and biosynthesis. Adv Drug Deliv Rev 55, 1531-46. 2003. doi: 10.1016/j.addr.2003.08.002

6. Kivirikko, K.I. & Myllyla, R. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods Enzymol 82 Pt A, 245-304. 1982. doi: 10.1016/0076-6879(82)82067-3

7. Kadler, K.E., Holmes, D.F., Trotter, J.A. & Chapman, J.A. Collagen fibril formation. Biochem J 316 (Pt 1), 1-11. 1996. doi: 10.1042/bj3160001

8. Bella, J. Collagen structure: new tricks from a very old dog. Biochemical Journal 473, 1001-1025. 2016. doi: 10.1042/bj20151169

9. Myllyharju, J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix Biol 22, 15-24. 2003. doi: 10.1016/S0945-053X(03)00006-4

10. Gorres, K.L. & Raines, R.T. Prolyl 4-hydroxylase. Crit Rev Biochem Mol Biol 45, 106-24. 2010. doi: 10.3109/10409231003627991

11. Yamauchi, M. & Sricholpech, M. Lysine post-translational modifications of collagen. Essays Biochem 52, 113-33. 2012. doi: 10.1042/bse0520113

12. Cummings, R.D. The repertoire of glycan determinants in the human glycome. Mol Biosyst 5, 1087-104. 2009. doi: 10.1039/b907931a

13. Hennet, T. Collagen glycosylation. Curr Opin Struct Biol 56, 131-138. 2019. doi: 10.1016/j.sbi.2019.01.015

14. 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. Biochem Biophys Res Commun 560, 93-98. 2021. doi: 10.1016/j.bbrc.2021.04.128

15. Tvaroska, I. Glycosylation Modulates the Structure and Functions of Collagen: A Review. Molecules 29. 2024. doi: 10.3390/molecules29071417

16. De Giorgi, F., Fumagalli, M., Scietti, L. & Forneris, F. Collagen hydroxylysine glycosylation: non-conventional substrates for atypical glycosyltransferase enzymes. Biochem Soc Trans 49, 855-866. 2021. doi: 10.1042/BST20200767

17. Salo, A.M. & Myllyharju, J. Prolyl and lysyl hydroxylases in collagen synthesis. Exp Dermatol. 2020. doi: 10.1111/exd.14197

18. Yamauchi, M., Terajima, M. & Shiiba, M. Lysine Hydroxylation and Cross-Linking of Collagen. Methods Mol Biol 1934, 309-324. 2019. doi: 10.1007/978-1-4939-9055-9_19

19. Kivirikko, K.I. & Myllyla, R. Collagen glycosyltransferases. Int Rev Connect Tissue Res 8, 23-72. 1979. doi: 10.1016/b978-0-12-363708-6.50008-4

20. Wang, C. et al. The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. Matrix Biol 21, 559-66. 2002. doi: 10.1016/S0945-053X(02)00071-9

21. Geister, K.A. et al. Loss of function of Colgalt1 disrupts collagen post-translational modification and causes musculoskeletal defects. Dis Model Mech 12. 2019. doi: 10.1242/dmm.037176

22. Terajima, M. et al. Role of Glycosyltransferase 25 Domain 1 in Type I Collagen Glycosylation and Molecular Phenotypes. Biochemistry 58, 5040-5051. 2019. doi: 10.1021/acs.biochem.8b00984

23. Webster, J.A. et al. Collagen beta (1-O) galactosyltransferase 1 (GLT25D1) is required for the secretion of high molecular weight adiponectin and affects lipid accumulation. Biosci Rep 37. 2017. doi: 10.1042/BSR20170105

24. Baumann, S. & Hennet, T. Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase. J Biol Chem 291, 18514-24. 2016. doi: 10.1074/jbc.M116.723379

25. Yang, J. et al. Collagen beta(1-O) galactosyltransferase 2 deficiency contributes to lipodystrophy and aggravates NAFLD related to HMW adiponectin in mice. Metabolism 120, 154777. 2021. doi: 10.1016/j.metabol.2021.154777

26. Kehayova, Y.S., Wilkinson, J.M., Rice, S.J. & Loughlin, J. Osteoarthritis genetic risk acting on the galactosyltransferase gene COLGALT2 has opposing functional effects in articulating joint tissues. Arthritis Res Ther 25, 83. 2023. doi: 10.1186/s13075-023-03066-y

27. Schegg, B., Hulsmeier, A.J., Rutschmann, C., Maag, C. & Hennet, T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. Mol Cell Biol 29, 943-52. 2009. doi: 10.1128/mcb.02085-07

28. Liefhebber, J.M., Punt, S., Spaan, W.J. & 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. doi: 10.1186/1471-2121-11-33

29. Kehayova, Y.S., Watson, E., Wilkinson, J.M., Loughlin, J. & Rice, S.J. Genetic and Epigenetic Interplay Within a COLGALT2 Enhancer Associated With Osteoarthritis. Arthritis Rheumatol 73, 1856-1865. 2021. doi: 10.1002/art.41738

30. Guo, T. et al. COLGALT2 is overexpressed in ovarian cancer and interacts with PLOD3. Clin Transl Med 11, e370. 2021. doi: 10.1002/ctm2.370

31. Wang, Y. et al. Exosomes Secreted by Adipose-Derived Mesenchymal Stem Cells Foster Metastasis and Osteosarcoma Proliferation by Increasing COLGALT2 Expression. Front Cell Dev Biol 8, 353. 2020. doi: 10.3389/fcell.2020.00353

32. Wang, S. et al. Upregulation of GLT25D1 in Hepatic Stellate Cells Promotes Liver Fibrosis via the TGF-beta1/SMAD3 Pathway In Vivo and In vitro. J Clin Transl Hepatol 11, 1-14. 2023. doi: 10.14218/JCTH.2022.00005

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. doi: 10.1371/journal.pone.0029390

34. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. Nucleic Acids Res 38, W545-9. 2010. doi: 10.1093/nar/gkq366

35. Luther, K.B. et al. Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J Biol Chem 286, 43701-9. 2011. doi: 10.1074/jbc.M111.309096

36. Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem 77, 521-55. 2008. doi: 10.1146/annurev.biochem.76.061005.092322

37. Taujale, R. et al. Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. elife 9. 2020. doi: 10.7554/eLife.54532

38. Miyatake, S. et al. Biallelic COLGALT1 variants are associated with cerebral small vessel disease. Ann Neurol 84, 843-853. 2018. doi: 10.1002/ana.25367

39. Teunissen, M.W.A. et al. Biallelic Variants in the COLGALT1 Gene Causes Severe Congenital Porencephaly: A Case Report. Neurol Genet 7, e564. 2021. doi: 10.1212/NXG.0000000000000564

40. Scietti, L. et al. Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. Nat Commun 9, 3163. 2018. doi: 10.1038/s41467-018-05631-5

41. Sadakierska-Chudy, A., Szymanowski, P., Lebioda, A. & Ploski, R. Identification and In Silico Characterization of a Novel COLGALT2 Gene Variant in a Child with Mucosal Rectal Prolapse. Int J Mol Sci 23. 2022. doi: 10.3390/ijms23073670

42. Salo, A.M. et al. Lysyl hydroxylase 3 (LH3) modifies proteins in the extracellular space, a novel mechanism for matrix remodeling. J Cell Physiol 207, 644-53. 2006. doi: 10.1002/jcp.20596

43. Wang, C., Ristiluoma, M.M., Salo, A.M., Eskelinen, S. & Myllyla, R. Lysyl hydroxylase 3 is secreted from cells by two pathways. J Cell Physiol 227, 668-75. 2012. doi: 10.1002/jcp.22774

44. Chen, Y.L. et al. Lysyl Hydroxylase 2 Is Secreted by Tumor Cells and Can Modify Collagen in the Extracellular Space. J Biol Chem 291, 25799-25808. 2016. doi: 10.1074/jbc.M116.759803

45. Banushi, B. et al. Regulation of post-Golgi LH3 trafficking is essential for collagen homeostasis. Nat Commun 7, 12111. 2016. doi: 10.1038/ncomms12111

46. Studier, F.W. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41, 207-34. 2005. doi: 10.1016/j.pep.2005.01.016

47. Gorrec, F. The MORPHEUS protein crystallization screen. J Appl Crystallogr 42, 1035-1042. 2009. doi: 10.1107/S0021889809042022

48. Kabsch, W. Xds. Acta Crystallogr D Biol Crystallogr 66, 125-32. 2010. doi: 10.1107/S0907444909047337

49. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69, 1204-14. 2013. doi: 10.1107/S0907444913000061

50. Pape, T. & Schneider, T.R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. Journal of Applied Crystallography 37, 843-844. 2004. doi: 10.1107/S0021889804018047

51. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr 62, 1002-11. 2006. doi: 10.1107/S0907444906022116

52. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213-21. 2010. doi: 10.1107/S0907444909052925

53. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67, 355-67. 2011. doi: 10.1107/S0907444911001314

54. Gore, S. et al. Validation of Structures in the Protein Data Bank. Structure 25, 1916-1927. 2017. doi: 10.1016/j.str.2017.10.009

55. Brennich, M.E., Round, A.R. & Hutin, S. Online Size-exclusion and Ion-exchange Chromatography on a SAXS Beamline. J Vis Exp. 2017. doi: 10.3791/54861

56. Panjkovich, A. & Svergun, D.I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small angle X-ray scattering data. Bioinformatics. 2017. doi: 10.1093/bioinformatics/btx846

57. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography 36, 1277-1282. 2003. doi: 10.1107/S0021889803012779

58. Panjkovich, A. & Svergun, D.I. SASpy: a PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data. Bioinformatics 32, 2062-4. 2016. doi: 10.1093/bioinformatics/btw071

59. Petoukhov, M.V. et al. New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography 45, 342-350. 2012. doi: 10.1107/S0021889812007662

60. Mattoteia, D. et al. Identification of Regulatory Molecular "Hot Spots" for LH/PLOD Collagen Glycosyltransferase Activity. Int J Mol Sci 24. 2023. doi: 10.3390/ijms241311213

61. Iacobucci, C. et al. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. Nat Protoc 13, 2864-2889. 2018. doi: 10.1038/s41596-018-0068-8

62. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26, 1367-72. 2008. doi: 10.1038/nbt.1511

63. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966-8. 2010. doi: 10.1093/bioinformatics/btq054

64. Laskowski, R.A. & Swindells, M.B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51, 2778-86. 2011. doi: 10.1021/ci200227u

65. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Crystallogr 50, 1212-1225. 2017. doi: 10.1107/S1600576717007786

66. Svergun, D., Barberato, C. & Koch, M.H.J. CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography 28, 768-773. 1995. doi: 10.1107/S0021889895007047

67. Rojas Echeverri, J.C. et al. A Workflow for Improved Analysis of Cross-Linking Mass Spectrometry Data Integrating Parallel Accumulation-Serial Fragmentation with MeroX and Skyline. Anal Chem 96, 7373-7379. 2024. doi: 10.1021/acs.analchem.4c00829

Chapter 6 STRUCTURAL INVESTIGATION OF PLOD1: A Cryo-EM SINGLE PARTICLE ANALYSIS

Contribution: Cryo-EM SPA data analysis and LH1 model building.

ABSTRACT

Lysine hydroxylases as described in the previous chapter is a modification that is particularly important and crucial for the collagen's cross-links. While the experimental structure of PLOD3/LH3 has been well characterized, few biochemical data are available for PLOD1/LH1. In this chapter the current results of LH1 structure are presented, obtained through SPA cryo-EM. These preliminary data show an unusual quaternary structure of LH1. It is a dimer of dimer with a pseudo C2 symmetry, and up to date no other PLOD isoforms have been described with such organization. The physiological role of this quaternary organization is still to be understood, as well as its role in collagen maturation and cross-links formation.

MATERIAL AND METHODS

RECOMBINANT PRODUCTION OF LH1

Recombinant his-TEV-PLOD1 plasmid DNA (pUPE.106.08-LH1) was used to make transient transfection in HEK293 human cells, for a detailed description of the methods for cells maintenance and transfection see (mettere articolo silvia). The cells were harvested 6 days after the transfection by centrifugation. The first step of centrifugation was done at 1000 g for 10 minutes at 4°C; the pellet was discarded and the supernatant recollected. The recovered supernatant was further centrifuged at 5000 g for 10 minutes, discarding again the pellet and recollecting the cell-free supernatant containing the recombinant LH1. The pH of the supernatant was adjusted adding buffer A [25 mM HEPES, 500 mM NaCl, pH 8.00] until was reached the pH point of 8.0. The solution was then filtered with a 0.8 µm glass fiber filter. For the IMAC chromatographic purification two buffers were utilized: buffer A [25 mM HEPES, 500 mM NaCl, pH 8.00] and buffer B [25 mM HEPES, 500 mM NaCl, 500 mM Imidazole, pH 8.00]. At the flow of 1 ml min-1 utilizing an Äkta Prime system (Cytiva) the supernatant was loaded into a 5 ml His Trap excel column (Cytiva) at 4°C. At a flow of 5 ml min-1 utilizing an Äkta Go system (Cytiva), a step gradient [25 mM imidazole and 250 mM imidazole] was applied to separate the contaminant proteins from the protein of interest. A volume of 10 uL was collected for all the fractions for further SDS-PAGE analysis. The eluate containing LH1 was then loaded on a 5 ml desalting column pre-equilibrated with buffer A, to remove the excess of imidazole. The protein was concentrated with a 30 kDa MWCO Vivaspin Turbo filters (Sartorius) till was reached the concentration of 1 mg ml-1. To remove the His-TEV tag, the His-tagged TEV protease was added to the eluate containing LH1 with a 1:20 v/v ratio at 4°C over night. The cleaved protein was then loaded into a 5 ml His Trap excel column, to separate the cleaved protein from the tag and/or additional contaminant material. The sample was then further concentrated to 1 mg ml-1 and loaded on a Superdex 200 increase 10/300 (Cytiva) pre-equilibrated with a buffer [25 mM HEPES, 150 mM NaCl, pH 8.00]. The aliquots containing LH1 were divided into aliquots of 10 uL, flash-freeze in liquid nitrogen and stored at -80°C.

CRYO-EM SAMPLE PREPARATION

Freshly prepared samples of LH1 were used to prepare cryo-EM grids. Quantifoil R 1.2/1.3 Cu 300 mesh grids (Quantifoil) were glow discharged with the following parameter: pressure 0.39 mBar, current 20 mA, time 45 s, negative polarity. The Vitrobot Mark IV System was used for the specimen vitrification process (Thermo Fisher), 4 L samples were applied on the grid at the temperature of 4 °C and with a 95% humidity. The blotting parameters were 2 s, 4 s, 6 s and 8 s as blotting time with no blotting force applied. Each grid was prepared

in duplicate. The grids were immediately loaded into the microscope or stored in liquid nitrogen at -80° C.

CRYO-EM DATA ACQUISITION AND DATA PROCESSING

Optimal grids suitable for high resolution data collection were screened in house with a Thermo Fischer Scientific Glacios, equipped with a high-brightness field emission gun X-FEG, 200 kV. The system is equipped with a CMOS detector Ceta 16M Camera, and a direct electron detector Falcon 3EC. The EPU software (Thermo Fisher Scientific) was used to handle the grids screening and data collection. The grids which displayed the best sample homogeneity and quality were shipped to the cryo-EM facility at ESRF, Grenoble. A high-resolution data collection was performed on a Titan Krios G4 (Thermo Fischer Scientific). For the data collection 3 acquisition per hole were set up with the following parameters: magnification 105 K, pixel size 0.42 Å/pixel, total dose per movie 42.87 electron/Å2, 40 frames per movies, 1.77 s of exposure, defocus spread from -0.8 to -2 µm. The dataset was then processed in-house with the software platform Cryosparc, for both pre-processing as well as post-processing tasks. The movies were imported to Cryosparc and the reference gain was provided to apply the proper corrections. After motion correction and CTF estimation, 1.114.155 particles were selected and extracted applying a Fourier cropping of 884 px→448px. After the first run of 2D classification 11 classes were identified as optimal, they were selected for the next steps of 2D classification leading to 149.596 particles for ab-initio reconstruction. Two initial volumes were obtained from the ab-initio reconstruction, then used as starting volume for a first step of heterogeneous refinement. The best volume obtained from the heterogeneous refinement job, reconstructed with 29.932 particles, was used as a template for a homogeneous refinement using all the 149.596 initial particles. C2 symmetry expansion was applied to the set of particles used in the homogeneous refinement and then were used for a 3D variability job. Manual masks were created on Chimera, imported in Cryosparc and used for Local refinement jobs. The final was feature-enhanced with the algorithm Locscale 2.0, to improve the interpretability of the volume. Global and directional FSC were calculated on the Remote 3DFSC Processing Server. A LH1 model was generated using LH3 structure as template, thus was docked into the LH1volume and refined in coot (Fig.1)

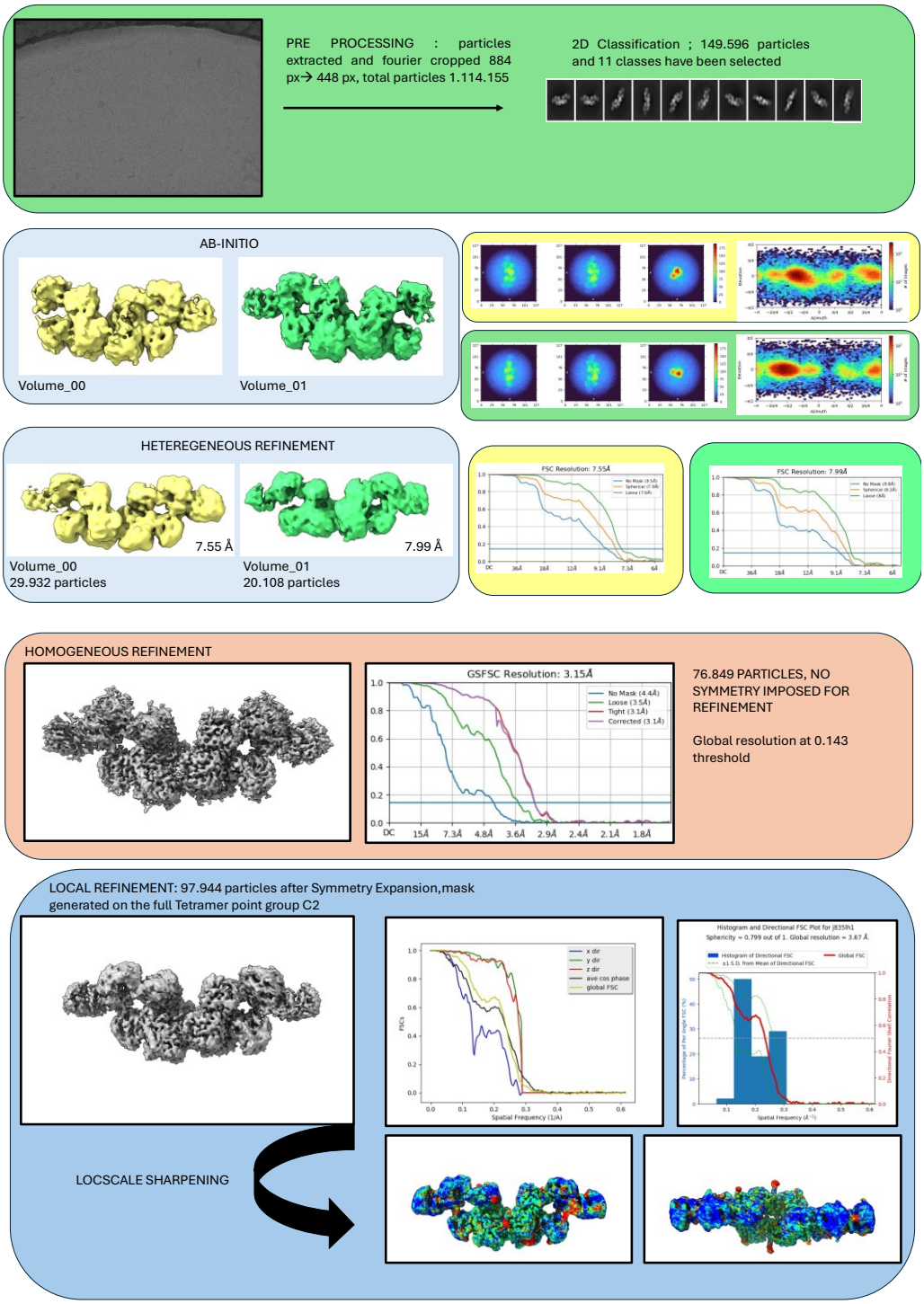


Figure 1: Schematic graphical representation of LH1 data processing: pre-processing, ab-initio reconstruction, volume post-processing and validation.

RESULT AND DISCUSSION

The recombinant protein LH1 has been successfully produced in HEK293 human cells and purified. The LH1 sample, compared to the isoform LH3, is particularly much more prone to aggregation events, especially at concentrations higher than 1 mg ml-1. Notwithstanding the complexity of the purification procedure and the low final yield, homogeneous and highly pure sample, suitable for SPA cryo-EM analysis, has been obtained. Grids of good quality have been obtained, with an ice quality optimal for high resolution data collection. After the preprocessing of the collected movies, only 13% of the total extracted particles have been used to start the initial LH1 volume reconstruction. This is because the 2D classes showing the LH1 dimer of dimers were just a sub-population in the whole dataset. The reconstructed volume clearly suffers from strong anisotropy, especially along the x axis. This makes the interpretability of LH1 volume especially difficult at the AC and GGT domain, where the intrinsic flexibility of the GGT domains enhances the complexity underneath the interpretation of this volume. However, a good starting volume has been obtained with a series of homogeneous reconstruction. The surprising feature that has been discovered in this processing, is the unusual quaternary structure of LH1 compared to the one of LH3. This isoform seems to be like two dimers of LH3 which associate together at the level of the LH domains; each dimer has two GGT domain, one of them is very flexible and distal from all the other domains, while the other is perfectly underneath the AC domain of the counter dimer which make it much more rigid. This arrangement makes the entire ensemble much more like a dimer of dimmers, with a pseudo C2-symmetry, than a perfectly symmetric C2 tetramer. Hence, imposing the C2 symmetry through the 3D refinement jobs clearly improve the overall quality of the reconstructed volume. It is worth mentioning that glycosylation on LH1 volume has been seen, and in more detail are clear the ones present on the LH domains and on the AC domains (Fig.2). To try to bust the high-resolution information at the GGT and AC domains, local refinement jobs have been performed. Thus, mask have been manually created using as a template the best volume obtained, different approaches have been tried: 1) the generation of masks covering only the GGT domains, thus pushing the volume sharpening only on the LH1 GGT tip applying different fulcrum of rotation; 2) the generation of masks covering isolated dimer, using as a fulcrum of rotation the centre of the oligomeric ensemble; 3) the generation of a mask covering the entire volume, to exclude the high information noise around the original sharpened volume. Hence, the final volume has been obtained, and through modern IA based software a local sharpening has been performed using LocScale 2.0.

The LH1 volume obtained is now currently used as a starting point for model building. However, after a first rigid docking of LH1 model into the volume and the following first refinement steps, the information of the AC and GGT domain are not still enough to precisely model these regions. Thus, trials to improve the overall sample quality, and hence the volume interpretability are still ongoing. The next goal of the project is to achieve detailed atomistic resolution of LH1 in complex with important interactors, such as Cyclophyllin B, which have a crucial physiological importance.

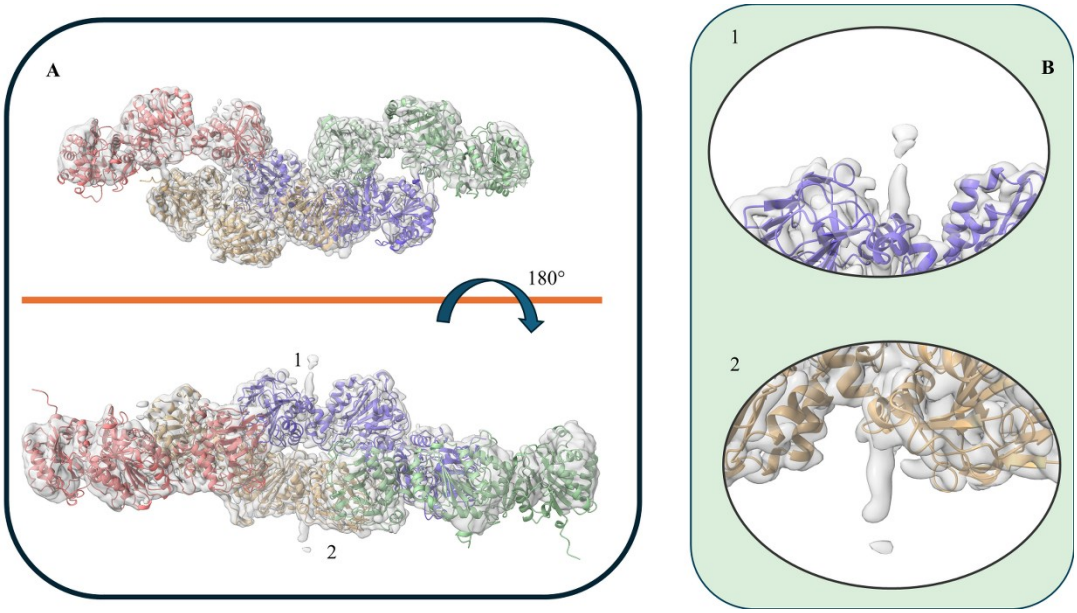


Figure 2. The enzyme LH1 is a dimer of dimers with the interface of oligomerization involving LH and GT domains (each monomer has a unique colour). Between the AC and LH domains is possible to see additional volume density (1, 2) suggesting PLOD1 glycosylation (section B zoom of the extra volume seen).

Chapter 7 CONCLUSION

The synergistic interplay between cells, small molecules as well as biological macromolecules, creates a unique environment which enables the formation of tissue and organs in multicellular organisms; in other words, this can be named as the extracellular matrix. The papers proposed in this thesis, however, are related to specific enzymes which participate in the post-translational modifications of collagens molecules; the lysine hydroxylase isoform 1 and 3 (PLOD1/PLOD3) and the procollagen galactosyl transferase GLT25D1.

The increasing number of the evidence correlating the excessive Lys→Hyl conversion to tumours progression and metastasis, make PLOD enzyme valuable targets for the development of antitumoral drugs. In chapter 4 has been presented the work done on LH3, to develop new antitumoral drugs. The 2-OG-competitive inhibitors or the structure-guided libraries are not effective to inhibit LH activity; thus, this moved us to explore other possibility to achieve this aim. Our experimental observations, gathered with MD simulations, revealed that while Fe^{2+} is essential for lysine hydroxylase activity, an excess of this metal ion leads LH3 into an allosteric self-inhibited state. This because the LH domain host two Fe^{2+} ions, one in the catalytic domain coordinated by the residues His595, Asp611 and His613 and the other one is coordinated by the residues Asp597 in proximity of the capping loop. Asp611 or Asp597 mutations are correlated with a significant improvement in the uncoupled conversion of 2-OG into succinate, suggesting a more accessible catalytic pocket. As well as Fe^{2+}, metal ions chelators, like 2,2'-bypyridine (BPY), show different effect on LH3 activity: at low concentrations they enhance LH3 activity, while at high concentrations they inhibit LH3 activity. The effect of low BPY concentrations can be explained by the chelating effect that it has toward the Fe^{2+} in proximity to the external capping loop, thus making the catalytic site much more accessible. These findings are essential to start the development of new LH3 inhibitors which could effectively interfere with PLOD activity, thus proposing a spectrum of candidates for the development of new antitumoral drugs.

Two of the works which have been presented (chapter 3 and 5) are crucial to understanding the structural and biochemical features of glycosyltransferase. These family of enzymes, despite the broad spectrum of substrates that they can process, have similar mechanism of action so that is possible to analysed them with structural comparisons. In the work about the molecular 'hot spot' of LH3 glycosyltransferases domain, has been revealed how, however the similarity in the PLOD GGT domain superfold,, PLOD 3 has unique features which enable high Glc-T transferase activity of this isoform compared to the others. A structured based guided mutagenesis has been used to map the critical residues involved in LH3 Glc-T activity. A first region identified as important for enzymatic activity is a flexible region named as glycoloop. In this region Val80 is a critical residue for PLOD3 activity (Lys68 in LH1 and Gly80 in LH2). When this residue is mutated the Glc-T activity of LH3

is drastically compromised, this as result of an increased degree of flexibility of the glycoloop: leading to an incorrect positioning of the residue toward the donor substrate. However, there are other residues in the glycoloop which synergistically coordinate the reaction's substrate. For example, Trp92 (Leu80 in LH1 and Leu92 in LH2) is critical to stabilize the LH3 glycoloop through its aromatic ring. However, its mutation does not completely abolish LH3 Glc-T activity. The second region which has been described as critical for PLOD Glc-T activity is the poly-Asp-helix [290, 291]. In LH3 the residue Asp190 when mutated to Ser (making LH3 like LH1) reduce drastically the Glc-T activity, while in LH1 the mutation Ser178Asp (hence making LH1 like LH3) surprisingly increase Lh1 Glc.T activity. In the Gt active site of LH3 there is a non-conserved loop with two aromatic residues, Trp145 and Trp148. In other GT-A glycosyltransferases, usually the Position of Trp145 is occupied by aromatic residues, and it seems a critical residue involved in the catalysis. Trp148 is more open and less conserved than Trp145 and its role is still not completely clear; it could be involved in the acceptor recognition through long range stabilizing interactions with collagen. Another residue identified near the poly-Asp-helix is Gln192, which lie away from the donor binding site, thus having a minor role into the direct catalysis of the substrate. The other residues which are strictly related to the processing of the donor glycan are Glu141 and Asn165, with their side chains pointing directly toward the donor sugar. The mutation Asn165Ala strongly reduce the Glc-T activity, because this residue is involved in the activation or in the transfer of the donor substrate. While mutations of the residue Glu141 lead to a complete absence of the Glc-T activity of LH3, making it as the main 'hot spot' for the LH3 glycosyltransferase activity.

This work has been of crucial importance since the crystal structure of the procollagen galactosyltransferase GLT25D1 was solved. The GGT domain of LH3 as well as the GT domains of GLT25D1 are GT-A glycosyltransferases, however there are deep difference among them which define their substrate specificity and affinity. The enzyme GLT25D1, is a head-to-head homodimer, each monomer is composed of two globular domains interconnected by a rigid loop. Has been demonstrated that the catalytically active domain is the C-terminal GT2 domain, while the N-terminal GT1 domain has a primary role into dimerization and in the interaction with LH3 [292]. While the GT1 domain is closely related to the LH3 AC domain, the topology of the GT2 domain is different. There is a high conservation of the residues around the Mn^{2+} and the UDP-α-galactose in the GT2 domain, the metal ion is coordinated by an EDD motif instead of the canonical DxD motif which usually characterises inverting GT-A glycosyltransferases. Between the apo and the donor-bound structure there are no big differences, aside from the rearrangement of the flexible C-terminal segment which acts as a gating loop. Thus, both LH3 and GLT25D1 show elongated, s-shaped, quaternary structure which could be a critical geometrical configuration for the recognition and processing of collagen molecules. However, they have catalytic domains with unique structural features which reflect the peculiarity of the collagenous molecules and their unusual set of post-translational modifications.

In the last chapter have been described the recent advancement in the LH1 structure determination and has been highlighted the unusual quaternary structure which this enzyme adopts. Is still not clear the reason of such organization and few experimental data related to this are present in literature. One putative speculation is the interaction of LH1 with the protein CYPB, thus making the interaction between these proteins critical for collagen

secretion[293, 294]. However, further data need to be collected to decipher the complex relationship between the structures of these enzymes with their substrate specificity and/or their intricated protein-protein interaction network.

Chapter 8 REFERENCES

1. Özbek S, Balasubramanian PG, Chiquet-Ehrismann R, Tucker R, Adams JC. The Evolution of Extracellular Matrix. Molecular biology of the cell. 2010. https://doi.org/10.1091/mbc.e10-03-0251.

2. Karamanos N, Theocharis A, Piperigkou Z, Manou D, Passi A, Skandalis S, et al. A guide to the composition and functions of the extracellular matrix. The FEBS Journal. 2021. https://doi.org/10.1111/febs.15776.

3. Hynes R. The evolution of metazoan extracellular matrix. The Journal of cell biology. 2012. https://doi.org/10.1083/jcb.201109041.

4. Subramanian A, Kanzaki LF, Galloway J, Schilling T. Mechanical force regulates tendon extracellular matrix organization and tenocyte morphogenesis through TGFbeta signaling. eLife. 2018. https://doi.org/10.7554/elife.38069.

5. Ishikawa Y, Ito S, Nagata K, Sakai LY, Bächinger HP. Intracellular mechanisms of molecular recognition and sorting for transport of large extracellular matrix molecules. Proc Natl Acad Sci USA. 2016;113. https://doi.org/10.1073/pnas.1609571113.

6. Graf F, Horn P, Ho A, Boutros M, Maercker C. The extracellular matrix proteins type I collagen, type III collagen, fibronectin, and laminin 421 stimulate migration of cancer cells. The FASEB Journal. 2021. https://doi.org/10.1096/fj.202002558rr.

7. Liao J, Chen R, Lin B, Deng R, Liang Y, Zeng J, et al. Cross-Talk between the TGF-β and Cell Adhesion Signaling Pathways in Cancer. International Journal of Medical Sciences. 2024;21:1307–20. https://doi.org/10.7150/ijms.96274.

8. Corda S, Samuel J, Rappaport L. Extracellular Matrix and Growth Factors During Heart Growth. Heart Failure Reviews. 2000;5:119–30. https://doi.org/10.1023/A:1009806403194.

9. Sack K, Teran M, Nugent M. Extracellular Matrix Stiffness Controls VEGF Signaling and Processing in Endothelial Cells. Journal of Cellular Physiology. 2016;231. https://doi.org/10.1002/jcp.25312.

10. Townsend S, Gannon M. Extracellular matrix-associated factors play critical roles in regulating pancreatic β-cell proliferation and survival. Endocrinology. 2019. https://doi.org/10.1210/en.2019-00206.

11. Taipale J, Keski-oja J. Growth factors in the extracellular matrix. The FASEB Journal. 1997;11:51–9. https://doi.org/10.1096/fasebj.11.1.9034166.

12. Hyytiäinen M, Penttinen C, Keski-oja J. Latent TGF-β Binding Proteins: Extracellular Matrix Association and Roles in TGF-β Activation. Critical Reviews in Clinical Laboratory Sciences. 2004;41:233–64. https://doi.org/10.1080/10408360490460933.

13. Nakamura A, Mashima T, Lee J, Inaba S, Kawata N, Morino S, et al. Intratumor transforming growth factor-β signaling with extracellular matrix-related gene regulation marks chemotherapy-resistant gastric cancer. Biochemical and biophysical research communications. 2024;721:150108. https://doi.org/10.1016/j.bbrc.2024.150108.

14. Saharinen J, Hyytiäinen M, Taipale J, Keski-oja J. Latent transforming growth factor-beta binding proteins (LTBPs)--structural extracellular matrix proteins for targeting TGF-beta action. Cytokine & growth factor reviews. 1999;10 2:99–117. https://doi.org/10.1016/S1359-6101(99)00010-6.

15. MacFarlane E, Haupt J, Dietz H, Shore E. TGF-β Family Signaling in Connective Tissue and Skeletal Diseases. Cold Spring Harbor perspectives in biology. 2017;9 11. https://doi.org/10.1101/cshperspect.a022269.

16. Roberts A, McCune B, Sporn M. TGF-β: Regulation of extracellular matrix. Kidney International. 1992;41:557–9. https://doi.org/10.1038/KI.1992.81.

17. Firmansyah Y, Sidharta V, Wijaya L, Tan ST. Unraveling the Significance of Growth Factors (TGF-β, PDGF, KGF, FGF, Pro Collagen, VEGF) in the Dynamic of Wound Healing. Asian Journal of Medicine and Health. 2024. https://doi.org/10.9734/ajmah/2024/v22i3992.

18. Exposito J, Valcourt U, Cluzel C, Lethias C. The Fibrillar Collagen Family. International Journal of Molecular Sciences. 2010;11:407–26. https://doi.org/10.3390/ijms11020407.

19. Mak K, Mei R. Basement Membrane Type IV Collagen and Laminin: An Overview of Their Biology and Value as Fibrosis Biomarkers of Liver Disease. The Anatomical Record. 2017;300. https://doi.org/10.1002/ar.23567.

20. Sipilä L, Ruotsalainen H, Sormunen R, Baker N, Lamandé S, Vapola M, et al. Secretion and Assembly of Type IV and VI Collagens Depend on Glycosylation of Hydroxylysines*. Journal of Biological Chemistry. 2007;282:33381–8. https://doi.org/10.1074/jbc.M704198200.

21. Gatseva A, Sin YY, Brezzo G, Van Agtmael T. Basement membrane collagens and disease mechanisms. Essays in Biochemistry. 2019;63:297–312. https://doi.org/10.1042/EBC20180071.

22. Fitzgerald J, Bateman J. A new FACIT of the collagen family: COL21A1. FEBS Letters. 2001;505. https://doi.org/10.1016/S0014-5793(01)02754-5.

23. Sadri G, Fischer A, Brittian K, Elliott E, Nystoriak M, Uchida S, et al. Collagen type XIX regulates cardiac extracellular matrix structure and ventricular function. Matrix biology : journal of the International Society for Matrix Biology. 2022;109:49–69. https://doi.org/10.1016/j.matbio.2022.03.007.

24. Grässel S, Bauer R. Collagen XVI in health and disease. Matrix biology : journal of the International Society for Matrix Biology. 2013;32 2:64–73. https://doi.org/10.1016/j.matbio.2012.11.001.

25. Käpylä J, Jäälinoja J, Tulla M, Ylöstalo J, Nissinen L, Viitasalo T, et al. The Fibril-associated Collagen IX Provides a Novel Mechanism for Cell Adhesion to Cartilaginous Matrix*. Journal of Biological Chemistry. 2004;279:51677–87. https://doi.org/10.1074/jbc.M409412200.

26. Gogiel T, Bańkowski E. [New collagenous proteins: FACIT collagens, transmembrane collagens and multiplexins]. Postepy higieny i medycyny doswiadczalnej. 2001;55 1:133–56.

27. Ayad S, Boot-Handford R, Humphries M, Kadler K, Shuttleworth A. Collagen type XII. 1998;:99–102. https://doi.org/10.1016/B978-012068911-8.50117-2.

28. Ayad S, Boot-Handford R, Humphries M, Kadler K, Shuttleworth A. Collagen type XIV. 1998;:106–8. https://doi.org/10.1016/B978-012068911-8.50119-6.

29. Ayad S, Boot-Handford R, Humphries M, Kadler K, Shuttleworth A. Collagen type XIX. 1998;:123–5. https://doi.org/10.1016/B978-012068911-8.50124-X.

30. Ayad S, Boot-Handford R, Humphries M, Kadler K, Shuttleworth A. Collagen type XVI. 1998;:112–4. https://doi.org/10.1016/B978-012068911-8.50121-4.

31. Snellman A, Keränen M, Hägg P, Lamberg A, Hiltunen J, Kivirikko K, et al. Type XIII Collagen Forms Homotrimers with Three Triple Helical Collagenous Domains and Its Association into Disulfide-bonded Trimers Is Enhanced by Prolyl 4-Hydroxylase*. The Journal of Biological Chemistry. 2000;275:8936–44. https://doi.org/10.1074/JBC.275.12.8936.

32. Watanabe M, Natsuga K, Nishie W, Kobayashi Y, Donati G, Suzuki S, et al. Type XVII collagen coordinates proliferation in the interfollicular epidermis. eLife. 2017;6. https://doi.org/10.7554/eLife.26635.

33. Banyard J, Bao L, Zetter B. Type XXIII Collagen, a New Transmembrane Collagen Identified in Metastatic Tumor Cells*. Journal of Biological Chemistry. 2003;278:20989–94. https://doi.org/10.1074/JBC.M210616200.

34. Izzi V, Heljasvaara R, Heikkinen A, Karppinen S, Koivunen J, Pihlajaniemi T. Exploring the roles of MACIT and Multiplexin collagens in stem cells and cancer. Seminars in cancer biology. 2020. https://doi.org/10.1016/j.semcancer.2019.08.033.

35. Harpaz N, Ordan E, Ocorr K, Bodmer R, Volk T. Multiplexin Promotes Heart but Not Aorta Morphogenesis by Polarized Enhancement of Slit/Robo Activity at the Heart Lumen. PLoS Genetics. 2013;9. https://doi.org/10.1371/journal.pgen.1003597.

36. Sorushanova A, Delgado L, Wu Z, Shologu N, Kshirsagar A, Raghunath R, et al. The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development. Advanced Materials. 2018;31. https://doi.org/10.1002/adma.201801651.

37. Ricard-Blum S. The collagen family. Cold Spring Harbor perspectives in biology. 2010;3 1. https://doi.org/10.1101/cshperspect.a004978.

38. Leonard A, Cumming M, Ali MA, Cabral J. Fish Collagen Cross-Linking Strategies to Improve Mechanical and Bioactive Capabilities for Tissue Engineering and Regenerative Medicine. Advanced Functional Materials. 2024;34. https://doi.org/10.1002/adfm.202405335.

39. Goldbloom-Helzner L, Hao D, Wang A. Developing Regenerative Treatments for Developmental Defects, Injuries, and Diseases Using Extracellular Matrix Collagen-Targeting Peptides. International Journal of Molecular Sciences. 2019;20. https://doi.org/10.3390/ijms20174072.

40. Antonio JS, Jacenko O, Fertala A, Orgel J. Collagen Structure-Function Mapping Informs Applications for Regenerative Medicine. Bioengineering. 2020;8. https://doi.org/10.3390/bioengineering8010003.

41. Benito-Martínez S, Pérez-Köhler B, Rodríguez M, Rivas-Santos C, Izco JM, Recalde J, et al. Assessing New Collagen Therapies for Wound Healing: A Murine Model Approach. International Wound Journal. 2025;22. https://doi.org/10.1111/iwj.70589.

42. Lin K, Zhang D, Macedo M, Cui W, Sarmento B, Shen G. Advanced Collagen-Based Biomaterials for Regenerative Biomedicine. Advanced Functional Materials. 2018;29. https://doi.org/10.1002/adfm.201804943.

43. Di Rosa L, De Pasquale A, Baldassano S, Marguglio N, Drid P, Proia P, et al. New Regenerative and Anti-Aging Medicine Approach Based on Single-Stranded Alpha-1 Collagen for Neo-Collagenesis Induction: Clinical and Instrumental Experience of a New Injective Polycomponent Formulation for Dermal Regeneration. Biomedicines. 2024;12. https://doi.org/10.3390/biomedicines12040916.

44. Lionetto F, Corcione E. Recent Applications of Biopolymers Derived from Fish Industry Waste in Food Packaging. Polymers. 2021;13. https://doi.org/10.3390/polym13142337.

45. Areniello M, Matassa S, Esposito G, Lens P. Biowaste upcycling into second-generation microbial protein through mixed-culture fermentation. Trends in biotechnology. 2022. https://doi.org/10.1016/j.tibtech.2022.07.008.

46. Salim N, Madhan B, Glattauer V, Ramshaw J. Comprehensive review on collagen extraction from food by-products and waste as a value-added material. International journal of biological macromolecules. 2024;:134374. https://doi.org/10.1016/j.ijbiomac.2024.134374.

47. Rigueto CVT, Rosseto M, Alessandretti I, De Oliveira R, Wohlmuth DAR, Menezes JF, et al. Gelatin films from wastes: A review of production, characterization, and application trends in food preservation and agriculture. Food research international. 2022;162 Pt B:112114. https://doi.org/10.1016/j.foodres.2022.112114.

48. Batista M, Fernández N, Gaspar F, Bronze M, Duarte A. Extraction of Biocompatible Collagen From Blue Shark Skins Through the Conventional Extraction Process Intensification Using Natural Deep Eutectic Solvents. Frontiers in Chemistry. 2022;10. https://doi.org/10.3389/fchem.2022.937036.

49. Seixas M, Martins E, Reis R, Silva T. Extraction and Characterization of Collagen from Elasmobranch Byproducts for Potential Biomaterial Use. Marine Drugs. 2020;18. https://doi.org/10.3390/md18120617.

50. Prajaputra V, Isnaini N, Maryam S, Ernawati E, Deliana F, Haridhi H, et al. Exploring Marine Collagen: Sustainable Sourcing, Extraction Methods, and Cosmetic Applications. South African Journal of Chemical Engineering. 2023. https://doi.org/10.1016/j.sajce.2023.11.006.

51. Qiang T, Chen L, Yan Z, Liu X. Evaluation of a Novel Collagenous Matrix Membrane Cross-Linked with Catechins Catalyzed by Laccase: A Sustainable Biomass. Journal of agricultural and food chemistry. 2019;67 5:1504–12. https://doi.org/10.1021/acs.jafc.8b05810.

52. Selvaraj V, Sekaran S, Dhanasekaran A, Warrier S. Type 1 collagen: Synthesis, structure and key functions in bone mineralization. Differentiation; research in biological diversity. 2024;136:100757. https://doi.org/10.1016/j.diff.2024.100757.

53. Ramachandran GN. Stereochemistry of collagen*. International Journal of Peptide and Protein Research. 1988;31:1–16. https://doi.org/10.1111/j.1399-3011.1988.tb00001.x.

54. Trelstad R, Hayashi K, Gross J. Collagen fibrillogenesis: intermediate aggregates and suprafibrillar order. Proceedings of the National Academy of Sciences of the United States of America. 1976;73 11:4027–31. https://doi.org/10.1073/PNAS.73.11.4027.

55. Wang C, Brisson B, Terajima M, Li Q, Hoxha K, Han B, et al. Type III Collagen is a Key Regulator of the Collagen Fibrillar Structure and Biomechanics of Articular Cartilage and

Meniscus. Matrix biology : journal of the International Society for Matrix Biology. 2020. https://doi.org/10.1016/j.matbio.2019.10.001.

56. Ishikawa Y, Lennon R, Forneris F, Myllyharju J, Salo AM. Collagen IV biosynthesis: Intracellular choreography of post-translational modifications. Matrix Biology. 2025;140:59–77. https://doi.org/10.1016/j.matbio.2025.07.002.

57. Grant M, Heathcote G, Orkin R. Current concepts of basement-membrane structure and function. Bioscience Reports. 1981;1:819–42. https://doi.org/10.1007/BF01114816.

58. Timpl R. Macromolecular organization of basement membranes. Current opinion in cell biology. 1996;8 5:618–24. https://doi.org/10.1016/S0955-0674(96)80102-5.

59. M M. Basement membrane proteins: structure, assembly, and cellular interactions. Critical reviews in biochemistry and molecular biology. 1992;27 1-2:93–127. https://doi.org/10.3109/10409239209082560.

60. Yurchenco P. Basement membranes: cell scaffoldings and signaling platforms. Cold Spring Harbor perspectives in biology. 2011;3 2. https://doi.org/10.1101/cshperspect.a004911.

61. Umumararungu T, Gahamanyi N, Mukiza J, Habarurema G, Katandula J, Rugamba A, et al. Proline, a unique amino acid whose polymer, polyproline II helix, and its analogues are involved in many biological processes: a review. Amino Acids. 2024. https://doi.org/10.1007/s00726-024-03410-9.

62. Persikov A, Ramshaw J, Kirkpatrick A, Brodsky B. Amino acid propensities for the collagen triple-helix. Biochemistry. 2000;39 48:14960–7. https://doi.org/10.1021/BI001560D.

63. Kirkness M, Lehmann K, Forde N. Mechanics and structural stability of the collagen triple helix. Current opinion in chemical biology. 2019;53:98–105. https://doi.org/10.1016/j.cbpa.2019.08.001.

64. Brodsky B, Ramshaw J. The collagen triple-helix structure. Matrix biology : journal of the International Society for Matrix Biology. 1997;15 8-9:545–54. https://doi.org/10.1016/S0945-053X(97)90030-5.

65. Bella J, Liu J, Kramer R, Brodsky B, Berman H. Conformational effects of Gly-X-Gly interruptions in the collagen triple helix. Journal of molecular biology. 2006;362 2:298–311. https://doi.org/10.1016/J.JMB.2006.07.014.

66. Arcoria P, Ware R, Makwana S, Troya D, Etzkorn F. Conformational Analysis of Fluoro-, Chloro-, and Proteo-Alkene Gly-Pro and Pro-Pro Isosteres to Mimic Collagen. The journal of physical chemistry B. 2021. https://doi.org/10.1021/acs.jpcb.1c09180.

67. Kobayashi M, Sim J, Sato H. Conformational analyses of collagen-like Co-Glycine/L-proline oligopeptides by quantum chemical calculation (QCC): Sequence effects on conformations and intra-molecular hydrogen bonds. Polymer Bulletin. 2021;79:6627–44. https://doi.org/10.1007/s00289-021-03805-8.

68. Li Y, Yang J, He X. Characterizing polyproline II conformational change of collagen superhelix unit on adsorption on gold surface. Nanoscale Advances. 2023;5:5322–31. https://doi.org/10.1039/d3na00185g.

69. Kreutzberger M, Yu L, Bui T, Hancu M, Purdy M, Osinski T, et al. A Collagen Triple Helix without the Superhelical Twist. ACS Central Science. 2025;11:331–45. https://doi.org/10.1021/acscentsci.5c00018.

70. Umashankara M, Sonar M, Bansode N, Ganesh K. Orchestration of Structural, Stereoelectronic, and Hydrogen-Bonding Effects in Stabilizing Triplexes from Engineered Chimeric Collagen Peptides (Pro(X)-Pro(Y)-Gly)6 Incorporating 4(R/S)-Aminoproline. The Journal of organic chemistry. 2015;80 17:8552–60. https://doi.org/10.1021/acs.joc.5b01032.

71. Etzkorn F, Ware R, Pester A, Troya D. Conformational Analysis of n→π* Interactions in Collagen Triple Helix Models. The journal of physical chemistry B. 2018;123 2:496–503. https://doi.org/10.1021/acs.jpcb.8b08384.

72. Suzuki H, Mahapatra D, Board A, Steel P, Dyer J, Gerrard J, et al. Sub-Ångstrom structure of collagen model peptide (GPO)10 shows a hydrated triple helix with pitch variation and two proline ring conformations. Food chemistry. 2020;319:126598. https://doi.org/10.1016/j.foodchem.2020.126598.

73. Adzhubei A, Sternberg M, Makarov A. Polyproline-II helix in proteins: structure and function. Journal of molecular biology. 2013;425 12:2100–32. https://doi.org/10.1016/j.jmb.2013.03.018.

74. Schweizer S, Bick A, Subramanian L, Krokidis X. Influences on the stability of collagen triple-helix. Fluid Phase Equilibria. 2014;362:113–7. https://doi.org/10.1016/J.FLUID.2013.09.033.

75. Chen C-C, Hsu W, Hwang K, Hwu J, Lin C, Horng J. Contributions of cation-π interactions to the collagen triple helix stability. Archives of biochemistry and biophysics. 2011;508 1:46–53. https://doi.org/10.1016/j.abb.2011.01.009.

76. Bella J, Brodsky B, Berman H. Hydration structure of a collagen peptide. Structure. 1995;3 9:893–906. https://doi.org/10.1016/S0969-2126(01)00224-6.

77. Chiang Y, Lin Y, Horng J. Stereoelectronic effects on the transition barrier of polyproline conformational interconversion. Protein Science. 2009;18:1967–77. https://doi.org/10.1002/pro.208.

78. DeTar DF, Luthra NP. Conformations of proline. J Am Chem Soc. 1977;99:1232–44. https://doi.org/10.1021/ja00446a040.

79. Orgel J, Persikov A, Antipova O. Variation in the Helical Structure of Native Collagen. PloS one. 2014. https://doi.org/10.1371/journal.pone.0089519.

80. Fechete I. Hermann Emil Fischer – The most outstanding chemist in history. Comptes Rendus Chimie. 2016;19:1143–9. https://doi.org/10.1016/j.crci.2016.08.002.

81. Ramshaw J, Shah N, Brodsky B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. Journal of structural biology. 1998. https://doi.org/10.1006/jsbi.1998.3977.

82. Pauling L, Niemann C. The Structure of Proteins.

83. Panasik N, Eberhardt E, Edison A, Powell D, Raines R. Inductive effects on the structure of proline residues. International journal of peptide and protein research. 2009. https://doi.org/10.1111/j.1399-3011.1994.tb00169.x.

84. Eberhardt ES, Panisik N, Raines RT. Inductive Effects on the Energetics of Prolyl Peptide Bond Isomerization: Implications for Collagen Folding and Stability. Journal of the American Chemical Society. 1996. https://doi.org/10.1021/ja9623119.

85. Bretscher L, Jenkins C, Taylor KM, Derider M, Raines R. Conformational stability of collagen relies on a stereoelectronic effect. Journal of the American Chemical Society. 2001. https://doi.org/10.1021/ja005542v.

86. Derider M, Wilkens SJ, Waddell M, Bretscher L, Weinhold F, Raines R, et al. Collagen stability: insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. Journal of the American Chemical Society. 2002. https://doi.org/10.1021/ja0166904.

87. Hentzen N, Islami V, Köhler M, Zenobi R, Wennemers H. A Lateral Salt Bridge for the Specific Assembly of an ABC-Type Collagen Heterotrimer. Journal of the American Chemical Society. 2020. https://doi.org/10.1021/jacs.9b13037.

88. Freudenberg U, Behrens S, Welzel P, Müller M, Grimmer M, Salchert K, et al. Electrostatic interactions modulate the conformation of collagen I. Biophysical journal. 2007;92 6:2108–19. https://doi.org/10.1529/BIOPHYSJ.106.094284.

89. Chiu H-S, Horng J. Modulating the Stability of Collagen Triple Helices by Terminal Charged Residues. The journal of physical chemistry B. 2021. https://doi.org/10.1021/acs.jpcb.1c01631.

90. Zheng H, Liu H, Hu J, Xu F. Stabilizing effects of pairwise salt bridges between acidic and basic residues in a collagen heterotrimer. Chemical Physics Letters. 2018. https://doi.org/10.1016/J.CPLETT.2018.10.025.

91. Fallas J, Dong J, Tao Y, Hartgerink J. Structural Insights into Charge Pair Interactions in Triple Helical Collagen-like Proteins*. The Journal of Biological Chemistry. 2011;287:8039–47. https://doi.org/10.1074/jbc.M111.296574.

92. Gurry T, Nerenberg P, Stultz C. The contribution of interchain salt bridges to triple-helical stability in collagen. Biophysical journal. 2009;98 11:2634–43. https://doi.org/10.1016/j.bpj.2010.01.065.

93. Okuyama K, Miyama K, Mizuno K, Bachinger HP. Crystal Structure of (Gly-Pro-Hyp)9: 3b0s. 2012. https://doi.org/10.2210/pdb3b0s/pdb.

94. Raote I, Saxena S, Malhotra V. Sorting and Export of Proteins at the Endoplasmic Reticulum. Cold Spring Harbor perspectives in biology. 2022. https://doi.org/10.1101/cshperspect.a041258.

95. Hutchings J, Stancheva V, Miller E, Zanetti G. Subtomogram averaging of COPII assemblies reveals how coat organization dictates membrane shape. Nature Communications. 2018;9. https://doi.org/10.1038/s41467-018-06577-4.

96. Saxena S, Foresti O, Liu A, Androulaki S, Rodriguez MP, Raote I, et al. Endoplasmic reticulum exit sites are segregated for secretion based on cargo size. bioRxiv. 2023. https://doi.org/10.1101/2023.12.07.570627.

97. Saito K, Maeda M, Katada T. Regulation of the Sar1 GTPase Cycle Is Necessary for Large Cargo Secretion from the Endoplasmic Reticulum. Frontiers in Cell and Developmental Biology. 2017;5. https://doi.org/10.3389/fcell.2017.00075.

98. Raote I, Ortega-Bellido M, Santos A, Foresti O, Zhang C, Garcia-Parajo M, et al. TANGO1 builds a machine for collagen export by recruiting and spatially organizing COPII, tethers and membranes. eLife. 2018;7. https://doi.org/10.7554/eLife.32723.

99. Bard F, Casano L, Mallabiabarrena A, Wallace E, Saito K, Kitayama H, et al. Functional genomics reveals genes involved in protein secretion and Golgi organization. Nature. 2006;439:604–7. https://doi.org/10.1038/nature04377.

100. Maeda M, Kurokawa K, Katada T, Nakano A, Saito K. COPII proteins exhibit distinct subdomains within each ER exit site for executing their functions. Scientific Reports. 2019;9. https://doi.org/10.1038/s41598-019-43813-3.

101. Maeda M, Katada T, Saito K. TANGO1 recruits Sec16 to coordinately organize ER exit sites for efficient secretion. The Journal of Cell Biology. 2017;216:1731–43. https://doi.org/10.1083/jcb.201703084.

102. Santos A, Nogueira C, Ortega-Bellido M, Malhotra V. TANGO1 and Mia2/cTAGE5 (TALI) cooperate to export bulky pre-chylomicrons/VLDLs from the endoplasmic reticulum. The Journal of Cell Biology. 2016;213:343–54. https://doi.org/10.1083/jcb.201603072.

103. Yasuda Y, Yoshida T, Oue M, Sengiku M, Ishikawa T, Saito S, et al. Tango1L but not Tango1S, Tali and cTAGE5 is required for export of type II collagen in medaka fish. Cell structure and function. 2025. https://doi.org/10.1247/csf.25001.

104. Clark E, Link B. Complementary and divergent functions of zebrafish Tango1 and Ctage5 in tissue development and homeostasis. Molecular Biology of the Cell. 2021;32:391–401. https://doi.org/10.1091/mbc.E20-11-0745.

105. Wenfu, Goldberg J. TANGO1/cTAGE5 receptor as a polyvalent template for assembly of large COPII coats. Proceedings of the National Academy of Sciences. 2016;113:10061–6. https://doi.org/10.1073/pnas.1605916113.

106. Cai H, Sasikumar P, Little G, Bihan D, Hamaia SW, Zhou A, et al. Identification of HSP47 Binding Site on Native Collagen and Its Implications for the Development of HSP47 Inhibitors. Biomolecules. 2021. https://doi.org/10.3390/biom11070983.

107. Ito S, Nagata K. Biology of Hsp47 (Serpin H1), a collagen-specific molecular chaperone. Seminars in Cell and Developmental Biology. 2017. https://doi.org/10.1016/j.semcdb.2016.11.005.

108. Arnolds O, Stoll R. Characterization of a fold in TANGO1 evolved from SH3 domains for the export of bulky cargos. Nat Commun. 2023;14. https://doi.org/10.1038/s41467-023-37705-4.

109. Raote I, Ortega-Bellido M, Santos AJ, Foresti O, Zhang C, Garcia-Parajo MF, et al. TANGO1 builds a machine for collagen export by recruiting and spatially organizing COPII, tethers and membranes. eLife. 2018;7. https://doi.org/10.7554/elife.32723.

110. Raote I, Chabanon M, Walani N, Arroyo M, Garcia-Parajo M, Malhotra V, et al. A physical mechanism of TANGO1-mediated bulky cargo export. eLife. 2019;9. https://doi.org/10.7554/eLife.59426.

111. Bunel L, Pincet L, Malhotra V, Raote I, Pincet F. A model for collagen secretion by intercompartmental continuities. Proceedings of the National Academy of Sciences of the United States of America. 2023;121. https://doi.org/10.1101/2023.09.21.558805.

112. Ishikawa Y, Rubin K, Bächinger HP, Kalamajski S. The endoplasmic reticulum–resident collagen chaperone Hsp47 interacts with and promotes the secretion of decorin, fibromodulin, and lumican. Journal of Biological Chemistry. 2018;293:13707–16. https://doi.org/10.1074/jbc.RA117.000758.

113. Sarohi V, Basak T. Perturbed post-translational modification (PTM) network atlas of collagen I during stent-induced neointima formation. Journal of proteomics. 2023;:104842. https://doi.org/10.1016/j.jprot.2023.104842.

114. Murthy AV, Sulu R, Lebedev A, Salo AM, Korhonen K, Venkatesan R, et al. Crystal structure of the collagen prolyl 4-hydroxylase (C-P4H) catalytic domain complexed with PDI: Toward a model of the C-P4H α2β2 tetramer. Journal of Biological Chemistry. 2022. https://doi.org/10.1016/j.jbc.2022.102614.

115. Kivirikko K, Prockop D. Enzymatic hydroxylation of proline and lysine in protocollagen. Proceedings of the National Academy of Sciences of the United States of America. 1967. https://doi.org/10.1073/pnas.57.3.782.

116. Sipilä K, Drushinin K, Rappu P, Jokinen J, Salminen TA, Salo AM, et al. Proline hydroxylation in collagen supports integrin binding by two distinct mechanisms. Journal of Biological Chemistry. 2018. https://doi.org/10.1074/jbc.ra118.002200.

117. Yamauchi M, Terajima M, Shiiba M. Lysine Hydroxylation and Cross-Linking of Collagen. Methods in molecular biology. 2019;1934:309–24. https://doi.org/10.1007/978-1-4939-9055-9_19.

118. Peng J, Li W, Yao D, Xia Y, Wang Q, Cai Y, et al. The structural basis for the human procollagen lysine hydroxylation and dual-glycosylation. Nature Communications. 2025;16. https://doi.org/10.1038/s41467-025-57768-9.

119. De Marco M, Rai SR, Scietti L, Mattoteia D, Liberi S, Pinnola A, et al. Molecular Structure and Enzymatic Mechanism of the Human Collagen Hydroxylysine Galactosyltransferase GLT25D1/COLGALT1. bioRxiv. 2024. https://doi.org/10.1101/2024.06.21.600124.

120. Scietti L, Chiapparino A, De Giorgi F, Fumagalli M, Khoriauli L, Nergadze S, et al. Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. Nature Communications. 2018;9. https://doi.org/10.1038/s41467-018-05631-5.

121. Zhang X, Wang Q, Wu J, Wang J, Shi Y, Liu M. Crystal structure of human lysyl oxidase-like 2 (hLOXL2) in a precursor state. Proceedings of the National Academy of Sciences of the United States of America. 2018;115:3828–33. https://doi.org/10.1073/pnas.1720859115.

122. Schmelzer C, Heinz A, Troilo H, Lockhart-Cairns M, Jowitt T, Marchand M, et al. Lysyl oxidase–like 2 (LOXL2)–mediated cross-linking of tropoelastin. The FASEB Journal. 2019;33:5468–81. https://doi.org/10.1096/fj.201801860RR.

123. López-Jiménez A, Basak T, Vanacore R. Proteolytic processing of lysyl oxidase–like-2 in the extracellular matrix is required for crosslinking of basement membrane collagen IV. The Journal of Biological Chemistry. 2017;292:16970–82. https://doi.org/10.1074/jbc.M117.798603.

124. Acevedo-Jake A, Ngo D, Hartgerink J. Control of Collagen Triple Helix Stability by Phosphorylation. Biomacromolecules. 2017;18 4:1157–61. https://doi.org/10.1021/acs.biomac.6b01814.

125. Qiu Y, Poppleton E, Mekkat A, Yu H, Banerjee S, Wiley S, et al. Enzymatic Phosphorylation of Ser in a Type I Collagen Peptide. Biophysical journal. 2018;115 12:2327–35. https://doi.org/10.1016/j.bpj.2018.11.012.

126. Zheng B, Zhao L, Chen L, Lai H, Wang C, Chen Y, et al. Phosphorylation of collagen fibrils enhances intrafibrillar mineralization and dentin remineralization. Nanoscale. 2024. https://doi.org/10.1039/d4nr00652f.

127. Roy A, Gauld J. Sulfilimine bond formation in collagen IV. Chemical communications. 2023. https://doi.org/10.1039/d3cc05715a.

128. Salminen A, Kauppinen A, Kaarniranta K. 2-Oxoglutarate-dependent dioxygenases are sensors of energy metabolism, oxygen availability, and iron homeostasis: potential role in the regulation of aging process. Cell Mol Life Sci. 2015;72:3897–914. https://doi.org/10.1007/s00018-015-1978-z.

129. He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science. 2011;333:1303–7. https://doi.org/10.1126/science.1210944.

130. Forneris F, Binda C, Vanoni MA, Battaglioli E, Mattevi A. Human Histone Demethylase LSD1 Reads the Histone Code. Journal of Biological Chemistry. 2005;280:41360–5. https://doi.org/10.1074/jbc.M509549200.

131. McNeill LA, Hewitson KS, Gleadle JM, Horsfall LE, Oldham NJ, Maxwell PH, et al. The use of dioxygen by HIF prolyl hydroxylase (PHD1). Bioorganic & Medicinal Chemistry Letters. 2002;12:1547–50. https://doi.org/10.1016/S0960-894X(02)00219-6.

132. Shin W-H, Kihara D. 55 Years of the Rossmann Fold. Methods of Molecular Biology. 2019. https://doi.org/10.1007/978-1-4939-9161-7_1.

133. Tiainen P, Myllyharju J, Koivunen P. Characterization of a Second Arabidopsis thaliana Prolyl 4-Hydroxylase with Distinct Substrate Specificity*. Journal of Biological Chemistry. 2005;280:1142–8. https://doi.org/10.1074/jbc.M411109200.

134. Culpepper M, Scott E, Limburg J. Crystal structure of prolyl 4-hydroxylase from Bacillus anthracis. Biochemistry. 2010;49 1:124–33. https://doi.org/10.1021/bi901771z.

135. Myllyharju J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix biology : journal of the International Society for Matrix Biology. 2003;22 1:15–24. https://doi.org/10.1016/S0945-053X(03)00006-4.

136. Koski M, Hieta R, Böllner C, Kivirikko K, Myllyharju J, Wierenga R. The Active Site of an Algal Prolyl 4-Hydroxylase Has a Large Structural Plasticity*. Journal of Biological Chemistry. 2007;282:37112–23. https://doi.org/10.1074/jbc.M706554200.

137. Mócsai R, Göritzer K, Stenitzer D, Maresch D, Strasser R, Altmann F. Prolyl Hydroxylase Paralogs in Nicotiana benthamiana Show High Similarity With Regard to Substrate Specificity. Frontiers in Plant Science. 2021;12. https://doi.org/10.3389/fpls.2021.636597.

138. Myllyharju J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix biology : journal of the International Society for Matrix Biology. 2003;22 1:15–24. https://doi.org/10.1016/S0945-053X(03)00006-4.

139. Rahman M, Sulu R, Adediran B, Tu H, Salo A, Murthy S, et al. Binding Differences of the Peptide-Substrate-Binding Domain of Collagen Prolyl 4-Hydroxylases I and II for Proline- and Hydroxyproline-Rich Peptides. Proteins. 2025. https://doi.org/10.1002/prot.26839.

140. Annunen P, Helaakoski T, Myllyharju J, Veijola J, Pihlajaniemi T, Kivirikko K. Cloning of the Human Prolyl 4-Hydroxylase α Subunit Isoform α(II) and Characterization of the Type II Enzyme Tetramer. The Journal of Biological Chemistry. 1997;272:17342–8. https://doi.org/10.1074/jbc.272.28.17342.

141. Annunen P, Helaakoski T, Myllyharju J, Veijola J, Pihlajaniemi T, Kivirikko K. Cloning of the human prolyl 4-hydroxylase alpha subunit isoform alpha(II) and characterization of the type II enzyme tetramer. The alpha(I) and alpha(II) subunits do not form a mixed alpha(I)alpha(II)beta2 tetramer. The Journal of biological chemistry. 1997;272 28:17342–8.

142. Kivirikko K, Myllyharju J. Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. Matrix biology : journal of the International Society for Matrix Biology. 1998;16 7:357–68. https://doi.org/10.1016/S0945-053X(98)90009-9.

143. Lu Y, Yang A, Zhao Z, Han Y, Wu D, Wu Y. Protein disulfide isomerase is essential for osteoblast differentiation in mice. Communications Biology. 2025;8. https://doi.org/10.1038/s42003-025-07824-3.

144. Salo A, Rappu P, Koski M, Karjalainen E, Izzi V, Drushinin K, et al. Collagen prolyl 4-hydroxylase isoenzymes I and II have sequence specificity towards different X-Pro-Gly triplets. Matrix biology : journal of the International Society for Matrix Biology. 2023. https://doi.org/10.1016/j.matbio.2023.12.001.

145. Hieta R, Kukkola L, Permi P, Pirilä P, Kivirikko K, Kilpeläinen I, et al. The Peptide-Substrate-binding Domain of Human Collagen Prolyl 4-Hydroxylases. Journal of Biological Chemistry. 2003;278:34966–74. https://doi.org/10.1074/JBC.M303624200.

146. Sarohi V, Basak T. Decoding the comprehensive substrate-specificity and evidence of altered site-specific collagen prolyl-3-hydroxylation, lysyl-hydroxylation, and lysyl O-glycosylation in P4ha1 and P4ha2 deleted mutant mice. bioRxiv. 2023. https://doi.org/10.1101/2023.06.28.546985.

147. Kiriakidis S, Hoer SS, Burrows N, Biddlecome G, Khan MN, Thinnes CC, et al. Complement C1q is hydroxylated by collagen prolyl 4 hydroxylase and is sensitive to off-target inhibition by prolyl hydroxylase domain inhibitors that stabilize hypoxia-inducible factor. Kidney International. 2017;92:900–8. https://doi.org/10.1016/j.kint.2017.03.008.

148. Hirsilä M, Koivunen P, Günzler V, Kivirikko K, Myllyharju J. Characterization of the Human Prolyl 4-Hydroxylases That Modify the Hypoxia-inducible Factor*. Journal of Biological Chemistry. 2003;278:30772–80. https://doi.org/10.1074/JBC.M304982200.

149. Onursal C, Knüppel L, Merl-Pham J, Lietman C, Hatz R, Behr J, et al. Prolyl-3-hydroxylase 4 (P3H4), a novel player in collagen synthesis and secretion in lung fibroblasts. Idiopathic interstitial pneumonias. 2021. https://doi.org/10.1183/13993003.congress-2021.pa3283.

150. Terajima M, Taga Y, Cabral W, Liu Y, Nagasawa M, Sumida N, et al. Cyclophilin B control of lysine post-translational modifications of skin type I collagen. PLoS Genetics. 2019;15. https://doi.org/10.1371/journal.pgen.1008196.

151. Tiainen P, Pasanen A, Sormunen R, Myllyharju J. Characterization of Recombinant Human Prolyl 3-Hydroxylase Isoenzyme 2, an Enzyme Modifying the Basement Membrane Collagen IV. Journal of Biological Chemistry. 2008;283:19432–9. https://doi.org/10.1074/jbc.M802973200.

152. Pokidysheva E, Boudko S, Vranka J, Zientek K, Maddox K, Moser M, et al. Biological role of prolyl 3-hydroxylation in type IV collagen. Proc Natl Acad Sci USA. 2014;111:161–6. https://doi.org/10.1073/pnas.1307597111.

153. Yamauchi M, Shiiba M. Lysine Hydroxylation and Cross-linking of Collagen. In: Kannicht C, editor. Post-translational Modifi cations of Proteins. Totowa, NJ: Humana Press; 2008. p. 95–108. https://doi.org/10.1007/978-1-60327-084-7_7.

154. Yang Z-N, Zhou Z, Yang M. Lysyl hydroxylase PLOD1 enhances actin network to promote confined migration of hepatocellular carcinoma cells via binding with Septin2. Journal of Clinical Oncology. 2022. https://doi.org/10.1200/jco.2022.40.16_suppl.e16111.

155. Van Der Slot A, Zuurmond A, Bardoel A, Wijmenga C, Pruijs H, Sillence D, et al. Identification of PLOD2 as Telopeptide Lysyl Hydroxylase, an Important Enzyme in Fibrosis*. Journal of Biological Chemistry. 2003;278:40967–72. https://doi.org/10.1074/jbc.M307380200.

156. Slot AJ van der, Zuurmond A, Bardoel A, Wijmenga C, Pruijs H, Sillence D, et al. Identification of PLOD2 as Telopeptide Lysyl Hydroxylase, an Important Enzyme in Fibrosis*. Journal of Biological Chemistry. 2003. https://doi.org/10.1074/jbc.m307380200.

157. Myllylä R, Wang C, Heikkinen J, Juffer A, Lampela O, Risteli M, et al. Expanding the lysyl hydroxylase toolbox: New insights into the localization and activities of lysyl hydroxylase 3 (LH3). Journal of Cellular Physiology. 2007;212. https://doi.org/10.1002/jcp.21036.

158. Schegg B, Hülsmeier A, Rutschmann C, Maag C, Hennet T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. 2018.

159. Vallet S, Ricard-Blum S. Lysyl oxidases: from enzyme activity to extracellular matrix cross-links. Essays in biochemistry. 2019. https://doi.org/10.1042/EBC20180050.

160. González-Santamaría J, Villalba M, Busnadiego Ó, López-Olañeta M, Sandoval P, Snabel J, et al. Matrix cross-linking lysyl oxidases are induced in response to myocardial infarction and promote cardiac dysfunction. Cardiovascular research. 2015;109 1:67–78. https://doi.org/10.1093/cvr/cvv214.

161. Piersma B, Bank RA. Collagen cross-linking mediated by lysyl hydroxylase 2: an enzymatic battlefield to combat fibrosis. Essays in Biochemistry. 2019;63:377–87. https://doi.org/10.1042/EBC20180051.

162. Paschalis EP, Verdelis K, Doty SB, Boskey AL, Mendelsohn R, Yamauchi M. Spectroscopic Characterization of Collagen Cross-Links in Bone. Journal of Bone and Mineral Research. 2001;16:1821–8. https://doi.org/10.1359/jbmr.2001.16.10.1821.

163. Yamauchi M, Taga Y, Hattori S, Shiiba M, Terajima M. Analysis of collagen and elastin cross-links. Methods in Cell Biology. 2018. https://doi.org/10.1016/bs.mcb.2017.08.006.

164. Aronson D. Cross-linking of glycated collagen in the pathogenesis of arterial and myocardial stiffening of aging and diabetes. Journal of hypertension. 2003. https://doi.org/10.1097/00004872-200301000-00002.

165. Pehrsson M, Mortensen J, Manon-Jensen T, Bay-Jensen A, Karsdal M, Davies M. Enzymatic cross-linking of collagens in organ fibrosis – resolution and assessment. Expert Review of Molecular Diagnostics. 2021;21:1049–64. https://doi.org/10.1080/14737159.2021.1962711.

166. Brinckmann J, Notbohm H, Tronnier M, Açil Y, Fietzek P, Schmeller W, et al. Overhydroxylation of lysyl residues is the initial step for altered collagen cross-links and fibril architecture in fibrotic skin. The Journal of investigative dermatology. 1999. https://doi.org/10.1046/j.1523-1747.1999.00735.x.

167. Hanson DA, Weis MAE, Bollen AM, Maslan SL, Singer FR, Eyre DR. A specific immunoassay for monitoring human bone resorption: quantitation of type I collagen cross-linked N-telopeptides in urine. Journal of Bone and Mineral Research. 2010. https://doi.org/10.1002/jbmr.5650071119.

168. Rodríguez-Pascual F, Slatter D. Collagen cross-linking: insights on the evolution of metazoan extracellular matrix. Scientific Reports. 2016;6. https://doi.org/10.1038/srep37374.

169. Zaker F, Qorbanpour A. PB1798: EVALUATION OF EXPRESSION OF PLODS FAMILY GENES (PLOD1, PLOD2, PLOD3) IN PATIENTS WITH ACUTE MYELOID LEUKEMIA. HemaSphere. 2022;6. https://doi.org/10.1097/01.HS9.0000850044.44072.08.

170. Scietti L, Moroni E, Mattoteia D, Fumagalli M, De Marco M, Negro L, et al. A Fe2+-dependent self-inhibited state influences the druggability of human collagen lysyl hydroxylase (LH/PLOD) enzymes. Frontiers in Molecular Biosciences. 2022;9. https://doi.org/10.3389/fmolb.2022.876352.

171. Hennet T. Collagen glycosylation. Current Opinion in Structural Biology. 2019. https://doi.org/10.1016/j.sbi.2019.01.015.

172. Peng J, Li W, Yao D, Xia Y, Wang Q, Cai Y, et al. The structural basis for the human procollagen lysine hydroxylation and dual-glycosylation. Nat Commun. 2025;16:2436. https://doi.org/10.1038/s41467-025-57768-9.

173. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: Structures, Functions, and Mechanisms. Annu Rev Biochem. 2008;77:521–55. https://doi.org/10.1146/annurev.biochem.76.061005.092322.

174. Liu J, Mushegian A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Science. 2003;12:1418–31. https://doi.org/10.1110/ps.0302103.

175. Busch C, Hofmann F, Selzer J, Munro S, Jeckel D, Aktories K. A Common Motif of Eukaryotic Glycosyltransferases Is Essential for the Enzyme Activity of Large Clostridial Cytotoxins. Journal of Biological Chemistry. 1998;273:19566–72. https://doi.org/10.1074/jbc.273.31.19566.

176. Götting C, Müller S, Schöttler M, Schön S, Prante C, Brinkmann T, et al. Analysis of the DXD Motifs in Human Xylosyltransferase I Required for Enzyme Activity. Journal of Biological Chemistry. 2004;279:42566–73. https://doi.org/10.1074/jbc.M401340200.

177. Liu J, Mushegian A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Science. 2003;12:1418–31. https://doi.org/10.1110/ps.0302103.

178. Albesa-Jove D, Giganti D, Jackson M, Alzari PM, Guerin ME. Structure-function relationships of membrane-associated GT-B glycosyltransferases. Glycobiology. 2014;24:108–24. https://doi.org/10.1093/glycob/cwt101.

179. Alexander JAN, Locher KP. Emerging structural insights into C-type glycosyltransferases. Current Opinion in Structural Biology. 2023;79:102547. https://doi.org/10.1016/j.sbi.2023.102547.

180. Zhang X, Guo L, Zhang X, Xu L, Tian Y, Fan Z, et al. GLT25D2 Is Critical for Inflammatory Immune Response to Promote Acetaminophen-Induced Hepatotoxicity by Autophagy Pathway. Front Pharmacol. 2020;11:01187. https://doi.org/10.3389/fphar.2020.01187.

181. Sipilä L, Ruotsalainen H, Sormunen R, Baker NL, Lamandé SR, Vapola M, et al. Secretion and Assembly of Type IV and VI Collagens Depend on Glycosylation of Hydroxylysines. Journal of Biological Chemistry. 2007;282:33381–8. https://doi.org/10.1074/jbc.M704198200.

182. Terajima M, Perdivara I, Sricholpech M, Deguchi Y, Pleshko N, Tomer KB, et al. Glycosylation and Cross-linking in Bone Type I Collagen. Journal of Biological Chemistry. 2014. https://doi.org/10.1074/jbc.m113.528513.

183. Visser DR, Loo TS, Norris GE, Parry DAD, Visser DR, Loo TS, et al. Potential implications of the glycosylation patterns in collagen α1(I) and α2(I) chains for fibril assembly and growth. Journal of Structural Biology. 2023. https://doi.org/10.1016/j.jsb.2023.107938.

184. Terajima M, Taga Y, Sricholpech M, Kayashima Y, Sumida N, Maeda N, et al. Role of Glycosyltransferase 25 Domain 1 in Type I Collagen Glycosylation and Molecular Phenotypes. Biochemistry. 2019. https://doi.org/10.1021/acs.biochem.8b00984.

185. Tang M, Wang X, Gandhi NS, Foley BL, Burrage K, Woods RJ, et al. Effect of hydroxylysine-O-glycosylation on the structure of type I collagen molecule: A computational study. Glycobiology. 2020. https://doi.org/10.1093/glycob/cwaa026.

186. Kadler K, Holmes D, Trotter J, Chapman J. Collagen fibril formation. The Biochemical journal. 1996;316 (Pt 1):1–11. https://doi.org/10.1042/BJ3160001.

187. Siadat SM, Ruberti J. Mechanochemistry of Collagen. Acta biomaterialia. 2023. https://doi.org/10.1016/j.actbio.2023.01.025.

188. Andriotis O, Nalbach M, Thurner P. Mechanics of isolated individual collagen fibrils. Acta biomaterialia. 2022. https://doi.org/10.1016/j.actbio.2022.12.008.

189. Asgari M, Latifi N, Heris HK, Vali H, Mongeau L. In vitro fibrillogenesis of tropocollagen type III in collagen type I affects its relative fibrillar topology and mechanics. Sci Rep. 2017;7:1392. https://doi.org/10.1038/s41598-017-01476-y.

190. Wenstrup R, Florer J, Brunskill E, Bell S, Chervoneva I, Birk D. Type V Collagen Controls the Initiation of Collagen Fibril Assembly*. Journal of Biological Chemistry. 2004;279:53331–7. https://doi.org/10.1074/jbc.M409622200.

191. Chowdhury AS, Oxford J. Collagen Alpha 1(XI) Amino-Terminal Domain Modulates Type I Collagen Fibril Assembly. Biochemistry. 2025;64:735–47. https://doi.org/10.1021/acs.biochem.4c00434.

192. Sun M, Luo E, Adams S, Adams T, Ye Y, Shetye S, et al. Collagen XI regulates the acquisition of collagen fibril structure, organization and functional properties in tendon. Matrix biology : journal of the International Society for Matrix Biology. 2020;94:77–94. https://doi.org/10.1016/j.matbio.2020.09.001.

193. Chen D, Smith L, Khandekar G, Patel P, Yu C, Zhang K, et al. Distinct effects of different matrix proteoglycans on collagen fibrillogenesis and cell-mediated collagen reorganization. Scientific Reports. 2020;10. https://doi.org/10.1038/s41598-020-76107-0.

194. Kalamajski S, Oldberg A. The role of small leucine-rich proteoglycans in collagen fibrillogenesis. Matrix biology : journal of the International Society for Matrix Biology. 2010;29 4:248–53. https://doi.org/10.1016/j.matbio.2010.01.001.

195. Pal GK, Suresh P. Comparative assessment of physico-chemical characteristics and fibril formation capacity of thermostable carp scales collagen. Materials science & engineering C, Materials for biological applications. 2016;70 Pt 1:32–40. https://doi.org/10.1016/j.msec.2016.08.047.

196. Herchenhan A, Uhlenbrock F, Eliasson P, Weis M, Eyre D, Kadler K, et al. Lysyl Oxidase Activity Is Required for Ordered Collagen Fibrillogenesis by Tendon Cells*. The Journal of Biological Chemistry. 2015;290:16440–50. https://doi.org/10.1074/jbc.M115.641670.

197. Varma S, Orgel J, Schieber J. Nanomechanics of Type I Collagen. Biophysical journal. 2016;111 1:50–6. https://doi.org/10.1016/j.bpj.2016.05.038.

198. Revell CK, Jensen OE, Shearer T, Lu Y, Holmes DF, Kadler KE. Collagen fibril assembly: New approaches to unanswered questions. Matrix Biology Plus. 2021;12:100079. https://doi.org/10.1016/j.mbplus.2021.100079.

199. Kubow K, Vukmirovic R, Zhe L, Klotzsch E, Smith M, Gourdon D, et al. Mechanical forces regulate the interactions of fibronectin and collagen I in extracellular matrix. Nature Communications. 2015;6. https://doi.org/10.1038/ncomms9026.

200. Li S, Van Den Diepstraten C, D'souza S, Chan B, Pickering J. Vascular smooth muscle cells orchestrate the assembly of type I collagen via alpha2beta1 integrin, RhoA, and fibronectin polymerization. The American journal of pathology. 2003;163 3:1045–56.

201. Saunders J, Schwarzbauer J. Fibronectin matrix as a scaffold for procollagen proteinase binding and collagen processing. Molecular Biology of the Cell. 2019;30:2218–26. https://doi.org/10.1091/mbc.E19-03-0140.

202. Barocas VH, Tranquillo RT. An Anisotropic Biphasic Theory of Tissue-Equivalent Mechanics: The Interplay Among Cell Traction, Fibrillar Network Deformation, Fibril Alignment, and Cell Contact Guidance. Journal of Biomechanical Engineering. 1997;119:137–45. https://doi.org/10.1115/1.2796072.

203. Meng Q, An S, Damion RA, Jin Z, Wilcox R, Fisher J, et al. The effect of collagen fibril orientation on the biphasic mechanics of articular cartilage. Journal of the Mechanical Behavior of Biomedical Materials. 2017;65:439–53. https://doi.org/10.1016/j.jmbbm.2016.09.001.

204. Busby GA, Grant MH, MacKay SP, Riches PE. Confined compression of collagen hydrogels. Journal of Biomechanics. 2013;46:837–40. https://doi.org/10.1016/j.jbiomech.2012.11.048.

205. Paten JA, Siadat SM, Susilo ME, Ismail EN, Stoner JL, Rothstein JP, et al. Flow-Induced Crystallization of Collagen: A Potentially Critical Mechanism in Early Tissue Formation. ACS Nano. 2016;10:5027–40. https://doi.org/10.1021/acsnano.5b07756.

206. Canty EG, Starborg T, Lu Y, Humphries SM, Holmes DF, Meadows RS, et al. Actin Filaments Are Required for Fibripositor-mediated Collagen Fibril Alignment in Tendon. Journal of Biological Chemistry. 2006;281:38592–8. https://doi.org/10.1074/jbc.M607581200.

207. Kapacee Z, Richardson S, Lu Y, Starborg T, Holmes D, Baar K, et al. Tension is required for fibripositor formation. Matrix Biology. 2008;27:371–5. https://doi.org/10.1016/j.matbio.2007.11.006.

208. Canty EG, Lu Y, Meadows RS, Shaw MK, Holmes DF, Kadler KE. Coalignment of plasma membrane channels and protrusions (fibripositors) specifies the parallelism of tendon. The Journal of Cell Biology. 2004;165:553–63. https://doi.org/10.1083/jcb.200312071.

209. Paten JA, Siadat SM, Susilo ME, Ismail EN, Stoner JL, Rothstein JP, et al. Correction to Flow-Induced Crystallization of Collagen: A Potentially Critical Mechanism in Early Tissue Formation. ACS Nano. 2017;11:8527–8527. https://doi.org/10.1021/acsnano.7b05275.

210. Dobson J, Kumar A, Willis LF, Tuma R, Higazi DR, Turner R, et al. Inducing protein aggregation by extensional flow. Proc Natl Acad Sci USA. 2017;114:4673–8. https://doi.org/10.1073/pnas.1702724114.

211. Dunderdale GJ, Davidson SJ, Ryan AJ, Mykhaylyk OO. Flow-induced crystallisation of polymers from aqueous solution. Nat Commun. 2020;11:3372. https://doi.org/10.1038/s41467-020-17167-8.

212. Zhao R, Liu Y, Ma Z. Flow-Induced Crystallization Starting from Self-Nucleated Precursors. Macromolecules. 2024;57:606–15. https://doi.org/10.1021/acs.macromol.3c02243.

213. Maassen S, Warner HM, Grijpstra P, Van Den Bogaart G. A quantitative in vitro collagen uptake assay. MethodsX. 2023;11:102288. https://doi.org/10.1016/j.mex.2023.102288.

214. Madsen DH, Ingvarsen S, Jürgensen HJ, Melander MC, Kjøller L, Moyer A, et al. The Non-phagocytic Route of Collagen Uptake. Journal of Biological Chemistry. 2011;286:26996–7010. https://doi.org/10.1074/jbc.M110.208033.

215. Iwahashi H, Kawashima Y, Masaki H. Decreased levels of endocytic collagen receptor Endo180 in dermal fibroblasts lead to decreased production of type I collagen and increased expression of matrix metalloproteinase-1. Photoderm Photoimm Photomed. 2022;38:150–7. https://doi.org/10.1111/phpp.12728.

216. Chang J, Pickard A, Herrera J, O'Keefe S, Garva R, Hartshorn M, et al. Endocytic recycling is central to circadian collagen fibrillogenesis and disrupted in fibrosis. eLife. 2025;13. https://doi.org/10.7554/eLife.95842.

217. Gerstenfeld L, Riva A, Hodgens K, Eyre D, Landis W. Post-translational control of collagen fibrillogenesis in mineralizing cultures of chick osteoblasts. Journal of Bone and Mineral Research. 1993;8. https://doi.org/10.1002/jbmr.5650080903.

218. Gjaltema R, Bank R. Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease. Critical Reviews in Biochemistry and Molecular Biology. 2017;52:74–95. https://doi.org/10.1080/10409238.2016.1269716.

219. Kent O, Casey E, Brown M, Bell S, Ehrman M, Flagler M, et al. New imaging tools reveal live cellular collagen secretion, fibril dynamics and network organisation. Scientific Reports. 2024;15. https://doi.org/10.1038/s41598-025-96280-4.

220. Roth J, Hoop C, Williams J, Nanda V, Baum J. Real-time single-molecule observation of incipient collagen fibrillogenesis and remodeling. Proceedings of the National Academy of Sciences of the United States of America. 2024;121. https://doi.org/10.1073/pnas.2401133121.

221. Musiime M, Chang J, Hansen U, Kadler K, Zeltz C, Gullberg D. Collagen Assembly at the Cell Surface: Dogmas Revisited. Cells. 2021;10. https://doi.org/10.3390/cells10030662.

222. Chang J, Garva R, Pickard A, Yeung C, Mallikarjun V, Swift J, et al. Circadian control of the secretory pathway maintains collagen homeostasis. Nature Cell Biology. 2020;22:74–86. https://doi.org/10.1038/s41556-019-0441-z.

223. Ireland J. Basement membrane. Journal of Clinical Pathology. 2020;31:59–66. https://doi.org/10.1136/jcp.31.Suppl_12.59.

224. Töpfer U. Basement membrane dynamics and mechanics in tissue morphogenesis. Biology Open. 2023;12. https://doi.org/10.1242/bio.059980.

225. Naylor R, Morais M, Lennon R. Complexities of the glomerular basement membrane. Nature Reviews Nephrology. 2020;17:112–27. https://doi.org/10.1038/s41581-020-0329-y.

226. Pozzi A, Yurchenco P, Iozzo R. The nature and biology of basement membranes. Matrix biology : journal of the International Society for Matrix Biology. 2016;57–58:1–11. https://doi.org/10.1016/j.matbio.2016.12.009.

227. Khalilgharibi N, Mao Y. To form and function: on the role of basement membrane mechanics in tissue development, homeostasis and disease. Open Biology. 2021;11. https://doi.org/10.1098/rsob.200360.

228. Sherwood D. Basement membrane remodeling guides cell migration and cell morphogenesis during development. Current opinion in cell biology. 2021;72:19–27. https://doi.org/10.1016/j.ceb.2021.04.003.

229. Weber M. Basement membrane proteins. Kidney international. 1992;41 3:620–8. https://doi.org/10.1038/KI.1992.95.

230. Randles M, Humphries M, Lennon R. Proteomic definitions of basement membrane composition in health and disease. Matrix biology : journal of the International Society for Matrix Biology. 2016;57–58:12–28. https://doi.org/10.1016/j.matbio.2016.08.006.

231. Iozzo R. Basement membrane proteoglycans: from cellar to ceiling. Nature Reviews Molecular Cell Biology. 2005;6:646–56. https://doi.org/10.1038/nrm1702.

232. Torricelli A, Singh V, Santhiago M, Wilson S. The corneal epithelial basement membrane: structure, function, and disease. Investigative ophthalmology & visual science. 2013;54 9:6390–400. https://doi.org/10.1167/iovs.13-12547.

233. Wilson S, Torricelli A, Marino G. Corneal epithelial basement membrane: Structure, function and regeneration. Experimental eye research. 2020;:108002. https://doi.org/10.1016/j.exer.2020.108002.

234. Timpl R, Wiedemann H, Delden V, Furthmayr H, Kuhn K. A Network Model for the Organization of Type IV Collagen Molecules in Basement Membranes. Eur J Biochem. 1981;120:203–11. https://doi.org/10.1111/j.1432-1033.1981.tb05690.x.

235. Kalluri R, Cosgrove D. Assembly of Type IV Collagen. Journal of Biological Chemistry. 2000;275:12719–24. https://doi.org/10.1074/jbc.275.17.12719.

236. Bailey AJ, Sims TJ, Light N. Cross-linking in type IV collagen. Biochemical Journal. 1984;218:713–23. https://doi.org/10.1042/bj2180713.

237. Añazco C, López-Jiménez AJ, Rafi M, Vega-Montoto L, Zhang M-Z, Hudson BG, et al. Lysyl Oxidase-like-2 Cross-links Collagen IV of Glomerular Basement Membrane. Journal of Biological Chemistry. 2016;291:25999–6012. https://doi.org/10.1074/jbc.M116.738856.

238. Than ME, Henrich S, Huber R, Ries A, Mann K, Kühn K, et al. The 1.9-Å crystal structure of the noncollagenous (NC1) domain of human placenta collagen IV shows stabilization via a novel type of covalent Met-Lys cross-link. Proc Natl Acad Sci USA. 2002;99:6607–12. https://doi.org/10.1073/pnas.062183499.

239. Vanacore RM, Friedman DB, Ham A-JL, Sundaramoorthy M, Hudson BG. Identification of S-Hydroxylysyl-methionine as the Covalent Cross-link of the Noncollagenous (NC1) Hexamer of the α1α1α2 Collagen IV Network. Journal of Biological Chemistry. 2005;280:29300–10. https://doi.org/10.1074/jbc.M502752200.

240. Birk DE, Silver FH. Molecular structure and physical properties of type IV collagen in solution. International Journal of Biological Macromolecules. 1987;9:7–10. https://doi.org/10.1016/0141-8130(87)90017-1.

241. Grant DS, Leblond CP. Immunogold quantitation of laminin, type IV collagen, and heparan sulfate proteoglycan in a variety of basement membranes. J Histochem Cytochem. 1988;36:271–83. https://doi.org/10.1177/36.3.2963856.

242. Lewis PN, Pinali C, Young RD, Meek KM, Quantock AJ, Knupp C. Structural Interactions between Collagen and Proteoglycans Are Elucidated by Three-Dimensional Electron Tomography of Bovine Cornea. Structure. 2010;18:239–45. https://doi.org/10.1016/j.str.2009.11.013.

243. Schoenenberger MS, Halfter W, Ferrand A, Halfter K, Tzankov A, Scholl HPN, et al. The biophysical and compositional properties of human basement membranes. The FEBS Journal. 2024;291:477–88. https://doi.org/10.1111/febs.17007.

244. Sanes J. The Basement Membrane/Basal Lamina of Skeletal Muscle*. The Journal of Biological Chemistry. 2003;278:12601–4. https://doi.org/10.1074/JBC.R200027200.

245. Myllyharju J, Kivirikko K. Collagens and collagen-related diseases. Annals of Medicine. 2001;33:21–7. https://doi.org/10.3109/07853890109002055.

246. Uitto J, Lichtenstein J. Defects in the biochemistry of collagen in diseases of connective tissue. The Journal of investigative dermatology. 1976;66 02:59–79. https://doi.org/10.1111/1523-1747.EP12481404.

247. Xu S, Xu H-X, Wang W, Li S, Li H, Li T, et al. The role of collagen in cancer: from bench to bedside. Journal of Translational Medicine. 2019;17. https://doi.org/10.1186/s12967-019-2058-1.

248. Kivirikko K. Collagens and their abnormalities in a wide spectrum of diseases. Annals of medicine. 1993;25 2:113–26. https://doi.org/10.3109/07853899309164153.

249. Jürgensen H, Van Putten S, Nørregaard K, Bugge T, Engelholm L, Behrendt N, et al. Cellular uptake of collagens and implications for immune cell regulation in disease. Cellular and Molecular Life Sciences. 2020;77:3161–76. https://doi.org/10.1007/s00018-020-03481-3.

250. Ricard-Blum S, Baffet G, Théret N. Molecular and tissue alterations of collagens in fibrosis. Matrix biology : journal of the International Society for Matrix Biology. 2018;68–69:122–49. https://doi.org/10.1016/j.matbio.2018.02.004.

251. Forlino A, Marini JC. Osteogenesis imperfecta. The Lancet. 2016;387:1657–71. https://doi.org/10.1016/S0140-6736(15)00728-X.

252. Malfait F, Castori M, Francomano CA, Giunta C, Kosho T, Byers PH. The Ehlers–Danlos syndromes. Nat Rev Dis Primers. 2020;6:64. https://doi.org/10.1038/s41572-020-0194-9.

253. Cosgrove D, Liu S. Collagen IV diseases: A focus on the glomerular basement membrane in Alport syndrome. Matrix biology : journal of the International Society for Matrix Biology. 2016;57–58:45–54. https://doi.org/10.1016/j.matbio.2016.08.005.

254. Abegglen O, Srikantharupan S, Zotter K, Marcionelli G, Ndarugendamwo T, Lim PJ, et al. Registry-Based Frequency of Molecularly Confirmed Osteogenesis Imperfecta in a Swiss Cohort of Individuals With Connective Tissue Disorders. American J of Med Genetics Pt A. 2025;197:e64016. https://doi.org/10.1002/ajmg.a.64016.

255. Marom R, Rabenhorst BM, Morello R. Management of Endocrine Disease: Osteogenesis imperfecta: an update on clinical features and therapies. European Journal of Endocrinology. 2020;183:R95–106. https://doi.org/10.1530/EJE-20-0299.

256. Budsamongkol T, Intarak N, Theerapanon T, Yodsanga S, Porntaveetus T, Shotelersuk V. A novel mutation in COL1A2 leads to osteogenesis imperfecta/Ehlers-Danlos overlap syndrome with brachydactyly. Genes & Diseases. 2019;6:138–46. https://doi.org/10.1016/j.gendis.2019.03.001.

257. Malfait F, Symoens S, Goemans N, Gyftodimou Y, Holmberg E, López-González V, et al. Helical mutations in type I collagen that affect the processing of the amino-propeptide result in an Osteogenesis Imperfecta/Ehlers-Danlos Syndrome overlap syndrome. Orphanet J Rare Dis. 2013;8:78. https://doi.org/10.1186/1750-1172-8-78.

258. Marini JC, Reich A, Smith SM. Osteogenesis imperfecta due to mutations in non-collagenous genes: lessons in the biology of bone formation. Current Opinion in Pediatrics. 2014;26:500–7. https://doi.org/10.1097/MOP.0000000000000117.

259. Tan Z, Chen P, Zhang J, Shek HT, Li Z, Zhou X, et al. Multi-omics analyses reveal aberrant differentiation trajectory with WNT1 loss-of-function in type XV osteogenesis imperfecta. Journal of Bone and Mineral Research. 2024;39:1253–67. https://doi.org/10.1093/jbmr/zjae123.

260. Joeng KS, Lee Y-C, Jiang M-M, Bertin TK, Chen Y, Abraham AM, et al. The swaying mouse as a model of osteogenesis imperfecta caused by WNT1 mutations. Human Molecular Genetics. 2014;23:4035–42. https://doi.org/10.1093/hmg/ddu117.

261. Laine CM, Joeng KS, Campeau PM, Kiviranta R, Tarkkonen K, Grover M, et al. *WNT1* Mutations in Early-Onset Osteoporosis and Osteogenesis Imperfecta. N Engl J Med. 2013;368:1809–16. https://doi.org/10.1056/NEJMoa1215458.

262. Scicluna K, Formosa MM, Farrugia R, Borg I. Hypermobile EHLERS–DANLOS syndrome: A review and a critical appraisal of published genetic research to date. Clinical Genetics. 2022;101:20–31. https://doi.org/10.1111/cge.14026.

263. Moreland LW, editor. Ehlers-Danlos syndrome. In: Rheumatology and Immunology Therapy. Berlin/Heidelberg: Springer-Verlag; 2004. p. 296–8. https://doi.org/10.1007/3-540-29662-X_936.

264. Germain DP. Clinical and Genetic Features of Vascular Ehlers-Danlos Syndrome. Annals of Vascular Surgery. 2002;16:391–7. https://doi.org/10.1007/s10016-001-0229-y.

265. Frank M, Adham S, Seigle S, Legrand A, Mirault T, Henneton P, et al. Vascular Ehlers-Danlos Syndrome. Journal of the American College of Cardiology. 2019;73:1948–57. https://doi.org/10.1016/j.jacc.2019.01.058.

266. Pepin M, Schwarze U, Superti-Furga A, Byers PH. Clinical and Genetic Features of Ehlers-Danlos Syndrome Type IV, the Vascular Type: Obstetrical & Gynecological Survey. 2000;55:469–71. https://doi.org/10.1097/00006254-200008000-00004.

267. Scietti L, Campioni M, Forneris F. SiMPLOD, a Structure-Integrated Database of Collagen Lysyl Hydroxylase (LH/PLOD) Enzyme Variants. Journal of Bone and Mineral Research. 2019;34:1376–82. https://doi.org/10.1002/jbmr.3692.

268. Martín-Martín M, Cortés-Martín J, Tovar-Gálvez MI, Sánchez-García JC, Díaz-Rodríguez L, Rodríguez-Blanque R. Ehlers–Danlos Syndrome Type Arthrochalasia: A Systematic Review. IJERPH. 2022;19:1870. https://doi.org/10.3390/ijerph19031870.

269. Malfait F, De Coster P, Hausser I, Van Essen AJ, Franck P, Colige A, et al. The natural history, including orofacial features of three patients with Ehlers–Danlos syndrome, dermatosparaxis type (EDS type VIIC). American J of Med Genetics Pt A. 2004;131A:18–28. https://doi.org/10.1002/ajmg.a.30299.

270. Barat-Houari M, Sarrabay G, Gatinois V, Fabre A, Dumont B, Genevieve D, et al. Mutation Update for *COL2A1* Gene Variants Associated with Type II Collagenopathies: HUMAN MUTATION. Human Mutation. 2016;37:7–15. https://doi.org/10.1002/humu.22915.

271. Mercuri E, Yuva Y, Brown SC, Brockington M, Kinali M, Jungbluth H, et al. Collagen VI involvement in Ullrich syndrome: A clinical, genetic, and immunohistochemical study. Neurology. 2002;58:1354–9. https://doi.org/10.1212/WNL.58.9.1354.

272. Di Martino A, Cescon M, D'Agostino C, Schilardi F, Sabatelli P, Merlini L, et al. Collagen VI in the Musculoskeletal System. International Journal of Molecular Sciences. 2023;24. https://doi.org/10.3390/ijms24065095.

273. Qi Y, Xu R. Roles of PLODs in Collagen Synthesis and Cancer Progression. Front Cell Dev Biol. 2018;6:66. https://doi.org/10.3389/fcell.2018.00066.

274. Sato K, Parag-Sharma K, Terajima M, Musicant AM, Murphy RM, Ramsey MR, et al. Lysyl hydroxylase 2-induced collagen cross-link switching promotes metastasis in head and neck squamous cell carcinomas. Neoplasia. 2021;23:594–606. https://doi.org/10.1016/j.neo.2021.05.014.

275. Terajima M, Taga Y, Nakamura T, Guo H-F, Kayashima Y, Maeda-Smithies N, et al. Lysyl hydroxylase 2 mediated collagen post-translational modifications and functional outcomes. Scientific Reports. 2022;12. https://doi.org/10.1038/s41598-022-18165-0.

276. Cheriyamundath S, Kumar A, Gavert N, Brabletz T, Ben-Ze'ev A. The Collagen-Modifying Enzyme PLOD2 Is Induced and Required during L1-Mediated Colon Cancer Progression. IJMS. 2021;22:3552. https://doi.org/10.3390/ijms22073552.

277. Eisinger-Mathason TSK, Zhang M, Qiu Q, Skuli N, Nakazawa MS, Karakasheva T, et al. Hypoxia-Dependent Modification of Collagen Networks Promotes Sarcoma Metastasis. Cancer Discovery. 2013;3:1190–205. https://doi.org/10.1158/2159-8290.CD-13-0118.

278. Kang H, Strong AL, Sun Y, Guo L, Juan C, Bancroft AC, et al. The HIF-1α/PLOD2 axis integrates extracellular matrix organization and cell metabolism leading to aberrant musculoskeletal repair. Bone Res. 2024;12:17. https://doi.org/10.1038/s41413-024-00320-0.

279. Wu X, Xiang H, Cong W, Yang H, Zhang G, Wang Y, et al. PLOD1, a target of miR-34c, contributes to cell growth and metastasis via repressing LATS1 phosphorylation and inactivating Hippo pathway in osteosarcoma. Biochemical and Biophysical Research Communications. 2020;527:29–36. https://doi.org/10.1016/j.bbrc.2020.04.052.

280. Shi J, Bao M, Wang W, Wu X, Li Y, Zhao C, et al. Integrated Profiling Identifies PLOD3 as a Potential Prognostic and Immunotherapy Relevant Biomarker in Colorectal Cancer. Front Immunol. 2021;12:722807. https://doi.org/10.3389/fimmu.2021.722807.

281. Vahidnezhad H, Youssefian L, Saeidian AH, Touati A, Pajouhanfar S, Baghdadi T, et al. Mutations in PLOD3, encoding lysyl hydroxylase 3, cause a complex connective tissue disorder including recessive dystrophic epidermolysis bullosa-like blistering phenotype with abnormal anchoring fibrils and type VII collagen deficiency. Matrix Biology. 2019;81:91–106. https://doi.org/10.1016/j.matbio.2018.11.006.

282. Bolshakova OI, Latypova EM, Komissarov AE, Slobodina AD, Ryabova EV, Varfolomeeva EYu, et al. Cellular and Molecular Effects of the Bruck Syndrome-Associated Mutation in the PLOD2 Gene. IJMS. 2024;25:13379. https://doi.org/10.3390/ijms252413379.

283. Deng X, Pan Y, Yang M, Liu Y, Li J. PLOD3 Is Associated with Immune Cell Infiltration and Genomic Instability in Colon Adenocarcinoma. BioMed Research International. 2021;2021:4714526. https://doi.org/10.1155/2021/4714526.

284. Baumann S, Hennet T. Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase. Journal of Biological Chemistry. 2016. https://doi.org/10.1074/jbc.m116.723379.

285. Qiu S, Han H, Zhang H, Yang M, Wang H, Li K, et al. The collagen-modifying enzyme GLT25D1 is a prognostic indicator related to immunosuppression and malignant phenotypes in hepatocellular carcinoma. Cancer Cell Int. 2025;25:84. https://doi.org/10.1186/s12935-025-03715-z.

286. Wang S, He L, Xiao F, Gao M, Wei H, Yang J, et al. Upregulation of GLT25D1 in Hepatic Stellate Cells Promotes Liver Fibrosis via the TGF-β1/SMAD3 Pathway *In Vivo* and *In vitro*. J Clin Transl Hepatol. 2022;000:000–000. https://doi.org/10.14218/JCTH.2022.00005.

287. 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 and Biophysical Research Communications. 2021;560:93–8. https://doi.org/10.1016/j.bbrc.2021.04.128.

288. Zou Y, Donkervoort S, Salo AM, Foley AR, Barnes AM, Hu Y, et al. P4HA1 mutations cause a unique congenital disorder of connective tissue involving tendon, bone, muscle and the eye. Human Molecular Genetics. 2017;26:2207–17. https://doi.org/10.1093/hmg/ddx110.

289. Tolonen J, Salo AM, Finnilä M, Aro E, Karjalainen E, Ronkainen V, et al. Reduced Bone Mass in Collagen Prolyl 4-Hydroxylase *P4ha1*$^{+/-}$; *P4ha2*$^{-/-}$ Compound Mutant Mice. JBMR Plus. 2022;6:e10630. https://doi.org/10.1002/jbm4.10630.

290. Wang C, Luosujärvi H, Heikkinen J, Risteli M, Uitto L, Myllylä R. The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. Matrix biology : journal of the International Society for Matrix Biology. 2002;21 7:559–66. https://doi.org/10.1016/S0945-053X(02)00071-9.

291. Wang Y, Xu A, Knight C, Xu L, Cooper G. Hydroxylation and Glycosylation of the Four Conserved Lysine Residues in the Collagenous Domain of Adiponectin. The Journal of Biological Chemistry. 2002;277:19521–9. https://doi.org/10.1074/JBC.M200601200.

292. Peng J, Li W, Yao D, Xia Y, Wang Q, Cai Y, et al. The structural basis for the human procollagen lysine hydroxylation and dual-glycosylation. Nat Commun. 2025;16:2436. https://doi.org/10.1038/s41467-025-57768-9.

293. Ishikawa Y, Taga Y, Zientek K, Mizuno N, Salo AM, Semenova O, et al. Type I and type V procollagen triple helix uses different subsets of the molecular ensemble for lysine posttranslational modifications in the rER. Journal of Biological Chemistry. 2021;296:100453. https://doi.org/10.1016/j.jbc.2021.100453.

294. Cabral WA, Perdivara I, Weis M, Terajima M, Blissett AR, Chang W, et al. Abnormal Type I Collagen Post-translational Modification and Crosslinking in a Cyclophilin B KO Mouse Model of Recessive Osteogenesis Imperfecta. PLoS Genet. 2014;10:e1004465. https://doi.org/10.1371/journal.pgen.1004465.

Chapter 9 Ringraziamenti

Ringrazio la Prof.ssa Alberta Pinnola e il Prof. Federico Forneris per l'opportunità e la fiducia che mi è stata data per conseguire il mio dottorato di ricerca. Un ringraziamento speciale per Anselmo Canciani (grazie delle immagini), Luigi Scietti (la primavera arriva sempre) e Silvia Faravelli (Venerdì neomelodico), che non solo sono stati punti di riferimento e guida in questo mio percorso di formazione, ma anche, e soprattutto, amici cari che mi sono stati vicini anche nei momenti più difficoltosi. Un ringraziamento speciale anche a Rai Sristi Raj, amica e collega, il cui supporto e collaborazione sono stati preziosissimi (bokachoda). Un grande ringraziamento alla mia famiglia per tutto il supporto durante questi anni di studi, soprattutto a mia madre Antonina Prometti e mia sorella Federica De Marco, ed ai mie cari amici Davide Raimondi, Aris Viglietti e Viola Rota.

Chapter 10 Paper collection

International Journal of
Molecular Sciences

MDPI

Article

Identification of Regulatory Molecular "Hot Spots" for LH/PLOD Collagen Glycosyltransferase Activity

Daiana Mattoteia [1,†], Antonella Chiapparino [1,†], Marco Fumagalli [1,‡], Matteo De Marco [1,‡], Francesca De Giorgi [1,‡], Lisa Negro [1], Alberta Pinnola [1], Silvia Faravelli [1], Tony Roscioli [2,3], Luigi Scietti [1,§] and Federico Forneris [1,4,*]

[1] The Armenise-Harvard Laboratory of Structural Biology, Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9A, 27100 Pavia, Italy
[2] NSW Health Pathology Randwick Genomics Laboratory, Prince of Wales Hospital, Sydney, NSW 2031, Australia
[3] Neuroscience Research Australia (NeuRA), Prince of Wales Clinical School, University of New South Wales, Sydney, NSW 2052, Australia
[4] Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Policlinico San Matteo, 27100 Pavia, Italy
* Correspondence: federico.forneris@unipv.it
† These authors contributed equally to this work.
‡ These authors contributed equally to this work.
§ Present address: Biochemistry and Structural Biology Unit, Department of Experimental Oncology, IRCCS European Institute of Oncology (IEO), Via Adamello 16, 20139 Milan, Italy.

Abstract: Hydroxylysine glycosylations are post-translational modifications (PTMs) essential for the maturation and homeostasis of fibrillar and non-fibrillar collagen molecules. The multifunctional collagen lysyl hydroxylase 3 (LH3/PLOD3) and the collagen galactosyltransferase GLT25D1 are the human enzymes that have been identified as being responsible for the glycosylation of collagen lysines, although a precise description of the contribution of each enzyme to these essential PTMs has not yet been provided in the literature. LH3/PLOD3 is thought to be capable of performing two chemically distinct collagen glycosyltransferase reactions using the same catalytic site: an inverting beta-1,O-galactosylation of hydroxylysines (Gal-T) and a retaining alpha-1,2-glucosylation of galactosyl hydroxylysines (Glc-T). In this work, we have combined indirect luminescence-based assays with direct mass spectrometry-based assays and molecular structure studies to demonstrate that LH3/PLOD3 only has Glc-T activity and that GLT25D1 only has Gal-T activity. Structure-guided mutagenesis confirmed that the Glc-T activity is defined by key residues in the first-shell environment of the glycosyltransferase catalytic site as well as by long-range contributions from residues within the same glycosyltransferase (GT) domain. By solving the molecular structures and characterizing the interactions and solving the molecular structures of human LH3/PLOD3 in complex with different UDP-sugar analogs, we show how these studies could provide insights for LH3/PLOD3 glycosyltransferase inhibitor development. Collectively, our data provide new tools for the direct investigation of collagen hydroxylysine PTMs and a comprehensive overview of the complex network of shapes, charges, and interactions that enable LH3/PLOD3 glycosyltransferase activities, expanding the molecular framework and facilitating an improved understanding and manipulation of glycosyltransferase functions in biomedical applications.

Keywords: collagen biosynthesis; extracellular matrix; post-translational modifications; collagen glycosylations; glycosyltransferase; lysyl hydroxylase

Citation: Mattoteia, D.; Chiapparino, A.; Fumagalli, M.; De Marco, M.; De Giorgi, F.; Negro, L.; Pinnola, A.; Faravelli, S.; Roscioli, T.; Scietti, L.; et al. Identification of Regulatory Molecular "Hot Spots" for LH/PLOD Collagen Glycosyltransferase Activity. *Int. J. Mol. Sci.* **2023**, *24*, 11213. https://doi.org/10.3390/ijms241311213

Academic Editor: Daniel Arcos

Received: 2 April 2023
Revised: 22 June 2023
Accepted: 5 July 2023
Published: 7 July 2023

1. Introduction

Collagens are the most abundant proteins in the human body and are highly conserved [1,2]. The different oligomeric architectures and roles of collagen molecules depend strongly on a variety of post-translational modifications (PTMs), including proline

and lysine hydroxylations, as well highly specific glycosylations of hydroxylated lysines (Hyl) [1,3]. The disaccharide present in Hyl contains a highly conserved glucosyl(α-1,2)-galactosyl(β-1,O) glycan moiety, which was discovered in the late sixties [4–6]. Monosaccaridic galactosyl-(β-1,O)-Hyl were identified as a result of catabolic reactions carried out by the collagen α-glucosidase, an enzyme highly specific for the disaccharide substrate found 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], the functional area inside tissues [12–14], the developmental stage [15,16], and on disease states [17–19]. However, although studied extensively, the precise mechanisms of collagen glycosylation and their biological relevance in collagen homeostasis remain poorly understood.

The stereochemistry of the Glc(α-1,2)-Gal(β-1,O)-Hyl-linked carbohydrate is consistent with at least two distinct enzyme types that are required for collagen molecule PTMs [20,21]. Since both donor substrates (UDP-galactose and UDP-glucose) exhibit an α-glycosidic bond, the first reaction requires an inverting-type galactosyltransferase (Gal-T) acting on Hyl, whereas the subsequent glucosylation is catalyzed by a retaining-type glucosylgalactosyltransferase (Glc-T). Multifunctional collagen lysyl hydroxylases (LHs/PLODs) possess both Fe^{2+}-dependent lysyl hydroxylase (LH) and Mn^{2+}-dependent glucosyltransferase activities [22–24]. The human isoenzyme 3 (LH3/PLOD3) was reported to also exhibit Gal-T activity in vitro [22]. In vivo studies have demonstrated that decreased LH3/PLOD3 protein levels and/or pathogenic mutations in the glycosyltransferase (GT) domain exclusively impair Glc-T activity [18,25,26]. This occurs secondarily to the LH3/PLOD3 p.(Asn223Ser), which introduces an additional glycosylation site within the enzyme's GT domain, leading to osteogenesis imperfecta [18], and in the recently identified LH3/PLOD3 p.(Pro270Leu), which results in a Stickler-like syndrome with vascular complications and variable features typical of Ehlers–Danlos syndromes and Epidermolysis Bullosa [26]. Mouse studies have also shown that only the LH3/PLOD3 Glc-T activity is indispensable for the biosynthesis of collagen IV and formation in the basement membrane during embryonic development [15,27], consistent with the presence of additional collagen galactosyltransferases. To date, two genes encoding for Mn^{2+}-dependent *O*-galactosyltransferases (GLT25D1 and GLT25D2) have been identified [28,29], and GLT25D1 has been proposed to act in concert with LH3/PLOD3 on collagen molecules [28,30,31]. Studies on osteosarcoma cell lines, which produce large amounts of fibrillar collagens, have shown that the simultaneous deletion of GLT25D1 and GLT25D2 results in growth arrest due to a lack of glycosylation, causing some to hypothesize that the Gal-T activity of LH3/PLOD3 might not be as essential as the Glc-T activity [32].

These data support the hypothesis that, in vivo, the entire collagen glycosylation machinery may involve distinct proteins and protein complexes for Gal-T and Glc-T reactions. This raises the intriguing question of how this highly conserved process is spatiotemporally regulated at the molecular level. However, our current understanding of collagen glycosyltransferases is restricted to three-dimensional structures of human LH3/PLOD3 in complex with UDP-sugar donor substrates [33] and the few mutagenesis studies focusing on the hallmarks of Mn^{2+}-dependent glycosyltransferase catalysis [22,34].

By combining structure-guided mutagenesis with conventional indirect activity assays and a new mass spectrometry-based direct assay to monitor the Lys-to-Glc-Gal-Hyl conversion altogether, we identified a broad group of amino acid residues in the LH3/PLOD3 GT domain that are cooperatively responsible for its glycosyltransferase activity. The remarkable conservation of the amino acid network, which shapes the GT domain in all human LH/PLOD isoforms, combined with recent reports of Glc-T activity for LH1/PLOD1 [23] and LH2/PLOD2 [24] extends the significance of the proposed network, providing a detailed molecular blueprint of the Glc-T activity hallmarks belonging to the entire enzyme family. The results obtained in this study on GLT25D1 illustrate that this enzyme is the sole responsible for Gal-T activity on Hyl residues and rule out the role of LH3/PLOD3. GLT25D1 generates the substrate for the LH/PLOD-mediated Glc-T reaction, ultimately

yielding Glc-Gal-Hyl. Finally, we also identified and characterized UDP-sugar substrate analogs acting as inhibitors of LH3/PLOD3 Glc-T activity, providing templates for the development of the first collagen Glc-T inhibitors.

2. Results

2.1. A Direct MS-Based Assay to Evaluate Lys-to-Glc-Gal-Hyl Conversion

The results from previous studies have consistently highlighted that the N-terminal GT domain is the only catalytic glycosyltransferase domain of LH/PLOD enzymes [21,23,24,33]. In this respect, the question of the controversial possible coexistence of two chemically opposite glycosyltransferase catalytic activities (i.e., the inverting Gal-T and the retaining Glc-T activities) within the same catalytic site is still open. The recent identification of the collagen galactosyltransferase family GLT25D1/2 has provoked questions about the initially observed LH3/PLOD3 Gal-T reactivity [33]. Therefore, we recombinantly produced both human LH3/PLOD3 and GLT25D1 and optimized the previously established MS-based assay used to monitor Fe^{2+}-dependent Lys-to-Hyl conversion [33] to directly detect the presence of Gal-Hyl and Glc-Gal-Hyl residues on synthetic collagen peptides (Figure 1A). Firstly, the LH activity of LH3/PLOD3 was tested for both enzymes in the presence of all of the necessary cofactors (Fe^{2+}, 2-OG and ascorbate) and the acceptor substrate (i.e., synthetic collagen peptide $(GIK)_4$), observing the expected introduction of a hydroxyl group (16 Da) on a single Lys residue (Figures 1B and S1). The same reaction mixture containing GLT25D1 and not LH3/PLOD3 did not produce any modifications to the substrate peptides (Figure S2), supporting the idea that GLT25D1 exclusively accepts Hyl residues as substrates (Figure 1A). To assess the putative inverting Gal-T activity of LH3/PLOD3, Mn^{2+} and UDP-Gal (i.e., cofactors of GalT reaction) were added to the reaction mixture. Surprisingly, no additional peaks were observed (Figure 1C). By replicating the same assay with the addition of GLT25D1, the MS spectrum revealed an additional peak compatible with the galactosylation of a single Hyl residue (Figure 1D). These results unambiguously establish that GLT25D1 (not LH3/PLOD3) is solely responsible for the galactosylation of collagen Hyl residues. Finally, the incorporation of UDP-Glc into the reaction mixture resulted in the occurrence of a fourth MS peak, corresponding to the expected mass of the nude peptide with a single Glc-Gal-Hyl residue (Figure 1E).

2.2. The Amino Acid Residues Shaping the UDP Binding Site Are Essential for LH3/PLOD3 Glc-T Activity

The LH3/PLOD3 N-terminal GT domain shares its fold with Mn^{2+}-dependent GT-A glycosyltransferases, encompassing a UDP-donor substrate binding cavity stretched towards a catalytic pocket [33]. To date, the Glc-T activity of homologous LH1/PLOD1 and LH2/PLOD2 enzymes has been poorly characterized. Using indirect luminescence-based assays, we previously reported the LH1/PLOD1 Glc-T activity to be ten-fold less than LH3/PLOD3 [23]. Other studies also reported on the possible role of the LH2/PLOD2 Glc-T activity in lung adenocarcinoma progression [24]. To identify key residues in the catalytic site and dissect their contribution to the LH3/PLOD3 Glc-T activity, we used previously generated homology models of LH1/PLOD1 and LH2/PLOD2 [35] and performed a comparative structural analysis between the GT domains of all human LH/PLOD variants (Figure S3). Overall, the models for LH1 and LH2 could superimpose well with the crystal structure of LH3/PLOD3, enabling an accurate comparison at the amino acid side chain level (Figure S3B).

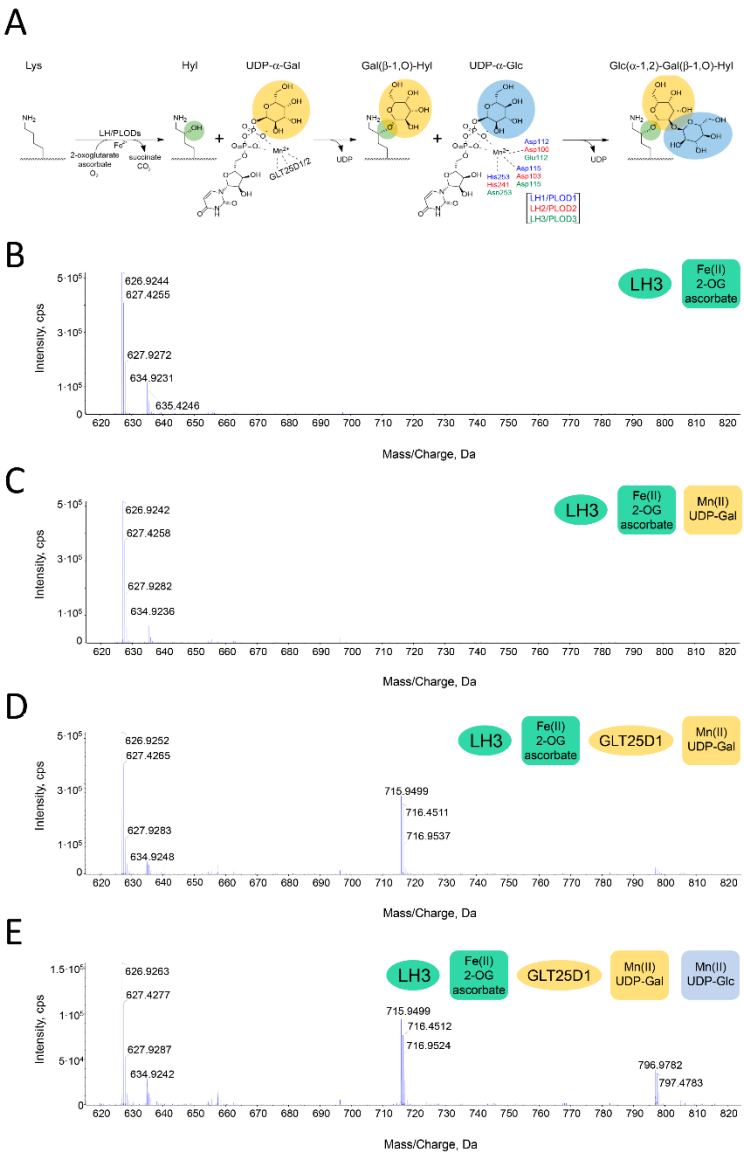


Figure 1. A direct assay to probe Lys-to-Glc-Gal-Hyl conversion. The reaction schematic monitored by the assay is depicted in (**A**). The MS spectra show the results obtained by incubating a synthetic

GIKGIKGIKGIK peptide (MW 1254 Da) with enzymes and cofactors, as shown in the figure legends. All peaks identified are doubly charged, resulting in nominal masses corresponding to half of the expected MW. (**B**) Using LH3/PLOD3 and LH activity cofactors (i.e., 2-OG and Fe^{2+}), MS peaks corresponding to a singly hydroyxylated Lys on the peptide (i.e., 635 Da) appear. (**C**) The addition of Gal-T activity cofactors (i.e., UDP-Gal and Mn^{2+}) to the same mixture as in (**B**) does not yield additional MS peaks. (**D**) When the mixture in (**C**) is incubated with GLT25D1, the MS peaks corresponding to Gal-Hyl are found (i.e., 716 Da). (**E**) When the same mixture as in (**D**) also contains UDP-Glc, the peaks corresponding to Glc-Gal-Hyl appear (i.e., 797 Da).

We inspected the UDP-binding cavity to identify distinguishing features by comparing the amino acids found in LH3/PLOD3 with those predicted for the LH1/PLOD1 and LH2/PLOD2 isoforms, which were described to have low Glc-T activity [23,24]. We observed that nearly all of the residues involved in Mn^{2+} and UDP binding are conserved in all paralogs, except for Val80, becoming Lys68 in LH1/PLOD1 and Gly80 in LH2/PLOD2 (Figures 2A and S3). 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/PLOD enzymes. In LH3/PLOD3, Val80 is located in the middle of a flexible "glycoloop" (Gly72-Gly87), not visible in the electron density of the ligand-free LH3/PLOD3 structure, and stabilized upon UDP-substrate binding [33]. Within the glycoloop, residue Val80 is in close proximity to the ribose ring of the UDP-sugar donor substrates. We hypothesized that the introduction of a large, positively charged residue such as Lys in LH1/PLOD1 or alterations due to the complete removal of side chain steric hindrance such as Gly in LH2/PLOD2 could interfere with the binding of donor substrates. Therefore, we generated the LH3/PLOD3 p.(Val80Lys) and p.(Val80Gly) mutants. These enzyme variants were found to be folded based on analytical gel filtration and differential scanning fluorimetry (DSF) and showed lysine hydroxylation activity comparable to wild-type LH3/PLOD3 (Figure S4A–D). Conversely, both mutations resulted in non-detectable Glc-T activity in MS-based assays (Figure 2B, Table 1). Moreover, since LH3/PLOD3 is capable of activating donor UDP-sugar substrates and releases UDP in the absence of the acceptor collagen substrate [33], we investigated the impact of these mutations in both the absence (uncoupled activity) and presence (coupled activity) of acceptor substrates by performing luminescence-based assays, obtaining very similar results using both enzyme variants. These experiments demonstrate that the Glc-T activity of LH3/PLOD3 p.(Val80Lys) and p.(Val80Gly) is mostly imputable to the uncoupled activity (Figure 2C, Table 1). Consistently, Val80 might be involved in the productive positioning of the donor substrate to enable the transfer of the glycan moiety to the collagen acceptor substrate rather than stabilizing the UDP moiety in the catalytic pocket.

To further investigate the influence of LH3/PLOD3 Val80 on Glc-T activity, we crystallized and solved the 3.0-Å resolution structure of the p.(Val80Lys) mutant in complex with Mn^{2+} and also obtained the 2.3-Å resolution structure of the same mutant in the presence of both Mn^{2+} and the UDP-Glc donor substrate (Figure 3A, Table S1). Overall, both structures superimpose almost perfectly with wild-type LH3/PLOD3 for all domains (Figure S5A). The structure of the LH3/PLOD3 p.(Val80Lys) mutant bound to Mn^{2+} is essentially identical to that of wild-type LH3/PLOD3 in complex with the same cofactors. In both structures, the glycoloop containing Val/Lys80 could not be modelled in the experimental electron density due to its high flexibility (Figure 3C). On the other hand, the side chain of the Lys80 residue could be modelled unambiguously in the experimental electron density of the UDP-donor substrate-bound structure. 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 was observed for wild-type LH3/PLOD3, the glycan moiety of UDP-Glc was not visible in the electron density (Figure 3A). Collectively, these data are consistent with the alteration introduced by the p.(Val80Lys) mutation impacting on the LH3/PLOD3 glycosyltransferase catalytic activities.

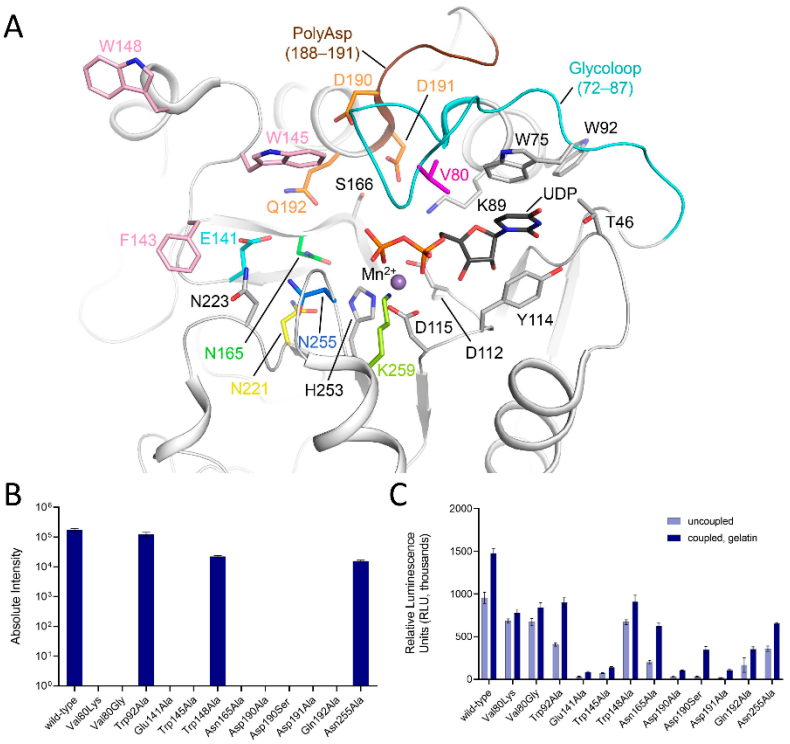


Figure 2. Structural and functional features of the LH3/PLOD3 glycosyltransferase (GT) domain. (**A**) Cartoon representation of the LH3/PLOD3 GT domain (PDB ID: 6FXR) showing the key residues shaping the catalytic site as sticks. The PolyAsp motif (brown) and the glycoloop (cyan) involved in the binding 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 gray have already been shown to be essential in Mn^{2+} (purple sphere) and UDP (black sticks) coordination. (**B**) Summary of the evaluation of the Glc-T activity of LH3/PLOD3 mutants compared to the wild-type using MS direct assays. (**C**) Evaluation of the Glc-T activity of LH3/PLOD3 mutants compared to wild-type using luminescence-based indirect assays. Each graph shows the enzymatic activity detected in the absence (i.e., "uncoupled", light blue) or presence of gelatin, which was used as the acceptor substrate. The plotted data are baseline-corrected, where the baseline was the background control. In both (**B**,**C**) panels, the error bars represent standard deviations from the averages of independent experiments (N > 3).

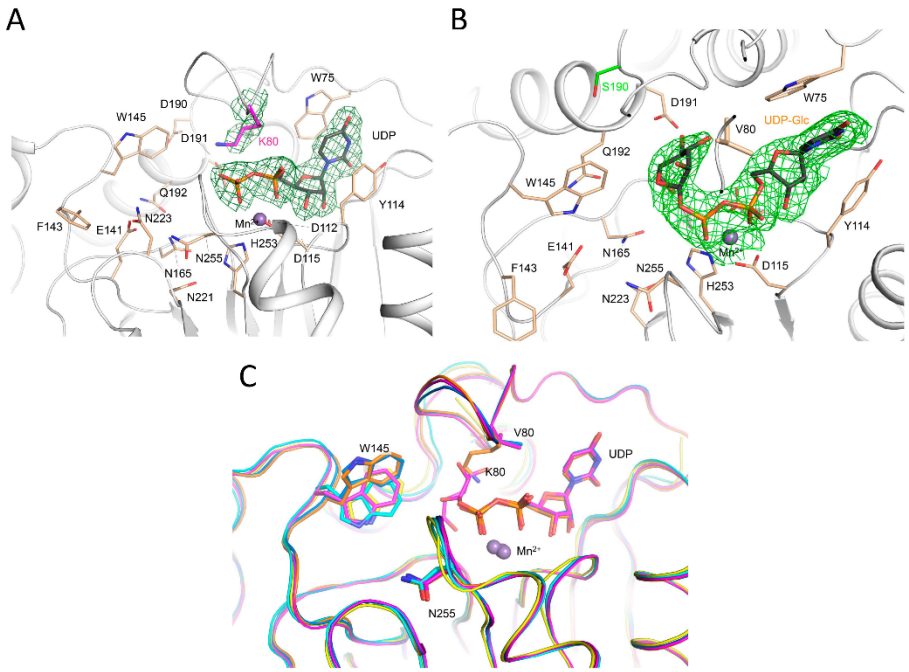


Figure 3. Structural characterization of LH3/PLOD3 mutants. (**A**) Crystal structure of the LH3/PLOD3 p.(Val80Lys) mutant in complex with UDP-glucose and Mn^{2+}. Electron density is visible for the mutated lysine and the UDP portion of the donor substrate (green mesh, $2F_o-F_c$ omit electron density map, contoured at 1.3 σ). Catalytic residues shaping the enzyme cavity are shown as sticks; Mn^{2+} is shown as a purple sphere. Consistent with what was observed in the crystal structure of wild-type LH3/PLOD3, the glucose moiety of the donor substrate is not visible in the experimental electron density. (**B**) Crystal structure of the LH3/PLOD3 p.(Asp190Ser) mutant in complex with UDP-glucose and Mn^{2+}. Electron density is visible for the mutated Serine and for the entire donor substrate, including the sugar moiety (green mesh, $2F_o-F_c$ omit electron density map, contoured at 1.3 σ). Colors and representations as in (**A**). (**C**) Superposition of wild-type, p.(Val80Lys), and p.(Asp190Ser) LH3/PLOD3 available crystal structures in substrate-free (cyan for wild-type, yellow for p.(Val80Lys), respectively) and with UDP-glucose bound (marine for wild-type, orange for p.(Val80Lys), magenta for p.(Asp190Ser), 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 only visible in the in UDP-sugar-bound structures.

2.3. *The LH3/PLOD3 Glc-T Activity Is Affected by the Long-Range Rearrangement of Trp92 and Trp75*

The glycoloop is a structural feature found exclusively in the GT domains of LH/PLOD enzymes. It incorporates Trp75, a residue whose aromatic side chain stabilizes the uridine moiety of the donor substrate through pi-stacking and, together with residue Tyr114 of the DxxD motif (a distinguishing feature of LH/PLOD GT domain, [33]), "sandwiches" the donor substrate in an aromatic stacking environment (Figure 2A). Both residues are critical

for LH3/PLOD3 Glc-T enzymatic activity [33]. However, the conformation adopted by the LH3/PLOD3 glycoloop in the presence of UDP-donor substrates is not accompanied by other significant structural changes in the surrounding amino acids, except for the minor rearrangements of distant residue Trp92—which is not conserved in other LH/PLOD isoenzymes (Figures 2A and S3)—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 nor its LH/PLOD enzymatic activity (Figure S4A–D). Conversely, the p.(Trp75Ala) mutant Glc-T activity is undetectable compared to wild-type LH3/PLOD3 (Figure 2B,C, Table 1), and the impact of the p.(Trp92Ala) mutation seemed to slightly affect Glc-T activity (Figure 2B,C, Table 1). These findings suggest that the presence of non-conserved residues within the LH3/PLOD3 GT domain that are distant from those involved in first-shell interactions with the donor and acceptor substrates may contribute to the productive conformations of the glycoloop in donor substrate-bound states.

Table 1. Results of activity assays (performed using either MS or luminescence) for all LH3/PLOD3 mutants described in this work compared to wild-type LH3/PLOD3.

Mutation	Localization	Folding State/ LH Activity	Glc-T Activity (MS) (%)	Glc-T Activity (Luminescence) (%)	
				Uncoupled	Coupled, Gelatin
wild-type	N/A	Yes	100	100	100
Val80Lys	glycoloop	Yes	N.D.	72 ± 2	53 ± 3
Val80Gly	glycoloop	Yes	N.D.	70 ± 4	54 ± 4
Trp92Ala	UDP-binding cavity	Yes	71 ± 14	43 ± 2	60 ± 4
Asp190Ala	poly-Asp helix	Yes	N.D.	3 ± 0.27	7 ± 0.5
Asp191Ala	poly-Asp helix	Yes	N.D.	2 ± 0.3	7 ± 0.6
Asp190Ser	poly-Asp helix	Yes	N.D.	2 ± 0.4	17 ± 2
Trp145Ala	acceptor substrate cavity and gates	Yes	N.D.	8 ± 0.5	9 ± 0.6
Trp148Ala	acceptor substrate cavity and gates	Yes	13 ± 2	70 ± 3	61 ± 5
Asn165Ala	region proximate UDP-sugar	Yes	N.D.	21 ± 3	42 ± 2
Gln192Ala	region proximate UDP-sugar	Yes	N.D.	17 ± 9	24 ± 2
Glu141Ala	region proximate UDP-sugar	Yes	N.D.	3 ± 0.6	6 ± 0.6
Asn255Ala	region proximate UDP-sugar	Yes	9 ± 0.5	38 ± 3	44 ± 0.7
Pro270Leu	interface of AC and GT domains	No	N.D.	N.D.	N.D.

2.4. A Poly-Asp Sequence near the Donor Sugar Binding Site Is Essential for Glc-T Activity in LH1/PLOD1 and in LH3/PLOD3

The poly-Asp sequence (Asp188-Asp191, Figure 2A) is visible in UDP-sugar-bound structures and in contact with the glycoloop. This sequence is partially conserved in LH/PLOD isoforms (Figure S3), and mutations of Asp190 and Asp191 have also been reported to affect the glycosyltransferase enzymatic activities of LH3/PLOD3 [34]. Based on LH3/PLOD3 crystal structures, such behavior is expected since residues Asp190 and Asp191 point towards the Glc-T catalytic cavity. We designed and generated individual alanine mutants for both LH3/PLOD3 Asp190 and Asp191 that showed that both variants were compatible with folded and functional LH/PLOD enzymes (Figure S4A–D). When tested for Glc-T activity, both these mutants resulted inactive, as confirmed by MS and the luminescence-based assay results (Figure 2B,C, Table 1). Collectively, these data are consistent with the involvement of the residues of the poly-Asp repeat, and in particular Asp190 and Asp191, in both the positioning and recognition of donor or acceptor substrates.

A recent report showed that the p.(Ser178Arg) mutation on LH1/PLOD1 (corresponding to LH3/PLOD3 Asp190 in the poly-Asp sequence) caused a heritable thoracic aortic disease [23]. The p.(Asp190Arg) mutation strongly affected the LH3/PLOD3 enzyme

function by causing a drastic drop in Glc-T catalytic activity. Aiming to further elucidate the significance of the differences caused by this specific residue in LH1/PLOD1 and LH3/PLOD3, we generated a LH3/PLOD3 p.(Asp190Ser) variant to mimic LH1/PLOD1 at that specific site (Figure S4A–D). MS and luminescence data showed that this single-point mutation completely abolishes the LH3/PLOD3 Glc-T activity (Figure 2B,C) without affecting the ability of this mutant to bind UDP-Glc donor substrates (Figure S4E). We also solved the 2.30-Å crystal structure of the full-length human LH3/PLOD3 p.(Asp190Ser) mutant in complex with Mn^{2+} and UDP-Glc. Overall, the structure appeared nearly identical to wild-type LH3/PLOD3 (Figure S5A). Analysis of the electron density at the GT catalytic site showed clear signals for the Glc moiety of the UDP-Glc donor substrate (Figure 3B). This feature was never observed when using wild-type LH3/PLOD3 [33], further supporting the importance of the Poly-Asp loop for Glc-T activity. The side chains of Asp191, Lys89, and Ser166, as well as the main chain carbonyl of Thr83, were found at distances and orientations possibly compatible with electrostatic contacts with the Glc moiety of the donor substrate, resulting in the trapping of the sugar in a portion of the catalytic cavity, leaving a large pocket shaped by residues Glu141, Phe143, Asn165, Asn223, and Asn255 available to likely host the Gal moiety of the acceptor substrate for catalysis (Figure 4).

To confirm the importance of this residue, we also investigated how this site could affect the Glc-T activity of the human LH1/PLOD1 counterpart by investigating a LH1/PLOD1 p.(Ser178Asp) mutant. Strikingly, this substitution resulted significantly rescued the LH1/PLOD1 Glc-T activity, which could be detected using both direct and indirect assays (Figure 5). Together, these data provide experimental evidence for in vitro LH1/PLOD1 Glc-T activity and further confirm the critical role for Asp190 in the activation of the donor substrate prior to its transfer to the collagen acceptor substrate.

2.5. Two Gating Trp Residues Modulate Glc-T Activity by Affecting Acceptor Substrate Binding

After investigating the amino acid residues involved in stabilizing 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 2A). Many LH3/PLOD3 residues shaping this part of the glycosyltransferase catalytic cavity matched the catalytic amino acids found in other GT-A type glycosyltransferases (Figure S6) [37,38]. In particular, LH3/PLOD3 Trp145, a residue located in one of the loops of the GT domain uniquely found in LH3/PLOD3, was previously suggested as a possible candidate for the modulation of Glc-T activity. This residue was found to adopt different side chain conformations in substrate-free and substrate-bound structures, affecting the shape and steric hindrance of the enzyme's catalytic cavity [33]. Interestingly, nearly identical conformational changes were also observed when comparing substrate-free and substrate-bound LH3/PLOD3 p.(Val80Lys) and p.(Asp190Ser) structures (Figures 3C and 4A). Mutating Trp145 residue into alanine completely abolishes Glc-T enzymatic activity (Figure 2B,C) without affecting protein folding stability and LH activity (Figure S4A–D). This supports previous the hypotheses of a "gating" role for Trp145 in GT catalytic cavity, assisting the productive positioning of sugar moieties of donor substrates for effective transfer during catalysis and/or shaping the pocket to host the Lys-O-Gal moiety of the collagen acceptor substrate [33]. A 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 that is similar to that of Trp145 in LH3/PLOD3. 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 S2, Figure S6). This is surprising considering that the loop encompassing residues 142–163 is a unique structural feature of LH/PLOD enzymes [33]. This further highlights the high versatility of glycosyltransferases, 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.

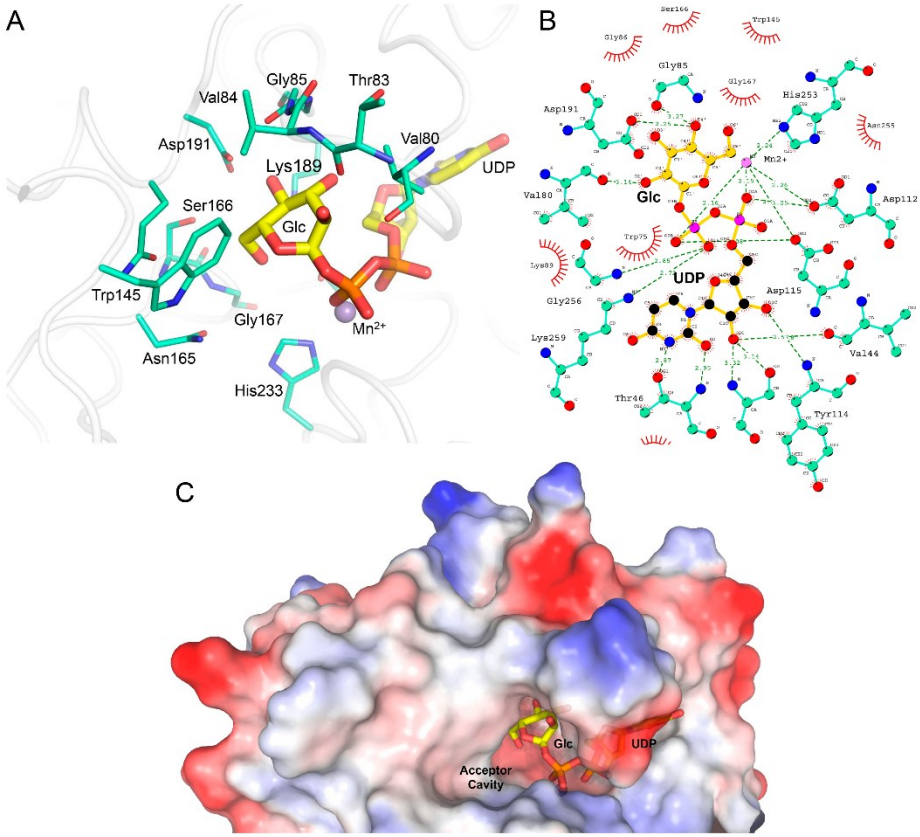


Figure 4. Binding mode of the glucose moiety of the UDP-Glc donor substrate observed in the crystal structure of LH3/PLOD3 p.(Asp190Ser). (**A**) Highlight of the amino acid network surrounding the Glc moiety of the donor substratein the crystal structure. UDP-Glc is shown as thick yellow sticks, whereas amino acids found at less than 5 Å distance from the Glc moiety are shown as thin blue/green sticks. (**B**) Overview of the interaction network surrounding the UDP-Glc donor substrate in the co-crystal structure with LH3/PLOD3 p.(Asp190Ser). Colors are as in (**A**). Figure made with LIGPLOT+ [36]. (**C**) The conformation adopted by the Glc moiety of the UDP-Glc substrate in the glycosyltransferase catalytic site of LH3/PLOD3 p.(Asp190Ser) leaves an empty cavity that is geometrically and sizably compatible with the Gal moiety of the acceptor substrate. Shown is a surface rendering of the GT domain of LH3/PLOD3 p.(Asp190Ser) colored by electrostatic potential, with highlights of the UDP-Glc donor substrate shown as sticks.

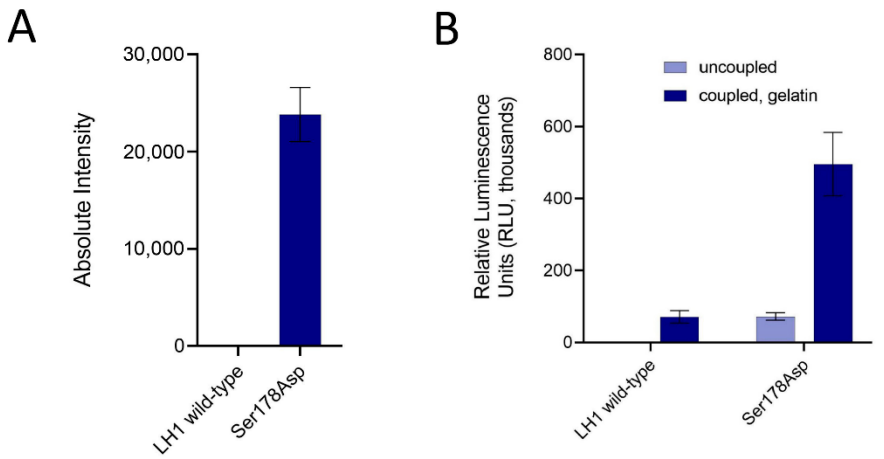


Figure 5. Ser178 in LH1/PLOD1, corresponding to Asp190 in LH3/PLOD3, is a key residue for Glc-T activity for both enzyme isoforms. (**A**) Direct MS-based assays comparing the signal associated with Glc-Gal-Hyl using wild-type and Ser178Asp LH1/PLOD1 variants. (**B**) Evaluation of the Glc-T activity of LH1 wild-type and Ser178Asp using luminescence-based indirect assays. The analysis of coupled and uncoupled enzymatic activities is as in Figure 2C. In both panels, the error bars represent standard deviations from the averages of independent experiments (N > 3).

Our previous structural comparisons of ligand-free and substrate-bound LH3/PLOD3 suggested the intriguing additional possibility of a concerted mechanism involving conformational changes in a non-conserved aromatic residue located on the enzyme's surface (Trp148, Figure 2A) together with Trp145 [33]. To investigate such a possibility, we mutated Trp148 into alanine. The resulting mutant enzyme was folded and displayed LH activity comparable to wild-type LH3/PLOD3 (Figure S4A–D). However, glycosyltransferase assays showed that this variant had reduced Glc-T activity compared to the wild-type in the presence of both donor and acceptor substrates (Figure 2B,C, Table 1). Despite the less pronounced alterations than those observed when mutating Trp145, these data allow us to speculate about 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.

2.6. Additional Residues Facing Both Donor and Acceptor Substrates Affect the LH3/PLOD3 Glc-T Activity

The molecular structures of LH3/PLOD3 in complex with UDP-Glc and Mn^{2+} showed very weak electron density near the UDP pyrophosphate group, likely representative of multiple conformations simultaneously trapped in the substrate binding cavity [33]. We explored the LH3/PLOD3 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, which would mean they were capable of acting as candidate catalytic nucleophiles for the formation of a (covalent) glycosyl-enzyme intermediate prior to the 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/PLOD3 structures, we noticed that, despite being distant from the donor substrate, residues Gln192, Asn165, and Glu141 point towards the cavity that accommodates the glycan moiety (Figure 2A). We generated

Ala mutations of all these residues, obtaining folded functional LH enzymes in all cases (Figure S4A–D). When probed for Glc-T activity, we found that both p.(Asn165Ala) and p.(Gln192Ala) mutants were still capable of activating UDP-donor substrates, albeit less efficiently (Figure 2C, Table 1); however, no sugar transfer to acceptor substrates could be detected (Figure 2B, Table 1). Conversely, the p.(Glu141Ala) mutant was completely inactive and incapable of UDP-donor substrate activation (Figure 2B,C, Table 1). These data suggest essential roles for Glu141 in catalysis, either through the initial binding of UDP-Glc donor substrates prior to their final positioning in the catalytic site or through the stabilization of water molecule networks in the large GT cavity, enabling donor substrate processing. The surrounding negatively charged pocket composed of Asp190, Asp191, Gln192, Asn165, and all residues found relevant to (but not essential for) catalysis likely aids glycosyltransferase activity. Proximate to Glu141 in the GlcT cavity, residue Asn255 is the closest amino acid to the UDP phosphate–sugar bond. Despite being fully conserved in LH/PLOD isoforms (Figure S3), to date, it has not been found in any GT-A-type glycosyltransferases with known structures. The side chain of Asn255 consistently points to a direction opposite the donor substrate in all LH3/PLOD3 structures (Figure 2A). We hypothesize that the side chain amide group might also be involved in catalysis, possibly through the recognition of collagen acceptor substrates given the conformation displayed by this side chain. However, the LH3/PLOD3 p.(Asn255Ala) mutant folded properly (Figure S4A–D) and showed that Glc-T enzymatic activity was only slightly reduced (Figure 2B,C, Table 1).

2.7. Pathogenic Mutations in the LH3/PLOD3 GT Domain Affect Protein Folding

Recently a pathogenic LH3/PLOD3 mutation, p.(Pro270Leu), has been identified and mapped at the interface of the AC and GT domains [26]. This proline residue localizes in a loop responsible for shaping the GT cavity; however, given its position, it is unlikely to play direct roles in catalysis. To better understand the impact of this pathogenic mutation on LH3/PLOD3 enzymatic activity, we attempted the recombinant production of a p.(Pro270Leu) LH3/PLOD3 mutant. In this case, due to the almost non-detectable protein levels, we could not reliably carry out any in vitro investigations. Considering the high reproducibility associated with the recombinant production of a large variety of LH3/PLOD3 point mutants, this result may indicate that this mutation is likely to severely impact the overall enzyme stability rather than its enzymatic activity, resulting in extremely low protein expression levels in vitro and, most likely, in vivo as well.

2.8. The Molecular Structures of LH3/PLOD3 in Complex with UDP-Sugar Analogs Provide Insights into the Processing of Glycan Moieties in the Catalytic Cavity

A frequent limitation associated with molecular characterizations of glycosyltransferases is the high flexibility of the donor substrate glycan moiety within the catalytic cavity. Such a limitation becomes even more relevant when the enzyme is capable of processing UDP-sugar molecules in the absence of acceptor substrates, such as in the case of LH3/PLOD3. Considering our previous [33] and current co-crystallization results, we considered whether free UDP, the product of the enzymatic reaction, could remain bound in the LH3/PLOD3 GT domain with the same efficiency as physiological donor substrates even after processing. Therefore, we compared the binding of free UDP and donor UDP-sugars using DSF and detected a thermal shift of 3–3.5 °C for free UDP compared to a 2–2.5 °C shift using UDP-sugar substrates (Figure 6A). These results were consistent with free UDP binding to LH3/PLOD3, likely with an even higher affinity than the UDP-glycan substrates, indicating that the Glc-T reaction may therefore be affected by product inhibition. Surprisingly, the increase in thermal stability did not correlate with the efficient trapping of the reaction product in wild-type LH3/PLOD3 molecular structures. Independently of the UDP concentration used in co-crystallization and soaking experiments, electron density for free UDP was not observed, yielding molecular structures completely identical to ligand-free enzymes (Figure S7).

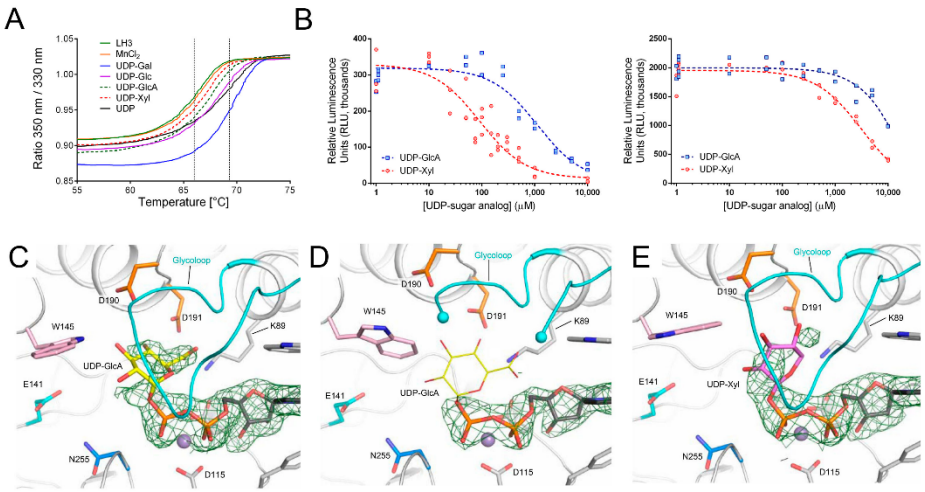


Figure 6. Characterization of UDP-sugar analogs. (**A**) Thermal stability of LH3 wild-type (solid green) using differential scanning fluorimetry (DSF) in the presence of various Mn^{2+} and several UDP-sugars. A prominent stabilization effect is achieved in the presence of the biological donor substrates 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**) Luminescence-based competition assays evaluating the binding of increasing concentrations of UDP-GlcA or UDP-Xyl to wild-type LH3/PLOD3 in the presence of either UDP-Gal (**left**) or UDP-Glc (**right**) and acceptor substrates (i.e., gelatin). (**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 the partial electron density. (**D**) Crystal structure of the LH3 Val80Lys mutant in complex with Mn^{2+} and UDP-glucuronic acid. While 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. Similar 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).

A report by Kivirikko and colleagues [43] suggested that UDP-glucuronic acid (UDP-GlcA) could act as a competitive inhibitor of collagen glycosyltransferases. Based on this, UDP-GlcA was used to isolate LH3/PLOD3 from chicken embryo preparation [22,44]. However, no follow-up biochemical studies were published. We used DSF and luminescence-based Glc-T activity assays to investigate how UDP-GlcA could affect LH3/PLOD3 enzymatic activity. DSF showed that UDP-GlcA indeed binds weakly, resulting in a thermal shift of 1–1.5 °C (Figure 6A), highlighting limited stabilization compared to the UDP-glycan substrates and free UDP. Enzymatic assays also confirmed the weak competitive inhibition displayed by this molecule (Figure 6B), with IC_{50} values in the millimolar range (Table 2). We also successfully co-crystallized and determined the 2.2-Å resolution crystal structure of wild-type LH3/PLOD3 in complex with Mn^{2+} and UDP-GlcA (Table S1) and found that the inhibitor could efficiently replace UDP-sugar donor substrates in the substrate cavity (Figure 6C). We observed additional experimental 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, it unambiguously showed that the sugar adopts a bent conformation resembling that observed for UDP-Glc in the LH3/PLOD3 p.(Asp190Ser) mutant crystal structure. 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 6C), thereby leaving the remaining space in the cavity to accommodate acceptor substrates.

Table 2. Evaluation of the competitive binding of UDP-GlcA and UDP-Xyl in wild-type LH3/PLOD3 in the presence of UDP-Gal or UDP-Glc and acceptor substrates.

Inhibitor	UDP-Gal (IC_{50}, µM)	UDP-Glc (IC_{50}, µM)
wild-type + UDP-GlcA	1130 ± 370	>10,000
wild-type + UDP-Xyl	91 ± 23	3170 ± 211

Considering the possible conformations adopted by the GlcA based on analysis of the electron density and the close proximity of the glucuronic acid moiety to LH3/PLOD3 Val80, we wondered whether the p.(Val80Lys) mutation could interfere with inhibitor binding. Therefore, we co-crystallized and solved the 2.7-Å resolution structure of the LH3/PLOD3 p.(Val80Lys) mutant in complex with UDP-GlcA (Table S1) and surprisingly observed the 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/PLOD3 structures (Figure 6D). At the same time, we could not observe improvements in the quality of the electron density for the glucuronic acid moiety, which was worse compared to that observed in wild-type LH3/PLOD3. This is consistent with the intrinsic flexibility of the sugar-like moiety not being influenced by specific conformations of the glycoloop, but rather by the lack of specific protein–ligand interactions that could stabilize the sugar ring in a unique structural arrangement.

The absence of a precise conformation for the glucuronic acid moiety observed in crystal structures prompted us to further investigate another UDP-sugar substrate analog, UDP-Xylose (UDP-Xyl), which is characterized by a lack of the carboxylic moiety of UDP-GlcA. Similar to UDP-GlcA, UDP-Xyl was shown to be a weak inhibitor of LH3/PLOD3 Glc-T activity (Figure 6A,B), with IC50 in the high micromolar range (Table 2). The 2.0-Å resolution structure of LH3/PLOD3 in complex with Mn^{2+} and UDP-Xyl also showed that the inhibitor was bound inside the enzyme's catalytic cavity, with weak electron density associated with the sugar moiety, suggesting multiple conformations of the xylose moiety attached to UDP, similar to what was observed for UDP-GlcA (Figure 6E). Taken together, these results suggest that the LH3/PLOD3 GT cavity can host 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, it may be expected 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 LH/PLOD glycosyltransferase activity.

3. Discussion

Glycosyltransferases are highly versatile enzymes with broad substrate specificity. When investigated carefully, they show a series of recurrent features that allow their comparative characterization even in the presence of low sequence/structure conservation. LH3/PLOD3 is widely known to be a multifunctional enzyme that is able to catalyze multiple steps of the Lys-to-Glc-Gal-Hyl PTM pathway [1]. Recent reports have also suggested similar multifunctional activities for the paralogs LH1/PLOD1 and LH2/PLOD2 [23,24], but direct evidence for the final Glc-Gal-Hyl product generated is still missing. Our in vitro investigations combining direct MS-based assays with the indirect detection of glucosyltransferase activity products shed light on the specific functions of this unique enzyme

family. In this work, for the first time, it has been shown that LH/PLOD enzymes are exclusively involved in the glucosylation of galactosyl hydroxylysines, whereas collagen galactosyltransferases, such as GLT25D1, are solely responsible for Hyl galactosylation. Thus, our results confirm that LH/PLOD enzymes are retaining-type glucosyltransferases. Furthermore, the direct MS-based assay allowed us to detect the formation of the first fully synthetic Glc-Gal-Hyl collagen peptide in vitro through one-pot total synthesis.

A structure-guided mutagenesis investigation was performed to fully investigate the LH/PLOD enzyme's GT domains by focusing on only non-matching residue present within the amino acid sequence directly surrounding the UDP portion of the donor substrate. This residue (Val80 in LH3/PLOD3, which corresponds to Lys68 in LH1/PLOD1 and Gly80 in LH2/PLOD2) is located in the middle of the glycoloop, in close proximity to the ribose ring of the UDP-sugar donor substrate(s). The biochemical data described here are consistent with the p.(Val80Lys) and p.(Val80Gly) substitutions impairing the LH3/PLOD3 catalytic activity. This result was corroborated by the disorder observed around the glycoloop of the LH3/PLOD3 p.(Val80Lys) crystal structure, thus providing a mechanistic explanation for the lack of Glc-T activity. It is likely that Val80 assists the positioning of the glycan moiety of the bound donor substrate. Alternatively, Val80 might be involved in the productive positioning of the donor substrate during the transfer of the glycan moiety to the collagen acceptor substrate.

By comparing the apo and UDP-Glc-bound LH3/PLOD3, we identified three residues in the second-shell environment that were adopting different conformations at the side chain level, somehow suggesting a possible involvement in the catalytic mechanism. The non-conserved LH3/PLOD3 Trp92 (Leu80 in LH1, Leu92 in LH2a/b) positions its aromatic side chain in a conformation that stabilizes the entire glycoloop to facilitate enzymatic reactions. However, the mutation of this residue did not lead to a full loss of Glc-T activity, suggesting that the folding topology of the LH/PLOD GT domain is highly versatile and adaptable to multiple amino acid alterations without abolishing Glc-T catalytic activity.

The sequence discrepancy in another critical site of the GT domain, the poly-Asp helix, which is known to be critical for catalysis, was explored [34]. It was found that LH3/PLOD3 Asp190, positioned at the entrance of the catalytic site, is critical for enzymatic activity. LH3/PLOD3 p.(Asp190Ser), matching the corresponding residue in LH1/PLOD1, resulted in strongly diminished Glc-T activity and yielded a crystal structure with observable electron density for the donor sugar moiety for the first time, enabling the identification of the amino acid residues directly involved in Glc binding and those shaping the portion of the cavity likely hosting the Gal moiety of the acceptor substrate. Strikingly, the reverse mutation in LH1/PLOD1 (Ser178Asp) strongly enhanced the Glc-T activity compared to wild-type, confirming this to be a key residue essential for the catalytic processing of the donor glucose prior to its transfer to the acceptor substrate.

At the entrance of the GT catalytic pocket, LH3/PLOD3 exhibits a non-conserved loop shaping the GT catalytic cavity bearing two aromatic residues: the Trp145 and the Trp148, which seem to act in a concerted way during catalysis, as suggested by comparisons between substrate-free and substrate-bound molecular structures (Figure 2A). Trp145 is indispensable for Glc-T activity; its conformational changes consistently respond to the presence and positioning of the donor substrate inside the catalytic cavity (Figures 3C and 4). Although located in a loop that is uniquely found in LH3/PLOD3, 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 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 less pronounced. This residue might respond to the rearrangements of its counterpart Trp145 in the catalytic site or contribute to long-range stabilizing interactions with collagen

molecules while they dock their Hyl or Gal-Hyl residues in the acceptor substrate site during the Glc-T reaction, 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 [40,45,46] (Figure S5). Conversely, extensive structural comparisons and mutagenesis experiments focusing on matching residue Gln189 have been performed in the *O*-galactosyltransferase LgtC from *Neisseria menengitidis* [47]. However, the location of this residue as a catalytic nucleophile in LH3/PLOD3 is inconsistent with its proposed function, as this site is occupied by Gln192. Such a 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. Mutagenesis results indicate that the removal of the Gln192 side chain has a strong impact on LH3/PLOD3 glycosyltransferase activity; however, indirect assays show that the enzyme is still partially active (Figure 2C). Nearby, we identified two other amino acid residues potentially involved in donor substrate activation or the transfer of sugar moieties to the acceptor molecule. Both Glu141 and Asn165 point directly towards the glycan moiety of the donor substrate (Figure 2A). Whilst the Asn165Ala was incapable of Glc transfer to acceptor substrates but still capable of UDP-Glc uncoupling (Figure 2B,C, Table 1), we found that the carboxylate moiety of Glu141 is essential for catalysis, as the Glu141Ala mutation yields a completely inactive LH3/PLOD3 glycosyltransferase. 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* (Figure S6, Table S2). As the removal of this carboxylate side chain completely abolishes LH3/PLOD3 glycosyltransferase activities, we suggest that the site corresponding to Glu141 might be a critical "hot spot" during catalysis in the glycosyltransferases characterized by the carboxylate same or amide side chains in similar positions.

Our work provides a comprehensive update to the biochemical landscape for the characterization of the Lys-to-Glc-Gal-Hyl conversion pathway based on reliable orthogonal in vitro assays and a new set of three-dimensional structures of LH3/PLOD3 and its mutants. The absence of experimental density in the co-crystal structures of LH3/PLOD3 with free UDP indicates that the retention of donor substrates in the GT domain of LH/PLODs requires the presence of both the UDP and the sugar moiety and that, regardless of the presence of glycan acceptor substrates, after hydrolysis, free UDP is rapidly released from the catalytic site, possibly through conformational changes involving the flexible glycoloop. In this respect, despite the high flexibility of the glycan moieties of the bound molecules, the new data obtained for LH3/PLOD3 in complex with UDP-sugar analogs acting as mild inhibitors (Figure 6) provide valuable insights for the structure-based drug development LH3/PLOD3 Glc-T enzymatic activity inhibitors. Whilst co-crystal structures with donor substrate consistently display no density for the sugar moiety because of uncoupled enzymatic hydrolysis or very low occupancy for a very flexible moiety, the inhibitors show density for the glycan part connected to UDP because hydrolysis is not taking place. This scenario is very similar to what we observed for the LH3/PLOD3 p.(Asp190Ser) mutant, where a significant reduction in catalytic activity (with no alterations in UDP-Glc binding affinity) results in a visible Glc moiety in the catalytic cavity (Figures 3B and 4). Therefore, the insights obtained using UDP-GlcA and UDP-Xyl molecules may offer opportunities for the development of innovative therapeutic strategies against pathological conditions characterized by excess collagen glycosylations, such as osteogenesis imperfecta [48]. 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/PLOD3 Glc-T activities.

4. Materials and Methods

4.1. Chemicals

All chemicals were purchased from Sigma-Aldrich (Merck) (St. Louis, MO, USA), unless otherwise specified.

4.2. Molecular Cloning and Site-Directed Mutagenesis

The LH3/PLOD3 (UniProt Q60568—obtained from Source Bioscience, Nottingham, UK), the LH1/PLOD1 (UniProt Q02809—obtained from Source Bioscience), and the GLT25D1 (UniProt Q8NBJ5—obtained as codon-optimized from GeneWiz, South Plainfield, NJ, USA) coding sequences, devoid of the N-terminal signal peptide, were amplified using oligonucleotides containing in-frame 5′-BamHI and 3′-NotI (Table S3) and cloned in a pCR8 vector, which was used as a template for the subsequent mutagenesis experiments. The LH/PLOD mutants were generated using the Phusion Site Directed Mutagenesis Kit (Thermo Fisher Scientific, Waltham, MA, USA). The entire plasmid was amplified using phosphorylated primers. For all mutants, the forward primer introduced the mutation of interest (Table S3). The linear mutagenized plasmid was phosphorylated using T4 polynucleotide kinase (Life Technologies, Carlsbad, CA, USA) prior to ligation. All plasmids were checked via Sanger sequencing prior to cloning into the expression vector. Wild-type and mutant LH3/PLOD3 and LH1/PLOD1 mature sequences were cloned into the pUPE.106.08 mammalian expression vector (U-protein Express BV) in frame with a 6xHis-tag followed by a Tobacco Etch Virus (TEV) protease cleavage site. The GLT25D1 was sub-cloned into a modified pET28b-SUMO vector (Novagen, Paris, France), yielding the final construct bearing an N-terminal 8xHis-SUMO fusion for recombinant production using *E. coli*.

4.3. LH/PLOD Recombinant Expression and Protein Purification

Suspension growing HEK293-F cells (Life Technologies, Paisley, 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, Warrington, PA, USA). Cells were harvested 6 days after transfection by centrifuging the medium for 15 min at 1000× *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 described elsewhere [33]. All proteins were isolated from the medium exploiting the affinity of the 6xHis tag for the HisTrap Excel (GE Healthcare, Chicago, IL, USA) affinity column. The purified proteins were 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}.

4.4. GLT25D1 Recombinant Expression and Protein Purification

The pET28b-SUMO-GLT25D1 plasmid was transformed in *E. coli* BL21-codonplus(DE3)-RP Plus (Agilent, Santa Clara, CA, USA), a single colony was picked and inoculated in 100 mL Luria–Bertani medium supplemented with 0.1 mg/mL Kanamycin (1:1000 *v*/*v*). This pre-culture was grown overnight at 37 °C in a shaking incubator set at 220 rpm. On the following day, this pre-culture was used to inoculate 6 L of autoinducing ZYP5052 medium [49] for large-scale production. The culture was grown for 4 h at 37 °C. Then, the culturing temperature was lowered to 20 °C for a total of 24 h prior to cell harvesting by centrifugation at 4000× *g* for 20 min. The resulting cell pellet was resuspended and homogenized in 100 mL of buffer A (25 mM HEPES/NaOH, 0.5 M NaCl, pH 8) in a 1:5 (*w*/*v*) wet cell pellet–buffer ratio and then disrupted by sonication (16 cycles, 9 s on, 6 s off pulses). The cell debris was removed via centrifugation (50,000× *g*, 40 min, 4 °C); the supernatant was then filtered through a1 µm syringe-driven filter (Minisart GF, Sartorius, Göttingen, Germany). The clarified lysate was loaded onto a His-Trap Excel 5 mL column (GE Healthcare) that was pre-equilibrated with buffer A using a NGC chromatography system (Bio-Rad, Hercules, CA, USA). Elution was carried out stepwise by adding imidazole up to a final 250 mM concentration. Fractions containing

His-SUMO-GLT25D1 were pooled, subjected to imidazole removal through passage on a HiTrap Desalting 5 mL column (GE Healthcare), supplemented with 3 µg/mL of His-tagged SUMO protease (1:300 v/v), and incubated for 2 h at room temperature. The sample was then loaded onto a His-Trap Excel 5 mL column (GE Healthcare) pre-equilibrated with buffer A. Cleaved GLT25D1 eluted in the flow-through fraction. The sample was concentrated to 5 mg mL^{-1} using a 30 kDa Amicon Ultra-15 centrifugal filter concentrator (Merck, Rahway, NJ, USA). The concentrated sample was the further polished via Size Exclusion Chromatography (SEC) through the use of a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated with SEC buffer (25 mM HEPES/NaOH, 200 mM NaCl, pH 8). Pure GLT25D1 peak fractions, as assessed by SDS-PAGE analysis, were further concentrated to 5 mg/mL, flash-frozen in liquid nitrogen, and stored at −80 °C until usage.

4.5. Direct Mass Spectrometry Activity Assays

Direct LH activity assays were performed using 5 µM LH/PLOD (wild-type or variants), 50 µM $FeCl_2$, 100 µM 2-OG, 500 µM Ascorbate, and 1 mM peptide substrate. For Gal-T activity measurements, 5 µM GLT25D1, 50 µM Mn, 100 µM UDP-Gal were added. For Glc-T activity measurements, 100 µM UDP-Glc was also added. Each experiment was performed in at least three technical repeats. Control experiments were performed by the selective removal of either the LH/PLOD or GLT25D1 enzyme, as described in the text and in the figure captions. All reactions were allowed to proceed for 3 h at 37 °C. Moreover, 10 µL of each reaction sample were mixed with 38 µL of Milli-Q water and acidified by the addition of 2 µL of formic acid (FA) to reach a total volume of 50 µL. These samples were placed into the cooled autosampler (10 °C) of a UHPLC system (AB Sciex, Framingham, MA, USA) connected to a high-resolution QTOF mass spectrometer (AB Sciex X500B) equipped with a Turbo V Ion source and a Twin-Sprayer electrospray ionization (ESI) probe, controlled by SCIEX OS 2.1 software. Peptides were separated via reverse phase (RP) HPLC on a Hypersil Gold (Thermo Fisher Scientific) C18 column (150 × 2.1 mm, 3 µm particle size, 175 Å pore size) using a linear gradient (2–50% solvent B in 15 min), with solvent A consisting of 0.1% aqueous formic acid (FA) and solvent B acetonitrile (ACN) supplemented with 0.1% FA. Flow rate was kept constant at 0.2 mL/min. Mass spectra were collected in positive polarity under constant instrumental conditions, which were as follows: ion spray voltage 4500 V, declustering potential 100 V, curtain gas 30 psi, ion source gas 1 40 psi, ion source gas 2 45 psi, temperature 350 °C, collision energy 10 V. Spectra analyses were performed using SCIEX OS 2.1 software.

4.6. Indirect Luminescence-Based Activity Assays

The LH and Glc-T activities were tested using luminescence-based enzymatic assays with a GloMax Discovery (Promega, Madison, WI, USA) as described in Scietti et al. [45]. Minor modifications were established for the Glc-T competitive inhibition assays, where 1 µL of a mixture of 250 µM $MnCl_2$, 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 triplicate. The control experiments were performed in the same conditions by selectively removing the LH/PLOD enzymes. Data were analyzed and plotted using GraphPad Prism 7 (Graphpad Software, San Diego, CA, USA).

4.7. Differential Scanning Fluorimetry (DSF)

DSF assays were performed on LH/PLOD wild-type and mutants using a Tycho NT.6 instrument (NanoTemper Technologies GmbH, München, Germany). The enzyme samples were kept 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/PLOD3 with 50 mM $MnCl_2$ and 5 mM free UDP or UDP-sugar donor substrates or their analogs. Data were analyzed and plotted using GraphPad Prism 7 (Graphpad Software, USA).

4.8. Crystallization, Data Collection, Structure Determination, and Refinement

Wild-type and mutant LH3/PLOD3 co-crystallization experiments were performed using the hanging-drop vapor-diffusion method protocols as described elsewhere [33], i.e., by mixing 0.5 µL of enzyme concentrated at 3.5 mg mL^{-1} and, depending on the experiment performed, pre-incubating with 500 µM $FeCl_2$, 500 µM $MnCl_2$, and supplementing with 1 mM of the appropriate UDP-sugar analogs (UDP, UDP-glucose, UDP-glucuronic acid, UDP-xylose) with 0.5 µL of reservoir solutions composed of 600 mM sodium formate, 12% poly-glutamic-acid (PGA-LM, Molecular Dimensions, Sheffield, UK), 100 mM HEPES/NaOH, pH 7.8. Crystals were cryo-protected with the mother liquor supplemented with 20% glycerol, harvested using MicroMounts Loops (Mitegen, Ithaca, NY, USA), 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 Table S1. 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/PLOD3 in complex with Fe^{2+}, 2-OG and Mn^{2+} (PDB ID: 6FXM) [33] as a search model using *PHASER* [53]. The structure was refined with successive steps of manual building in *COOT* [54], and automated refinement was carried out using *phenix.refine* [55]. Model validation was performed using *MolProbity* [56]. The refinement statistics for the final models are reported in Table S1.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms241311213/s1.

Author Contributions: A.C. and A.P. generated LH3/PLOD3, LH1/PLOD1, and GLT25D1 recombinant expression constructs. D.M. and A.C. purified LH3/PLOD3 mutants with support from S.F., carried out biochemical investigations with support from L.S. and L.N., and crystallized LH3/PLOD3 mutants with support from L.S.; L.S. purified and crystallized wild-type LH3/PLOD3 in complex with UDP-sugar analogs; D.M. and L.N. purified LH1/PLOD1 wild-type and mutants. M.D.M. purified and characterized GLT25D1. M.F. established and performed MS-based assays with support from L.S., D.M. and M.D.M.; D.M., M.D.M., F.D.G. and L.N. performed luminescence-based assays; L.S. and F.F. carried out structural refinement and analyses; T.R. was part of the study group focusing on LH3/PLOD3 pathogenic variants. F.F. designed and supervised the study with help from A.C. and L.S.; A.C., D.M., L.S. and F.F. wrote the paper, with contributions from all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Curie grant agreement COTETHERS—n. 745934 to A.C., the Italian Association for Cancer Research (AIRC, "My First AIRC Grant", id. 20075 and "Bridge Grant", id. 27004 to F.F.), the Ehlers-Danlos Society (Rarer Types Grant 2023 to F.F.), Fondazione Giovanni Armenise-Harvard (CDA2013 to F.F.), and the Mizutani Foundation for Glycoscience (grant No. 200039 to F.F.). Some of the instrumentation used for this research was acquired through funding from Regione Lombardia, regional law n° 9/2020, resolution n° 3776/2020. None of the funders had roles in the design of the study; in the collection, analysis, and interpretation of data; or in the writing of the article and the decision to submit it for publication.

Data Availability Statement: Coordinates and Structure Factors of the new molecular structures presented in this study have been deposited in the Protein Data Bank under accession codes 6TE3, 6TES, 6TEC, 6TEU, 6TEX, 6TEZ, 8ONE. Other data are available from the corresponding author upon reasonable request.

Acknowledgments: We thank Sristi Raj Rai for the assistance they provided during the purification of LH/PLOD and GLT25D1 enzymes, M. Campioni for partaking in useful discussions, and M. Miao for their support in the crystallization experiments. We thank the European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS) for providing the synchrotron radiation facilities.

We thank the Centro Grandi Strumenti (CGS) of the University of Pavia for providing HR-MS instrumentation and excellent support.

Conflicts of Interest: Authors declare no conflict of interest.

References

1. Myllyharju, J.; Kivirikko, K.I. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* **2004**, *20*, 33–43. [CrossRef]
2. Luther, K.B.; Hulsmeier, A.J.; Schegg, B.; Deuber, S.A.; Raoult, D.; Hennet, T. Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. *J. Biol. Chem.* **2011**, *286*, 43701–43709. [CrossRef]
3. Cummings, R.D. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **2009**, *5*, 1087–1104. [CrossRef]
4. Spiro, R.G. Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. *J. Biol. Chem.* **1969**, *244*, 602–612. [CrossRef]
5. Spiro, R.G. The structure of the disaccharide unit of the renal glomerular basement membrane. *J. Biol. Chem.* **1967**, *242*, 4813–4823. [CrossRef]
6. Spiro, M.J.; Spiro, R.G. Studies on the biosynthesis of the hydroxylsine-linked disaccharide unit of basement membranes and collagens. II. Kidney galactosyltransferase. *J. Biol. Chem.* **1971**, *246*, 4910–4918. [CrossRef]
7. Hamazaki, H.; Hamazaki, M.H. Catalytic site of human protein-glucosylgalactosylhydroxylysine glucosidase: Three crucial carboxyl residues were determined by cloning and site-directed mutagenesis. *Biochem. Biophys. Res. Commun.* **2016**, *469*, 357–362. [CrossRef]
8. Sternberg, M.; Spiro, R.G. 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. *Ren. Physiol.* **1980**, *3*, 1–3. [CrossRef]
9. Sternberg, M.; Grochulski, A.; Peyroux, J.; Hirbec, G.; Poirier, J. 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. *Coll. Relat. Res.* **1982**, *2*, 495–506. [CrossRef]
10. Bornstein, P.; Sage, H. Structurally distinct collagen types. *Annu. Rev. Biochem.* **1980**, *49*, 957–1003. [CrossRef]
11. Terajima, M.; Perdivara, I.; Sricholpech, M.; Deguchi, Y.; Pleshko, N.; Tomer, K.B.; Yamauchi, M. Glycosylation and cross-linking in bone type I collagen. *J. Biol. Chem.* **2014**, *289*, 22636–22647. [CrossRef]
12. Moro, L.; Romanello, M.; Favia, A.; Lamanna, M.P.; Lozupone, E. Posttranslational modifications of bone collagen type I are related to the function of rat femoral regions. *Calcif. Tissue Int.* **2000**, *66*, 151–156. [CrossRef]
13. Schofield, J.D.; Freeman, I.L.; Jackson, D.S. The isolation, and amino acid and carbohydrate composition, of polymeric collagens prepared from various human tissues. *Biochem. J.* **1971**, *124*, 467–473. [CrossRef]
14. Toole, B.P.; Kang, A.H.; Trelstad, R.L.; Gross, J. Collagen heterogeneity within different growth regions of long bones of rachitic and non-rachitic chicks. *Biochem. J.* **1972**, *127*, 715–720. [CrossRef]
15. Rautavuoma, K.; Takaluoma, K.; Sormunen, R.; Myllyharju, J.; Kivirikko, K.I.; Soininen, R. Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14120–14125. [CrossRef]
16. Sipila, L.; Ruotsalainen, H.; Sormunen, R.; Baker, N.L.; Lamande, S.R.; Vapola, M.; Wang, C.; Sado, Y.; Aszodi, A.; Myllyla, R. Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. *J. Biol. Chem.* **2007**, *282*, 33381–33388. [CrossRef]
17. Lehmann, H.W.; Wolf, E.; Roser, K.; Bodo, M.; Delling, G.; Muller, P.K. Composition and posttranslational modification of individual collagen chains from osteosarcomas and osteofibrous dysplasias. *J. Cancer Res. Clin. Oncol.* **1995**, *121*, 413–418. [CrossRef]
18. Salo, A.M.; Cox, H.; Farndon, P.; Moss, C.; Grindulis, H.; Risteli, M.; Robins, S.P.; Myllyla, R. A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. *Am. J. Hum. Genet.* **2008**, *83*, 495–503. [CrossRef]
19. Tenni, R.; Valli, M.; Rossi, A.; Cetta, G. Possible role of overglycosylation in the type I collagen triple helical domain in the molecular pathogenesis of osteogenesis imperfecta. *Am. J. Med. Genet.* **1993**, *45*, 252–256. [CrossRef]
20. Hennet, T. Collagen glycosylation. *Curr. Opin. Struct. Biol.* **2019**, *56*, 131–138. [CrossRef] [PubMed]
21. De Giorgi, F.; Fumagalli, M.; Scietti, L.; Forneris, F. Collagen hydroxylysine glycosylation: Non-conventional substrates for atypical glycosyltransferase enzymes. *Biochem. Soc. Trans.* **2021**, *49*, 855–866. [CrossRef] [PubMed]
22. Wang, C.; Luosujarvi, H.; Heikkinen, J.; Risteli, M.; Uitto, L.; Myllyla, R. The third activity for lysyl hydroxylase 3: Galactosylation of hydroxylysyl residues in collagens in vitro. *Matrix Biol.* **2002**, *21*, 559–566. [CrossRef] [PubMed]
23. Koenig, S.N.; Cavus, O.; Williams, J.; Bernier, M.; Tonniges, J.; Sucharski, H.; Dew, T.; Akel, M.; Baker, P.; Madiai, F.; et al. New mechanistic insights to PLOD1-mediated human vascular disease. *Transl. Res.* **2022**, *239*, 1–17. [CrossRef] [PubMed]
24. Guo, H.F.; Bota-Rabassedas, N.; Terajima, M.; Leticia Rodriguez, B.; Gibbons, D.L.; Chen, Y.; Banerjee, P.; Tsai, C.L.; Tan, X.; Liu, X.; et al. A collagen glucosyltransferase drives lung adenocarcinoma progression in mice. *Commun. Biol.* **2021**, *4*, 482. [CrossRef]
25. Savolainen, E.R.; Kero, M.; Pihlajaniemi, T.; Kivirikko, K.I. Deficiency of galactosylhydroxylysyl glucosyltransferase, an enzyme of collagen synthesis, in a family with dominant epidermolysis bullosa simplex. *N. Engl. J. Med.* **1981**, *304*, 197–204. [CrossRef]

26. Ewans, L.J.; Colley, A.; Gaston-Massuet, C.; Gualtieri, A.; Cowley, M.J.; Mccabe, M.J.; Anand, D.; Lachke, S.A.; Scietti, L.; Forneris, F.; et al. Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. *J. Med. Genet.* **2019**, *56*, 629–638. [CrossRef]
27. Ruotsalainen, H.; Sipila, L.; Vapola, M.; Sormunen, R.; Salo, A.M.; Uitto, L.; Mercer, D.K.; Robins, S.P.; Risteli, M.; Aszodi, A.; et al. Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. *J. Cell. Sci.* **2006**, *119*, 625–635. [CrossRef]
28. Schegg, B.; Hulsmeier, A.J.; Rutschmann, C.; Maag, C.; Hennet, T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. *Mol. Cell. Biol.* **2009**, *29*, 943–952. [CrossRef]
29. Perrin-Tricaud, C.; Rutschmann, C.; Hennet, T. Identification of domains and amino acids essential to the collagen galactosyltransferase activity of GLT25D1. *PLoS ONE* **2011**, *6*, e29390. [CrossRef]
30. Sricholpech, M.; Perdivara, I.; Nagaoka, H.; Yokoyama, M.; Tomer, K.B.; Yamauchi, M. Lysyl hydroxylase 3 glucosylates galactosylhydroxylysine residues in type I collagen in osteoblast culture. *J. Biol. Chem.* **2011**, *286*, 8846–8856. [CrossRef]
31. Yamauchi, M.; Sricholpech, M. Lysine post-translational modifications of collagen. *Essays Biochem.* **2012**, *52*, 113–133. [PubMed]
32. Baumann, S.; Hennet, T. Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase. *J. Biol. Chem.* **2016**, *291*, 18514–18524. [CrossRef] [PubMed]
33. Scietti, L.; Chiapparino, A.; De Giorgi, F.; Fumagalli, M.; Khoriauli, L.; Nergadze, S.; Basu, S.; Olieric, V.; Cucca, L.; Banushi, B.; et al. Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. *Nat. Commun.* **2018**, *9*, 3163. [CrossRef] [PubMed]
34. Wang, C.; Risteli, M.; Heikkinen, J.; Hussa, A.K.; Uitto, L.; Myllyla, R. Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). *J. Biol. Chem.* **2002**, *277*, 18568–18573. [CrossRef]
35. Scietti, L.; Campioni, M.; Forneris, F. SiMPLOD, a structure-integrated database of collagen lysyl hydroxylase (LH/PLOD) enzyme variants. *J. Bone Miner. Res.* **2019**, *34*, 1376–1382. [CrossRef]
36. Laskowski, R.A.; Swindells, M.B. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **2011**, *51*, 2778–2786. [CrossRef]
37. Lairson, L.L.; Henrissat, B.; Davies, G.J.; Withers, S.G. Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* **2008**, *77*, 521–555. [CrossRef]
38. Ardevol, A.; Iglesias-Fernandez, J.; Rojas-Cervellera, V.; Rovira, C. The reaction mechanism of retaining glycosyltransferases. *Biochem. Soc. Trans.* **2016**, *44*, 51–60. [CrossRef]
39. Gloster, T.M. Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **2014**, *28*, 131–141. [CrossRef]
40. Albesa-Jove, D.; Sainz-Polo, M.A.; Marina, A.; Guerin, M.E. Structural Snapshots of alpha-1,3-Galactosyltransferase with Native Substrates: Insight into the Catalytic Mechanism of Retaining Glycosyltransferases. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 14853–14857. [CrossRef]
41. Coutinho, P.M.; Deleury, E.; Davies, G.J.; Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **2003**, *328*, 307–317. [CrossRef] [PubMed]
42. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef] [PubMed]
43. Kivirikko, K.I.; Myllyla, R. Collagen glycosyltransferases. *Int. Rev. Connect. Tissue Res.* **1979**, *8*, 23–72. [PubMed]
44. Myllyla, R.; Anttinen, H.; Risteli, L.; Kivirikko, K.I. Isolation of collagen glucosyltransferase as a homogeneous protein from chick embryos. *Biochim. Biophys. Acta* **1977**, *480*, 113–121. [CrossRef]
45. Gomez, H.; Lluch, J.M.; Masgrau, L. Essential role of glutamate 317 in galactosyl transfer by alpha3GalT: A computational study. *Carbohydr. Res.* **2012**, *356*, 204–208. [CrossRef]
46. Patenaude, S.I.; Seto, N.O.; Borisova, S.N.; Szpacenko, A.; Marcus, S.L.; Palcic, M.M.; Evans, S.V. The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nat. Struct. Biol.* **2002**, *9*, 685–690. [CrossRef]
47. Lairson, L.L.; Chiu, C.P.; Ly, H.D.; He, S.; Wakarchuk, W.W.; Strynadka, N.C.; Withers, S.G. Intermediate trapping on a mutant retaining alpha-galactosyltransferase identifies an unexpected aspartate residue. *J. Biol. Chem.* **2004**, *279*, 28339–28344. [CrossRef]
48. Raghunath, M.; Bruckner, P.; Steinmann, B. Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. *J. Mol. Biol.* **1994**, *236*, 940–949. [CrossRef]
49. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005**, *41*, 207–234. [CrossRef]
50. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132. [CrossRef]
51. Evans, P.R.; Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204–1214. [CrossRef]
52. Tickle, I.J.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. *STARANISO*; Global Phasing Ltd.: Cambridge, UK, 2018.
53. Mccoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674. [CrossRef]
54. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D* **2010**, *66*, 486–501. [CrossRef]

55. Adams, P.D.; Afonine, P.V.; Bunkoczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213–221. [CrossRef]
56. Chen, V.B.; Arendall, W.B., 3rd; Headd, J.J.; Keedy, D.A.; Immormino, R.M.; Kapral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12–21. [CrossRef]

Supplementary Data for

Identification of regulatory molecular hot spots for LH/PLOD collagen glycosyltransferase activity

D. Mattoteia, A. Chiapparino, *et al.*

Table S1: crystallographic statistics for data collection, structure solution and refinement.

[a] Values in parentheses are for reflections in the highest resolution shell.

	LH3/PLOD3 + Fe^{2+} + Mn^{2+} + UDP	LH3/PLOD3 + Fe^{2+} + Mn^{2+} + UDP-GlcA	LH3/PLOD3 + Fe^{2+} + Mn^{2+} + UDP-Xyl	LH3/PLOD3 Val80Lys + Fe^{2+} + Mn^{2+}	LH3/PLOD3 Val80Lys + Fe^{2+} + Mn^{2+} + UDP-Glc	LH3/PLOD3 Val80Lys + Fe^{2+} + Mn^{2+} + UDP-GlcA	LH3/PLOD3 Asp190Ser + Fe^{2+} + Mn^{2+} + UDP-Glc
Data Collection[a]							
X-ray source	ESRF ID30A-3	SLS X06SA	ESRF ID23-EH2	SLS X06SA	SLS X06SA	SLS X06SA	ESRF ID23-EH2
Processing programs	XDS, AIMLESS, STARANISO	XDS, AIMLESS, STARANISO	XDS, AIMLESS, STARANISO	XDS, AIMLESS, STARANISO	XDS, AIMLESS, STARANISO	XDS, AIMLESS, STARANISO	XDS, AIMLESS, STARANISO
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Cell parameters	a = 97.0 Å; α = 90° b = 100.0 Å; β = 90° c = 225.2 Å; γ = 90°	a = 98.2 Å; α = 90° b = 100.5 Å; β = 90° c = 224.7 Å; γ = 90°	a = 97.2 Å; α = 90° b = 100.0 Å; β = 90° c = 224.0 Å; γ = 90°	a = 98.0 Å; α = 90° b = 100.8 Å; β = 90° c = 225.7 Å; γ = 90°	a = 98.1 Å; α = 90° b = 100.4 Å; β = 90° c = 225.2 Å; γ = 90°	a = 98.0 Å; α = 90° b = 99.8 Å; β = 90° c = 224.5 Å; γ = 90°	a = 97.1 Å; α = 90° b = 100.2 Å; β = 90° c = 223.8 Å; γ = 90°
Wavelength (Å)	0.968	1.000	0.873	1.000	1.000	1.000	0.775
Resolution (Å)	48.84-2.30 (2.38-2.30)	49.10-2.20 (2.26-2.20)	48.79-2.40 (2.49-2.40)	49.14-3.00 (3.18-3.00)	49.02-2.30 (2.38-2.30)	49.00-2.70 (2.83-2.70)	48.88-2.30 (2.38-2.30)
Total reflections	230588 (21551)	430059 (32322)	237715 (25682)	93694 (15275)	251674 (15330)	197473 (22212)	322199 (25283)
Unique reflections	48288 (4455)	56604 (4537)	43005 (4500)	22529 (3578)	49068 (4243)	30477 (3911)	48804 (4443)
CC1/2[b]	0.997 (0.512)	0.999 (0.860)	0.992 (0.586)	0.986 (0.490)	0.998 (0.461)	0.996 (0.580)	0.995 (0.587)
Redundancy	4.8 (4.8)	7.6 (7.1)	5.5 (5.7)	4.2 (4.3)	5.1 (3.6)	6.5 (5.7)	6.6 (5.7)
Mean I/σ(I)	8.4 (0.9)	11.8 (0.5)	5.5 (0.7)	4.5 (0.7)	6.7 (0.8)	9.4 (1.3)	5.7 (0.5)
Completeness (%)	98.7 (99.7)	99.8 (98.6)	99.9 (99.9)	99.0 (98.5)	98.7 (94.0)	99.5 (98.3)	99.9 (99.9)
R_{sym}[b]	0.118 (1.476)	0.104 (2.566)	0.209 (1.831)	0.227 (3.045)	0.100 (1.351)	0.138 (1.312)	0.175 (1.827)
R_{pim}[c]	0.086 (1.107)	0.060 (1.554)	0.147 (1.274)	0.186 (2.479)	0.071 (1.155)	0.088 (0.902)	0.107 (1.255)

[b] $R_{sym} = | \Sigma_{hkl} \Sigma_j | I_{hkl,j} - \langle I_{hkl} \rangle | | / | \Sigma_{hkl} \Sigma_j I_{hkl,j} |$, where I is the observed intensity for a reflection and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[c] $R_{pim} = [\Sigma_{hkl} (1/(n-1))^{1/2} \Sigma_j | I_{hkl,j} - \langle I_{hkl} \rangle |] / [\Sigma_{hkl} \Sigma_j I_{hkl,j}]$ where I is the observed intensity for a reflection and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

	LH3/PLOD3 + Fe^{2+} + Mn^{2+} + UDP	LH3/PLOD3 + Fe^{2+} + Mn^{2+} + UDP-GlcA	LH3/PLOD3 + Fe^{2+} + Mn^{2+} + UDP-Xyl	LH3/PLOD3 Val80Lys + Fe^{2+} + Mn^{2+}	LH3/PLOD3 Val80Lys + Fe^{2+} + Mn^{2+} + UDP-Glc	LH3/PLOD3 Val80Lys + Fe^{2+} + Mn^{2+} + UDP-GlcA	LH3/PLOD3 Asp190Ser + Fe^{2+} + Mn^{2+} + UDP-Glc
Refinement							
R_{work}/R_{free}[c]	0.1943/0.2458	0.1982/0.2411	0.1852/0.2159	0.2079/0.2412	0.1701/0.2268	0.1895/0.2278	0.2043/0.2288
Number of atoms:	5874	6293	6019	5732	5938	5893	6125
Protein	5646	5754	5754	5665	5736	5716	5693
Ligands	73	100	105	67	92	104	109
Solvent	155	439	160	-	110	73	323
Average B-factor ($Å^2$)	35.83	36.51	31.45	42.33	44.80	43.73	35.71
Protein	35.73	35.91	31.16	42.07	44.54	42.99	35.33
Ligands	56.66	58.25	53.94	63.99	64.67	91.24	54.65
Solvent	29.78	39.40	26.84	-	41.54	33.94	36.11
Structure quality							
RMS bond lengths (Å)	0.006	0.005	0.003	0.004	0.009	0.003	0.002
RMS bond angles (°)	0.78	0.84	0.58	0.70	1.05	0.71	0.50
Ramachandran stats							
Favored (%)	97.5	97.1	97.0	95.9	96.3	96.5	96.5
allowed (%)	2.5	2.7	2.9	3.8	3.6	3.3	3.0
outliers (%)	0.0	0.2	0.1	0.3	0.1	0.2	0.5
PDB ID	6TE3	6TES	6TEC	6TEU	6TEX	6TEZ	8ONE

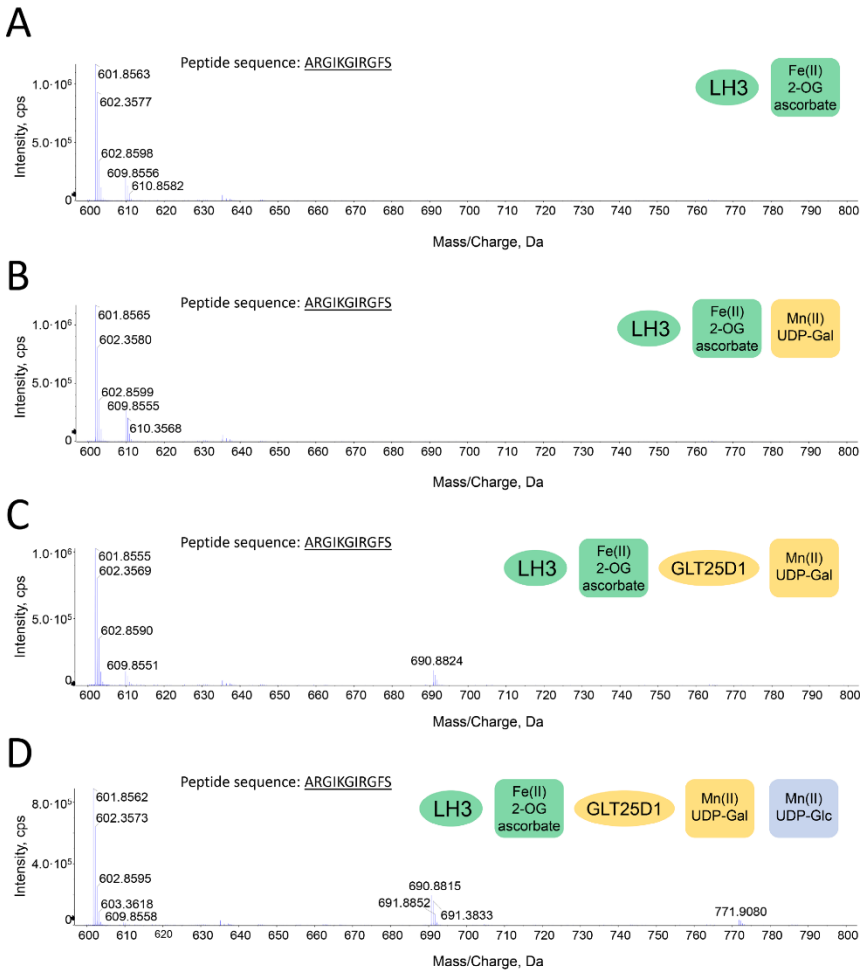
[c] R_{free} values are calculated based on 5% randomly selected reflections, selected prior to STARANISO correction as recommended by the software developers (Tickle et al, 2018).

Table S2: List of glycosyltransferase enzymes used for comparisons with human LH3/PLOD3. 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/PLOD3 based on structural superpositions.

Protein Name	Type	PDB ID	catalytic base residue	nucleophile acceptor residue	corresp. LH3/PLOD3 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

Table S3: list of oligonucleotides used for mutagenesis

Oligonucleotide Name	Sequence (5'→3')
Forward LH3/PLOD3 Val80Lys	AAGGCTCGAACAGTTGGTGGAGGAC
Reverse LH3/PLOD3 Val80Lys	ATCACCCCCTCGCCACTCCTC
Forward LH3/PLOD3 Val80Gly	GAGCTCGAACAGTTGGTGGAGGAC
Reverse LH3/PLOD3 Val80Gly	CATCACCCCCTCGCCACTCC
Forward LH3/PLOD3 Trp92Ala	GCATTAAAGAAGGAAATGGAGAAATACG
Reverse LH3/PLOD3 Trp92Ala	CCGGACCTTCTGTCCTCCACC
Forward LH3/PLOD3 Glu141Ala	CGAGCTTCTGCTGGCCCGAGTG
Reverse LH3/PLOD3 Glu141Ala	CTGCAGAGAAGAGCAGGCGGCTG
Forward LH3/PLOD3 Trp145Ala	GCACCCGAGTGGGGGCTGGC
Reverse LH3/PLOD3 Trp145Ala	GCAGAAGCTCTCTGCAGAGAAGAGC
Forward LH3/PLOD3 Trp148Ala	GCAGGGCTGGCGGAGCAGTAC
Reverse LH3/PLOD3 Trp148Ala	CTCGGGCCAGCAGAAGCTCTC
Forward LH3/PLOD3 Asn165Ala	GCTTCTGGTGGATTCATCGGTTTTGC
Reverse LH3/PLOD3 Asn165Ala	GAGGAAGCGCTTCCCCGTGC
Forward LH3/PLOD3 Asp190Ala	CTGACCAGCTGTTCTACACACGGC
Reverse LH3/PLOD3 Asp190Ala	CGTCATCATCCTTGTACTTCCACTGG
Forward LH3/PLOD3 Asp190Ser	AGCGACCAGCTGTTCTACACACGGCT
Reverse LH3/PLOD3 Asp190Ser	GTCATCATCCTTGTACTTCCACTGGCG
Forward LH3/PLOD3 Asp191Ala	CTCAGCTGTTCTACACACGGCTC
Reverse LH3/PLOD3 Asp191Ala	CGTCGTCATCATCCTTGTACTTCCAC
Forward LH3/PLOD3 Gln192Ala	GCGCTGTTCTACACACGGCTCTAC
Reverse LH3/PLOD3 Gln192Ala	GTCGTCGTCATCATCCTTGTACTTCC
Forward LH3/PLOD3 Asn255Ala	GCCGGTCCCACTAAGCTGCAGC
Reverse LH3/PLOD3 Asn255Ala	TCCATGGACCACAATGGGGAGCGTG
Forward LH3/PLOD3 Pro270Leu	TCAATGGCTGGACTCCTGAGGG
Reverse LH3/PLOD3 Pro270Leu	GGACGTAGTTTCCCAGGTAGTTGAG
Forward LH1/PLOD1 Ser178Asp	GACGATCAGCTGTTTTACACCAAGATC
Reverse LH1/PLOD1 Ser178Asp	GTCGCTGTCCTGGCCCTCCCACTCGG


A
Peptide sequence: ARGIKGIRGFS
LH3
Fe(II) 2-OG ascorbate
Intensity, cps
Mass/Charge, Da
B
Peptide sequence: ARGIKGIRGFS
LH3
Fe(II) 2-OG ascorbate
Mn(II) UDP-Gal
Intensity, cps
Mass/Charge, Da
C
Peptide sequence: ARGIKGIRGFS
LH3
Fe(II) 2-OG ascorbate
GLT25D1
Mn(II) UDP-Gal
Intensity, cps
Mass/Charge, Da
D
Peptide sequence: ARGIKGIRGFS
LH3
Fe(II) 2-OG ascorbate
GLT25D1
Mn(II) UDP-Gal
Mn(II) UDP-Glc
Intensity, cps
Mass/Charge, Da

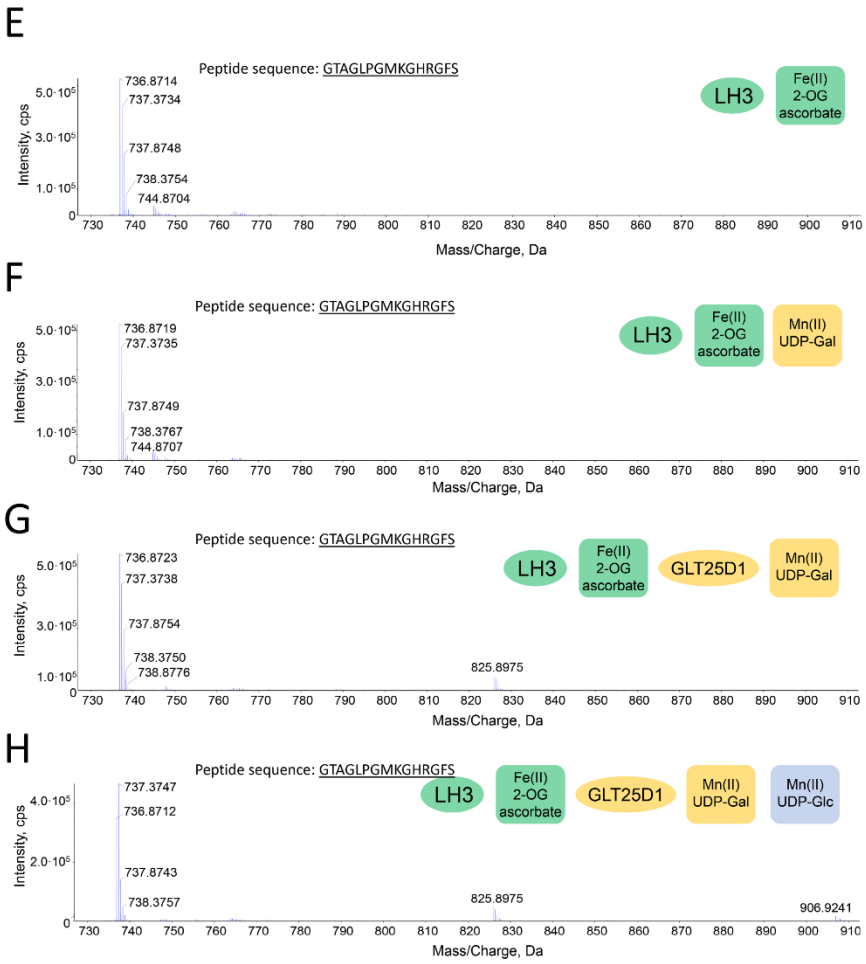


Figure S1: Results of the direct MS activity assays with two different synthetic peptides other than GIKGIKGIKGIK shown in figure 1. Panels (A) to (D): peptide sequence ARGIKGIRGFS; panels (E) to (H): peptide sequence GTAGLPGMKGHRGFS.

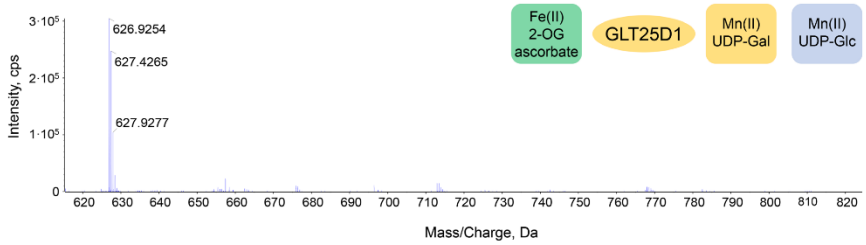


Figure S2: Direct MS activity assay to evaluate Lys-to-Hyl conversion in absence of LH/PLOD enzymes. The experiment shown was performed using the GIKGIKGIKGIK peptide. Results exclusively show the peak corresponding to the unmodified peptide (627 Da, doubly charged).

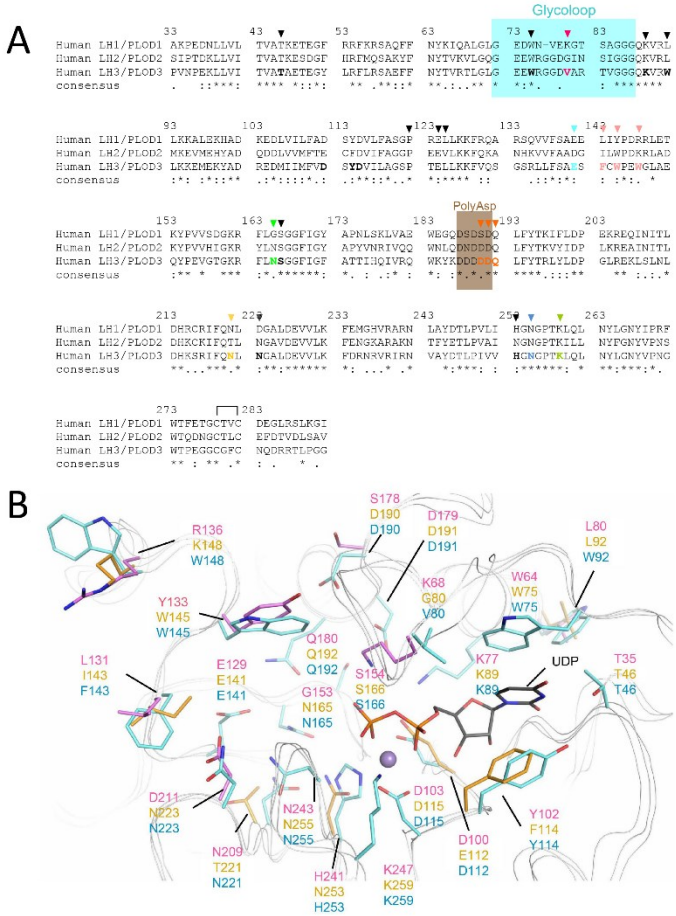


Figure S3: Conserved features of the GT domains in human LH/PLOD isoenzymes. (A) sequence alignment of human LH1/PLOD1, LH2/PLOD2 and LH3/PLOD3 GT domains, highlighting similarities and differences in the amino acid residues within the active site. Residues highlighted in Figure 2A are indicated here with a triangle. Colored boxes indicate the PolyAsp motif (brown) and the Glycoloop (cyan). (B) Superposition of the GT site of LH3/PLOD3 in complex with UDP-glucose (PDB ID: 6FXT) and LH1/PLOD1 and LH2/PLOD2 homology models (Scietti et al, 2019) (available at http://fornerislab.unipv.it/SiMPLOD/). Aminoacids identified as part of the LH3/PLOD3 catalytic site are shown as light blue sticks, while the LH1/PLOD1 and LH2/PLOD2 residues which differs from LH3/PLOD are shown in pink and orange respectively. UDP moiety (black lines) and Mn^{2+} cofactor (purple sphere) are also shown from the LH3/PLOD3 structure.

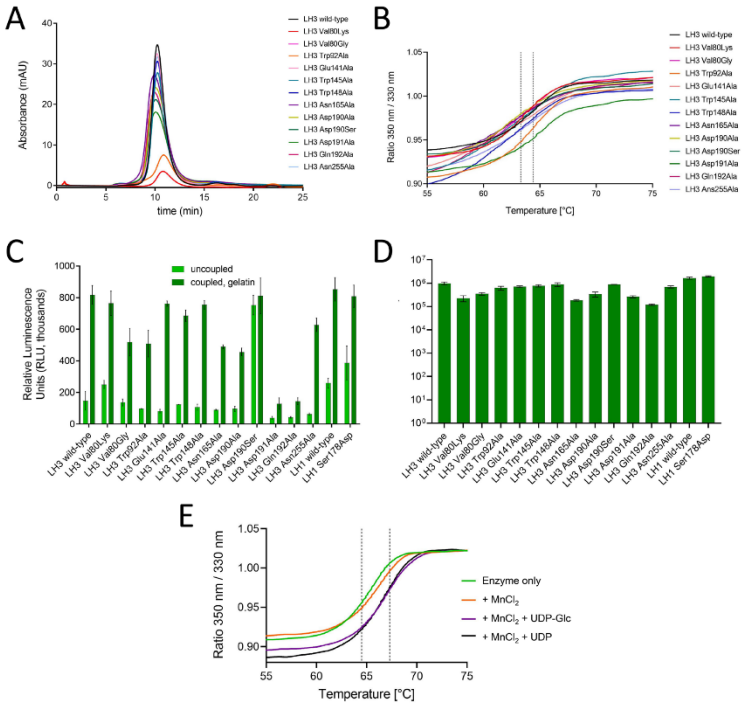


Figure S4: biochemical evaluation of LH3/PLOD3 wild-type and mutants. (A) Analytical size-exclusion chromatography analysis comparing wild-type and mutants LH3/PLOD3. (B) DSF comparing wild-type and mutants LH3/PLOD3. The dashed lines indicate the temperature range incorporating the calculated T_m values for all curves. (C) LH activity comparison for wild-type and mutants LH3/PLOD3 using indirect luminescence-based assays. The measurements were performed in absence ("uncoupled") and in presence ("coupled") of gelatin acceptor substrates, as described in (Scietti et al, 2018). Error bars represent standard deviations from average of triplicate independent experiments. (D) Assessment of LH enzymatic activity using direct MS-based assays for all LH3/PLOD3 mutants described in this work. (E) DSF analysis of UDP and UDP-Glc binding by LH3/PLOD3 Asp190Ser mutant.

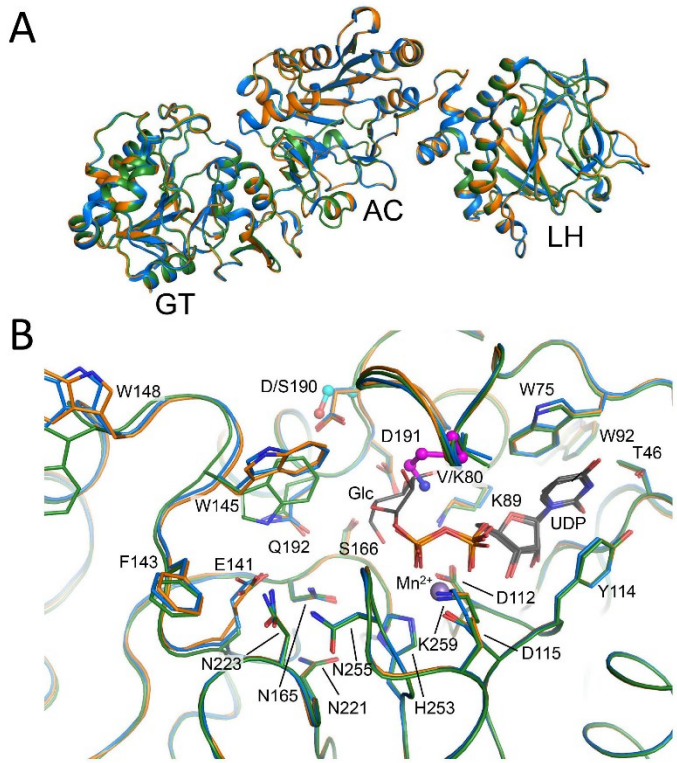
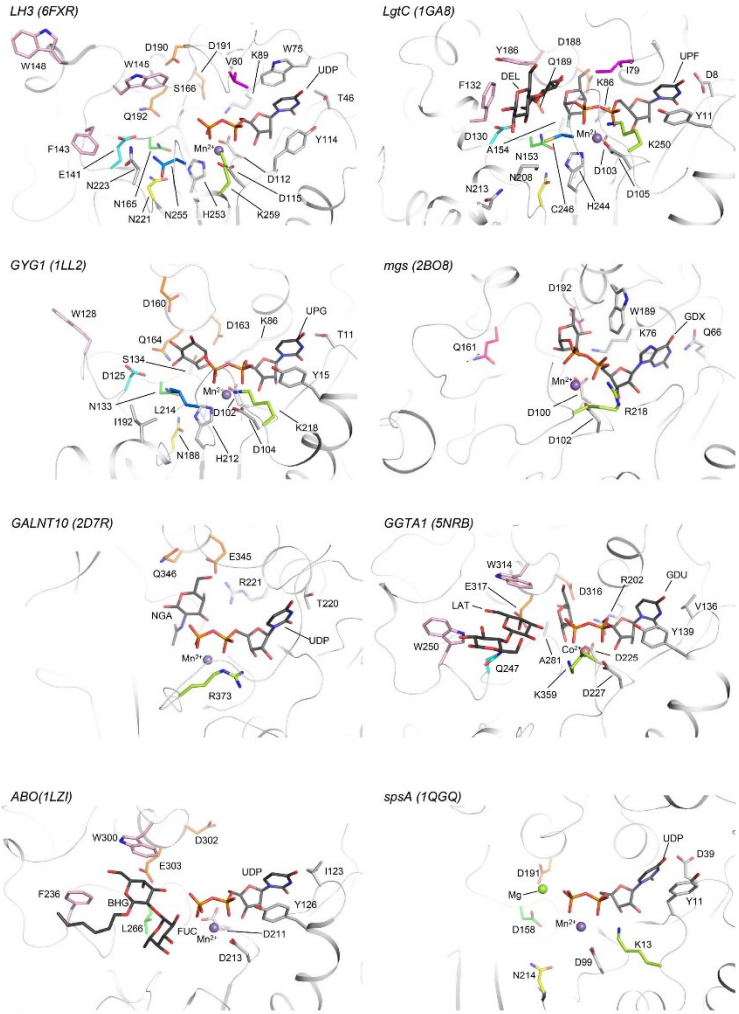


Figure S5: Structural comparison of the LH3/PLOD3 wild-type and mutant structures in complex with UDP-glucose. (A) The overall structure superposition of the asymmetric units of LH3/PLOD3 wild-type (blue), Val80Lys (orange), and Asp190Ser (green) shows that the two structures are almost identical. (B) Zoomed view of the GT active site of the superimposed structures shown in (A) with the same overall coloring, indicating that the sole differences are constituted by the mutated Val80 to Lys in the Val80Lys structure (shown in pink ball-and-stick) and the mutated Asp190 to Ser in the Asp190Ser structure (shown in cyan ball-and stick).


LH3 (6FXR)
W148
D190
D191
K89
W75
W145
V80
UDP
S166
T46
Q192
Y114
F143
Mn2+
E141
D112
N223
D115
N165
N221
N255
H253
K259
LgtC (1GA8)
D188
Y186
Q189
UPF
I79
D8
F132
DEL
K86
Y11
D130
Mn2+
K250
A154
N153
D103
N208
D105
N213
C246
H244
GYG1 (1LL2)
D160
W128
D163
K86
UPG
T11
Q164
S134
Y15
D125
Mn2+
N133
L214
D102
I192
K218
N188
H212
D104
mgs (2BO8)
D192
W189
GDX
K76
Q66
Q161
Mn2+
D100
R218
D102
GALNT10 (2D7R)
E345
Q346
R221
T220
NGA
UDP
Mn2+
R373
GGTA1 (5NRB)
W314
E317
D316
R202
GDU
LAT
V136
Y139
W250
Co2+
A281
D225
Q247
K359
D227
ABO(1LZI)
W300
D302
E303
UDP
I123
F236
BHG
Y126
L266
FUC
Mn2+
D211
D213
spsA (1QGQ)
UDP
D39
D191
Mg
Y11
D158
Mn2+
K13
D99
N214

S12

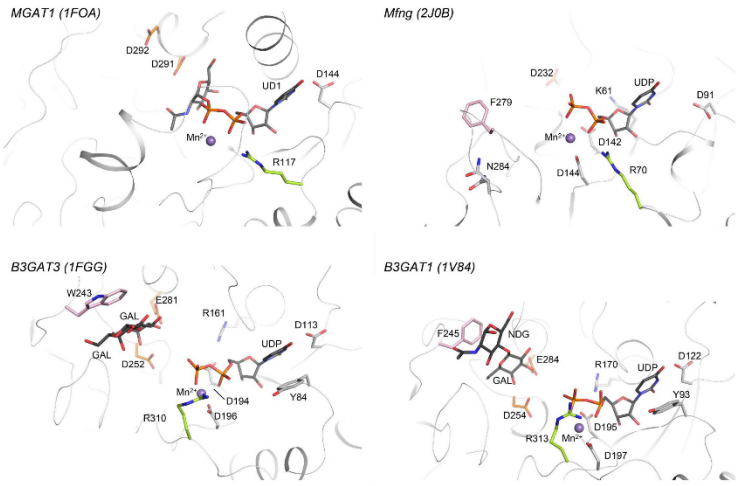


Figure S6: structural comparison between LH3/PLOD3 and other glycosyltranferases. The conserved aminoacids are shown as sticks; where present, Mn^{2+} cofactor and UDP-sugar are shown as purple sphere and black sticks respectively. Protein name and related PDB ID are indicated for each panel. Colour coding is as in Figure 2A and is maintained throughout the structures allowing to compare the LH3/PLOD3 critical residues with the other structures. For additional information on the proteins indicated in this panel refer to Table S2.

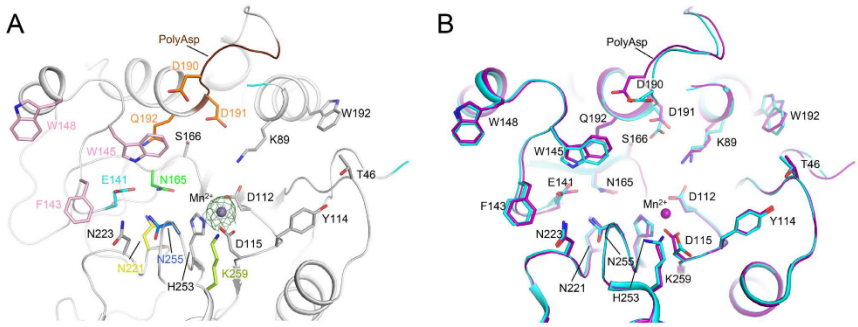


Figure S7: comparison of the GT domains of wild-type LH3/PLOD3 in ligand-bound and ligand-free states. (A) The structure of wild-type LH3/PLOD3 crystallized in presence of Mn^{2+} and UDP shows clear electron density for Mn^{2+} ($2F_o$-F_c omit electron density map shown as green mesh, contour level 2 σ), but unexpectedly no density is present for UDP. Residue highlight and colouring as in Figure 1A. (B) The comparative superposition of wild-type LH3/PLOD3 GT domain co-crystallized in presence of Mn^{2+} and UDP (purple), and without cofactors (cyan – PDB ID: 6FXK) shows no differences in the conformations of the side chains for the residues delimiting the Glc-T catalytic site.

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fmolb.2022.876352

OPEN ACCESS

EDITED BY
Gianluca Molla,
University of Insubria, Italy

REVIEWED BY
Aiwu Zhou,
Shanghai Jiao Tong University, China
Teresita Padilla-Benavides,
Wesleyan University, United States
Gabi U. Dachs,
University of Otago, Christchurch, New Zealand

*CORRESPONDENCE
Luigi Scietti,
luigi.scietti@unipv.it
Federico Forneris,
federico.forneris@unipv.it

†Present address
Luigi Scietti, Biochemistry and Structural Biology Unit, Department of Experimental Oncology, IRCCS European Institute of Oncology (IEO), Milan, Italy.

‡These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 15 February 2022
ACCEPTED 11 July 2022
PUBLISHED 25 August 2022

CITATION
Scietti L, Moroni E, Mattoteia D, Fumagalli M, De Marco M, Negro L, Chiapparino A, Serapian SA, De Giorgi F, Faravelli S, Colombo G and Forneris F (2022), A Fe^{2+}-dependent self-inhibited state influences the druggability of human collagen lysyl hydroxylase (LH/PLOD) enzymes. *Front. Mol. Biosci.* 9:876352. doi: 10.3389/fmolb.2022.876352

A Fe^{2+}-dependent self-inhibited state influences the druggability of human collagen lysyl hydroxylase (LH/PLOD) enzymes

Luigi Scietti[1*†‡], Elisabetta Moroni[2‡], Daiana Mattoteia[1‡], Marco Fumagalli[1], Matteo De Marco[1], Lisa Negro[1], Antonella Chiapparino[1], Stefano A. Serapian[3], Francesca De Giorgi[1], Silvia Faravelli[1], Giorgio Colombo[3] and Federico Forneris[1*]

[1]The Armenise-Harvard Laboratory of Structural Biology, Department of Biology and Biotechnology, University of Pavia, Pavia, Italy, [2]Consiglio Nazionale delle Ricerche, Istituto di Scienze e Tecnologie Chimiche "Giulio Natta" (SCITEC-CNR), Milano, Italy, [3]Department of Chemistry, University of Pavia, Pavia, Italy

Multifunctional human collagen lysyl hydroxylase (LH/PLOD) enzymes catalyze post-translational hydroxylation and subsequent glycosylation of collagens, enabling their maturation and supramolecular organization in the extracellular matrix (ECM). Recently, the overexpression of LH/PLODs in the tumor microenvironment results in abnormal accumulation of these collagen post-translational modifications, which has been correlated with increased metastatic progression of a wide variety of solid tumors. These observations make LH/PLODs excellent candidates for prospective treatment of aggressive cancers. The recent years have witnessed significant research efforts to facilitate drug discovery on LH/PLODs, including molecular structure characterizations and development of reliable high-throughput enzymatic assays. Using a combination of biochemistry and *in silico* studies, we characterized the dual role of Fe^{2+} as simultaneous cofactor and inhibitor of lysyl hydroxylase activity and studied the effect of a promiscuous Fe^{2+} chelating agent, 2,2′-bipyridil, broadly considered a lysyl hydroxylase inhibitor. We found that at low concentrations, 2,2′-bipyridil unexpectedly enhances the LH enzymatic activity by reducing the inhibitory effect of excess Fe^{2+}. Together, our results show a fine balance between Fe^{2+}-dependent enzymatic activity and Fe^{2+}-induced self-inhibited states, highlighting exquisite differences between LH/PLODs and related Fe^{2+}, 2-oxoglutarate dioxygenases and suggesting that conventional structure-based approaches may not be suited for successful inhibitor development. These insights address outstanding questions regarding druggability of LH/PLOD lysyl hydroxylase catalytic site and provide a solid ground for upcoming drug discovery and screening campaigns.

KEYWORDS

collagen, lysyl hydroxylase (LH), Fe2+/2-oxoglutarate-dependent dioxygenases, structure-based drug design, molecular dynamics simulations, cancer metastasis, 2-2′-bipyridyl

Introduction

The supramolecular organization of collagen in the extracellular matrix (ECM) depends on various post-translational modifications (PTMs) that occur during its biosynthesis. Among the different PTMs, lysine (Lys) hydroxylation is key for proper collagen fibril formation, thus defining the overall physiochemical properties of ECM (Yamauchi and Sricholpech, 2012). The collagen lysyl hydroxylase (LH/PLOD) enzyme family comprises the three isoforms LH1/PLOD1, LH2/PLOD2 and LH3/PLOD3 (encoded by the procollagen-lysine, 2-oxoglutarate 5-dioxygenase (*PLOD*) genes) and is the sole enzyme capable of hydroxylating collagen Lys in humans (Scietti and Forneris, 2020). These enzymes use Fe^{2+}, 2-oxoglutarate (2-OG), ascorbate and O_2 to catalyze the addition of a hydroxyl group in position 5 of collagen Lys, yielding 5-hydroxylysine (Hyl) with the release of succinate and CO_2 (Figure 1). Unmodified collagen Lys and modified Hyl are both substrates of collagen lysyl oxidases (LOX), which catalyze the oxidative deamination of Lys and Hyl forming highly reactive aldehydes (Lys^{ald} and Hyl^{ald}, respectively) that spontaneously rearrange to form Lys-derived collagen cross-links (LCC) and Hyl-derived collagen cross-links (HLCC) in the ECM. A physiological ratio between LCC and HLCC is essential to establish and maintain a proper ECM functionality. Conversely, excess HLCC in the tumor microenvironment has been linked to biomechanical alterations and increased ECM tension and stiffness (Levental et al., 2009; Pankova et al., 2016). The deposition of ordered thicker collagen fibers, as consequence of the abnormal LCC/HLCC ratio, is a characteristic of severe tissue fibrosis, one of the hallmarks of cancer (Van Der Slot et al., 2004; Chen et al., 2015). Cancer cells take advantage of these "collagen highways" to migrate toward blood vessels that sustain metastatic progression (Provenzano et al., 2006; Du et al., 2017b; Gkretsi and Stylianopoulos, 2018). Over the last years, multiple studies correlated both hypoxia-dependent and independent overexpression and mislocalization of LH enzymes with increased propensity to metastatization in a wide variety of solid tumors, recognizing these enzymes as markers of adverse prognosis (Chen et al., 2015; Chen et al., 2016; Pankova et al., 2016; Sato et al., 2021).

Initially, the hypoxia-driven overexpression of *PLOD2* was the first identified prognostic factor in several tumors, as hepatocellular carcinoma (Noda et al., 2012; Du et al., 2017a), sarcoma (Eisinger-Mathason et al., 2013), lung and colon cancer (Du et al., 2017a), renal carcinoma (Kurozumi et al., 2016), glioma (Song et al., 2017; Xu et al., 2017), oral squamous cell and endometrial carcinoma (Saito et al., 2019; Wan et al., 2020) bone and breast metastasis (Blanco et al., 2012; Gilkes et al., 2013; Du et al., 2017b) and cervical cancer (Li et al., 2021). Interestingly, the downregulation of LH2/PLOD2 isoform in renal cell carcinoma via tumor suppressing miRNA significantly inhibited cell migration and invasion (Kurozumi et al., 2016), confirming the importance of LH in tumor progression. Later on, the LH1/PLOD1 and the LH3/PLOD3 isoforms were also identified as biomarkers in many different types of solid tumors. High *PLOD1* expression levels were found in gastrointestinal carcinoma (Wang et al., 2018), osteosarcoma (Jiang et al., 2020), glioma (Tian et al., 2021; Wang et al., 2021) and bladder cancer (Yamada et al., 2019). The LH3/PLOD3 isoform was identified to be upregulated in glioma (Tsai et al., 2018; Baek et al., 2019), gastric cancer (Wang et al., 2019) and colorectal cancer (Deng et al., 2021; Shi et al., 2021), acting as a promoter of metastatization in different cancer types (Gong et al., 2021). In agreement with the observations on *PLOD2*, also *PLOD3* knockdown suppressed the malignant phenotype in renal cell carcinoma (Xie et al., 2020). Furthermore, the entire LH/PLOD family was found correlated with metastatization of solid tumors as

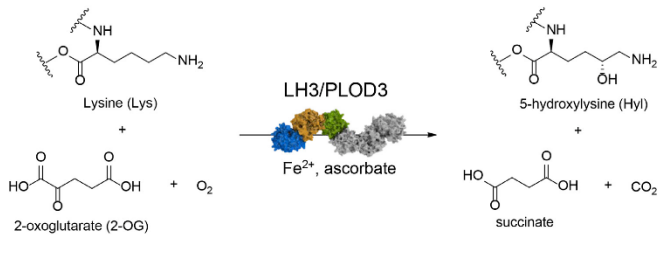


FIGURE 1
Reaction scheme for collagen lysine hydroxylation by LH3/PLOD3. The enzyme requires Fe^{2+} as an essential catalytic cofactor, O_2, and 2-oxoglutarate (2-OG) as co-substrates, and ascorbate to maintain the redox state of Fe^{2+} to enable catalytic turnover. Upon lysine hydroxylation, 2-OG is converted into succinate with release of one molecule of CO_2.

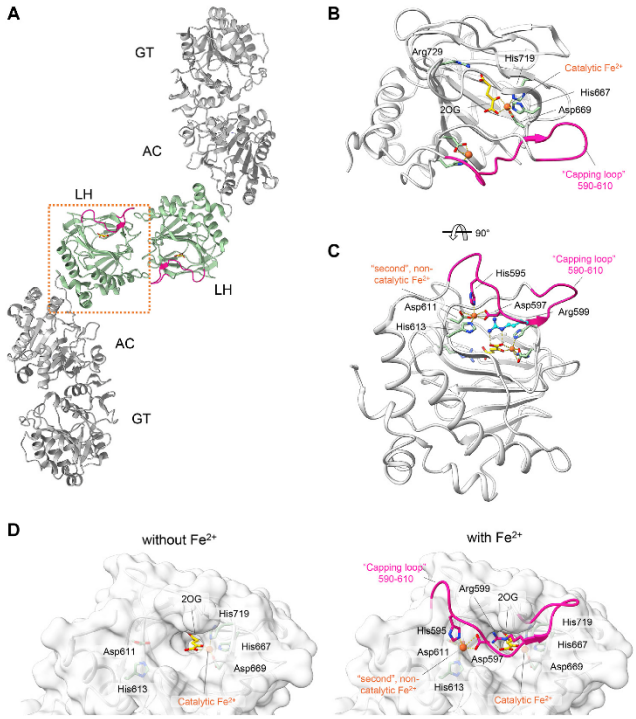


FIGURE 2
Structural features of the LH domain and catalytic site in LH3/PLOD3. **(A)** Cartoon representation of the dimeric assembly of multidomain LH3/PLOD3. The labels indicate the domain organization of LH3/PLOD3, showing the positions of the GT, AC, and LH domains along the quaternary structure. The box highlights the position of the LH domain (green) at the dimer interface. **(B)** Details of the LH catalytic site. Amino acids surrounding the Fe^{2+} cofactor and 2-OG are shown as sticks. The capping loop is shown in magenta, indicating the substrate entrance site. **(C)** Details of the second Fe^{2+} binding site interlocking the capping loop into the self-inhibitory state. Residues interacting with the second Fe^{2+} as shown as sticks. Residue Arg599, mimicking the collagen lysine substrate subject to hydroxylation, is shown in blue. **(D)** Side-by-side surface representations of the accessibility to the LH active site in absence (left) and in presence (right) of the capping loop. Residues involved in catalysis and in maintaining the self-inhibitory state of the capping loop are shown as in panels **(B,C)**.

hepatocellular and renal cell carcinomas (Xu et al., 2019; Yang et al., 2020), gliomas (Zhao et al., 2021), gastric (Li et al., 2020), and ovarian (Guo T. et al., 2021) cancers. Taken together, these observations strongly point these enzymes as a hot topic in cancer research: LH/PLODs are not only widely recognized prognostic markers of cancer metastatization with poor outcome, but also very promising druggable targets for anticancer therapy.

The lack of a structural templates of LH/PLODs have hampered for many years targeted drug discovery campaigns. Only recently, our group and others determined molecular structures suitable as templates for *in silico* drug discovery (Guo et al., 2018; Scietti et al., 2018). In particular, the crystal structure of the full-length LH3/PLOD3, the first of a human collagen lysyl hydroxylase, provided key insights on catalytic pockets. LH3/PLOD3 is a multifunctional enzyme capable of performing collagen lysine hydroxylation (as the other two isoforms) and glycosylation (Scietti and Forneris, 2020; De Giorgi et al., 2021). Indeed, LH3/PLOD3 can additionally catalyze the galactosylation and further glucosylation of Hyl to

form α-(1,2)-glucosyl-β-(1,O)-galactosyl-5-hydroxylysine *in vitro*. The identification of specific genes (*COLGALT1/2*) encoding for collagen galactosyltrasferases (GLT25D1/2) makes the physiological relevance of the galactosyltransferase activity of LH3/PLOD3 under debate. In addition, there are increasing indications that also LH1/PLOD1 and LH2/PLOD2 possess glycosyltransferase activity, although less pronounced (Ewans et al., 2019; Guo H. F. et al., 2021).

LH/PLOD enzymes belong to the Fe^{2+}, 2-OG-dependent dioxygenase superfamily (Martinez and Hausinger, 2015), a widespread class of enzymes that catalyzes oxidative reactions such as epimerization, demethylation, and hydroxylation (Hausinger, 2004; Flashman and Schofield, 2007; Loenarz and Schofield, 2008; Loenarz and Schofield, 2011). Despite the broad range of functions carried out, all Fe^{2+}, 2-OG-dependent dioxygenases display a common double-stranded β-helix folding (DSBH) topology with highly conserved binding sites and catalytic mechanisms (Costas et al., 2004; Clifton et al., 2006). The catalytic domain responsible for Lys hydroxylation is located at the C-terminus of the LH/PLOD structure and is essential for its unique dimeric quaternary structure, a fundamental prerequisite for collagen lysyl hydroxylase activity (Guo et al., 2018; Scietti et al., 2018). Within the enzymatic pocket of human LH3/PLOD3, a catalytic Fe^{2+} is coordinated by His667, Asp669 and His719 (Figures 2A,B). In a LH/PLOD viral homolog, this Fe^{2+} is a fundamental structural element of LH/PLOD enzymes and its chelation from the active site completely disrupt protein folding, dimer formation and catalytic activity (Guo et al., 2018). Pioneering work (Kivirikko and Prockop, 1967; Myllylä et al., 1979; Puistola et al., 1980) highlighted the complexity and binding promiscuity of LH/PLODs towards binding of different metal ions, and the associated impact on the LH enzymatic activity. Structural studies also revealed the presence of a possible second Fe^{2+} bound within the LH domain, shaping a unique site never observed in other Fe^{2+}, 2-OG-dependent dioxygenases. This second Fe^{2-} is coordinated by two Asp and two His residues (in human LH3/PLOD3, His595, Asp597, Asp611 and His613), whose side chain interactions with the metal ion induce a well-defined conformation of a "capping loop", a stretch comprising residues Gly590-Glu610 that closes the entrance of the catalytic site, mimicking the collagen Lys substrate by positioning Arg599 exactly in front of the 2-OG donor substrate (Scietti et al., 2018) (Figures 2C,D). The same region was characterized by pronounced flexibility in absence of Fe^{2+} (Guo et al., 2018; Scietti et al., 2018), but did not allow to unambiguously rule out crystallization-induced stabilization of the unique conformation observed in the presence of a second Fe^{2+} bound.

The simultaneous presence of features common to all Fe^{2+}, 2-OG-dependent dioxygenases and unique, distinguishing elements exclusively present in the LH/PLOD family, makes these enzymes ideal targets for structure-based drug discovery campaigns. In this study, we combined MD simulations with structure-guided mutagenesis of LH3/PLOD3 and used biochemical assays to elucidate the role of the capping loop in the accessibility of the active site. Our work sets the grounds for successful drug discovery campaigns on LH/PLOD enzymes to fight cancer metastasis.

TABLE 1 List of oligonucleotides used to generate LH3/PLOD3 mutants.

Oligonucleotide name	Oligonucleotide sequence
D597A-Fw	CTTCAAGGCTGGCTGGAGGCTAC
D597A-Rv	CCTCATGCCCGGCCGCCTGAC
D611A-Fw	CCATCCACATGAAGCAGGTGGGG
D611A-Rv	CCACGGTGGGCACATTCTCGTAG

Materials and methods

Chemicals

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

Molecular cloning and site-directed mutagenesis

The coding sequence for wild-type human *PLOD3* gene (GenBank accession number BC011674.2) was obtained from Source Bioscience. Oligonucleotides containing in-frame 5′-BamHI and 3′-NotI were designed and used to sub-clone the coding sequence devoid of the N-terminal signal peptide into a pCR8 vector, that was also used as a template for subsequent experiments. Single-point mutations were generated using Phusion Site Directed Mutagenesis (Invitrogen) with the oligonucleotides listed in Table 1. The linear mutagenized plasmids were phosphorylated using T4 polynucleotide kinase (Invitrogen) prior to ligation using T4 DNA ligase (Invitrogen). All plasmids were checked by Sanger sequencing prior to cloning into the pUPE.106.08 expression vector. This expression vector, kindly provided by U-protein Express, BV (U-PE, Netherlands) provides the N-terminal cystatin signal peptide, followed by a N-terminal 6xHis-tag and a recognition site for Tobacco Etch Virus (TEV) protease prior to the in-frame BamHI-NotI restriction cassette, followed by an in-frame stop codon.

Production of recombinant LH3/PLOD3 expression using transiently-transfected HEK293F cells

Recombinant tagged LH3/PLOD3 were produced using suspension cultures of HEK293F (Invitrogen) cells, maintained and transfected according to (Faravelli et al., 2021). Cells were not authenticated and not tested for *mycoplasma* contamination. Briefly, cells were transfected at cell densities of 1 million/ml using 3 µg of polyethyleneimine (PEI; Polysciences) for 1 µg of pUPE.106.08-LH3/PLOD3 plasmid DNA per mL of cells. Cultures were supplemented with 0.6% Primatone RL 4 h after transfection. The cell medium containing secreted LH3/PLOD3 was collected 6 days after transfection by centrifugation at 1,000 × g for 15 min.

Purification of recombinant LH3/PLOD3 enzymes

The LH3/PLOD3-containing medium from HEK293F cell cultures was filtered through a syringe 0.8 µm filter (Sartorius). The pH and ionic strength of the filtered medium were adjusted using a 5X concentrated buffer stock to reach a final concentration of 25 mM 4-(2-hydroxyethyl)−1-piperazineethanesulfonic acid (HEPES)/NaOH, 500 mM NaCl, 30 mM imidazole, pH 8.0. Recombinant LH3/PLOD3 was purified using a combination of affinity and size-exclusion chromatography on Äkta systems (GE Healthcare) according to (Scietti et al., 2018). The filtered supernatant was first loaded onto a 20 ml His-Prep FF column (GE Healthcare) and eluted using 250 mM imidazole. The eluate was then loaded onto a 5 ml HiPrep desalting FF column (GE Healthcare) equilibrated in 25 mM HEPES/NaOH, 500 mM NaCl, pH 8.0. The N-terminal His-tag was cleaved using overnight His-tagged TEV protease digestion at 4°C followed by affinity-based removal of TEV protease and the cleaved His-tag using a 5 ml HisTrap FF (GE Healthcare). Recombinant LH3 was concentrated to 5 mg/ml using 30,000 MWCO Vivaspin Turbo centrifugal filters (Sartorius), then loaded onto a Superdex 200 10/300 GL (preparative scale) or onto a Superdex 200 5/150 GL (analytical scale) columns (GE Healthcare) equilibrated with 25 mM HEPES/NaOH, 200 mM NaCl, pH 8.0. LH3/PLOD3-containing fractions as assessed from SDS-PAGE analysis were pooled, concentrated, and stored at −80°C until further usage.

LH assays using LC-MS and analysis of 2,2′-bipyridil effects on enzymatic activity

Synthetic collagen peptides were purchased from China Peptides. Peptides tested were ARGIKGIRGFS and GIKGIKGIKGIK sequences (Scietti et al., 2018). 5 µM LH3/PLOD3 was incubated with 50 µM $FeCl_2$, 100 µM 2-OG, 500 µM ascorbate, 1 mM peptide substrate and 0–500 µM 2,2′-bipyridine. Reactions were allowed to proceed for 3 h at 37°C. 10 µl of each sample were supplemented with 38 µl of Milli-Q water and acidified by addition of 2 µl of formic acid (FA) to reach a final volume of 50 µl, then analyzed on an UHPLC-HRMS/MS system (AB Sciex, United States). LC unit (ExionLC AD) consists of a column oven thermostated at 40°C, an autosampler cooled at 10°C and a binary gradient pump system. MS instrument consists of a high resolution QTOF mass spectrometer (AB Sciex X500B) equipped with a Turbo V Ion source and a Twin Sprayer ESI (electrospray ionization) probe, controlled by SCIEX OS 2.1 software. Peptides were separated by reverse phase (RP) HPLC on a Hypersil Gold C18 column (150 × 2.1 mm, 3 µm particle size, 175 Å pore size, Thermo Fisher Scientific) using a linear gradient (2–50% solvent B in 15 min) in which solvent A consisted of 0.1% aqueous FA and solvent B of acetonitrile (CAN) containing 0.1% FA. Flow rate was 0.2 ml/min. Mass spectra were generated in positive polarity under constant instrumental conditions: ion spray voltage 4,500 V, declustering potential 100 V, curtain gas 30 psi, ion source gas 1 40 psi, ion source gas 2 45 psi, temperature 350°C, collision energy 10 V. Spectra analyses were performed using SCIEX OS 2.1 software. Statistical evaluations based on pair sample comparisons between uncoupled and coupled assay values using Student's *t*-test in Prism 7 (Graphpad software).

Luminescence-based LH assays

Reaction mixtures (5 µl total volume) were prepared according to (Scietti et al., 2018) by sequentially adding LH3/PLOD3 at 0.2 mg/ml, 0–1 mM peptide substrate or 4 mg/ml gelatin in water (solubilized through heating denaturation at 95°C for 10 min), 500 µM ascorbate, 100 µM 2-OG, variable concentrations of $FeCl_2$ (0–200 µM) and let incubate for 1 h at 37°C. Reactions were stopped by heating samples at 95°C for 2 min prior to transfer into Proxiplate white 384-well plates (Perkin-Elmer), then 5 µl of the Succinate-Glo reagent I (Promega) were added and let incubate 1 h at 25°C, after that 10 µl of the Succinate-Glo reagent II (Promega) were added and let incubate 10 min at 25°C. The plates were then transferred into a GloMax Discovery plate reader (Promega) configured according to manufacturer's instructions for luminescence detection. All experiments were performed in triplicates. Control experiments were performed using identical conditions by selectively removing LH3/PLOD3, 2-OG or peptide substrates. Data were analyzed and plotted using Prism 7 (Graphpad Software). Statistical evaluations based on pair sample comparisons between uncoupled and coupled

assay values using Student's t-test in Prism 7 (Graphpad software).

Differential scanning fluorimetry assays

DSF assays were performed on LH3/PLOD3 wild-type using a Tycho NT.6 instrument (NanoTemper Technologies). LH3/PLOD3 samples at a concentration of 1 mg/ml in a buffer composed of 25 mM HEPES/NaOH, 200 mM NaCl, pH 8. Binding assays were performed by incubating LH3/PLOD3 with variable $FeCl_2$ and 2,2′-bipyridil concentrations. Data were analyzed and plotted using GraphPad Prism 7 (Graphpad Software).

Fe^{2+} binding assays

Recombinant LH3/PLOD3 was subject to labeling using the NHS-RED kit (NanoTemper Technologies) according to manufacturer's instructions. Labeled LH3/PLOD3 at a concentration of 50 nM was incubated in a buffer composed of 25 mM TRIS/HCl, 100 mM NaCl, pH 7.5 with variable concentrations of $FeCl_2$ for 40 min. The samples were then transferred into Dianthus 384-well plates for Temperature-Related Intensity Change (TRIC) (NanoTemper Technologies), and centrifuged at 1,000 g for 2 min. TRIC measurements were performed immediately after centrifugation using a Dianthus NT.23 instrument (NanoTemper Technologies). The samples were first measured for 1 s without heating and for 5 s with the IR-laser turned on. Normalized fluorescence values (F_{norm}), described as ratios between fluorescence values after and prior to infrared laser activation were collected and plotted as a function of ligand concentration (Schulte et al., 2021). Determination of binding affinities was carried out using the DI.Screening Analysis software (NanoTemper Technologies). Data were then exported and plotted using GraphPad Prism 7 (Graphpad Software).

Molecular dynamics simulations

Simulations were started from the extrapolation of the dimeric LH domain of human LH3/PLOD3 from its experimental crystal structure (PDB 6FXR) (Scietti et al., 2018) using *COOT* (Emsley et al., 2010); two systems were prepared: in one system both Fe^{2+} cations were left ($LH3_{Fe2}$), while in the other one only the catalytic Fe^{2+} was left in ($LH3_{Fe1}$). Residues were modeled in their standard protonation states at physiological pH, as predicted by *PROPKA*, version 3.1 (Sondergaard et al., 2011): this resulted in one disulfide bridge (between Cys 563 and 698), histidines 546, 586, 643, 681, 717 being protonated on Nε2, and histidines 595, 613, 667, 711, 719 protonated on Nδ1 (comprising histidines in both Fe^{2+} binding sites). Crystallographic waters were taken from the published crystal structure of LH3/PLOD3 (PDB: 6FXR) (Scietti et al., 2018). Hydrogen atoms were introduced using the *tleap* utility in *AmberTools* (version 19) (Case et al., 2005), as well as $-NH_3^+$ and $-COO^-$ caps at the *N*- and *C*-termini, respectively. All molecular dynamics simulations (MD) were carried out with the *AMBER* software package (version 18) (Case et al., 2005; Case et al., 2018), using its GPU-accelerated (Salomon-Ferrer et al., 2013) *pmemd.cuda* utility during equilibration and production, and *sander* otherwise; three independent MD replicas (different random seeds) were carried out for $LH3_{Fe1}$ and $LH3_{Fe2}$ alike. A 8.0 Å cutoff was applied for the calculation of Lennard-Jones and Coulomb interactions between nonbonded atoms; beyond this limit, only Coulomb interactions were computed, using the particle mesh Ewald approach (Darden et al., 1993). Each replica's production stage was 1 μs in length, and conducted with a 2 fs time-step in the *NpT* ensemble (with a temperature of 300 K enforced *via* Langevin's thermostat (Loncharich et al., 1992); collision frequency 1 ps ^{1}, and a 1 atm pressure enforced by Berendsen's barostat) (Berendsen et al., 1984). Preproduction stages for each replica consisted in minimization (10 steps of steepest descent + 290 steps of conjugate gradient); heating (20 ps; *NpT*; 25–300 K; increasingly softer harmonic restraints on Cα atoms; k = 5.0 kcal mol^{-1} Å, collision frequency 0.75 ps^{-1}, with 2 fs time-step); and equilibration (1.0 ns; *NpT*; 300 K collision frequency 1 ps ^{1}, with 2 fs time-step). Analyses of MD trajectories were carried out with the *CPPTRAJ* program distributed within the *AmberTools* suite (version 19) (Case et al., 2005) or with code written in-house.

Distance fluctuation

To characterize the impact of the second Fe^{2+} on the internal dynamics of LH domain of LH3, we made use of the previously introduced distance fluctuation (DF) analysis (Morra et al., 2012; Moroni et al., 2018). For each MD trajectory of the two systems, we computed on the combined meta trajectory the matrix of distance fluctuations, in which each element of the matrix corresponds to the DF parameter. DF is defined, for a couple of amino acids i and j, as the variance of the time-dependent distance d_{ij} of the C α atoms:

$$DF_{ij} = \langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle$$

where the brackets indicate the time-average over the trajectory. This parameter is invariant under translations and rotations of the molecules and, unlike the covariance matrix, does not depend on the choice of a particular protein reference structure. The resulting DF matrix can be used to assess the intrinsic flexibility of proteins. This parameter characterizes residues that move in a

highly coordinated fashion, and it is actually able to reflect the presence of specific coordination patterns and quasi-rigid domains motion in the protein of interest. In particular, pairs of amino acids belonging to the same quasi-rigid domain are associated with small distance fluctuations and vice versa.

Forcefield and parametrization of cofactor 2-OG and Fe^{2+} binding sites

Lennard-Jones and intramolecular bonded parameters for the 2-oxoglutaric acid cofactor (2-OG) in the catalytic site of LH3/PLOD3 were assigned according to the *generalized Amber forcefield* (GAFF) (Wang et al., 2004), using *AmberTools' antechamber* and *parmchk2* utilities (Case et al., 2005), after adding methyl hydrogens using the *reduce* tool (Case et al., 2005). Assignment of charges and (intermolecular) 2-OG-Fe^{2+} bonded parameters were performed as discussed below. Parametrization of both Fe^{2+} binding sites was carried out using the *MCPB.py* utility (Li and Merz, 2016) in conjunction with density functional theory calculations (DFT) using the *Gaussian09* suite (Frisch et al., 2009). Intermolecular bonded parameters for all residues in both Fe^{2+}-binding sites and for the 2-OG cofactor were derived by *MCPB.py* (Li and Merz, 2016) applying the Seminario method (Seminario, 1996) on a DFT-derived Hessian matrix. More specifically, this Hessian matrix was calculated at the B3LYP (Lee et al., 1988; Becke, 1993)/6-31G(*d*) level of theory after optimizing a "small" model of both binding sites at the same level to a confirmed minimum (no imaginary frequencies). Such "small" models (generated by *MCPB.py* (Li and Merz, 2016)) included: 1) the Fe^{2+} cation in that particular binding site; 2) binding site residue sidechains up to their Cβ, with a shorter Cβ-H bond replacing Cα-Cβ; and 3) in the catalytic site, the entire 2-OG cofactor. Residues' Cβ atoms and 2-OG's carboxylate oxygens farthest from Fe^{2+} were frozen during optimization and excluded from frequency calculations. Atomic point charges on both Fe^{2+} cations on individual binding site residues and on the entire 2-OG cofactor were fitted by *MCPB.py* (Li and Merz, 2016) using the *RESP* method (Bayly et al., 1993) based on the outcome of (single-point) *ESP* charge fitting calculations (Besler et al., 1990) at a higher DFT level, on a "large" version of each Fe^{2+} binding site. More specifically, the chosen level of DFT for *ESP* charge fitting (Besler et al., 1990) was B3LYP (Lee et al., 1988; Becke, 1993)/6-31G(*d*)/def2-SV(P), (Weigend and Ahlrichs, 2005) with the def2-SV(P) basis set specifically applied to Fe^{2+}. "Large" binding site models (generated by *MCPB.py* (Li and Merz, 2016)) comprised Fe^{2+}, 2-OG, binding site residues in their entirety (i.e., with their backbone) as well as backbones of Glu596 and His668, contiguous to His595/Asp597 and His667/Asp669, respectively. All contiguous backbone fragments in the large models are capped by acetyl and *N*-methyl moieties at their *N*- and *C*-termini, respectively. *Gaussian09* (Frisch et al., 2009) was programmed to perform *ESP* charge fitting (Besler et al., 1990) over 10 spherical shells around each atom, with 17 grid points per square Bohr. As *per MCPB.py*'s default (Li and Merz, 2016), the atomic radius of both Fe^{2+} cations was taken to be 1.409 Å. In all DFT calculations, Fe^{2+} centers in both the catalytic and noncatalytic binding sites were modeled in their quintet state, after comparative optimizations of each site with Fe^{3+} in the triplet and singlet state confirmed—at the B3LYP (Lee et al., 1988; Becke, 1993)/6-31G(*d*) level of theory—that the quintet state is the most energetically stable in both cases (data not published). All remaining LH3/PLOD3 residues—including intra-residue bonded parameters for residues in both Fe^{2+}-binding sites—were treated with the *ff14SB* forcefield (Maier et al., 2015), whereas Na^+ countercations were modeled with parameters by Joung and Cheatham (Joung and Cheatham, 2008): these are compatible with the TIP3P model in use for water (Jorgensen et al., 1983).

TABLE 2 List of 2-OG analogs tested. For each compound, the chemical formula of the associated free acid is reported. None of the compounds yielded detectable binding to LH3/PLOD3 or inhibition of the conversion of 2-OG into succinate in presence or absence of acceptor substrate.

Compound	Formula
Formate	CH_2O_2
Oxalate	$C_2H_2O_4$
Malonate	$C_3H_4O_4$
Tartronate	$C_3H_4O_5$
Mesoxalate	$C_3H_2O_5$
Aminomalonate	$C_3H_5NO_4$
Fumarate	$C_4H_4O_4$
Oxalacetate	$C_4H_4O_5$
Malate	$C_4H_6O_5$
Aspartate	$C_4H_7NO_4$
Tartrate	$C_4H_6O_6$
Glutamate	$C_5H_9NO_4$
Glutarate	$C_5H_8O_4$
Acetonedicarboxylate	$C_5H_6O_5$
2-hydroxyglutarate	$C_5H_8O_5$
Adipate	$C_6H_{10}O_4$

Results

The LH catalytic site does not accommodate competitive inhibitors

The presence of amino acid networks shared by Fe^{2+}, 2-OG-dependent dioxygenase enzymes in their catalytic sites provides a general template for structure-based design of potential inhibitors. The analysis of the residues surrounding the

catalytic Fe^{2+} and the 2-OG indeed supported the possibility that 2-OG analogs may act as competitive inhibitors of the co-substrate molecule (Rose et al., 2011). In human LH3/PLOD3, the catalytic Fe^{2+} is strongly coordinated by the side chains of His667, Asp669, and His719 of the DSBH fold, and by the 2-OG co-substrate, capped towards the outer solvent by a flexible capping loop defined by residues 590-610 (Figure 2). Using a combination of nano-differential scanning fluorimetry (nanoDSF) and luminescence-based activity assays, we screened a small, focused library of 2-OG analogs (Table 2) searching for compounds capable of inhibiting LH activity. After thorough testing, none of the compounds tested showed binding/folding stabilization in nanoDSF, nor inhibition of enzymatic conversion of 2-OG into succinate. Likewise, a custom-designed library of compounds selected through *in silico* virtual screening of specific candidate binders of the LH3/PLOD3 catalytic site did not provide suitable hints for inhibitors of LH activity using these assays.

A second Fe^{2+} binding site on the capping loop modulates accessibility to the LH catalytic site

Intrigued by the recalcitrance to inhibition of the LH catalytic site by 2-OG analogs, we focused our attention to the distinguishing features displayed by this domain when compared to homologous Fe^{2+}, 2-OG-dependent dioxygenases, and in particular to the capping loop and the stable conformation adopted in the presence of excess [Fe^{2+}]. This interlocked state may indeed constitute an obstacle when dealing with inhibition of the LH catalytic site, and the relative positioning of the capping loop is crucial for inhibitor accessibility to the active pocket. We decided to perform a thorough investigation of the flexibility of this loop *in silico* and with site-directed mutagenesis *in vitro* to validate the possible significance of the second Fe^{2+} binding site prior to attempting to quantitatively probe the specific metal ion binding to the two distinct LH3/PLOD sites within the LH domain.

To investigate the impact of the second Fe^{2+} on the structural conformation of the capping loop 590-610, the C-terminal LH domain involved in LH/PLOD dimerization underwent all-atom MD simulation in explicit solvent, with and without the second Fe^{2+} ion. In the following, we refer to the simulated system with the second Fe^{2+} as $LH3_{Fe2}$, while the system simulated without it has been named $LH3_{Fe1}$. Three independent replicate simulations were carried out for the two systems, each 1 μs long. In each independent replicate we used identical simulation parameters (see Material and Methods), varying only the initial velocities of atoms via random assignments from a Maxwell distribution consistent with the required temperature. Visual inspection of MD simulations shows that both systems are characterized by minimal atom fluctuations, suggesting that this fragment of the enzyme is stiff and allows for minimal protein motions away from the starting (crystal) structure, except for the gate loop and some protein region at the dimer interface.

The first step in the analysis of the dynamics of LH3/PLOD3 C-terminus and its potential variation as a function of the presence/absence of the second Fe^{2+}, was the identification of protein regions displaying higher levels of local flexibility during dynamics, through the computation of the local fluctuations (LF). This parameter detects the flexibility of a given residue with respect to the neighboring amino acids; the comparison of this calculation for the two systems (i.e., $LH3_{Fe1}$ and $LH3_{Fe2}$), allows to extrapolate variations in protein regions more affected by the presence/absence of the second Fe^{2+}. LF is calculated for residue *i*, as the mean of the variances of the distance d_{ij} of the C_α (i) and C_α (j) atoms, for j = i-2, i-2, i+1, i+2. In the $LH3_{Fe1}$ system residues 593-611 of the gate loop of both monomers displayed higher flexibility as compared to the $LH3_{Fe2}$ system (Figures 3A,B). The peak at residues 640-646, constituted by residues of one monomer which contact the gate loop of the other monomer, is more defined in the $LH3_{Fe1}$ system (Figures 3A,B). To quantify the observed differences in the dynamics of the enzyme during MD in the two simulated systems, we calculate the root-mean-square deviation (RMSD) distributions of MD trajectories, after optimal rigid body superposition of backbone atoms of each snapshot with the corresponding atoms of the crystal structure, where the gate loop is in the closed conformation.

The RMSD distributions of the loop 590-610 in the two monomers of the single trajectories are reported in Figures 3C,D, showing the different behavior of the two systems and the variations in the single trajectories. A threshold of 3 Å was used to distinguish the closed from the open conformation, based on visual inspection of MD simulations and superposition of snapshots representing the two states. These plots show that the gate loop is stabilized in the closed conformation in the $LH3_{Fe2}$, while the absence of second Fe^{2+} impacts the stability of the loop, shifting the population of $LH3_{Fe1}$ towards the open conformation, even though both systems can visit the two states. To understand the impact of the second Fe^{2+} on the internal dynamics of LH3/PLOD3, we performed the distance fluctuation analysis on the meta trajectories of the two systems, obtained concatenating the three MD replicas of each system. The DF parameter, which is the variance of the inter-residue distance d_{ij}, was calculated for any pair of residues during the trajectory (see Material and Methods). The resulting DF matrices calculated for $LH3_{Fe1}$ and $LH3_{Fe2}$ shown in Figures 3E,F, report on the fluctuation of the inter-residue distance in the corresponding residue pairs and matrix regions, describing the intrinsic flexibility and coordination of the LH domains. Relatively low DF values identify protein regions that move together, in coordination. The comparison of these matrices can be used to evaluate possible changes of the internal

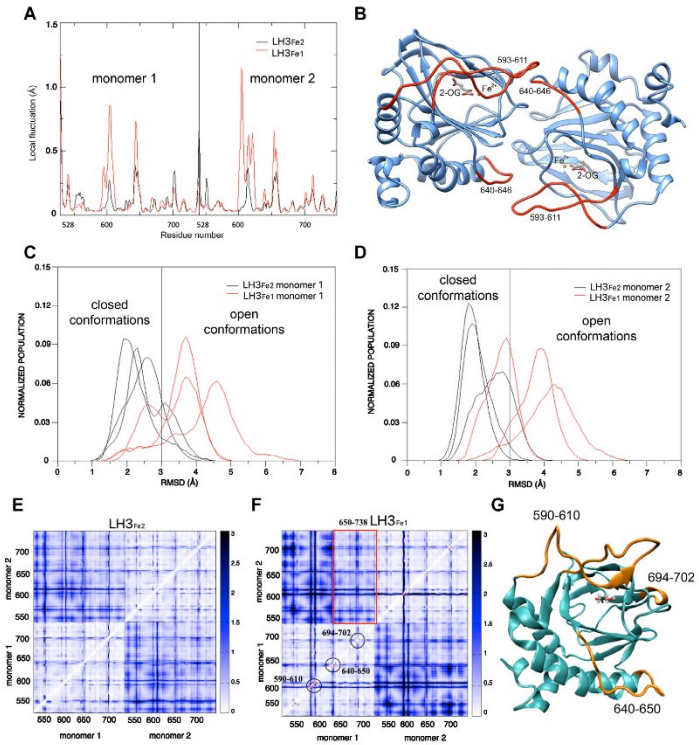


FIGURE 3
Computational analysis of the dynamic behavior of LH capping loop in modulating accessibility to the LH catalytic site. **(A)** Mean values of the local fluctuations for each residue of the dimerization LH domains of LH3, calculated over the three replicate simulations for $LH3_{Fe1}$ (red profile) and $LH3_{Fe2}$ (black profile) systems. **(B)** Structure of the dimerization LH domains of LH3: residues corresponding to the highest peaks in the local fluctuations profile of $LH3_{Fe1}$ (593-611 of the gate loop and residues 640–646) are colored in red. **(C,D)** RMSD distributions (loop 590–610) of the $LH3_{Fe2}$ (black lines) and $LH3_{Fe1}$ systems, for both LH domains. **(E,F)** Distance Fluctuations analysis. Color-coded DF matrices resulting from the DF analysis for the $LH3_{Fe2}$ **(E)** and $LH3_{Fe1}$ **(F)** systems. Pairs of amino acids that move in a coordinated way, that is with a quasi-rigid motion, are characterized by small DF values (white dots) and vice versa (blue dots). Red dots in the $LH3_{Fe2}$ matrix correspond to residues which show significant difference in DF values compared to the $LH3_{Fe2}$ matrix, according to the F-test. Red rectangle encloses residues 650-738 of one monomer, corresponding to the beta-sheet core of the LH domain, that move in a coordinated way with the other monomer. **(G)** Structure of the LH domain of LH3: residues which show significant differences in protein motion for the two systems are colored in orange.

dynamics and coordination due to the presence/absence of the second Fe^{2+}.

Overall, the matrices for both systems turned out to be similar, with largely overlapping patterns of small and large fluctuations of inter-residue distances. In both systems, the internal dynamics of the two domains is characterized by small atomic fluctuations, confirming that protein structure is stiff. Indeed, the highest DF parameters in both systems correspond to protein regions that do not adopt well-ordered 3D structures, which are intrinsically more flexible. Generally, the inter-domain motion is characterized by higher fluctuations than the internal domain motion. In particular, in the $LH3_{Fe1}$ system, residues 650-738 of one monomer, corresponding to the beta-sheet core, move in a more

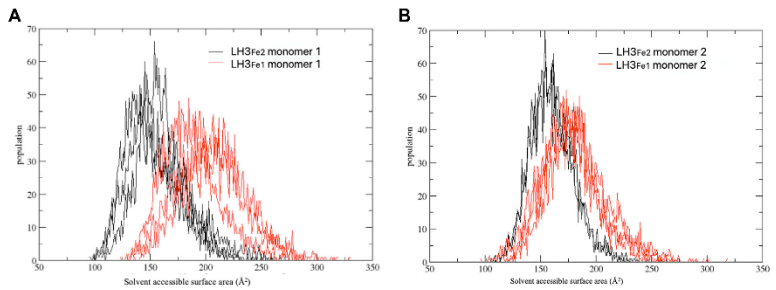


FIGURE 4
Distribution of the solvent accessible surface area calculated for the three replicate simulations. The simulations have been carried out separately for each computed system ($LH3_{Fe1}$ in red, $LH3_{Fe2}$ in black) and monomer **(A,B)**.

coordinated way with the other monomer than the rest of the protein (Figure 3F).

The significance of the differences observed between the two DF matrices was evaluated with a statistical analysis based on F-test. We used $LH3_{Fe2}$ as a reference state for comparing the two matrices. Red dots in the $LH3_{Fe1}$ system (Figure 3F) correspond to residues which show significant difference in DF values in respect to the corresponding DF values in $LH3_{Fe2}$, according to this test. This analysis highlights that relevant differences in protein motion for the two systems concern residues of the gate loop 590-610 and residues 640-650 forming the loop that connects the alpha helix 618-639 with the beta-sheet core of the protein. This loop partially forms the dimerization interface of the two domains, as well as residues 694-702, which displays higher DF values in the $LH3_{Fe1}$ system (Figures 3F,G).

To further investigate the role of the relative positioning of the capping loop for the accessibility to the active pocket depending on the second Fe^{2+} ion, we computed the part of the van der Waals surface of residues forming the catalytic site (protein residues within 6 Å of 2-OG and Fe^{2+}) that are accessible to solvent, that is the solvent accessible surface area, across each frame of MD trajectories, using the *VMD* program (Humphrey et al., 1996). This analysis gives a measure of the capability of the loop in regulating the access of the substrate to the catalytic site, depending on the presence of the second Fe2+. Figure 4 shows the distribution of the solvent accessible surface area calculated for the three replicas, for each system and monomer. In both monomers the distributions of the $LH3_{Fe1}$ system are shifted towards slightly higher values compared to the $LH3_{Fe2}$, suggesting that the presence of the Fe^{2+} has an impact on the access to the active site by modulating the opening/closure of the loop, even though both systems can visit the two distinct states, in agreement with the RMSD analysis discussed above.

To inspect the effect of the second Fe^{2+} on the dynamics of the capping loop at a finer level, we evaluated the persistence over the simulation time of the native and new interactions established by this loop with the rest of the protein. The results of this analysis show that most of the native contacts are maintained during the simulation time in both systems (Supplementary Figure S1A,B), while in $LH3_{Fe1}$ some residues of the loop tend to weaken some of these interactions compared to $LH3_{Fe2}$. As for new stable contacts, we observed that in $LH3_{Fe1}$ the loop can't form new interactions with the rest of the protein due to its higher mobility, while in the $LH3_{Fe2}$ the loop establishes new steady interactions with residues 643-648 of the other monomer (Supplementary Figure S1C).

As mentioned above, in the crystal structure of the LH domain obtained in the presence of excess [Fe^{2+}], residue Arg599 belonging to the loop 590-610 forms a salt bridge with the 2-OG co-substrate, yielding a conformation that mimics the collagen lysine substrate (Scietti et al., 2018) (Figures 2C,D). To check the impact of the presence/absence of the second Fe^{2+} on the stabilization of this bond, we tracked the hydrogen bonds between Arg599 and 2-OG co-substrate over the course of the trajectories. This analysis showed that in most of the conformations visited during the dynamics, Arg599 forms a salt-bridge with 2-OG in both systems, with a in slight prevalence in $LH3_{Fe2}$ (Supplementary Figure S1D–F). Taken together, the results from the *in silico* analysis corroborate the observation that the second Fe^{2+} constrains the LH domain in a tightly interlocked conformation with reduced conformational motility in proximity to the enzyme's catalytic site.

To better examine the impact of the second Fe^{2+}-binding site on enzymatic activity and substrate accessibility, we generated the Asp597Ala and Asp611Ala variants of human LH3/PLOD3 using site-directed mutagenesis. In particular, we

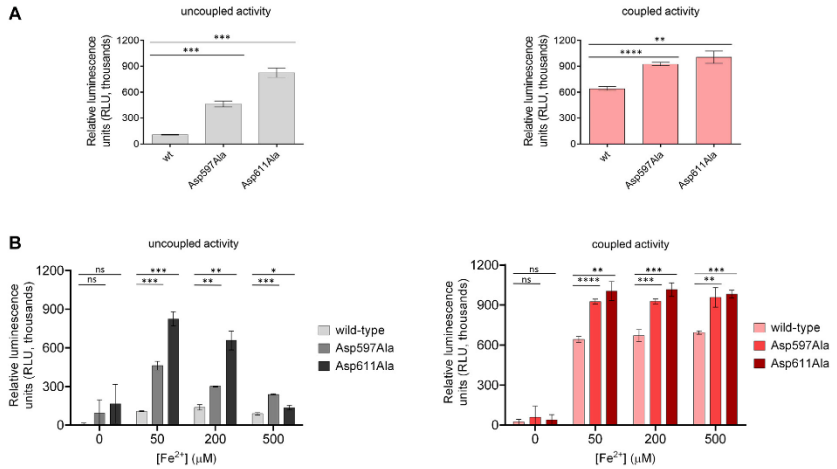


FIGURE 5
Comparison of enzymatic activity of LH3/PLOD3 wild-type and mutants affecting the second Fe^{2+} binding site. **(A)** Luminescence-based uncoupled (i.e., with *no* acceptor substrate, left) and coupled (i.e., using gelatin as acceptor substrate, right) enzymatic activity assays detecting succinate formation. Error bars represent standard deviations from average of triplicate independent experiments. Statistical evaluations based on pair sample comparisons between luminescence values for wild-type LH3 and mutants using Student's *t*-test. ns, non-significant; *, *p*-value<0.05; **, *p*-value<0.01; ***, *p*-value<0.001; ****, *p*-value < 0.0001. **(B)** Evaluation of the uncoupled (left) and coupled (right) enzymatic activities [as in panel **(A)**] in presence of increasing concentrations of Fe^{2+}. Error bars and statistical analyses as in panel **(A)**.

focused on Asp597 as a critical capping loop residue involved in the coordination of the second Fe^{2+}, whereas Asp611 could represent its counterpart as central to the Fe^{2+} coordination platform on the surface of the LH domain (Figures 2C,D). Both mutants could be expressed and purified to homogeneity, and showed yields, folding stability and oligomerization states comparable to wild-type LH3/PLOD3 (Supplementary Figure S2). We probed their Fe^{2+} binding affinity using Temperature-Related Intensity Change (TRIC) and found that both mutants showed 4-times lower affinity compared to wild-type LH3/PLOD3 (Supplementary Figure S3). We were not surprised of the weak binding observed: the catalytic Fe^{2+} is known to have structural roles within the LH domain and cannot be removed to probe its binding (Guo et al., 2018; Scietti et al., 2018), therefore we assumed that the binding data collected exclusively refer to possible additional Fe^{2+} binding sites. When tested for their enzymatic activity in the presence of acceptor substrates such as gelatin (i.e., coupled activity), the mutants showed slightly higher (i.e., 1.5X) activity than their wild-type counterpart (Figure 5A, right). Nevertheless, we consistently observed a much higher (i.e., 3-5X) degree of 2-OG conversion into succinate without the need of an acceptor substrate (i.e., uncoupled activity, Figure 5A, left). Prompted by this observation, we decided to investigate the modulatory effect of $[Fe^{2+}]$ on both uncoupled and coupled enzymatic activities of these mutants. We found that the uncoupled activities of both LH3/PLOD3 Asp597Ala and Asp611Ala mutants were less inhibited by high $[Fe^{2+}]$ compared to wild-type LH3/PLOD3 (Figure 5B, left), whereas the modulation of the coupled enzymatic activities was almost unaffected by the presence of the point mutations (Figure 5B, right). Based on the combined results obtained from the MD simulations and mutagenesis, we reasoned that binding of a second metal ion on the surface of the LH domain induces a conformationally stable, self-inhibited state which in turn reduces the ability of LH/PLOD enzymes to process 2-OG into succinate via uncoupled activity.

Fe^{2+} chelating agents produce unexpected effects on LH3/PLOD3 enzymatic activity

Given the unique presence of two distinct Fe^{2+} binding sites in close proximity, but with opposite effects on enzymatic

10.3389/fmolb.2022.876352

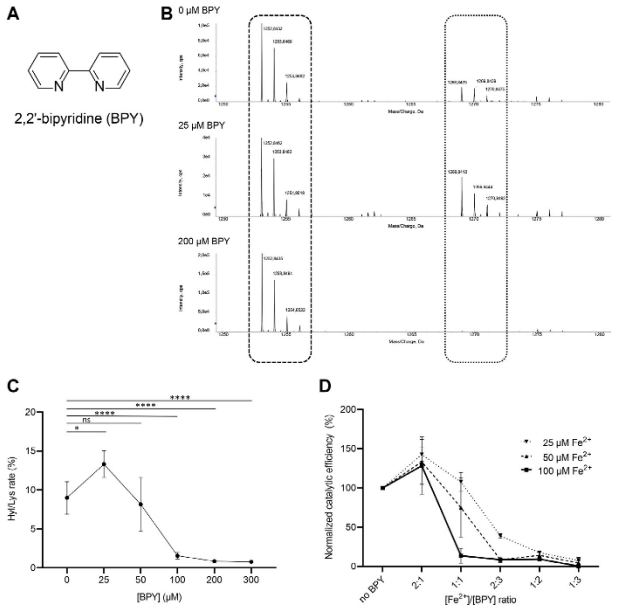


FIGURE 6
Evaluation of the effect of 2,2'-bipyridil (BPY) on LH3/PLOD3 enzymatic activity. **(A)** Chemical structure of BPY. **(B)** Detection of Hyl formation using direct mass spectrometry (MS) assays, and evaluation of the effect of BPY on Lys-to-Hyl conversion. The scheme shows the comparison between three MS spectra, showing the peaks consistently detected for unmodified Lys in the synthetic peptides (dashed box) and the presence of additional peaks (dotted box) indicating Hyl formation. **(C)** Results of MS analysis of LH3/PLOD3 enzymatic activity as a function of BPY concentration. The plot shows the presence of an unexpected increase of enzymatic activity at low BPY concentrations, followed by a drop consistent with inhibition due to sequestration of catalytic Fe^{2+}. Error bars represent standard deviations from average of triplicate independent experiments. Statistical evaluations based on pair sample comparisons between data points collected in absence or in presence of $[Fe^{2+}]$ using Student's *t*-test. ns, non-significant; *, *p*-value<0.05; **, *p*-value<0.01; ***, *p*-value<0.001; ****, *p*-value < 0.0001. **(D)** Evaluation of LH3/PLOD3 enzymatic activity as a function of varying [BPY]/$[Fe^{2+}]$ ratios. The analysis shows that the enzymatic activity increase, already observed at low BPY concentration, depends on the residual Fe^{2+} concentration available for catalysis. Error bars represent standard deviations from average of triplicate independent experiments. Statistical analyses conducted by comparing data points collected at the same $[Fe^{2+}]$/[BPY] ratio using one-way ANOVA analysis highlight non-significant differences for experiments performed using different concentrations of Fe^{2+} in the assay. Additional statistical evaluations for each data point are provided in Supplementary Figure S3.

activity, we wondered whether small-molecule inhibitors acting through metal ion chelation could efficiently modulate substrate processing in LH/PLOD enzymes. To date, despite large scale screenings identified potential hits (Devkota et al., 2019), specific inhibitors of the lysyl hydroxylase activity are missing. The only known inhibitor of LH/PLOD enzymes is the 2,2'-bipyridine (BPY), a non-specific inhibitor that act as chelating agent. BPY chemical structure is characterized by two pyridyl rings, heterocyclic chemical moieties containing nitrogens (Figure 6A). Although the mechanism of action of BPY on LH/PLOD enzymes is not well characterized, being a metal chelator, it likely acts on the Fe^{2+} ion present in the lysyl hydroxylase domain (Ikeda et al., 1994; Rose et al., 2011; Vasta and Raines, 2015; Jover et al., 2018). We therefore decided to explore the actual impact of BPY inhibition on human LH3/PLOD3 enzymatic activity. Firstly, we attempted to evaluate concentration-dependent inhibition using luminescence-based assays, but we realized that usage of BPY was not compatible with the assay setup, likely due to inhibition of the Mg^{2+}-dependent luciferase reaction by the reagent. We therefore focused on a strategy to directly investigate hydroxylysine formation on synthetic collagen peptides using

mass spectrometry (MS) by evaluating the relative ratios between non-hydroxylated and hydroxylated lysine side chains (Figure 6B). Using this method, we could perform accurate quantitation of the dose-dependent effects of BPY on LH3/PLOD3 enzymatic activity.

In absence of BPY, collagen peptide substrate processing was consistent with previously reported data (Scietti et al., 2018). Given a BPY:Fe^{2+} chelation stoichiometry of 3:1, and a half-maximal BPY concentration required to form complexes with 20 μM Fe^{2+}(Fe_{20}–FC_{50}) experimentally determined in about 40 μM (Vasta and Raines, 2015), we would have expected an initial reduction of LH activity at a [Fe^{2+}]/[BPY] ratio of 2:1. Surprisingly, assays performed with concentrations of BPY up to 25 μM unexpectedly revealed increasing Hyl/Lys ratios, suggesting enhancement of enzymatic activity, whereas only higher BPY concentrations caused the expected dramatic decay in substrate processing (Figures 6B,C). Considering that standard assays are carried out by supplementing a fixed concentration of Fe^{2+} (i.e., 50 μM), we wondered whether the activity boost observed at relatively low BPY concentrations could be associated to Fe^{2+} sequestration from the non-catalytic site on capping loop, whereas inhibition could be caused by chelation of the catalytic Fe^{2+}.

We therefore carried out additional assays, in which we simultaneously varied [Fe^{2+}] and [BPY]. Our results (Figure 6D, Supplementary Figure S4) showed that the LH3/PLOD3 catalytic activity always reaches its maximum at [Fe^{2+}]/[BPY] ratio of 2:1, supporting a fine balance between sequestration of excess inhibitory metal ions bound to the capping loop and inhibition through chelation of Fe^{2+} in the catalytic site. These results suggest that usage of BPY as a lead compound for the development of inhibitors of LH/PLOD enzymatic activity demands particular care, as unexpected concentration-dependent opposite effects may be produced through release of the self-inhibitory, Fe^{2+}-dependent capping loop.

Discussion

Human LH/PLOD enzymes are becoming a hot topic in cancer research, due to their involvement in fibrotic conversion of collagens in the tumor microenvironment and the association to higher risk of metastasis. As cancer metastatization has been clearly correlated to excess Lys-Hyl enzymatic conversion, development of highly specific LH/PLOD inhibitors is desirable. With the release of the molecular structure of full-length human LH3/PLOD3, the challenge of securing detailed atomic structures for at least one human LH/PLOD isoform has been overcome (Scietti et al., 2018), however as of yet no LH/PLOD inhibitors are available. In this work, we carried out a small molecule screening aiming at finding hits to be used for the development of LH/PLOD inhibitors. To achieve this goal, we used structural and mechanistic insights from related Fe^{2+}, 2-OG-dependent dioxygenases, and found that the challenge of developing specific LH/PLOD inhibitors may present additional obstacles for which extra care is needed.

Despite the high resemblance of the LH3/PLOD3 catalytic pocket with structurally-related Fe^{2+}, 2-OG dioxygenases that could be inhibited by 2-OG analogs (Rose et al., 2011), our initial campaign focused on development of small-molecule inhibitors based on their ability to compete with 2-OG in the LH catalytic site were not successful. Likewise, dedicated libraries of compounds developed *in silico* based on high-resolution structural templates of the enzyme cavity did not provide any useful leads towards LH/PLOD inhibition. Using MD simulations, we could interpret the systematic recalcitrance to inhibition with limited accessibility to the catalytic site, caused by a very stable, self-inhibited state generated by specific conformations adopted by the capping loop in the presence of excess [Fe^{2+}]. Previous structural studies demonstrated that the capping loop folding was strictly dependent on the iron coordination by four residues, three of which laying on the surface of the LH domain (i.e., His595, Asp611, His613), and one (Asp597) being part of the capping loop. By mutagenizing either Asp611 or Asp597 to alanine, we could observe a decrease in the binding affinity for non-catalytic Fe^{2+} and a much-increased uncoupled conversion of 2-OG into succinate which can be explained with improved accessibility to the catalytic site, supporting the physiological role of the Fe^{2+}-dependent self-inhibited conformation adopted by the capping loop.

Taken together, these results highlight the extremely delicate balance of Fe^{2+} concentration needed in LH/PLOD enzymes to enable their function: while too little [Fe^{2+}] hampers catalysis due to lack of an essential component in the catalytic site, even a little excess can instead interlock the LH domain into a self-inhibited state. This behavior is unique for LH/PLODs and different from related Fe^{2+}, 2-OG dioxygenases, and is likely responsible for the differential responses observed during treatment with Fe^{2+} chelators. Indeed, when testing BPY as candidate inhibitor of LH activity, we found that at low concentrations this compound is capable of enhancing the enzymatic activity rather than blocking it. We interpreted this boost in enzymatic activity with sequestration of the Fe^{2+} bound on the LH domain surface trapping the capping loop, and we could demonstrate that such effect depends on the ratio between [Fe^{2+}] and [BPY].

Collectively, the results obtained provide interesting perspectives regarding the mechanisms of substrate processing by LH/PLOD enzymes, as well as guidance for future inhibitor design. The strong stability of the Fe^{2+}-induced conformation adopted by the capping loop, together with the self-inhibited state obtained through substrate mimicry by LH3/PLOD3 Arg599 implies that physiological substrate processing may depend on the release of the second Fe^{2+} interlock through long-range interactions on the surface of the LH domain. Given the extended conformation of collagen polypeptide substrates, this is a likely option and demands further investigation, in

particular considering the possible roles that metal ions bound to collagen substrates may have upon/during post-translational modification of lysines, but also after the processed collagen has detached from the LH/PLOD enzymes. As for inhibitor design, the presence of this metal ion-dependent self-inhibited state represents an additional challenge, which may require the development of synergistic strategies acting simultaneously on the release of the capping loop and on the competition with either 2-OG in the catalytic site, or the Lys side chain subjected to hydroxylation.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

LS and FF carried out preliminary structural analyses, designed and supervised the research work. LS, DM, and AC produced wild-type LH3, with support from SF. DM, MDM, LN, and FDG performed enzymatic activity assays and analyzed results. DM performed TRIC studies to assess Fe^{2+} binding. AC carried out mutagenesis experiments and purified the LH3/PLOD3 mutants, with support from SF. MF carried out mass spectrometry experiments and analyzed data. SAS carried out the calculations for the parametrization of Fe^{2+} binding sites. EM, SS and GC carried out molecular simulations and analysis. LS, EM, DM and FF wrote the manuscript, with contributions from all authors.

Funding

This project has received funding from the Italian Association for Cancer Research (AIRC, "My First AIRC Grant" id. 20075 to FF), by the Mizutani Foundation for Glycoscience (grant id. 200039 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 Department of Biology and Biotechnology "L. Spallanzani," University of Pavia). AC carried out research while recipient of a Marie Curie Individual Fellowship (MSCA-IF) from the European Union's Horizon 2020 research and innovation program (grant agreement COTETHERS—n. 745934). The TRIC instrumentation used for this research was acquired through funding by Regione Lombardia, regional law n° 9/2020, resolution n° 3776/2020. 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.

Acknowledgments

We thank the PASS-BioMed Facility (Centro Grandi Strumenti) of the University of Pavia for provision of TRIC instrumentation and Sristi Raj Rai for help with enzymatic assays. SS wishes to acknowledge Prof. Ulf Ryde (*Lund University*, Sweden) for his kind advice on using the def2-SV(P) basis set for Fe^{2+}.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.876352/full#supplementary-material

References

Baek, J. H., Yun, H. S., Kwon, G. T., Lee, J., Kim, J. Y., Jo, Y., et al. (2019). PLOD3 suppression exerts an anti-tumor effect on human lung cancer cells by modulating the PKC-delta signaling pathway. *Cell Death Dis.* 10, 156. doi:10.1038/s41419-019-1405-8

Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* 97, 10269–10280. doi:10.1021/j100142a004

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi:10.1063/1.464913

Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. V., Dinola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118

Besler, B. H., Merz, K. M., Jr., and Kollman, P. A. (1990). Atomic charges derived from semiempirical methods. *J. Comput. Chem.* 11, 431–439. doi:10.1002/jcc.540110404

Blanco, M. A., Leroy, G., Khan, Z., Aleckovic, M., Zee, B. M., Garcia, B. A., et al. (2012). Global secretome analysis identifies novel mediators of bone metastasis. *Cell Res.* 22, 1339–1355. doi:10.1038/cr.2012.89

Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., Cruzeiro, V. W. D., et al. (2018). *Amber 2018*. San Francisco: University of California.

Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., et al. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688. doi:10.1002/jcc.20290

Chen, Y. L., Guo, H. F., Terajima, M., Banerjee, P., Liu, X., Yu, J., et al. (2016). Lysyl hydroxylase 2 is secreted by tumor cells and can modify collagen in the extracellular space. *J. Biol. Chem.* 291, 25799–25808. doi:10.1074/jbc.M116.759803

Chen, Y., Terajima, M., Yang, Y., Sun, L., Ahn, Y. H., Pankova, D., et al. (2015). Lysyl hydroxylase 2 induces a collagen cross-link switch in tumor stroma. *J. Clin. Invest.* 125, 1147–1162. doi:10.1172/JCI74725

Clifton, I. J., Mcdonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., and Schofield, C. J. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* 100, 644–669. doi:10.1016/j.jinorgbio.2006.01.024

Costas, M., Mehn, M. P., Jensen, M. P., and Que, L., Jr. (2004). Dioxygen activation at mononuclear nonheme iron active sites: Enzymes, models, and intermediates. *Chem. Rev.* 104, 939–986. doi:10.1021/cr020628n

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi:10.1063/1.464397

De Giorgi, F., Fumagalli, M., Scietti, L., and Forneris, F. (2021). Collagen hydroxylysine glycosylation: Non-conventional substrates for atypical glycosyltransferase enzymes. *Biochem. Soc. Trans.* 49 (2), 855–866. doi:10.1042/BST20200767

Deng, X., Pan, Y., Yang, M., Liu, Y., and Li, J. (2021). PLOD3 is associated with immune cell infiltration and genomic instability in colon adenocarcinoma. *Biomed. Res. Int.* 2021, 4714526. doi:10.1155/2021/4714526

Devkota, A. K., Veloria, J. R., Guo, H. F., Kurie, J. M., Cho, E. J., and Dalby, K. N. (2019). Development of a high-throughput lysyl hydroxylase (LH) assay and identification of small-molecule inhibitors against LH2. *SLAS Discov.* 24, 484–491. doi:10.1177/2472555218817057

Du, H., Chen, Y., Hou, X., Huang, Y., Wei, X., Yu, X., et al. (2017a). PLOD2 regulated by transcription factor FOXA1 promotes metastasis in NSCLC. *Cell Death Dis.* 8, e3143. doi:10.1038/cddis.2017.553

Du, H., Pang, M., Hou, X., Yuan, S., and Sun, L. (2017b). PLOD2 in cancer research. *Biomed. Pharmacother.* 90, 670–676. doi:10.1016/j.biopha.2017.04.023

Eisinger-Mathason, T. S., Zhang, M., Qiu, Q., Skuli, N., Nakazawa, M. S., Karakasheva, T., et al. (2013). Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. *Cancer Discov.* 3, 1190–1205. doi:10.1158/2159-8290.CD-13-0118

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 486–501. doi:10.1107/S0907444910007493

Ewans, L. J., Colley, A., Gaston-Massuet, C., Gualtieri, A., Cowley, M. J., Mccabe, M. J., et al. (2019). Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. *J. Med. Genet.* 56, 629–638. doi:10.1136/jmedgenet-2019-106019

Faravelli, S., Campioni, M., Palamini, M., Canciani, A., Chiapparino, A., and Forneris, F. (2021). Optimized recombinant production of secreted proteins using human embryonic kidney (HEK293) cells grown in suspension. *Bio. Protoc.* 11, e3998. doi:10.21769/BioProtoc.3998

Flashman, E., and Schofield, C. J. (2007). The most versatile of all reactive intermediates? *Nat. Chem. Biol.* 3, 86–87. doi:10.1038/nchembio0207-86

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09 revision A.2*. Wallingford, CT: Gaussian, Inc.

Gilkes, D. M., Bajpai, S., Wong, C. C., Chaturvedi, P., Hubbi, M. E., Wirtz, D., et al. (2013). Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. *Mol. Cancer Res.* 11, 456–466. doi:10.1158/1541-7786.MCR-12-0629

Gkretsi, V., and Stylianopoulos, T. (2018). Cell adhesion and matrix stiffness: Coordinating cancer cell invasion and metastasis. *Front. Oncol.* 8, 145. doi:10.3389/fonc.2018.00145

Gong, S., Duan, Y., Wu, C., Osterhoff, G., Schopow, N., and Kallendrusch, S. (2021). A human pan-cancer system Analysis of procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3 (PLOD3). *Int. J. Mol. Sci.* 22, 9903. doi:10.3390/ijms22189903

Guo, H. F., Bota-Rabassedas, N., Terajima, M., Leticia Rodriguez, B., Gibbons, D. L., Chen, Y., et al. (2021). A collagen glucosyltransferase drives lung adenocarcinoma progression in mice. *Commun. Biol.* 4, 482. doi:10.1038/s42003-021-01982-w

Guo, H. F., Tsai, C. L., Terajima, M., Tan, X., Banerjee, P., Miller, M. D., et al. (2018). Pro-metastatic collagen lysyl hydroxylase dimer assemblies stabilized by Fe(2+)-binding. *Nat. Commun.* 9, 512. doi:10.1038/s41467-018-02859-z

Guo, T., Gu, C., Li, B., and Xu, C. (2021). PLODs are overexpressed in ovarian cancer and are associated with gap junctions via connexin 43. *Lab. Invest.* 101, 564–569. doi:10.1038/s41374-021-00533-5

Hausinger, R. P. (2004). FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39, 21–68. doi:10.1080/10409230490440541

Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5

Ikeda, H., Ogata, I., and Fujiwara, K. (1994). Evidence that impaired intracellular collagen synthesis reduces proliferation in cultured rat hepatocytes. *Biochem. Biophys. Res. Commun.* 200, 1701–1707. doi:10.1006/bbrc.1994.1648

Jiang, H., Guo, W., Yuan, S., and Song, L. (2020). PLOD1 is a prognostic biomarker and mediator of proliferation and invasion in osteosarcoma. *Biomed. Res. Int.* 2020, 3418398. doi:10.1155/2020/3418398

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869

Joung, I. S., and Cheatham, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* 112, 9020–9041. doi:10.1021/jp8001614

Jover, E., Silvente, A., Marin, F., Martinez-Gonzalez, J., Orriols, M., Martinez, C. M., et al. (2018). Inhibition of enzymes involved in collagen cross-linking reduces vascular smooth muscle cell calcification. *FASEB J.* 32, 4459–4469. doi:10.1096/fj.201700653R

Kivirikko, K. I., and Prockop, D. J. (1967). Enzymatic hydroxylation of proline and lysine in protocollagen. *Proc. Natl. Acad. Sci. U. S. A.* 57, 782–789. doi:10.1073/pnas.57.3.782

Kurozumi, A., Kato, M., Goto, Y., Matsushita, R., Nishikawa, R., Okato, A., et al. (2016). Regulation of the collagen cross-linking enzymes LOXL2 and PLOD2 by tumor-suppressive microRNA-26a/b in renal cell carcinoma. *Int. J. Oncol.* 48, 1837–1846. doi:10.3892/ijo.2016.3440

Lee, C., Yang, W., and Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* 37, 785–789. doi:10.1103/physrevb.37.785

Levental, K. R., Yu, H., Kass, L., Lakins, J. N., Egeblad, M., Erler, J. T., et al. (2009). Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* 139, 891–906. doi:10.1016/j.cell.2009.10.027

Li, G., Wang, X., and Liu, G. (2021). Corrigendum to "PLOD2 is a potent prognostic marker and associates with immune infiltration in cervical cancer". *Biomed. Res. Int.* 2021, 9762405. doi:10.1155/2021/9762405

Li, P., and Merz, K. M., Jr. (2016). MCPB.py: A Python based metal center parameter builder. *J. Chem. Inf. Model.* 56, 599–604. doi:10.1021/acs.jcim.5b00674

Li, S. S., Lian, Y. F., Huang, Y. L., Huang, Y. H., and Xiao, J. (2020). Overexpressing PLOD family genes predict poor prognosis in gastric cancer. *J. Cancer* 11, 121–131. doi:10.7150/jca.35763

Loenarz, C., and Schofield, C. J. (2008). Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* 4, 152–156. doi:10.1038/nchembio0308-152

Loenarz, C., and Schofield, C. J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends biochem. Sci.* 36, 7–18. doi:10.1016/j.tibs.2010.07.002

Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992). Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide. *Biopolymers* 32, 523–535. doi:10.1002/bip.360320508

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi:10.1021/acs.jctc.5b00255

Martinez, S., and Hausinger, R. P. (2015). Catalytic mechanisms of Fe(II)- and 2-Oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290, 20702–20711. doi:10.1074/jbc.R115.648691

Moroni, E., Agard, D. A., and Colombo, G. (2018). The structural asymmetry of mitochondrial Hsp90 (Trap1) determines fine tuning of functional dynamics. *J. Chem. Theory Comput.* 14, 1033–1044. doi:10.1021/acs.jctc.7b00766

Morra, G., Potestio, R., Micheletti, C., and Colombo, G. (2012). Corresponding functional dynamics across the Hsp90 chaperone family: Insights from a multiscale analysis of MD simulations. *PLoS Comput. Biol.* 8, e1002433. doi:10.1371/journal.pcbi.1002433

Myllylä, R., Schubotz, L. M., Weser, U., and Kivirikko, K. I. (1979). Involvement of superoxide in the prolyl and lysyl hydroxylase reactions. *Biochem. Biophys. Res. Commun.* 89, 98–102. doi:10.1016/0006-291x(79)90948-3

Noda, T., Yamamoto, H., Takemasa, I., Yamada, D., Uemura, M., Wada, H., et al. (2012). PLOD2 induced under hypoxia is a novel prognostic factor for hepatocellular carcinoma after curative resection. *Liver Int.* 32, 110–118. doi:10.1111/j.1478-3231.2011.02619.x

Pankova, D., Chen, Y., Terajima, M., Schliekelman, M. J., Baird, B. N., Fahrenholtz, M., et al. (2016). Cancer-associated fibroblasts induce a collagen cross-link switch in tumor stroma. *Mol. Cancer Res.* 14, 287–295. doi:10.1158/1541-7786.MCR-15-0307

Provenzano, P. P., Eliceiri, K. W., Campbell, J. M., Inman, D. R., White, J. G., and Keely, P. J. (2006). Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Med.* 4, 38. doi:10.1186/1741-7015-4-38

Puistola, U., Turpeenniemi-Hujanen, T. M., Myllyla, R., and Kivirikko, K. I. (1980). Studies on the lysyl hydroxylase reaction. II. Inhibition kinetics and the reaction mechanism. *Biochim. Biophys. Acta* 611, 51–60. doi:10.1016/0005-2744(80)90041-8

Rose, N. R., Mcdonough, M. A., King, O. N., Kawamura, A., and Schofield, C. J. (2011). Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* 40, 4364–4397. doi:10.1039/c0cs00203h

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* 9, 3878–3888. doi:10.1021/ct400314y

Saito, T., Uzawa, K., Terajima, M., Shiiba, M., Amelio, A. L., Tanzawa, H., et al. (2019). Aberrant collagen cross-linking in human oral squamous cell carcinoma. *J. Dent. Res.* 98, 517–525. doi:10.1177/0022034519828710

Sato, K., Parag-Sharma, K., Terajima, M., Musicant, A. M., Murphy, R. M., Ramsey, M. R., et al. (2021). Lysyl hydroxylase 2-induced collagen cross-link switching promotes metastasis in head and neck squamous cell carcinomas. *Neoplasia* 23, 594–606. doi:10.1016/j.neo.2021.05.014

Schulte, C., Khayenko, V., Nordblom, N. F., Tippel, F., Peck, V., Gupta, A. J., et al. (2021). High-throughput determination of protein affinities using unmodified peptide libraries in nanomolar scale. *iScience* 24, 101898. doi:10.1016/j.isci.2020.101898

Scietti, L., Chiapparino, A., De Giorgi, F., Fumagalli, M., Khoriauli, L., Nergadze, S., et al. (2018). Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. *Nat. Commun.* 9, 3163. doi:10.1038/s41467-018-05631-5

Scietti, L., and Forneris, F. (2020). "Full-length human collagen lysyl hydroxylases," in *Encyclopedia of inorganic and bioinorganic chemistry*. Editor R. A. Scott (Hoboken, NJ: Wiley), 1–12.

Seminario, J. M. (1996). Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* 60, 1271–1277. doi:10.1002/(sici)1097-461x(1996)60:7<1271::aid-qua8>3.0.co;2-w

Shi, J., Bao, M., Wang, W., Wu, X., Li, Y., Zhao, C., et al. (2021). Integrated profiling identifies PLOD3 as a potential prognostic and immunotherapy relevant biomarker in colorectal cancer. *Front. Immunol.* 12, 722807. doi:10.3389/fimmu.2021.722807

Sondergaard, C. R., Olsson, M. H., Rostkowski, M., and Jensen, J. H. (2011). Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* 7, 2284–2295. doi:10.1021/ct200133y

Song, Y., Zheng, S., Wang, J., Long, H., Fang, L., Wang, G., et al. (2017). Hypoxia-induced PLOD2 promotes proliferation, migration and invasion via PI3K/Akt signaling in glioma. *Oncotarget* 8, 41947–41962. doi:10.18632/oncotarget.16710

Tian, L., Zhou, H., Wang, G., Wang, W. Y., Li, Y., and Xue, X. (2021). The relationship between PLOD1 expression level and glioma prognosis investigated using public databases. *PeerJ* 9, e11422. doi:10.7717/peerj.11422

Tsai, C. K., Huang, L. C., Tsai, W. C., Huang, S. M., Lee, J. T., and Hueng, D. Y. (2018). Overexpression of PLOD3 promotes tumor progression and poor prognosis in gliomas. *Oncotarget* 9, 15705–15720. doi:10.18632/oncotarget.24594

Van Der Slot, A. J., Zuurmond, A. M., Van Den Bogaerdt, A. J., Ulrich, M. M., Middelkoop, E., Boers, W., et al. (2004). Increased formation of pyridinoline cross-links due to higher telopeptide lysyl hydroxylase levels is a general fibrotic phenomenon. *Matrix Biol.* 23, 251–257. doi:10.1016/j.matbio.2004.06.001

Vasta, J. D., and Raines, R. T. (2015). Selective inhibition of prolyl 4-hydroxylases by bipyridinedicarboxylates. *Bioorg. Med. Chem.* 23, 3081–3090. doi:10.1016/j.bmc.2015.05.003

Wan, J., Qin, J., Cao, Q., Hu, P., Zhong, C., and Tu, C. (2020). Hypoxia-induced PLOD2 regulates invasion and epithelial-mesenchymal transition in endometrial carcinoma cells. *Genes Genomics* 42, 317–324. doi:10.1007/s13258-019-00901-y

Wang, B., Xu, I., Ge, Y., Cai, X., Li, Q., Yu, Z., et al. (2019). PLOD3 is upregulated in gastric cancer and correlated with clinicopathologic characteristics. *Clin. Lab.* 65, 2901. doi:10.7754/Clin.Lab.2018.180541

Wang, D., Zhang, S., and Chen, F. (2018). High expression of PLOD1 drives tumorigenesis and affects clinical outcome in gastrointestinal carcinoma. *Genet. Test. Mol. Biomarkers* 22, 366–373. doi:10.1089/gtmb.2018.0009

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi:10.1002/jcc.20035

Wang, Z., Shi, Y., Ying, C., Jiang, Y., and Hu, J. (2021). Hypoxia-induced PLOD1 overexpression contributes to the malignant phenotype of glioblastoma via NF-κB signaling. *Oncogene* 40, 1458–1475. doi:10.1038/s41388-020-01635-y

Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7, 3297–3305. doi:10.1039/b508541a

Xie, D., Li, J., Wei, S., Qi, P., Ji, H., Su, J., et al. (2020). Knockdown of PLOD3 suppresses the malignant progression of renal cell carcinoma via reducing TWIST1 expression. *Mol. Cell. Probes* 53, 101608. doi:10.1016/j.mcp.2020.101608

Xu, W. H., Xu, Y., Wang, J., Tian, X., Wu, J., Wan, F. N., et al. (2019). Procollagen-lysine, 2-oxoglutarate 5-dioxygenases 1, 2, and 3 are potential prognostic indicators in patients with clear cell renal cell carcinoma. *Aging (Albany NY)* 11, 6503–6521. doi:10.18632/aging.102206

Xu, Y., Zhang, L., Wei, Y., Zhang, X., Xu, R., Han, M., et al. (2017). Procollagen-lysine 2-oxoglutarate 5-dioxygenase 2 promotes hypoxia-induced glioma migration and invasion. *Oncotarget* 8, 23401–23413. doi:10.18632/oncotarget.15581

Yamada, Y., Kato, M., Arai, T., Sanada, H., Uchida, A., Misono, S., et al. (2019). Aberrantly expressed PLOD1 promotes cancer aggressiveness in bladder cancer: A potential prognostic marker and therapeutic target. *Mol. Oncol.* 13, 1898–1912. doi:10.1002/1878-0261.12532

Yamauchi, M., and Sricholpech, M. (2012). Lysine post-translational modifications of collagen. *Essays Biochem.* 52, 113–133. doi:10.1042/bse0520113

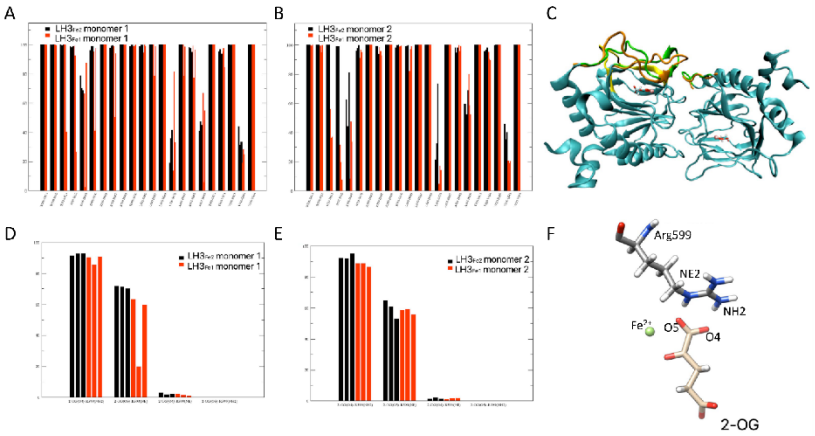
Yang, B., Zhao, Y., Wang, L., Zhao, Y., Wei, L., Chen, D., et al. (2020). Identification of PLOD family genes as novel prognostic biomarkers for hepatocellular carcinoma. *Front. Oncol.* 10, 1695. doi:10.3389/fonc.2020.01695

Zhao, Y., Zhang, X., and Yao, J. (2021). Comprehensive analysis of PLOD family members in low-grade gliomas using bioinformatics methods. *PLoS One* 16, e0246097. doi:10.1371/journal.pone.0246097

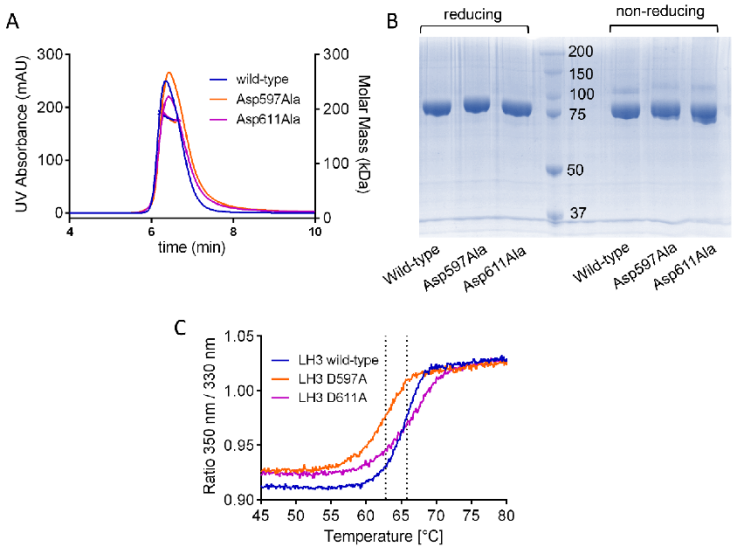
Supplementary Information for:

A Fe^{2+}-dependent self-inhibited state influences the druggability of human collagen lysyl hydroxylase (LH/PLOD) enzymes

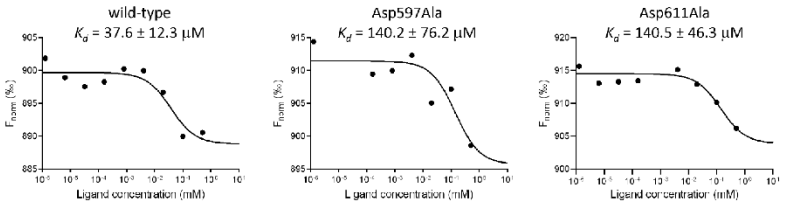
Luigi Scietti, Elisabetta Moroni, Daiana Mattoteia *et al.*



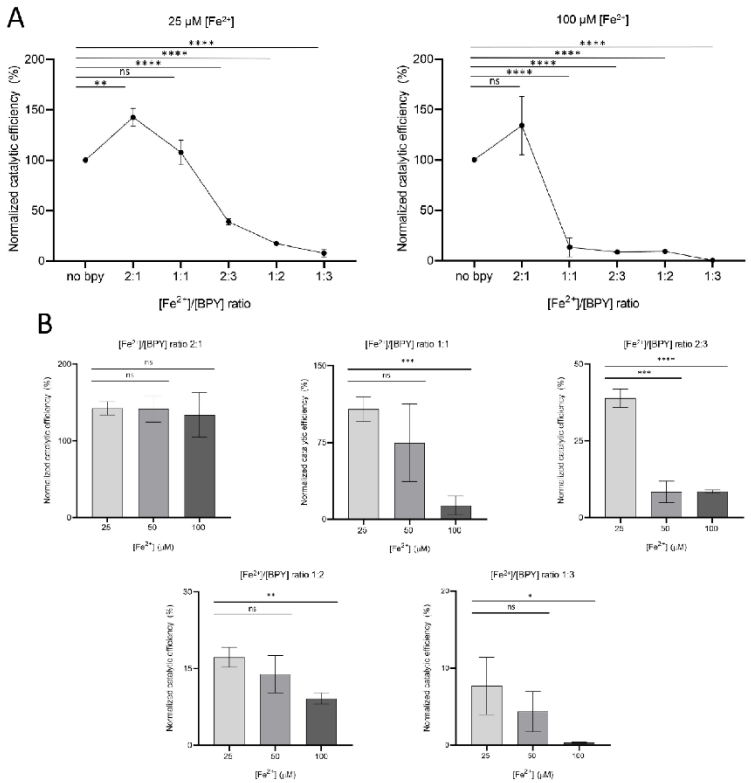
Supplementary Figure 1. Results from computational investigations. (A-B) Persistence of the native contacts of the loop 590-610 obtained from MD simulations, calculated for the three replicate simulations, for each system and monomer. The persistence of a contact is the number of frames in which this contact is present, divided by the total number of frames of MD simulation. A contact is considered as native if it is present in the crystal structure. A contact is considered as "present" when the distance of one of the C_α atoms of residues of the loop and one of the C_α atoms of residues not belonging to the loop is lower than 7.0 Å. (C) Snapshot from MD simulation of **LH3$_{Fe1}$** (cyan) superimposed to the crystal structure (orange). Residue of the loop 590-610 and residues 643-648 which form stable contacts with this loop during the dynamics, are colored in green, while the corresponding residues in the crystal structure are colored in orange. Yellow residues correspond to the protein portion where the native contacts are less stable in the **LH3$_{Fe1}$** system. (D-E) Persistence of hydrogen bonds between Arg599 and 2-OG co-substrate in both monomers over the course of the MD trajectories, calculate for the three replicate simulations. Hydrogen bonds are determined using a geometric criteria: the donor to acceptor heavy atom distance (distance of 3.0 Å) and the donor-hydrogen-acceptor angle (angle cutoff of 135°). (F) Atom names used to define donors and acceptors for Arg599 and 2-OG.



Supplementary Figure 2. Biochemical characterization of LH3/PLOD3 mutants and comparison with wild-type enzyme. (A) Comparative size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) analysis. The dots indicate the computed molar mass associated to each sample. (B) SDS-PAGE analysis. Shown are the results obtained in reducing (left) and non-reducing (right) conditions. (C) Differential scanning fluorimetry analysis. The dashed lines indicate the temperature range incorporating the calculated unfolding temperature values for all curves.



Supplementary Figure 3. Analysis of Fe^{2+} binding in wild-type and mutant LH3. Normalized fluorescence values (F_{norm}) extrapolated from TRIC curves obtained from the Dianthus NT.23 (NanoTemper Technologies) were plotted as as a function of Fe^{2+} concentration and used to compute the K_d values reported with the associated standard deviations as obtained from the binding affinity analysis performed using the using the DI.screening analysis software (NanoTemper Technologies).



Supplementary Figure 4. Detailed statistical evaluation of the data shown in Figure 6. (A) Results of the MS analysis of LH3/PLOD3 enzymatic activity as a function of the $[Fe^{2+}]/[BPY]$ ratio in assays performed using 25 µM Fe^{2+} (left) or 100 µM Fe^{2+} (right). Error bars represent standard deviations from average of triplicate independent experiments. Statistical evaluations based on pair sample comparisons between data points collected in absence or in presence of $[Fe^{2+}]$ using Student's t-test. ns, non-significant; *, P-value<0.05; **, P-value<0.01; ***, P-value<0.001; ****, P-value < 0.0001. (B) Pairwise statistical evaluation of each individual $[Fe^{2+}]/[BPY]$ ratio as a function of different $[Fe^{2+}]$ used in the assay.

nature communications

Article

https://doi.org/10.1038/s41467-025-59017-5

Molecular structure and enzymatic mechanism of the human collagen hydroxylysine galactosyltransferase GLT25D1/COLGALT1

Received: 21 June 2024

Accepted: 8 April 2025

Published online: 16 April 2025

Check for updates

Matteo De Marco[1,9], Sristi Raj Rai[1,9], Luigi Scietti[1,8,9], Daiana Mattoteia[1], Stefano Liberi[1], Elisabetta Moroni[2], Alberta Pinnola[3], Alice Vetrano[4], Claudio Iacobucci[4], Carlo Santambrogio[5], Giorgio Colombo[6] & Federico Forneris[1,7]

During collagen biosynthesis, lysine residues undergo extensive post-translational modifications through the alternate action of two distinct metal ion-dependent enzyme families (i.e., LH/PLODs and GLT25D/COLGALT), ultimately producing the highly conserved α-(1,2)-glucosyl-β-(1,O)-galactosyl-5-hydroxylysine pattern. Malfunctions in these enzymes are linked to developmental pathologies and extracellular matrix alterations associated to enhanced aggressiveness of solid tumors. Here, we characterized human GLT25D1/COLGALT1, revealing an elongated head-to-head homodimeric assembly. Each monomer encompasses two domains (named GT1 and GT2), both unexpectedly capable of binding metal ion cofactors and UDP-α-galactose donor substrates, resulting in four candidate catalytic sites per dimer. We identify the catalytic site in GT2, featuring an unusual Glu-Asp-Asp motif critical for Mn^{2+} binding, ruling out direct catalytic roles for the GT1 domain, but showing that in this domain the unexpectedly bound Ca^{2+} and UDP-α-galactose cofactors are critical for folding stability. Dimerization, albeit not essential for GLT25D1/COLGALT1 activity, provides a critical molecular contact site for multi-enzyme assembly interactions with partner multifunctional LH/PLOD lysyl hydroxylase-glycosyltransferase enzymes.

Collagens are highly conserved proteins, from invertebrates to higher vertebrates. Their amino acid composition, post-translational modification patterns, and supramolecular assembly confer a versatile spectrum of physical and mechanical properties, resulting in the most abundant protein family in the animal kingdom constituting essential components of several animal tissues including bones, cartilage, tendons, and dermis[1–3]. During biosynthesis, collagens undergo extensive post-translational modifications to fold into right-handed triple

[1]The Armenise-Harvard Laboratory of Structural Biology, Dept. Biology and Biotechnology, University of Pavia, Via Ferrata 9A, 27100 Pavia, Italy. [2]SCITEC-CNR, via Mario Bianco 9, I-20131 Milano, Italy. [3]BioPhotoLab, Dept. Biology and Biotechnology, University of Pavia, Via Ferrata 9A, 27100 Pavia, Italy. [4]Department of Physical and Chemical Sciences, University of L'Aquila, 67100 L'Aquila, Italy. [5]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy. [6]Department of Chemistry, University of Pavia, Via Taramelli 12, Pavia, Italy. [7]Fondazione IRCCS Policlinico San Matteo, Pavia, Italy. [8]Present address: Biochemistry and Structural Biology Unit, Department of Experimental Oncology, IRCCS European Institute of Oncology (IEO), Via Adamello 16, 20139 Milan, Italy. [9]These authors contributed equally: Matteo De Marco, Sristi Raj Rai, Luigi Scietti. e-mail: federico.forneris@unipv.it

helices. The repetitive unit consists of the Gly-Xaa-Yaa sequence motif, where positions Xaa and Yaa are frequently occupied by proline and hydroxyproline residues, respectively[4–7]. The role of proline hydroxylation in the global process of collagen biosynthesis is the most studied and understood to date, and it is essential to the winding process of the collagen right-handed triple helix[8–10].

Lysine residues are less abundant than proline residues in collagen polypeptides, and are typically found at Yaa positions of the Gly-Xaa-Yaa motif[11]. The role of lysine residues and their post-translational modifications is mainly linked to the cross-linking processes that occur between collagen molecules once secreted into the extracellular space. In collagens and collagen-like protein segments, lysine residues are modified sequentially by specific biosynthetic enzymes which catalyze the formation of a simple but very conserved post-translational modification pattern: the α-(1,2)-glucosyl-β-(1,O)-galactosyl-5-hydroxylysine (Glc-Gal-Hyl)[12–16]. The path to the formation of Glc-Gal-Hyl starts with the formation of 5-hydroxylysine (Hyl) operated by Fe^{2+}-dependent collagen lysyl hydroxylases (LH, also known as procollagen, lysyl 2-oxoglutarate dioxygenases (PLOD)). The LH/PLOD enzyme family is characterized in humans by three different isoforms (LH1/PLOD1, LH2/PLOD2, and LH3/PLOD3)[17–19]. The LH reaction occurs in the C-terminal dimeric lysyl hydroxylase domain of LH/PLODs. Hyl residues are then glycosylated by Mn^{2+}-dependent galactosyltransferase enzymes, through an inverting reaction that adds the galactose moiety of the uridine diphosphate-α-galactose (UDP-α-Gal) donor substrate to the Hyl hydroxyl group, generating β-(1,O)-5-hydroxylysine (Gal-T activity). Gal-Hyl is further glycosylated by the multifunctional LH/PLOD enzymes through a retaining reaction, that requires Mn^{2+} and uridine diphosphate-α-glucose (UDP-α-Glc) donor substrate (Glc-T activity), ultimately generating the Glc-Gal-Hyl. Initially, the multifunctional LH/PLOD enzyme(s) were thought to be responsible for the entire Lys–to–Glc-Gal-Hyl biosynthetic pathway[20,21]. The subsequent identification of a specific family of collagen galactosyltranfserases (named GLT25D or COLGALT), underpinned their essential roles in collagen biosynthesis in vivo[22–27]. Additional biochemical studies on human LH/PLOD and GLT25D/COLGALT enzymes ultimately ruled out the single enzyme hypothesis. The resulting picture shows how the concerted action of two enzyme systems (i.e., LH/PLOD and GLT25D/COLGALT), alternatively acting on the same substrate and possibly directly interacting with each other[28], leads to complete Glc-Gal-Hyl collagen post-translational modifications (Fig. 1a).

While multifunctional lysyl hydroxylase and glucosyltransferase LH/PLOD enzymes have been the subject of extensive molecular and structural characterizations, much less is known regarding the GLT25D/COLGALT enzymes, which are responsible for the central galactosyltransferase reaction. Human GLT25D/COLGALT are soluble proteins anchored to the endoplasmic reticulum through a C-terminal Arg-Asp-Glu-Leu (RDEL) sequence, and have been found to co-localize with LH/PLOD enzymes[28,29]. Three human GLT25D/COLGALT isoforms have been identified: GLT25D1/COLGALT1, constitutively present in all human tissues, GLT25D2/COLGALT2, expressed mostly in nervous tissue, and GLT25D3/COLGALT3, found in all secretory tissues such as salivary glands, pancreas, placenta, as well as in the nervous system. However, only the first two isoforms are active, whereas isoform 3 (also called CERCAM) is inactive in vitro and potentially not related at all to collagen lysine modification.

Of considerable relevance is the involvement of the *colgalt* genes and their protein products GLT25D/COLGALT enzymes in numerous pathological conditions, including cerebral small vessel disease, liver pathologies, osteoarthritis, musculoskeletal defects, and cancer[26,30–32]. Liver fibrosis is a chronic pathological condition that sees an accumulation of proteins at the level of the extracellular matrix, whose evolution often leads affected subjects to develop cirrhosis and hepatocellular carcinoma. Recent work has observed high levels of GLT25D1/COLGALT1 in the plasma of cirrhosis patients compared to unaffected patients, and one of the proposed hypotheses is the stimulatory role of the enzyme against hepatic stellate cells[33].

Efforts to decipher the structure-function relationships underlying GLT25D/COLGALT activity on collagen molecules have yielded interesting insights, but highlighted the presence of multiple Asp-x-Asp (DxD) motifs along the polypeptide sequence potentially relevant for enzymatic activity, thus demanding more in-depth structure-based investigations.

In this work, we merged computational and experimental structural biology methods with biochemistry and biophysics to reveal the quaternary assembly of GLT25D1/COLGALT1. The X-ray crystal structures of full-length human GLT25D1/COLGALT1 in its substrate-free state as well as in complex with the donor substrate UDP-α-Gal were determined using experimental phasing. Small-angle X-ray scattering, negative stain electron microscopy and mass photometry jointly revealed an elongated quaternary structure reminiscent of that of LH/PLOD enzymes, characterized by a head-to-head dimer with each monomer composed of two similar domains with Rossman fold architecture, both of which support binding of metal ions and the donor substrate. Structure-guided site-directed mutagenesis, native mass spectrometry and molecular dynamics simulations ultimately identified the C-terminal domain as the functional site for enzyme catalysis, while the N-terminal domain retains UDP-α-Gal unusually bound to Ca^{2+} exclusively for folding stability purposes. Cross-linking mass spectrometry analysis showed that the enzyme's homodimerization is crucial for interaction with partner LH/PLOD enzymes. Collectively, our results provide insights into the molecular architecture and functions of a key human enzyme of collagen biosynthesis.

Results

GLT25D1/COLGALT1 is a Mn^{2+} dependent galactosyltransferase

We established methods to recombinantly produce and purify human full-length GLT25D1/COLGALT1 in *E. coli* and obtained enzyme preparations suitable for evaluation of galactosyltransferase activity in vitro (Supplementary Fig. 1), through direct detection of Gal-T activity on synthetic collagen peptides using high-resolution mass spectrometry (HRMS), as well as by monitoring UDP generation from consumption of the UDP-α-Gal donor substrate in the presence of collagen hydroxylysines. Using luminescence-based detection of UDP as reaction product, we established a biochemical assay to evaluate steady-state enzyme kinetics (Supplementary Fig. 2). The results obtained from steady-state Michaelis-Menten analysis (Supplementary Table 5, Supplementary Fig. 2) were comparable to previously published investigations performed using radiolabeled UDP-α-Gal donor substrate[34]. Using these assays, we firstly evaluated the enzyme's strong preference for the Mn^{2+} metal ion cofactor, highlighting complete absence of enzymatic activity in presence of other divalent metal ions, with the exception of Mg^{2+}, which nevertheless yielded lower enzymatic activity compared to Mn^{2+} (Supplementary Fig. 3a, b). Given the ability of closely-related multifunctional LH/PLOD enzymes to process multiple donor substrates also in absence of acceptor substrates, we probed GLT25D1/COLGALT1 and found that differently from LH/PLOD, GLT25D1/COLGALT1 is specific for UDP-α-Gal and displays limited processing of UDP-α-Gal in absence of acceptor substrates (i.e., uncoupled activity, Supplementary Fig. 3c).

GLT25D1/COLGALT1 features an elongated multi-domain architecture

The 2.8 Å crystal structure of human wild-type GLT25D1/COLGALT1 was obtained by experimental phasing, exploiting single wavelength anomalous dispersion (SAD) of a Hg^{2+} derivative (Supplementary Tables 2 and 3). The quality of the electron density was excellent for the globular region of the two monomers of enzyme found in the asymmetric unit (Supplementary Fig. 4), whereas flexible regions

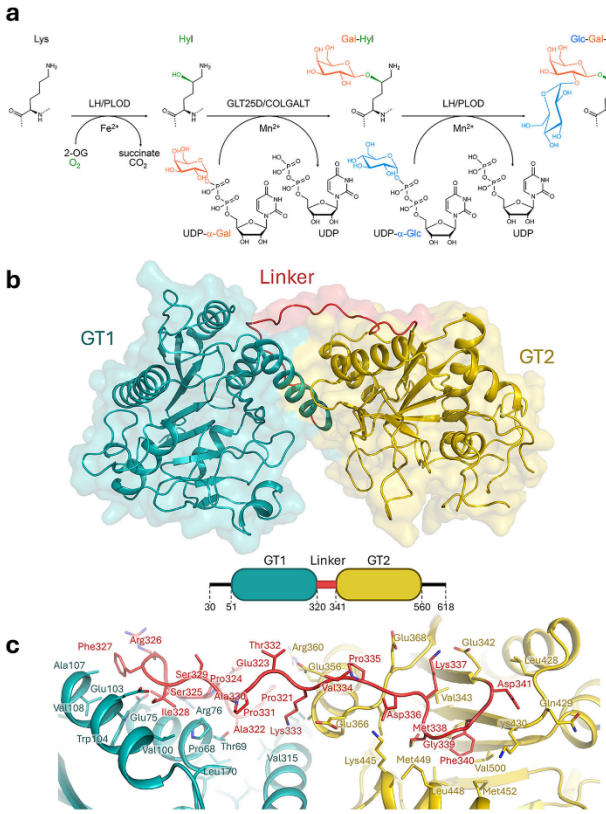


Fig. 1 | Biochemical and structural features of human GLT25D1/COLGALT1. **a** Reaction scheme showing the collagen Lys-to-Glc-Gal-Hyl conversion post-translational modification process mediated by the alternate action of LH/PLOD and GLT25D1/COLGALT1, with color highlight of donor substrates and associated products. **b** Cartoon representation of the crystal structure of human GLT25D1/COLGALT1, showing the domain architecture of the enzyme with GT1 and GT2 domains connected via a rigid linker element, anchored to both domains via hydrophobic anchorages shown in details in (**c**).

located at the N- (residues 30–41) and C-termini (residues 572–618) could not be modeled.

Each GLT25D1/COLGALT1 polypeptide is characterized by two globular domains, which we named GT1 and GT2 (Fig. 1b), both characterized by similar glycosyltransferase architectures, but different topological organization (Supplementary Fig. 5a). In GLT25D1/COLGALT1, the two GT domains contact each other through electrostatics and hydrophobic contacts along helix α7 (GT1) and α8 (GT2). A 25-amino acid linker interconnects the two domains by shielding otherwise surface-exposed hydrophobic patches with two phenylalanine residues (Phe327 and Phe340, respectively, Fig. 1c).

Although both domains are structurally divergent from previously determined Rossman fold topologies, the GT1 domain appears more similar to known glycosyltransferases, yielding Z-scores higher than 15 with few bacterial and human GT-A glycosyltransferases (Supplementary Fig. 5b) according to DALI[35], despite very limited sequence identity conservation. The closest structural homolog with an experimentally determined structure is human LH3/PLOD3, whose central accessory (AC) domain shares only 24% sequence identity, but shows very similar structural arrangement with RMSD of 1.2 Å (Supplementary Fig. 5c). This was further confirmed by extensive protein BLAST searches using the GLT25D1/COLGALT GT1 domain sequence as search query, which returned results including vertebrates, invertebrates, and giant viruses (Supplementary Fig. 6). The latter were of particular interest, as previous discoveries reported that *Acanthamoeba polyphaga mimivirus* is decorated with post-translationally modified collagens with high homology to mammalian orthologs[36]. Nevertheless, BLAST searches excluding LH/PLOD homologs identified almost exclusively the additional enzyme isoforms GLT25D2/COLGALT2 and CERCAM in vertebrates and GLT25D/COLGALT orthologs in insects (Supplementary Fig. 6). Collectively, these data highlight the conservation of the GT-A fold architecture of the GT1 domain, yielding closely matching folding topologies despite very limited amino acid sequence conservation (Supplementary Fig. 5b).

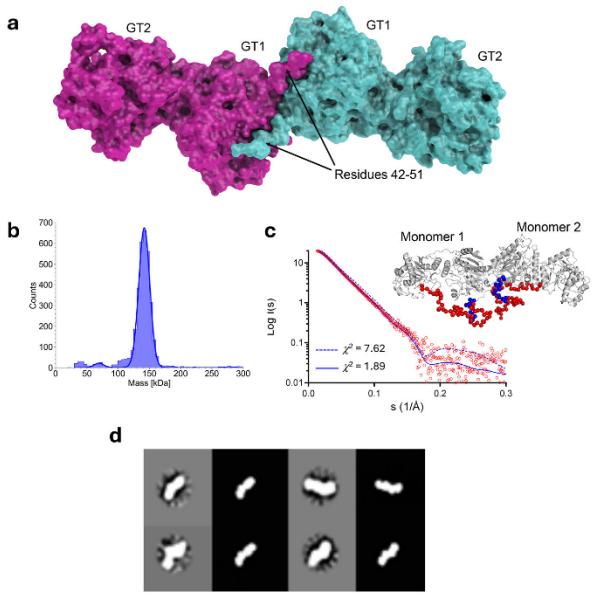


Fig. 2 | GLT25D1/COLGALT1 forms stable elongated dimers. a Structure of dimeric full-length GLT25D1/COLGALT1 enzyme, showing two molecules interconnected through their N-terminal domains. **b** Mass photometry analysis of recombinant wild-type GLT25D1/COLGALT1 samples. **c** Comparison of the GLT25D1/COLGALT1 dimer observed in the crystal structure with solution SAXS data (red circles) yields a χ^2 of 7.62 (dashed line). Modeling of the N- and C-terminal flexible loops (shown as blue and spheres, respectively) yields a structural model of the dimeric enzyme (represented as a white cartoon) with improved fitting of the experimental solution SAXS data, resulting in χ^2 of 1.89 (solid line). **d** Comparison of negative stain electron microscopy single particle 2D classes (gray background) with selected orientations of computed projections of the GLT25D1/COLGALT1 crystal structure (black background) shows matching molecular projections.

The GLT25D1/COLGALT1 GT2 domain shares only distant structural homology with known glycosyltransferases, and the resulting superpositions are less clear compared those obtained for GT1 (Supplementary Fig. 5d). Likewise, when superimposed to LH3/PLOD3, the GLT25D1/COLGALT1 GT2 domain shows only limited structural similarity with the N-terminal LH3/PLOD3 GT domain, with superposition RMSD values of 4.7 Å and Z-score lower than 10 according to DALI[35] (Supplementary Fig. 5e). When subject to BLAST search, the sequence of this domain yielded more results compared to the GT1 domain, including vertebrates, invertebrates and giant viruses, but also bacteria (Supplementary Fig. 7). The presence of additional kingdoms in the BLAST alignment supports possible evolutionary ancestral gene fusion events ultimately leading to the two-domain GLT25D/COLGALT assembly.

GLT25D1/COLGALT1 is a dimeric enzyme

The asymmetric unit of the GLT25D1/COLGALT1 crystal structure is compatible with two enzyme molecules aligned in an antiparallel fashion through contacts mediated via their GT1 domain and wrapped around the unstructured N-terminal loop comprising residues 42–51 (Fig. 2a). This quaternary structure is characterized by a buried surface area of 1250 Å^2, with 22 hydrogen bonds and 5 salt bridges out of 35 residues per monomer involved in the dimeric interface. Such assembly was supported by mass photometry (MP) data, showing that GLT25D1/COLGALT1 in solution is present with a molecular weight of 140 kDa, equivalent to the expected molecular weight for enzyme dimers (Fig. 2b), with minimal dispersion. Consistent results of 131 ± 10 kDa were obtained using size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) and small-angle X-ray scattering (SEC-SAXS) (Supplementary Table 6, Supplementary Fig. 8a, b), the latter producing ab initio sphere models of elongated molecular ensembles matching the size and overall envelope shape of crystallographic GLT25D1/COLGALT1 dimers with 95% correlation (Supplementary Fig. 8c). Orthogonal comparison of the unmodified experimental crystal structure with solution SAXS data yielded initial χ^2 of 7.62, which further improved to 1.89 after modeling of flexible N- and C-termini (Fig. 2c). Furthermore, low-resolution 2D classes from a sample of GLT25D1/COLGALT1 analyzed using single particle negative staining electron microscopy revealed elongated objects highly comparable to the SEC-SAXS solution data and the crystallographic dimers (Fig. 2d).

Probing the role of GLT25D1/COLGALT1 dimerization

To evaluate the importance of dimerization on GLT25D1/COLGALT1 function, we searched for amino acid residues located at the crystallographic dimer interface: Trp158, whose side chain is embedded into a cavity shaped by Arg53, Ala85, His113, Met157 of the other enzyme monomer; and Asp160, forming a hydrogen bonding network with Arg53 from both monomers (Fig. 3a). Mutation Trp158Arg was enough to disrupt the dimer interface of GLT25D1, resulting in monomeric species in solution as assessed by mass photometry and SEC-MALS (Fig. 3b, c). When prompted for enzymatic activity using either indirect

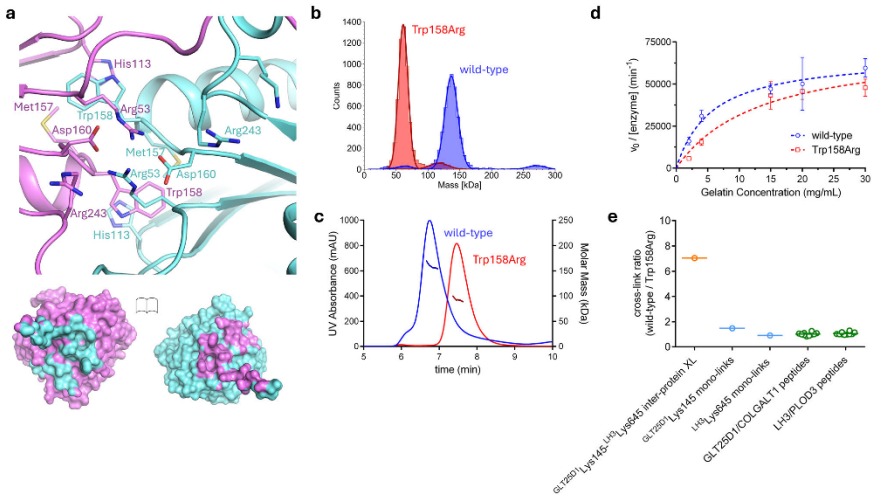


Fig. 3 | Probing GLT25D1/COLGALT1 dimeric architecture and its functional significance. a Details of the dimeric interface, showing amino acid residues involved in the direct contacts between the two monomers. The bottom panel shows an open-book rendering of the two GLT25D1/COLGALT1 monomers, with the colored footprint of the dimerization partner on the molecular surface. **b** Results of the mass photometry comparison of wild-type GLT25D1/COLGALT1 and the Trp158Arg mutant, designed to disrupt the dimer interface. **c** Results of the SEC-MALS comparison of wild-type GLT25D1/COLGALT1 and the Trp158Arg mutant, showing alterations of the elution volume as well as molar mass. **d** Comparison of the Gal-T enzymatic activity of GLT25D1/COLGALT1 wild-type and Trp158Arg mutant using luminescence. The Michaelis-Menten curves show the steady-state kinetics characterization, described as apparent turnover number (v_0/[enzyme]) as a function of gelatin substrate concentration. Quantitation of K_M and k_{cat} values for both wild-type and Trp158Arg mutant is provided in Supplementary Table 5. Error bars represent standard deviations from average of triplicate independent experiments. **e** Abundance ratios of wild-type and Trp158Arg GLT25D1/COLGALT1 ^{GLT25D1}Lys145-^{LH3}Lys645 inter-protein cross-links (orange), ^{GLT25D1}Lys145 and ^{LH3}Lys645 mono-links (cyan), and total GLT25D1/COLGALT1 and LH3/PLOD3 peptides (green) detected in the XL-MS/MS experiments. The relative abundance of LH3/PLOD3 and GLT25D1/COLGALT1 proteins was calculated from the 10 most abundant peptides per protein. Data are reported as mean. Individual peptide and cross-link abundance ratios are shown as dots. (Mean ± SD: LH3/PLOD3 1.07 ± 0.11, GLT25D1/COLGALT1 1.02 ± 0.12); median values are 1.02 for LH3/PLOD3 and 1.00 for GLT25D1/COLGALT1, with interquartile ranges of 1.01–1.12 and 0.95–1.09, respectively. The minimum and maximum values observed were 0.96–1.30 for LH3/PLOD3 and 0.83–1.27 for GLT25D1/COLGALT1.

luminescence-based detection of free UDP in the presence of gelatin, as well as direct evaluation of galactosyl-hydroxylysine modification of synthetic peptides using high-resolution mass spectrometry (HRMS), this mutant showed activities comparable to wild-type GLT25D1/COLGALT1 (Fig. 3d, Supplementary Table 5), supporting the hypothesis that enzyme dimerization does not constitute an essential requirement for the enzyme's catalytic activity in vitro.

Intrigued by this observation, we used cross-linking mass spectrometry to explore whether GLT25D1/COLGALT1 dimerization could contribute to molecular interactions with partner multifunctional lysyl hydroxylase-glucosyltransferase LH3/PLOD3. We cross-linked a 1:1 stoichiometric mixture of wild-type and Trp158Arg GLT25D1/COLGALT1 with disuccinimidyl dibutyric urea (DSBU), followed by in-solution digestion with trypsin. The resulting peptide mixtures were analyzed using nano-liquid chromatography-tandem mass spectrometry (LC-MS/MS), revealing in both cross-linked samples a clear inter-protein cross-link between Lys145 of GLT25D1/COLGALT1 (^{GLT25D1}Lys145) and Lys645 of LH3/PLOD3 (^{LH3}Lys645) (Supplementary Fig. 9). However, the abundance of this ^{GLT25D1}Lys145-^{LH3}Lys645 inter-protein cross-link was found to be approximately seven times higher using wild-type GLT25D1/COLGALT1 complex compared to the Trp158Arg mutant (Fig. 3e). Notably, this difference was not attributable to changes in protein concentration or lysine reactivity, as the overall abundance of both proteins, and the individual reactivity of ^{GLT25D1}Lys145 and ^{LH3}Lys645remained constant in both experiments (Fig. 3e). A nearly identical abundance of the ^{GLT25D1}Lys145 and ^{LH3}Lys645 mono-links, which occur when a lysine residue is covalently modified by DSBU without linkage to another residue, were measured in the two experiments (Fig. 3e). Collectively, these results show that GLT25D1/COLGALT1 dimerization critically contributes to macromolecular complex formation with multifunctional LH3/PLOD3 enzymes.

Both GT1 and GT2 domains bind donor substrates and metal ions

Analysis of the experimental electron density of GLT25D1/COLGALT1 revealed intense difference peaks in the GT1 domain of both enzyme monomers proximate to the DxD motif identified by residues Asp166 and Asp168, in a loop interconnecting strand β4 and helix α4, that could be modeled as UDP-α-Gal bound to a metal ion (Fig. 4a). The carboxylate group of Asp166 is not directly involved in the coordination of the metal ion. The donor substrate is surrounded by an extended amino acid network, characterized by a H-bonded interaction involving the side chains of Tyr126 with the C^4 carbonyl of the uridine moiety, which is trapped by a pi-Sigma interaction with Arg61 and a pi-Alkyl interaction with Val143, in a hydrophobic cavity also shaped by residues Leu59, Asp91, Arg139. The ribose moiety is kept into a well-defined conformation by Trp135, and by hydrogen bonds with the main chain carbonyl and amino groups of residues Leu59, Arg61, and Ala167. Additional interactions involve the side chains of

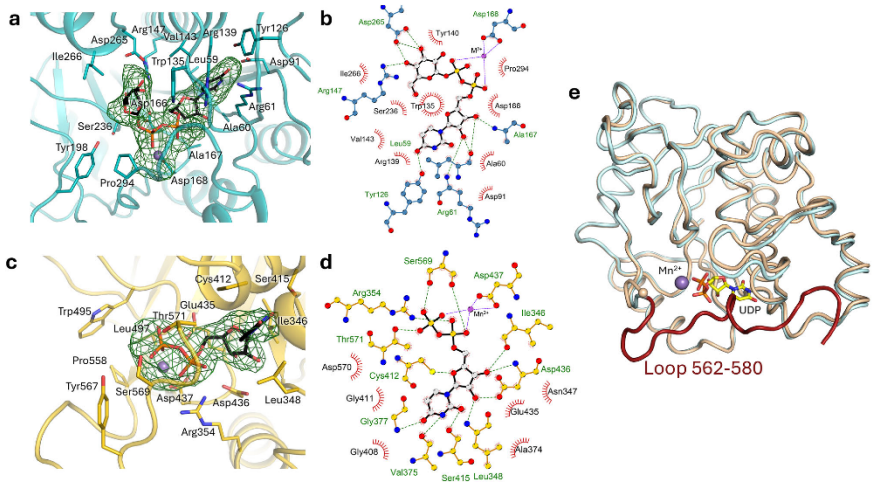


Fig. 4 | Structural insights into GLT25D1/COLGALT1 enzymatic activity and assembly. a The GT1 domain is constitutively populated by a divalent metal ion (labeled as M^{2+}, shown with a purple sphere) and UDP-α-Gal, as shown in the F_o–F_c omit electron density map (contour level 3.2 σ). Residues involved in coordination of the metal ion and interaction with the donor substrate are shown as sticks. **b** LIGPLOT+[72] diagram illustrating the interaction network for donor substrates and cofactors observed in the GT1 domain. **c** When co-crystallized with excess Mn^{2+} and UDP-α-Gal, also the electron density of the GT2 domain (F_o–F_c omit map, contour level 3.2 σ) shows presence of donor substrates and cofactors. In this domain, the galactose moiety of the donor substrate is not visible, hence only UDP has been modeled (shown as sticks). Residues involved in coordination of the metal ion and interaction with the donor substrate are shown as sticks. **d** LIGPLOT+[72] diagram illustrating the interaction network for donor substrates and cofactors observed in the GT2 domain. **e** Binding of Mn^{2+} and UDP-α-Gal induces conformational changes in the GT2 domain, by stabilizing the otherwise flexible C-terminal loop through direct interactions with the cofactor and the donor substrate. Shown is the superposition of GLT25D1 GT2 domains in substrate-free (light orange cartoon) *versus* substrate-bound (light blue cartoon) states, with highlight of the loop comprising residues 560–574 (brown), stabilized only when Mn^{2+} and UDP-α-Gal are present.

Arg147 and Asp265 with hydroxyl groups of the galactose moiety (Fig. 4a, b). Notably, Trp135 matches Trp130 in the GLT25D1/COLGALT1 mouse protein sequence, whose mutation (named *fosse*) has been associated to skeletal and muscular defects due to severely reduced expression of the mutated gene[22]. When co-crystallized with Mn^{2+} and UDP-α-Gal, the crystal structure of GLT25D1/COLGALT1 unexpectedly revealed electron density matching cofactors and donor substrates also in the GT2 domain of one of the two monomers in the asymmetric unit (Fig. 4c). In this domain the metal ion cofactor was found bound to a Glu-Asp-Asp motif adopting a conformation topologically equivalent to that found in Mn^{2+}-dependent glycosyltransferases, with the carboxyl groups of Glu435 and Asp437 directly chelating the metal ion, while the side chain of Asp436 was interacting through H-bonded interactions with the hydroxyl groups of the ribose moiety of the cofactor (Fig. 4c, d). The uridine ring of the donor substrate was constrained by Gly408 and Gly411 of helix α10, in a hydrophobic groove defined by Cys412 of the same helix, plus Leu348 and Ala374 on the opposite side (Fig. 4c, d). No electron density could be observed for the galactose moiety, suggesting multiple conformations for the sugar possibly due to the absence of acceptor substrates in the wide crevice defined by residues Trp495, Leu497, Tyr567, and Pro558 (Fig. 4c, d).

Comparison of GT2 domains with and without donor substrates highlighted a conformational rearrangement of the otherwise flexible C-terminus of the domain for residues 562–580, wrapping around the donor substrate binding cavity and positioning the backbone carbonyl of residue Ser569 and Thr571 in a productive conformation for hydrogen bonding with both phosphates of the UDP donor substrate (Fig. 4e). The presence of donor substrates and cofactors only in one of the two crystallographic monomers could be explained with the different crystal packing network surrounding the GLT25D1/COLGALT1 molecules, which prevents the rearrangement of the GT2 C-terminal loop in one of the two molecules due to crystal contacts (Supplementary Fig. 10).

Mutations in the GT1 domain impact on enzyme folding stability

Structure-guided point mutants designed to interfere with metal ion or UDP-α-Gal binding in the GT1 domain (Fig. 4a, b) such as Asp166Ala (abolishing interaction with the bound metal ion), Asp265Ala (removing H-bonding contacts with the galactose moiety), Ile266Gln (preventing rigid positioning of the galactose moiety in the GT1 cavity), did not produce any recombinant enzyme samples suitable for biochemical studies, and were consistent with previously published attempts describing a Asp166Ala-Asp168Ala double mutant, designed to abolish the GT1 DxD motif[34]. We also attempted production of a Trp135Arg mutant, matching the homologous Trp130Arg *fosse* mutation in mouse[22]. The recombinant production yields for this mutant were significantly lower compared to wild-type GLT25D1/COLGALT1, consistent with the observed reduced expression levels reported for the *fosse* mutation[22]. SDS-PAGE analysis showed significant differences compared to wild-type GLT25D1/COLGALT1, suggesting severe folding stability issues (Supplementary Fig. 1). Control mutants such as Ile267Gln (in close proximity to Ile266, but not affecting UDP-α-Gal binding or positioning) instead resulted in well folded and active enzyme samples (Supplementary Fig. 1), matching the previously published Pro292Asn mutant, originally designed to generate a

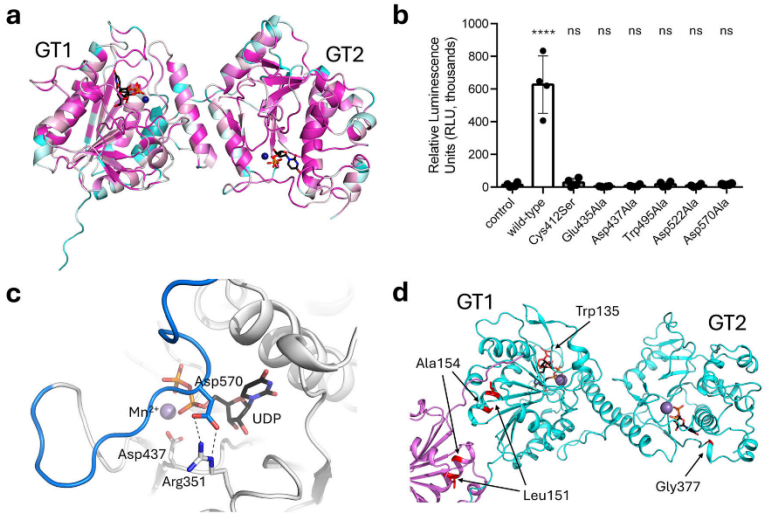


Fig. 5 | Structure-guided interpretation of GLT25D1/COLGALT1 functional features. a Amino acid conservation diagram for different GLT25D/COLGALT homologs. The colors represent the different degree of amino acid conservation from dark blue (not conserved at all) to dark purple (fully conserved). The complete list used for the representation and the associated alignment is shown in Supplementary Figs. 6 and 7. **b** Luminescence-based comparison of the Gal-T enzymatic activity of wild-type GLT25D1/COLGALT1 and mutants designed to interfere with cofactor and substrate binding in the GT2 domain. The histograms show the average result of triplicate independent measurements (whose results are individually shown as overlayed dots). Error bars represent standard deviations from average of four independent experiments; Statistical evaluations based on pair sample comparisons between control and assay values using Student's t-test. ****P-value < 0.0001. **c** A salt bridge involving Arg351 and Asp570 from the flexible C-terminal loop (blue) shapes the GLT25D1/COLGALT1 GT2 catalytic site when enzyme cofactors are bound. Mn^{2+} is shown as a purple sphere coordinated to Glu437. UDP as well as amino acid side chains directly involved in the described interactions are shown as sticks. **d** Mapping of the sites affected by known pathogenic mutations on the GLT25D1/COLGALT1 structure. Residues found mutated in the homologous *fosse* mouse phenotype (Trp135) and in cerebral small vessel disease (Leu151Arg, Ala154Pro, Gly377Arg) are shown in red and highlighted by arrows.

chimeric GLT25D1-CERCAM inactive galactosyltransferase, but still retaining enzymatic activity[34].

The GT2 domain is responsible for the catalytic activity

Our experimental structure investigations revealed that the GT2 domain of GLT25D1/COLGALT1 binds metal ions and donor substrates transiently compared to GT1, inducing conformational changes of the extended C-terminus of the enzyme that could have an impact on catalysis (Fig. 4). The identification of an unusual EDD motif rather than the canonical DxD motif coordinating the metal ion cofactor within this domain, together with the presence of a UDP-α-Gal donor substrate with a flexible sugar moiety (Fig. 4c, d), may ultimately provide insights for the identification of the enzyme's catalytic site. Conservation maps for the amino acid residues shaping the two putative ligand binding sites in homologous vertebrate enzymes highlighted that the amino acids involved in metal ion and UDP-α-Gal binding are highly conserved in both GT1 and GT2 domains (Fig. 5a). Prompted by this observation, and by our insights suggesting structural roles for the enzyme's N-terminal GT1 domain, we generated a series of point mutations affecting metal ion and donor substrate binding in the GT2 domain, as well as residues potentially involved in acceptor substrate binding during transfer of the galactose moiety. All mutants generated could be purified and showed biophysical features compatible with well-folded proteins (Supplementary Fig. 1). Mutants Glu435Ala and Asp437Ala, within the EDD sequence responsible for metal ion coordination, were found completely inactive in both luminescence and HRMS assays (Fig. 5b, Supplementary Fig. 3, Supplementary Table 5). Similar results could be obtained with mutant Cys412Ser, affecting productive positioning of the uridine-ribose moieties in a hydrophobic cavity (Fig. 5b, Supplementary Fig. 3, Supplementary Table 5). Inspired by the potential roles found for aromatic and charged residues located at precise positions in the catalytic cavity of inverting glycosyltransferase enzymes[37,38], we also generated mutants Trp495Ala and Asp522Ala, which both resulted in complete loss of GLT25D1/COLGALT1 Gal-T activity (Fig. 5b, Supplementary Fig. 3, Supplementary Table 5).

The conformational rearrangement of the GT2 C-terminus suggested a potential role also for Arg351, rearranging its side chain to form a hydrogen bond with Asp570 when cofactors and donor substrates are bound (Fig. 5c). We therefore generated mutants Arg351Ala and Asp570Ala. While production of mutant Arg351Ala resulted in poor quality enzyme preparations and was not suitable for subsequent biochemical studies, production of mutant GLT25D1/COLGALT1 mutant Asp570Ala was successful (Supplementary Fig. 1). Consistent with our hypothesis, this mutant was inactive (Fig. 5b, Supplementary Table 5).

The structural data obtained also provided a rationale for previously investigated mutations: Asp336, located at the C-terminus of the linker segment connecting GT1 and GT2 domains, whose hydrogen bond with GT2 Lys445 does not seem to be dramatically impacted when being mutated into a serine residue; and residues Asp461 and Asp433, the latter being involved in a hydrogen bonding network with

the side chains of Lys508 and His547, likely critical for GT2 domain folding stability. Taken together, our molecular structures and site-directed mutagenesis results enabled the unambiguous identification of GLT25D1/COLGALT1 catalytic site and its unusual features.

Molecular significance of GLT25D1/COLGALT1 pathogenic mutations

The molecular structure and the biochemical characterizations of several GLT25D1/COLGALT1 mutants enable correlating the possible impact of disease-causing mutations associated with this enzyme. The *fosse* mutation Trp130Arg, associated to a mouse phenotype with developmental defects, localizes on Trp135 of the human ortholog. Trp135 shapes the UDP-α-Gal binding cavity in the GT1 domain of the enzyme by interlocking its side chain between the uridine and the galactose moieties of the cofactor (Fig. 4a). We evaluated the impact of this mutation experimentally and observed significantly reduced expression levels for the recombinant enzyme variant in vitro, as well as stability issues that hampered subsequent biochemical characterizations (Supplementary Fig. 1a). Mapping of pathogenic mutations associated to cerebral small vessel disease[39,40] on the GLT25D1/COLGALT1 structure provided a rationale for the significance of these pathogenic enzyme variants (Fig. 5d). Mutations Leu151Arg and Ala154Pro localize at the C-terminus of the α3 helix in the GT1 domain. The hydrophobic side chains of Leu151 is buried into a hydrophobic pocket, therefore the introduction of a bulky Arg side chain at this position would likely induce severe perturbations to the stability of the GT1 domain. Likewise, replacing Ala154 with a Pro residue would likely impact on the structural stability of helix α3 near the homo-dimer interface. The possible effect of the Gly377Arg mutation[39] could be inferred by Gly377 shaping an alpha turn segment connecting the strand β14 with helix α9 in close proximity to the binding site for the uridine ring of the catalytic UDP-α-Gal donor substrate. Introduction of an Arg side chain at this position may severely perturb GT2 domain conformational stability, resulting in loss of function.

Mn^{2+} binds transiently in the GT2 domain during catalysis

Intrigued by the observation of a dimeric ensemble potentially characterized by four aligned catalytic sites, we wondered whether metal ions and donor substrates could possibly have structural, rather than functional roles. We therefore purified recombinant GLT25D1/COLGALT1 without any cofactor supplementation and attempted to characterize the possible bound metal ions. DSF experiments performed by incubating the enzyme with metal ion solutions showed significant stabilization of GLT25D1/COLGALT1 only in the presence of Mn^{2+} (Fig. 6a), supporting the biochemical data indicating Mn^{2+} as the metal ion cofactor transiently bound during catalysis. Comparison with DSF data collected on mutants Glu435Ala and Asp437Ala incubated with Mn^{2+} did not result in stabilization (Fig. 6a), consistent with the structural data showing transient binding of this catalytic metal ion in the GT2 domain (Fig. 4).

Roles of Ca^{2+} and UDP-α-Gal donor substrate in the GT1 domain

Parallel to Mn^{2+} binding experiments, we tested the effect of removal of the tightly bound metal ion in the GT1 domain by incubating GLT25D1/COLGALT1 with different concentrations of ethylenediaminetetraacetic Acid (EDTA). We consistently observed destabilization, with a differential scanning fluorimetry (DSF) thermal shift of 3.0–3.5 °C compared to untreated samples (Fig. 6b). We observed nearly identical destabilization effects by treating GLT25D1/COLGALT1 mutants Glu435Ala and Asp437Ala with EDTA (Fig. 6b). As these mutants were designed to interfere with transiently bound metal ions in the GT2 domain, we concluded that EDTA treatment impacts on protein stability by stripping the metal ion bound the GT1 domain. Nevertheless, EDTA-treated samples still showed DSF profiles compatible with a partially folded protein sample, thus we wondered whether removal of metal ions through EDTA treatment could impact on the enzyme's conformational stability. EDTA-treated wild-type GLT25D1/COLGALT1 samples were therefore subject to mass photometry analysis and size exclusion chromatography. While mass photometry did not show significant alterations in the oligomeric state of the enzyme upon EDTA treatment (Fig. 6c), stripping of the metal ion altered the elution profile of the enzyme, compatible with alterations of the enzyme's hydrodynamic radius (Fig. 6d). Surprisingly, ICP-MS experiments aimed at identifying Mn bound to enzyme in the GT1 domain did not reveal the presence of this transition metal ion. We therefore decided to further investigate the ligands consistently bound in the GLT25D1/COLGALT1 GT1 domain. Native mass spectrometry (Native-MS) analysis of recombinant wild-type and Trp158Arg GLT25D1/COLGALT1 samples under controlled partially unfolding conditions revealed the coexistence of free- and ligand bound-monomers, leading to the identification of Ca^{2+} as the most probable bound metal ion in the GT1 domain (Fig. 6e); ICP-MS confirmed the identification, indicating a 1:1 Ca^{2+}:GLT25D1/COLGALT1 stoichiometry. Prompted by this unexpected finding, we performed DSF analysis of GLT25D1/COLGALT1 samples in the presence of the highly selective Ca^{2+} chelating agent ethyleneglycol-*bis*(β-aminoethyl)-N,N,N′,N′-tetraacetic acid (EGTA), obtaining a folding destabilization profile matching that observed using EDTA (Fig. 6b).

Taken together, these data highlight the presence of Ca^{2+} bound in the GLT25D1/COLGALT1 GT1 domain, supporting structural roles for this metal ion cofactor and the UDP-α-Gal donor substrate in this domain. The absence of direct functional roles during catalysis for the GT1 domain is consistent with previous hypotheses derived from chimeric studies of GLT25D1/COLGALT1 and the CERCAM inactive paralog[34], and is further supported by the absence of the essential negatively charged amino acid side chains near the galactose moiety of the UDP-α-Gal donor substrate (Fig. 5a, b) required for the enzymatic activity of inverting GT-A glycosyltransferases[47,58].

MD simulations support distinct roles for GT1 and GT2 domains

To better characterize the molecular features observed in GLT25D1/COLGALT1 GT1 and GT2 domains, we carried out a comparative molecular dynamics study of the individual domains in the presence or in absence of metal ions and donor substrates. Since the focus of this study is to investigate how UDP-Gal ligand and metal ion cofactor binding influence the structural stabilization of these domains rather than to observe conformational transitions, conventional MD simulations are sufficient to capture the relevant dynamics. To this end, three independent 1-microsecond all-atom MD simulations per system in explicit solvent were perfomed, yielding sufficient accuracy to describe atomic interactions and solvation effects on nanosecond to sub-microsecond timescales, typical of domain stabilization and conformational flexibility. To assess the impact of cofactors binding on GT1 and GT2 stability and dynamics, analyses focused on conformational ensembles rather than time-dependent trajectories, to ensure a statistically robust characterization of the different simulated states. Equilibration and independence from the initial configurations were evaluated by analyzing the root mean square deviations of the protein backbone in each independent simulation replicate separately. After verifying the convergence, metatrajectories were built by concatenating the equilibrated part of the three replicates, on which ensemble-based analyses were performed. Further details on equilibration and the assessment of initial configurations independence are provided in the Methods section and in the Supplementary Methods. Simulations indicated that no large conformational changes occur in the GT1 and GT2 domains upon ligand binding. However, analyses of internal dynamics reveal that the binding of donor substrates and metal ions in the GT1 domain significantly stabilizes its structure, while a less pronounced effect is observed for the GT2 domain. Principal component analysis (PCA) was used to identify the dominant modes of

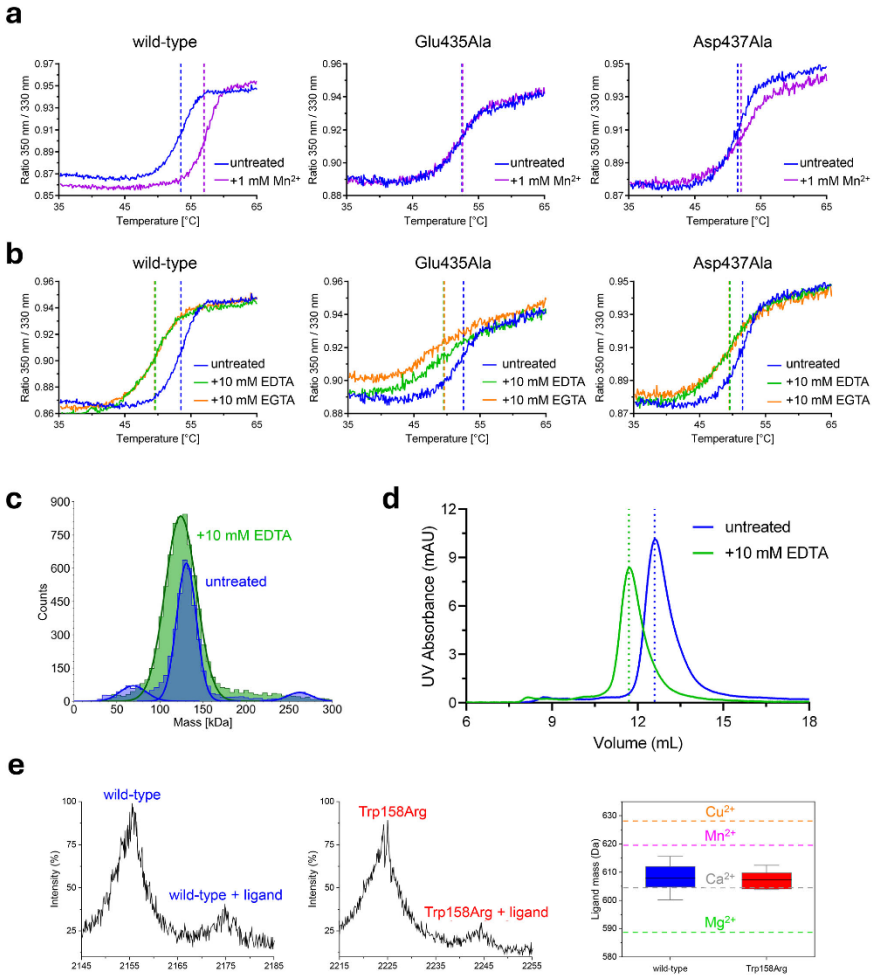


Fig. 6 | Probing the structural role of metal ions and cofactors in GLT25D1/COLGALT1 GT1 and GT2 domains. a DSF analysis of wild-type GLT25D1 (left) incubated with Mn^{2+} (purple trace) shows significant stabilization induced by binding of the metal ion when compared to untreated samples (blue trace); removal of amino acid side chains critical for binding of metal ions in the GT2 domain such as Glu435 (center) and Asp437 (right) results in loss of the metal ion-induced stabilization effect. **b** DSF analysis of GLT25D1/COLGALT1 subject to EDTA (green trace) and EGTA (orange trace) treatment, showing the destabilization induced by metal ion chelation when compared to untreated samples (blue trace); Comparison between wild type (left) and mutants in the GT2 domain highlight very similar destabilization effects. **c** EDTA-treated samples show alterations in dispersity when tested using mass photometry. **d** EDTA-treated samples show altered elution profiles when tested using size-exclusion chromatography. **e** Native-MS spectra under partially-denaturing conditions of wild-type 30+ charge state (left) and Trp158Arg 31+ charge state (center), with the signals corresponding to the free protein and protein-ligand complex labeled; box plot (bounds: percentile 25–75%; mean value: black line; whiskers: maxima-minima) of ligand mass calculation obtained from the entire charge state distribution ($n = 6$ for wild-type, $n = 7$ for Trp158Arg) of Native-MS spectra (right); reference lines report the theoretical ligand mass as a function of distinct divalent cation cofactors.

motion in simulations by capturing directions of the largest variance in the data. More specifically, for each domain, GT1 and GT2, the bound and unbound trajectories were combined and subjected to PCA. The first two principal components (PC1 and PC2), representing the majority of conformational variance were calculated, and snapshots of the bound and unbound trajectories were projected onto these components. Histograms of normalized projection frequencies provided a visual representation of the conformational space explored by each system, highlighting their dynamic differences (Supplementary Fig. 12a–d). For the GT1 domain, projections of bound and unbound conformations onto PC1 exhibit overlapping distributions, indicating access to similar conformational spaces. However, the unbound state displays broader and less pronounced peaks, reflecting greater conformational diversity and flexibility. In contrast, the bound state shows sharper and higher peaks, suggesting that cofactor binding restricts the GT1 domain to a narrower set of conformations (Supplementary Fig. 12a). Similarly, along PC2, the bound state maintains a more confined distribution compared to the broader, multi-peaked profile of the unbound state (Supplementary Fig. 12b). For the GT2 domain, projections of bound and unbound conformations along PC1 demonstrate that the unbound state predominantly samples a single conformational state, indicated by a single peak centered at zero, reflecting minimal movement and inherent structural stability. In contrast, the bound state exhibits a bimodal distribution with one peak overlapping the unbound state and a second, sharper peak, suggesting that cofactors binding induces an additional conformation (Supplementary Fig. 12c). Along PC2, the unbound state shows a narrow distribution centered around zero, whereas the bound state shows a slight shift toward negative values, indicating subtle conformational adjustments upon cofactors binding (Supplementary Fig. 12d). These results suggest that binding of metal ions and donor substrates contribute to the overall GT1 domain stabilization by limiting its structural dynamics. In contrast, the GT2 domain remains structurally stable also in the unbound state, while cofactor binding induces an additional conformational state, resulting in minor alterations to its internal dynamics without causing large-scale structural changes.

Hydrogen bond persistence analysis along the simulation time corroborates the PCA results. In the GT1 domain, the bound state exhibits more persistent hydrogen bonds (60–90% of the simulation time) compared to the unbound state (<60%) (Supplementary Fig. 12e, f), indicating greater rigidity and stability. Conversely, the GT2 domain shows minimal differences in hydrogen bond persistence between bound and unbound states, maintaining a dynamic hydrogen bond network with only slight variations (Supplementary Fig. 12g, h). This result suggests that cofactor binding has limited impact on the structural dynamics of the GT2 domain.

Distance fluctuation (DF) analysis, which characterizes the coordinated internal dynamics of the GT1 and GT2 domains, further supports these findings. In the GT1 domain unbound state, the loop hosting residues 130–140, the C-terminal portion of helix a7 of the GT1 domain and the linker between GT1 and GT2 domains are highly dynamic and less restricted to a well-defined conformation, losing their coordination with the rest of the domain, (Supplementary Fig. 13a). Ligand binding enhances internal coordination of these regions, restricting their dynamics to more stable conformations. In contrast, the GT2 domain shows minimal differences in DF between bound and unbound states, maintaining similar dynamic coordination patterns with only slight variations in flexible regions (Supplementary Fig. 13b). These observations suggest that cofactor binding increases the internal coordination of the GT1 domain while exerting a less pronounced effect on the GT2 domain. In particular, without the stabilizing effect induced by donor substrate and metal ion, the loop hosting residues 130–140, the C-terminal portion of helix a7 of the GT1 domain and the linker between GT1 and GT2 domains become highly dynamic and less restricted to a well-defined conformation (Supplementary Fig. 13c). GT1 and GT2 domains respond differently to ligand and cofactor binding within the simulated timescales, confirming that the relevant structural stabilization processes are well-characterized by the simulations performed.

Overall results from MD simulations and analysis support the hypothesis that cofactor binding significantly enhances the structural stability of the GT1 domain while exerting a less pronounced effect on the structural dynamics of the GT2 domain.

Discussion

Collagen hydroxylysine galactosylation is a fundamental step in the complex post-translational modification pathway that leads to Glc-Gal-Hyl, the most abundant O-linked glycosylation found in the animal kingdom[12,13,16]. Following decades of intensive biochemical investigations addressing the enzymes involved in this process and their associated functions, recent years have witnessed increasing efforts to produce a comprehensive molecular description of these reactions and the actors involved, with particular attention to multifunctional LH/PLOD lysyl hydroxylases-glucosyltransferases[17,41]. However, the lack of experimental structure data about GLT25D/COLGALT galactosyltransferases has limited our understanding of the molecular mechanisms of these glycosyltransferases, despite previous significant efforts[16,28,34]. In this work, we have determined the experimental structure of full-length human GLT25D1/COLGALT1, and used a combination of biochemistry, biophysics and site-directed mutagenesis to characterize its functionality in vitro.

Our results reveal an elongated, multi-domain dimeric quaternary structure characterized by two Rossmann fold-type domains per monomer (named GT1 and GT2, respectively) interconnected through a rigid linker region, and with dimer contacts wrapping around the N-terminus of each monomer (Fig. 1b, c). Surprisingly, both GT1 and GT2 domains were found indispensable for catalysis and were capable of binding metal ions and UDP-α-Gal donor substrate (Fig. 4), with the ligand binding regions of both domains showing the highest degree of sequence conservation (Supplementary Figs. 6 and 7). The N-terminal domain GT1 is closely related to the inactive central accessory (AC) domain of LH/PLOD enzymes and shares features more similar those typical of inverting GT-A glycosyltransferases, whereas the GT2 domain features a distinctive topology not matched by well-characterized glycosyltransferase enzymes.

Previous work addressing the putative functional roles of GLT25D1/COLGALT1 DxD motifs showed that point mutations affecting residues involved in metal ion coordination Asp166 and Asp168 in the GT1 domain completely abolished the activity[34]. The same work also indicated that replacing the entire N-terminus of the enzyme with that of inactive CERCAM did not impact on the ability of the enzyme to glycosylate collagen substrates[34]. These results are consistent with the experimental identification of Ca^{2+} and UDP-α-Gal bound in the GT1 domain of the enzyme (Fig. 6b–e). Contrary to the frequent presence of Mn^{2+} and Mg^{2+} metal ion cofactors[37,38], we could not find evidence of other Rossman fold GT-A glycosyltransferases with bound Ca^{2+} in their catalytic site. Combined with the absence of negatively charged amino acid residues at key positions essential for inverting glycosyltransferase catalysis[37,38], the presence of a stably bound Ca^{2+} ion interacting with UDP-α-Gal rules out direct catalytic roles for the GLT25D1/COLGALT1 GT1 domain. In addition, MD simulations performed on ligand-free and ligand-bound GLT25D1/COLGALT1 GT1 domain further supported the structural roles for the bound metal ion and donor substrate, highlighting increased molecular flexibility around residues 130–140 and 310–330 of this domain in absence of metal ions and UDP-α-Gal (Supplementary Figs. 12 and 13). Notably, these regions encompass residues proximate to UDP-α-Gal binding such as Trp135, matching the site affected by the loss-of-function mouse *fosse* mutation[22] and the main contact site of the GT1 domain with neighbor GT2, where mutations associated to cerebral small vessel disease can

be mapped (Fig. 5d). Increased molecular flexibility in this portion of the GLT25D1/COLGALT1 polypeptide may significantly perturb the globular organization of this multi-domain enzyme. Such hypothesis is supported by our data showing that after EDTA/EGTA treatment, GLT25D1/COLGALT1 is characterized by folding instability and conformational alterations in solution without significant changes in the enzyme's oligomeric state (Fig. 6b–d).

The GT2 domain displays amino acid sequence conservation for the residues surrounding the site transiently occupied by the Mn^{2+} cofactor and UDP-α-Gal during catalysis (Fig. 5a, Supplementary Fig. 7). Structural investigations proved essential to identify the GLT25D1/COLGALT1 catalytic site. Previous attempts to map the catalytic residues critical for GLT25D1/COLGALT1 enzymatic activity were not conclusive, likely due to the unusual EDD motif that we found responsible for binding Mn^{2+} in the GT2 domain, and the multiple DxD motifs within the enzyme polypeptide sequence (Supplementary Figs. 6 and 7). To ensure correct mapping of the GT2 catalytic network, we designed several point mutations surrounding the Mn^{2+} cofactor and the donor substrate observed in the crystal structures and found an extensive network of charged and hydrophobic side chains critically involved in UDP-α-Gal substrate recognition, activation, and catalysis (Fig. 5b, c).

Overall, substrate-free and substrate-bound structures for the GT2 catalytic cavity are almost identical (Fig. 4), except for the conformational rearrangement of the otherwise flexible C-terminal segment of the domain. This region is characterized by negatively charged amino acids, whose direct involvement in binding and processing of the donor substrate UDP-α-Gal has been demonstrated by our site-directed mutagenesis experiments. In this respect, our insights also support previous in silico hypotheses[42].

The observed dimeric quaternary structure, confirmed by multiple orthogonal experiments in solution and by site-directed mutagenesis, does not appear to be a requirement for GLT25D1/COLGALT1 enzymatic activity in vitro (Fig. 3). Using cross-linking mass spectrometry we found that the dimeric architecture of GLT25D1/COLGALT1 shapes a direct molecular contact site for GLT25D1/COLGALT1-LH3/PLOD3 molecular interactions, providing experimental evidence to previous hypotheses regarding formation of multimeric biosynthetic hetero-complexes capable of fully converting collagen Lys residues into Glc-Gal-Hyl[29,31]. These results are consistent with the recently reported cryo-electron microscopy structure of a GLT25D1/COLGALT1-LH3/PLOD3 complex[43]. Notably, our biophysical data consistently indicate that a small fraction of our recombinant enzyme preparations is dispersed, possibly suggesting the need for further stabilizing homo- or hetero- protein-protein interactions with physiological relevance in vivo.

The elongated shape of the GLT25D1/COLGALT1 dimer resembles that of the tail-to-tail dimer observed for LH3/PLOD3, possibly facilitating extensive interactions among the two enzymes responsible for the subsequent processing of collagen lysines into Glc-Gal-Hyl and posing questions regarding the possible evolution of this two-enzyme system from a common ancestor. This is further supported by the structural similarity observed between the two non-functional domains of these enzymes (Supplementary Fig. 5c): the GLT25D1/COLGALT1 GT1 domain, unexpectedly found bound to Ca^{2+} and hosting UDP-α-Gal but only to preserve the conformational integrity of the domain, and the LH/PLOD AC domain, which cannot bind metal ions and/or co-substrates[41]. Such hypothesis prompts for extensive future work, also considering the unusual trafficking of multifunctional LH/PLOD enzymes through the secretory pathway due to absence of specific ER retention sequences[44–47].

Methods

Chemicals

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

DNA constructs

The sequence encoding for human GLT25D1/COLGALT1, deprived of the N-terminal signal peptide and the C-terminal RDEL retention sequence (UniProt Q8NBJ5, residues 30 to 618), was synthesized by Genewiz and sub-cloned into a pCR8 cloning vector with in-frame 5′-BamHI and 3′-NotI restriction sites. GLT25D1/COLGALT1 mutants were generated using the Phusion Site Directed Mutagenesis Kit (Thermo Fisher Scientific). The entire plasmid was amplified using the primers listed in Supplementary Table 1. The linear mutagenized plasmids were then phosphorylated using T4 polynucleotide kinase (Invitrogen) prior to ligation. All plasmids were checked by Sanger sequencing (Microsynth) prior to cloning into a modified pET28b-SUMO vector (Novagen). The recombinant plasmids include an N-terminal 8xHis-tag followed by a small ubiquitin-like modifier (SUMO) tag preceding the in-frame 5′-BamHI restriction site, as well as an in-frame stop codon after the 3′-NotI restriction site.

Production of recombinant GLT25D1/COLGALT1

Chemically competent *E. coli* BL21 cells (Invitrogen) were transformed with the pET28b-SUMO-GLT25D1 recombinant plasmid using heat shock. A single colony was inoculated into Lysogeny broth (LB) medium supplemented with 100 μg ml^{-1} of kanamycin and grown overnight at 37 °C in a shaking incubator (New Brunswick). This pre-culture was then inoculated at a 1:50 ratio in 1 L of ZYP5052 autoinducing medium[48] in a 5 L Erlenmeyer flask. The culture was kept at 37 °C for 3 h at 180 r.p.m. shaking speed. The temperature was then lowered to 17 °C, and further incubated overnight. Bacterial cells were harvested by centrifugation at 5000 × *g*, resuspended at 1:10 *w/v* ratio in a lysis buffer composed of 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)/NaOH, 500 mM NaCl, 10 μM leupeptin, 10 μM pepstatin, 0.3 mg ml^{-1} chicken egg white lysozyme, 500 μM $MnSO_4$, pH 8.0. The suspension was subject to sonication (16 cycles, 9 s on, 6 s off pulses). Cell debris was removed by centrifuging the cell lysate at 60,000 × *g* for 45 min, at 4 °C in an Avanti J26 super centrifuge (Beckman Coulter). The supernatant containing soluble His-SUMO-GLT25D1/COLGALT1 protein was filtered through a MiniSart GF 0.8 μm filter (Sartorius), loaded onto a 5 mL Ni Sepharose Excel column (Cytiva) and eluted using a stepwise gradient of an elution buffer composed of 25 mM HEPES/NaOH, 500 mM NaCl, 500 mM imidazole, pH 8.0. The fractions of the eluate containing His-SUMO-GLT25D1/COLGALT1, as assessed by SDS-PAGE analysis, were pooled and dialyzed overnight against 2 L of 25 mM HEPES/NaOH, 500 mM NaCl, pH 8.0. The N-terminal 8xHis-SUMO tag was simultaneously cleaved by incubating the protein with 1 mg mL^{-1} His-tagged SUMO protease (1:300 *v/v*). Affinity-based removal of SUMO protease and 8xHis-SUMO-tag was achieved by passing the purified sample through a 5 mL Ni Sepharose Excel column (Cytiva) and collecting the flow-through fraction. The resulting sample was subject to buffer exchange in 30 kDa MWCO Vivaspin Turbo centrifugal filters (Sartorius) against 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0 and loaded into a HiScreen Capto Q column (Cytiva) pre-equilibrated with the same buffer. A linear NaCl gradient was then applied, resulting in GLT25D1/COLGALT1 elution with approximately 250 mM NaCl. The sample was concentrated using 30 kDa MWCO Vivaspin Turbo centrifugal filters (Sartorius) and subject to final polishing into a Superdex 200 Increase 10/300 GL (Cytiva), equilibrated with 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0. The final sample quality was assessed by reducing and non-reducing SDS-PAGE analysis. GLT25D1/COLGALT1-containing fractions were pooled, concentrated to 4 mg mL^{-1}, and stored at −80 °C prior to usage. Results of SDS-PAGE sample quality controls for all proteins used in this study are shown in Supplementary Fig. 1.

Protein crystallization

GLT25D1/COLGALT1 spherulites were initially found in various conditions of nanoliter-dispensed droplets (0.1 μL protein at

4 mg mL^{-1} + 0.1 µL reservoirs), dispensed using a Gryphon crystallization robot (Art Robbins) using commercial crystallization screens in sitting-drop vapor diffusion drop plates (SwissSci). In particular, small microcrystals were initially identified after few days at 4 °C in conditions of the Morpheus crystallization screen[49] (Molecular Dimensions), indicating a preference for poly-ethylene-glycol (PEG) as precipitant and pH between 6.0 and 7.0. Crystal optimization was carried out by systematically varying the precipitating agent, buffering agent, pH and additives composition of the initial hits from the commercial screen. Diffraction-quality crystals were eventually obtained by manually mixing 0.5 µL of protein concentrated at 3.5 mg mL^{-1} and 0.5 µL of an optimized reservoir solution composed of 8% PEG MW 4000, 100 mM 2-(N-morpholino)ethanesulfonic acid (MES)/NaOH, 20% glycerol, pH 6.5. Co-crystallization experiments were performed by setting up the same conditions and supplementing the protein solution with 500 µM $MnCl_2$ and 1 mM UDP-α-Gal. Crystals were harvested using mounted Litholoops (Molecular Dimensions), flash-cooled and stored in liquid nitrogen prior to data acquisition. Heavy atom derivatives were prepared by soaking the GLT25D1/COLGALT1 crystals in mother liquor conditions containing 1 mM K_2HgBr_4. Crystals were incubated with the heavy atom solution for at least 5 h at 4 °C prior to cryoprotection, harvesting and flash-cooling in liquid nitrogen.

X-ray data collection, structure determination and refinement

X-ray diffraction data from single crystals were collected at 100 K at various beamlines of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France and the Swiss Light Source (SLS) in Villigen, Switzerland. Experimental phasing was achieved by collecting X-ray diffraction data at the mercury inflection point at the ID23-1 beamline of the ESRF synchrotron equipped with an Eiger 2 16 M (Dectris) detector. Data were indexed and integrated with *XDS*[50], followed by scaling and merging using *AIMLESS*[51]. Data collection statistics are given in Supplementary Table 2. Single wavelength anomalous dispersion (SAD) phasing was performed with the *SHELXC/D/E* pipeline and *HKL2MAP*[52]. A total of 13 heavy atom sites were identified at 4.0 Å resolution in *SHELXD*. Phasing with *SHELXE* resulted in non-ambiguous identification of the map handedness immediately after the initial rounds of density modification with phase extension to 2.80 Å, with 62% solvent content. The density modified experimental electron density map was used for tracing the initial model by combining multiple rounds of *ARP/wARP* and *BUCCANEER*[53]. The model was completed and the structure was refined by alternating steps of manual building in *COOT* and automated refinement with *phenix.refine*[54] and *REFMACS*[55]. Crystal structures of GLT25D1/COLGALT1 in complex with donor substrates were subsequently determined using molecular replacement in *PHASER*, using the experimental structure of the enzyme as search model. Validation was carried out using *MolProbity* and the validation tools available on the Protein Data Bank server[56]. Final refinement statistics are summarized in Supplementary Table 3. Structural figures were prepared with *PyMol* (http://www.pymol.org). Superpositions were performed using the "*super*" command in *PyMol*. All-atom root mean square deviation (RMSD) values were computed accordingly.

Mass photometry

Mass photometry measurements were carried out on a Refeyn Two Mass Photometer (Refeyn) using 24 × 50 mm^2 glass coverslips (Refeyn). Initial 200 nM stocks of purified human GLT25D1/COLGALT1 wild-type and mutants were diluted 1:10 using 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0, and analyzed under a field of view of 4 µm × 11 µm at a frame rate of 500 Hz. The data were collected using *AcquireMP* (Refeyn), then processed and plotted using *DiscoverMP* (Refeyn). Results of the data analysis are summarized in Supplementary Table 4.

Size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS)

30 µl of 4 mg mL^{-1} recombinant GLT25D1/COLGALT1 were injected into a Protein KW-802.5 analytical size-exclusion column (Shodex) and separated with a flow rate of 1 mL min^{-1} in phosphate buffer saline using a Prominence high-pressure liquid chromatography (HPLC) system (Shimadzu). For molecular weight characterization, light scattering was measured with a miniDAWN multi-angle light scattering detector (Wyatt), connected to a RID-20A differential refractive index detector (Shimadzu) for quantitation of the total mass and to a SPD-20A UV detector (Shimadzu) for evaluation of the sole protein content. Chromatograms were collected and analyzed using the *ASTRA7* software (Wyatt), using an estimated *dn/dc* value of 0.185 ml/g. The calibration of the instrument was verified by injection of 10 µL of 2.5 mg mL^{-1} monomeric BSA.

Differential scanning fluorimetry (DSF)

DSF assays on recombinant GLT25D1/COLGALT1 samples (wild-type and mutants) at a concentration of 1 mg mL^{-1} in 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0 were performed using a Tycho NT.6 instrument (NanoTemper Technologies GmbH). Data were analyzed and plotted using *Graphpad Prism 7*[57].

Evaluation of Gal-T activity through luminescence

Reaction mixtures (5 µL total volume in 25 mM HEPES/NaOH, 100 mM NaCl, pH 8.0) were prepared by initially incubating GLT25D1/COLGALT1 (final concentration 1 mg mL^{-1}) at 37 °C for 3 h with variable amounts of bovine skin gelatin, previously solubilized through heating denaturation at 95 °C for 10 min. After incubation, 100 µM of UDP-α-sugar and 50 µM $MnCl_2$ or $MgCl_2$ were added to start the enzymatic reactions. After a precise timing between 1 and 120 min the reactions were stopped by heating the samples at 95 °C for 2 min, then the samples were transferred into Proxiplate white 384-well plates (Perkin-Elmer). 5 µL of the UDP-Glo luminescence detection reagent (Promega) were added and let incubate for 1 h at 25 °C. The plates were then transferred into a GloMax Discovery plate reader (Promega) configured according to manufacturer's instructions for luminescence detection. Initial velocities were measured through the determination of slopes from linear fit interpolation of individual time points using *GraphPad Prism 7*[57]. After visual inspection of double reciprocal plots, which were linear (Supplementary Fig. 2c), the initial velocity values, expressed as apparent mean turnover values and associated errors, were directly fitted to the Michaelis-Menten equation using *GraphPad Prism 7*[57], obtaining the values of apparent k_{cat} and K_m accounting for propagation of statistical errors[58]. All experiments were performed in triplicates. Control experiments were performed using identical conditions by selectively removing GLT25D1, donor or acceptor substrates. Results of the data analysis are summarized in Supplementary Table 5.

Evaluation of Gal-T activity through HR-LCMS

5 µM recombinant human LH3/PLOD3 (obtained in house as described in previous work) was incubated with 5 µM GLT25D1/COLGALT1 (wild-type or mutants), 50 µM $FeCl_2$, 100 µM 2-oxoglutarate (2-OG), 500 µM sodium ascorbate, 50 µM $MnCl_2$, 100 µM UDP-α-sugar, and 1 mM peptide (Ac-GIKGIKGIKGIK-COOH) substrate. The reactions were allowed to proceed for 3 h at 37 °C. 5 µL of each reaction were diluted with 43 µL of Milli-Q water and acidified by addition of 2 µL of formic acid, to reach a final volume of 50 µL, and then analyzed on an Exion LC AD UHPLC (AB Sciex) coupled to a X500B ESI-HRMS/MS system (AB Sciex) by a full scan acquisition. The column oven was kept at 40 °C, while the autosampler was cooled at 10 °C. Peptides were separated by reverse phase HPLC on a Hypersil Gold C18 column (150 × 2.1 mm, 3 µm particle size, 175 Å pore size, Thermo Fisher Scientific) using a linear gradient (2–50% solvent B in 15 min), with solvent A consisting of 0.1%

aqueous formic acid and solvent B of acetonitrile (ACN) containing 0.1% formic acid. Flow rate was kept constant at 0.2 mL min^{-1}. Mass spectra were generated in positive polarity under constant instrumental conditions: ion spray voltage 4500 V, declustering potential 100 V, curtain gas 30 psi, ion source gas 1 40 psi, ion source gas 2 45 psi, temperature 350 °C, collision energy 10 V. Spectra analyses were performed using the *SCIEX OS 2.1* software (AB Sciex). Results of the data analysis are summarized in Supplementary Table 5.

Size exclusion chromatography coupled with small-angle X-ray scattering (SEC-SAXS)

Solution scattering data were collected at ESRF BM29 using a sec^{-1} frame rate on Pilatus 1 M detector located at a fixed distance of 2.87 m from the sample, allowing a global q range of 0.01–4.00 nm. SEC-SAXS experiments were carried out using Nexera High Pressure Liquid/Chromatography (Shimadzu) system connected online to SAXS sample capillary[59]. For these experiments, 50 µL of GTL2SD1/COLGALT1 concentrated at 4 mg mL^{-1} were injected into a Superdex 200 PC 3.2/300 Increase column (GE Healthcare), pre-equilibrated with 25 mM HEPES/NaOH, 200 mM NaCl, pH 8.0. For SEC-SAXS data, frames corresponding to GTL25D1/COLGALT1 protein peak were identified, blank subtracted, and averaged using *CHROMIXS*[60].

SEC-SAXS modeling and data analysis

Radii of gyration (*Rg*), molar mass estimates and distance distribution functions *P(r)* were computed using *PRIMUS*[61] within the *ATSAS* package[62]. Comparison of experimental SAXS data and 3D models from crystal structures was performed using *CRYSOL*[63]. A summary of SAXS data collection and analysis results is reported in Supplementary Table 6. Ab initio sphere models were generated from the SAXS data using *GASBOR*, by providing the software with *P(r)* functions derived from experimental data and the number of residues defining GLT25D1/COLGALT1 monomers and without imposing any internal symmetry. 20 independent ab initio *GASBOR* runs were performed and subject to similarity analysis using *SUPCOMB* and *DAMSEL* from *ATSAS*, allowing selection of a representative individual model based on normalized spatial discrepancy. Superpositions were carried out using the *SUPALM* program available in the *saspy ATSAS* plugin in *PyMol*[64]. For SAXS rigid-body modeling, the model of a GLT25D1/COLGALT1 monomer from the experimental crystal structure was subject to loop modeling using *CORAL*[65]. Evaluation of the agreement between molecular models and experimental SAXS profiles was carried out using *CRYSOL*[63].

Single particle negative staining electron microscopy

200 mesh, Cu grids (Ted Pella) grids were negatively charged using a PELCO EasiGlow (Ted Pella) with the following parameters: pressure 40 mBar, power 15 mA, time 30 s. 2 µL of 1 mg mL^{-1} GLT25D1/COLGALT1 were applied on each grid and left for 60 s. The excess sample was blotted using Whatman paper and then a 25 µL drop of 2% (w/v) uranyl acetate solution was applied, gently stirring for 1 min. The excess stain was blotted with filter paper and the grids were left dry overnight. 350 EM micrographs were acquired on a JEM 1200EXIII (JEOL) electron microscope equipped with a Mega View III CCD camera (Olympus). Data processing was carried out using *RELION*, including auto-picking and multiple rounds of 2D classification. Comparison of electron microscopy single particle analysis 2D classes and GLT25D1/COLGALT1 molecular models from crystal structures was carried out using *AlignProjections* within *the COSMIC2* webserver[66].

GLT25D1/COLGALT1-LH3/PLOD3 cross-linking and enzymatic proteolysis

5 µM GLT25D1/COLGALT1 (wild-type and Trp158Arg) were mixed with 5 µM human LH3/PLOD3[67] (total protein concentration 10 µM) in 50 mM HEPES/NaOH, 100 mM NaCl, pH 8.0. The samples were incubated at room temperature for 1 h with 0.5 mM DSBU. The cross-linker was dissolved in neat dimethyl sulfoxide (DMSO) immediately before adding it to the peptide solution. The reaction was quenched by adding 2 µL of 1 M Tris/HCl solution. Cross-linked samples were loaded onto S-Trap microcolumns (Protifi) according to the manufacturer's instructions. In brief, after loading, samples were washed with 90:10% methanol/50 mM ammonium bicarbonate. The samples were then digested with trypsin (Promega) (1:20 trypsin/protein) for 1.5 h at 47 °C. The digested peptides were eluted using 50 mM ammonium bicarbonate. Two additional elutions were made using 0.2% formic acid and 0.2% formic acid in 50% ACN. The three fractions were pooled together and vacuum-centrifuged to dryness. The resulting peptides were resuspended in 5% ACN, 0.1% trifluoroacetic acid for subsequent LC-MS/MS analyses.

Liquid chromatography coupled to mass spectrometry for cross-linking peptide analysis

Proteolyzed cross-linked samples were analyzed via LC-MS/MS on an UltiMate 3000 RSLC nano-HPLC system (Thermo Fisher Scientific) coupled to a timsTOF Pro mass spectrometer equipped with CaptiveSpray source (Bruker Daltonics). Peptides were trapped on a C18 column (precolumn Acclaim PepMap 100, 300 µm × 5 mm, 5 µm, 100 Å, Thermo Fisher Scientific) at 50 °C and separated on a self-packed Picofrit (New Objective) nanospray emitter (360 µm o.d. × 75 µm i.d. × 400 mm L, 15 µm Tip i.d.) with C18-stationary phase (3.0 µm, 120 Å, Dr. Maisch GmbH). After trapping, peptides were eluted by a linear 90 min water–ACN gradient from 3% (v/v) to 50% (v/v) ACN. For the timsTOF Pro MS settings, the following parameters were adopted. The values for mobility-dependent collision energy ramping were set to 95 eV at an inverse reduced mobility ($1/k_0$) of 1.6 V s/cm^2 and 23 eV at 0.73 V s/cm^2. Collision energies were linearly S-5 interpolated between these two $1/k_0$ values and kept constant above or below. No merging of TIMS scans was performed. Target intensity per individual PASEF precursor was set to 20,000. The scan range was defined between 0.6 and 1.6 V s/cm^2 with a ramp time/accumulation time of 166 ms. 14 PASEF MS/MS scans were triggered per cycle (2.57 s) with a maximum of seven precursors per mobilogram. Precursor ions in an m/z range between 100 and 1700 with charge states ≥3+ and ≤8+ were selected for fragmentation. Active exclusion was enabled for 0.4 min (mass width 0.015 Th, $1/k_0$ width 0.015 V s/cm^2).

XL-MS data analysis

Cross-linked peptides were identified using MeroX[68] (version 2.0.1.7, https://stavrox.com). The search parameters were set as follows: tryptic cleavage C-terminal to Lys and Arg, allowing up to four missed cleavages; peptide length range 3–30 amino acids; fixed modification: Cys alkylation (iodoacetamide); variable modification: Met oxidation. The cross-linker was specified to react with Lys and N-termini. Cross-link identification was performed in RISEUP mode, allowing a maximum of three missing ions. The precursor and fragment mass tolerances were set to 10 ppm and 20 ppm, respectively. The signal-to-noise ratio was required to be >2.0. False discovery rate (FDR) filtering was applied at 1% at the peptide-spectrum match (PSM) level, with a minimum score cut-off of 100. To improve FDR estimation, a contaminant database (CRAP database) was included.

Proteomic data analysis was conducted using MaxQuant[69] (version 2.6.2.0) and Skyline[70] (version 24.1.1.339). Bruker raw files were processed using data-dependent acquisition (DDA) searches against the UniProt Human Reference Proteome (UP000005640), which includes 20,406 canonical proteins. Contaminants were filtered using the default MaxQuant contaminant database. The search parameters included tryptic cleavage C-terminal to Lys and Arg, allowing up to two missed cleavages. The peptide length was set to 6–25 amino acids. Fixed modifications included Cys alkylation (iodoacetamide), while variable modifications included Met oxidation and N-terminal acetylation. For label-free quantification (LFQ), a minimum of three unique peptides per

protein was required. The MS1 and MS2 tolerances were set to 10 ppm and 20 ppm, respectively. The Match Between Runs (MBR) feature was enabled, with a retention time alignment window of 0.4 min and an ion mobility window of 0.05. PSM and protein-level FDR were controlled at 1%. MaxQuant and MeroX results were imported into Skyline for MS/MS spectrum inspection and peptide quantification at MS1 level. Protein quantification was based on the ten most abundant validated peptides per protein. MS/MS spectra for the PLOD3-GLT25D1 inter-protein cross-link and relevant dead-end products were imported into Skyline, following previously established workflows[71], and quantified.

ICP-MS measurements

To quantify the possible presence and concentration of metal ions in GLT25D1/COLGALT1 recombinant preparations, 900 µL of a sample containing 4 mg mL^{-1} GLT25D1/COLGALT1 (purified without usage of divalent ions) were treated with 300 µL 65% ultrapure HNO_3 and 200 µL 30% *w/w* H_2O_2, made up to 5 mL with ultrapure water and analyzed on a single quadrupole inductively coupled plasma mass spectrometer (SQ-ICP-MS, iCAP RQ Thermo Fisher Scientific), equipped with a quartz cyclonic chamber cooled at 3 °C, a MicroMist nebulizer 179 (400 µL min^{-1}), a quartz torch, a Ni sampler, skimmer cones, and a QCell pressurized with helium (3 V, KED mode). Quantitative determinations for manganese and calcium were obtained by the external standard calibration with four standards (0, 1, 5, 10 µg L^{-1}) prepared daily in the same buffer used for sample preparation, at the same dilution and HNO_3 concentration. While manganese could not be detected, calcium was detected with a 1:1 stoichiometry.

Native mass spectrometry (Native-MS)

Native-MS was employed to identify metal cofactors of GLT25D1/COLGALT1. Recombinant wild-type and Trp158Arg GLT25D1/COLGALT1 enzymes (20 µM) were analyzed under partially denaturing conditions (50 mM ammonium bicarbonate, 40% acetonitrile), in order to improve droplets desolvation and to minimize the adducts formation during electrospray. Samples were directly injected into an Orbitrap Fusion instrument (Thermo Fisher Scientific) using metal-coated borosilicate emitters (~1 µm internal diameter). The main instrumental parameters were set as follows: resolving power at *m/z* 200, 120,000; IRM pressure, 3 mTorr (intact protein mode); ion spray voltage, 1.1–1.2 kV; ion-transfer tube temperature, 275 °C; in-source fragmentation, 75 V; AGC target, 4×10^5; maximum injection time, 100 ms. Final spectra were obtained by averaging the signal over 60 s acquisition. Ligand mass was calculated as the difference between detected bound- and free-protein signals over the entire charge state distribution of the spectra (Supplementary Fig. 10).

Molecular dynamics simulations

Molecular dynamics simulations were carried out using the *AMBER* software package (version 24) with the *ff19SB* force field, using its GPU-accelerated *pmemd.cuda* utility during equilibration and production, and sander otherwise. Comprehensive descriptions of the preparation of the simulated systems and molecular dynamics simulation protocols can be found in the Supplementary Information. Analyses of MD trajectories were carried out with the *cpptraj* program distributed within the *ambertools* suite (version 24) and in-house scripts.

Principal component analysis (PCA) was performed to identify the major modes of motion in the molecular dynamics simulations of the GT1 and GT2 domains, combining the trajectories of the ligand-free and ligand-bound states of each domain. The covariance matrix of the backbone atoms of the simulated systems was calculated, and diagonalized to obtain the principal components (eigenvectors) and their corresponding eigenvalues. The eigenvalues corresponding to each principal components were used to evaluate the percentage of the variance in the motion of the protein represented by each component. Since the first two components, PC1 and PC2, account for most of the variance in all the simulated systems, the snapshots of the trajectories were projected only onto these principal components, and histograms of the normalized frequency of the conformational states along PC1 and PC2 were generated.

Hydrogen bond persistence analysis was performed using the *hbond* command in the *cpptraj* module of *AMBER24*. A hydrogen bond was considered present based on geometric criteria: a donor–acceptor distance cutoff of 3.5 Å and a hydrogen-donor-acceptor angle cutoff of 135°. The persistence of each hydrogen bond was monitored throughout the simulation time, and the hydrogen bond frequencies were normalized based on the total number of hydrogen bonds to represent the percentage of time each bond was present during the trajectory. These results were then grouped into 10% frequency intervals to generate histograms, illustrating the persistence of hydrogen bonds over the course of the simulation.

The distance fluctuation (DF) parameter for each pair of residues i and j in a protein system is defined as $D_{ij} = \langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle$, where d_{ij} is the time-dependent distance between the Cα atoms of residues i and j, and the brackets indicates the time-averaged distance over the entire trajectory. This calculation results in a matrix that characterizes the extent of movement or flexibility between all residue pairs.

Reporting summary

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

Data availability

Unless otherwise stated, all data supporting the results of this study can be found in the article, supplementary, and source data files. Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 9EVJ, 9EVK, and 9EVL. In solution SEC-SAXS data have been deposited in the SASBDB under accession code SASDVZ2. XL-MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the project accession PXD061021. Native MS data have been deposited in the MassIVE database with dataset identifier MSV000097322 [https://doi.org/10.25345/C5TH8C06M]. MD simulation data, including initial coordinate files, simulation input files, and final output coordinate files have been deposited in Zenodo [https://doi.org/10.5281/zenodo.15020231]. The numerical source data for Supplementary Fig. 9 can be found in the raw data deposited on PRIDE under accession number PXD061021. The numerical source data for Supplementary Fig. 11 can be found in the raw data deposited on MassIVE under accession number MSV000097322 [https://doi.org/10.25345/C5TH8C06M]. Source data are provided with this paper.

References

1. Bourhis, J. M. et al. Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat. Struct. Mol. Biol.* **19**, 1031–1036 (2012).
2. Fidler, A. L., Boudko, S. P., Rokas, A. & Hudson, B. G. The triple helix of collagens—an ancient protein structure that enabled animal multicellularity and tissue evolution. *J. Cell Sci.* **131**, jcs203950 (2018).
3. Myllyharju, J. & Kivirikko, K. I. Collagens and collagen-related diseases. *Ann. Med.* **33**, 7–21 (2001).
4. Ishikawa, Y. & Bachinger, H. P. A molecular ensemble in the rER for procollagen maturation. *Biochim. Biophys. Acta* **1833**, 2479–2491 (2013).
5. Gelse, K., Poschl, E. & Aigner, T. Collagens-structure, function, and biosynthesis. *Adv. Drug Deliv. Rev.* **55**, 1531–1546 (2003).
6. Kivirikko, K. I. & Myllyla, R. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol.* **82**, 245–304 (1982).
7. Kadler, K. E., Holmes, D. F., Trotter, J. A. & Chapman, J. A. Collagen fibril formation. *Biochem. J.* **316**, 1–11 (1996).

8. Bella, J. Collagen structure: new tricks from a very old dog. *Biochem. J.* **473**, 1001–1025 (2016).
9. Myllyharju, J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **22**, 15–24 (2003).
10. Gorres, K. L. & Raines, R. T. Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **45**, 106–124 (2010).
11. Yamauchi, M. & Sricholpech, M. Lysine post-translational modifications of collagen. *Essays Biochem.* **52**, 113–133 (2012).
12. Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5**, 1087–1104 (2009).
13. Hennet, T. Collagen glycosylation. *Curr. Opin. Struct. Biol.* **56**, 131–138 (2019).
14. 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. *Biochem. Biophys. Res. Commun.* **560**, 93–98 (2021).
15. Tvaroska, I. Glycosylation modulates the structure and functions of collagen: a review. *Molecules* https://doi.org/10.3390/molecules29071417 (2024).
16. De Giorgi, F., Fumagalli, M., Scietti, L. & Forneris, F. Collagen hydroxylysine glycosylation: non-conventional substrates for atypical glycosyltransferase enzymes. *Biochem. Soc. Trans.* **49**, 855–866 (2021).
17. Scietti, L. & Forneris, F. Full-length human collagen lysyl hydroxylases. in *Encyclopedia of Inorganic and Bioinorganic Chemistry* (ed. Scott, R. A.) 1–12 (Wiley, 2020).
18. Salo, A. M. & Myllyharju, J. Prolyl and lysyl hydroxylases in collagen synthesis. *Exp. Dermatol.* https://doi.org/10.1111/exd.14197 (2020).
19. Yamauchi, M., Terajima, M. & Shiiba, M. Lysine hydroxylation and cross-linking of collagen. *Methods Mol. Biol.* **1934**, 309–324 (2019).
20. Kivirikko, K. I. & Myllyla, R. Collagen glycosyltransferases. *Int. Rev. Connect Tissue Res.* **8**, 23–72 (1979).
21. Wang, C. et al. The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. *Matrix Biol.* **21**, 559–566 (2002).
22. Geister, K. A. et al. Loss of function of Colgalt1 disrupts collagen post-translational modification and causes musculoskeletal defects. *Dis. Model Mech.* https://doi.org/10.1242/dmm.037176 (2019).
23. Terajima, M. et al. Role of glycosyltransferase 25 domain 1 in type I collagen glycosylation and molecular phenotypes. *Biochemistry* **58**, 5040–5051 (2019).
24. Webster, J. A. et al. Collagen beta (1-O) galactosyltransferase 1 (GLT25D1) is required for the secretion of high molecular weight adiponectin and affects lipid accumulation. *Biosci. Rep.* https://doi.org/10.1042/BSR20170105 (2017).
25. Baumann, S. & Hennet, T. Collagen accumulation in osteosarcoma cells lacking GLT25D1 collagen galactosyltransferase. *J. Biol. Chem.* **291**, 18514–18524 (2016).
26. Yang, J. et al. Collagen beta(1-O) galactosyltransferase 2 deficiency contributes to lipodystrophy and aggravates NAFLD related to HMW adiponectin in mice. *Metabolism* **120**, 154777 (2021).
27. Kehayova, Y. S., Wilkinson, J. M., Rice, S. J. & Loughlin, J. Osteoarthritis genetic risk acting on the galactosyltransferase gene COLGALT2 has opposing functional effects in articulating joint tissues. *Arthritis Res. Ther.* **25**, 83 (2023).
28. Schegg, B., Hulsmeier, A. J., Rutschmann, C., Maag, C. & Hennet, T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. *Mol. Cell Biol.* **29**, 943–952 (2009).
29. Liefhebber, J. M., Punt, S., Spaan, W. J. & 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).
30. Kehayova, Y. S., Watson, E., Wilkinson, J. M., Loughlin, J. & Rice, S. J. Genetic and epigenetic interplay within a COLGALT2 enhancer associated with osteoarthritis. *Arthritis Rheumatol.* **73**, 1856–1865 (2021).
31. Guo, T. et al. COLGALT2 is overexpressed in ovarian cancer and interacts with PLOD3. *Clin. Transl. Med.* **11**, e370 (2021).
32. Wang, Y. et al. Exosomes secreted by adipose-derived mesenchymal stem cells foster metastasis and osteosarcoma proliferation by increasing COLGALT2 expression. *Front Cell Dev. Biol.* **8**, 353 (2020).
33. Wang, S. et al. Upregulation of GLT25D1 in hepatic stellate cells promotes liver fibrosis via the TGF-beta1/SMAD3 pathway in vivo and in vitro. *J. Clin. Transl. Hepatol.* **11**, 1–14 (2023).
34. 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).
35. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
36. Luther, K. B. et al. Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. *J. Biol. Chem.* **286**, 43701–43709 (2011).
37. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem* **77**, 521–555 (2008).
38. Taujale, R. et al. Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. *Elife* https://doi.org/10.7554/eLife.54532 (2020).
39. Miyatake, S. et al. Biallelic COLGALT1 variants are associated with cerebral small vessel disease. *Ann. Neurol.* **84**, 843–853 (2018).
40. Teunissen, M. W. A. et al. Biallelic variants in the COLGALT1 gene causes severe congenital porencephaly: a case report. *Neurol. Genet.* **7**, e564 (2021).
41. Scietti, L. et al. Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. *Nat. Commun.* **9**, 3163 (2018).
42. Sadakierska-Chudy, A., Szymanowski, P., Lebioda, A. & Ploski, R. Identification and in silico characterization of a novel COLGALT2 gene variant in a child with mucosal rectal prolapse. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms23073670 (2022).
43. Peng, J. et al. The structural basis for the human procollagen lysine hydroxylation and dual-glycosylation. *Nat. Commun.* **16**, 2436 (2025).
44. Salo, A. M. et al. Lysyl hydroxylase 3 (LH3) modifies proteins in the extracellular space, a novel mechanism for matrix remodeling. *J. Cell Physiol.* **207**, 644–653 (2006).
45. Wang, C., Ristiluoma, M. M., Salo, A. M., Eskelinen, S. & Myllyla, R. Lysyl hydroxylase 3 is secreted from cells by two pathways. *J. Cell Physiol.* **227**, 668–675 (2012).
46. Chen, Y. L. et al. Lysyl hydroxylase 2 is secreted by tumor cells and can modify collagen in the extracellular space. *J. Biol. Chem.* **291**, 25799–25808 (2016).
47. Banushi, B. et al. Regulation of post-Golgi LH3 trafficking is essential for collagen homeostasis. *Nat. Commun.* **7**, 12111 (2016).
48. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
49. Gorrec, F. The MORPHEUS protein crystallization screen. *J. Appl. Crystallogr.* **42**, 1035–1042 (2009).
50. Kabsch, W. Xds. *Acta Crystallogr D. Biol. Crystallogr* **66**, 125–132 (2010).
51. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D. Biol. Crystallogr* **69**, 1204–1214 (2013).
52. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).
53. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D. Biol. Crystallogr* **62**, 1002–1011 (2006).

54. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D. Biol. Crystallogr* **66**, 213–221 (2010).
55. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D. Biol. Crystallogr* **67**, 355–367 (2011).
56. Gore, S. et al. Validation of structures in the protein data bank. *Structure* **25**, 1916–1927 (2017).
57. Graphpad Software, L.J. *Graphpad Prism 7* (Graphpad Software, USA).
58. Bevington, P. H. *Data Reduction and Error Analysis for the Physical Sciences*, 56–62 (McGraw-Hill, 1969).
59. Brennich, M. E., Round, A. R. & Hutin, S. Online size-exclusion and ion-exchange chromatography on a SAXS beamline. *J. Vis. Exp.* https://doi.org/10.3791/54861 (2017).
60. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small angle X-ray scattering data. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btx846 (2017).
61. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
62. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr* **50**, 1212–1225 (2017).
63. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
64. Panjkovich, A. & Svergun, D. I. SASpy: a PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data. *Bioinformatics* **32**, 2062–2064 (2016).
65. Petoukhov, M. V. et al. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
66. Cianfrocco, M. A., Wong-Barnum, M., Youn, C., Wagner, R. & Leschziner, A. COSMIC2: a science gateway for cryo-electron microscopy structure determination. in *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact* Article 22 (Association for Computing Machinery, 2017).
67. Mattoteia, D. et al. Identification of regulatory molecular "hot spots" for LH/PLOD collagen glycosyltransferase activity. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms241311213 (2023).
68. Iacobucci, C. et al. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nat. Protoc.* **13**, 2864–2889 (2018).
69. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
70. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
71. Rojas Echeverri, J. C. et al. A workflow for improved analysis of cross-linking mass spectrometry data integrating parallel accumulation-serial fragmentation with MeroX and Skyline. *Anal. Chem.* **96**, 7373–7379 (2024).
72. Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* **51**, 2778–2786 (2011).

Acknowledgements

We thank Ms. Lisa Negro and Ms. Martina Soffientini for assistance with production of recombinant GLT25D1/COLGALT1, Dr. Marco Fumagalli for support with HRMS assays, and Dr. Gianluca Santoni for support with SAD X-ray data collection setup. We thank Dr. Federica Marasca and Prof. Antonella Profumo for support with ICP-MS analysis. We also would like to thank Prof. Giulia F. Mancini and Prof. M. Freccero for inspiring discussions during manuscript revision. We thank Prof. Andrea Sinz for providing access to the Center for Structural Mass Spectrometry at Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. We thank the European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS) for the provision of synchrotron radiation facilities. We thank Centro Grandi Strumenti (University of Pavia) for the provision of HRMS and NS EM instrumentation. This work was supported by the Italian Association for Cancer Research (AIRC, Grants MFAG 20075 and Bridge 27004 to F.F.), the Mizutani Foundation for Glycoscience (Grant 200039 to F.F.), the Ehlers-Danlos Society (Rarer Types EDS Grant 2022 to F.F.), the Giovanni Armenise-Harvard Foundation (CDA 2013 to F.F.), the NextGenerationEU-MUR PNRR program through Italian Ministry for University and Research (grant PRIN PNRR 2022 P20224WAME to C.I. and F.F.), the University of Pavia (InROAD fellowship to L.S.). The Forneris Lab also acknowledges support from EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT), and by the Italian Ministry of Health (Piano Operativo Salute, IMMUNO-HUB). Some of the instrumentation used for this research was acquired through funding by Regione Lombardia, regional law n° 9/2020, resolution n° 3776/2020. 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

F.F. conceived the project and designed research. M.D.M., L.S. and S.R.R. produced recombinant GLT25D1/COLGALT1 and performed biochemical characterizations, with support from D.M. A.P. generated all mutants. L.S. crystallized GLT25D1/COLGALT1. F.F. and L.S. solved the structure and carried out structural refinements, with support from M.D.M. D.M. and M.D.M. performed HRMS assays. S.L. carried out mass photometry experiments and associated data analysis. A.V. and C.I. carried out XL-MS experiments and associated data analysis. E.M. and G.C. carried out molecular dynamics simulations and associated analyses. C.S. performed Native-MS experiments for the identification of cofactors bound in the GT1 domain. M.D.M., S.R.R., L.S. and F.F. analyzed the data, prepared the figures and wrote the paper, with contributions from all authors.

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/s41467-025-59017-5.

Correspondence and requests for materials should be addressed to Federico Forneris.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Supporting Information for:

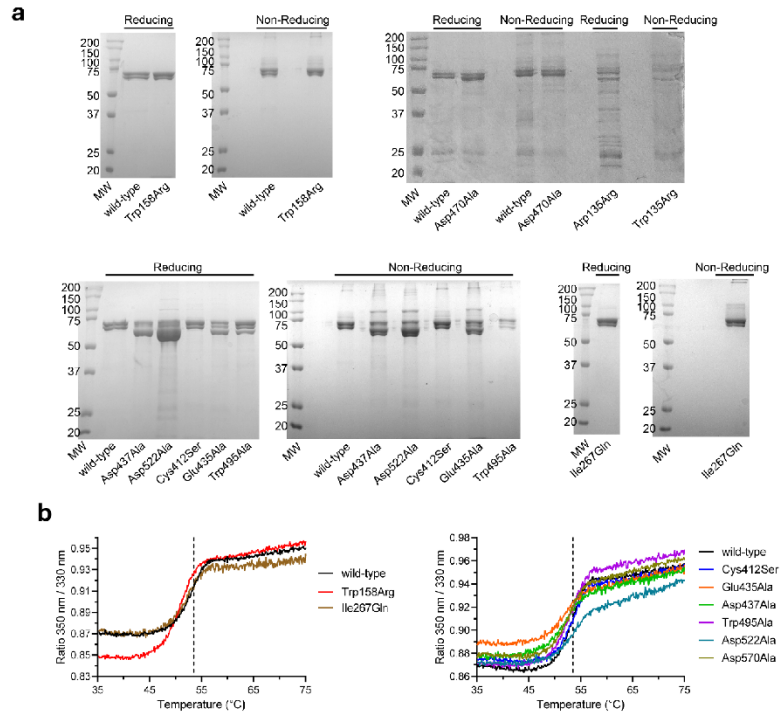
Molecular Structure and Enzymatic Mechanism of the Human Collagen Hydroxylysine Galactosyltransferase GLT25D1/COLGALT1

by Matteo De Marco, Sristi Raj Rai, Luigi Scietti, et al.

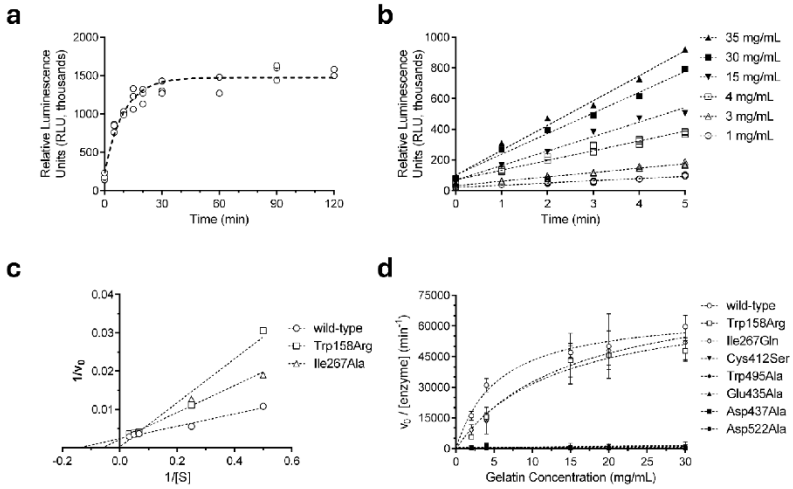
Correspondence to: Federico Forneris, e-mail: federico.forneris@unipv.it

Table of Contents:

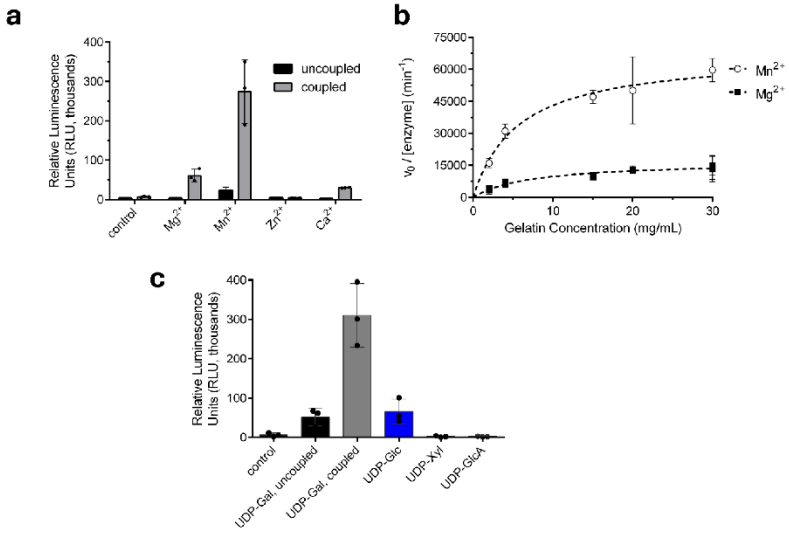
SUPPLEMENTARY FIGURES



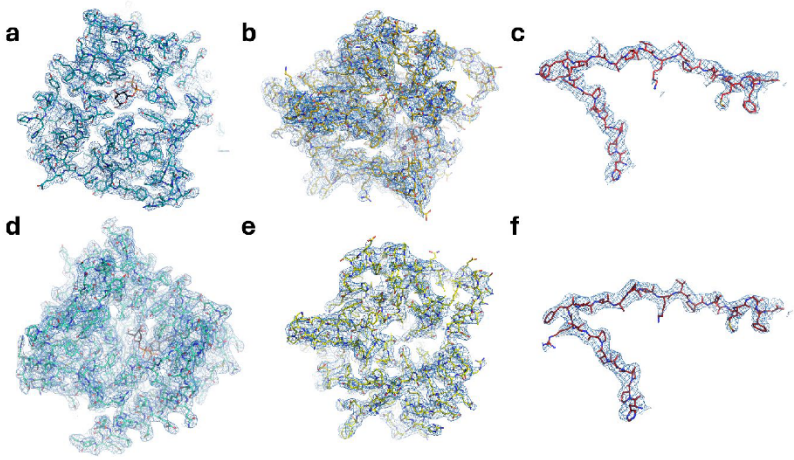
Supplementary Figure 1. Quality assessment of enzyme sample preparations. (a) All wild-type and mutant GLT25D1/COLGALT1 that could be generated underwent testing for correct folding using reducing and non-reducing SDS-PAGE. MW, molecular weight marker. (b) DSF analysis of the recombinant wild-type and mutant GLT25D1/COLGALT1 samples shown in panel (a). Mutant Trp135Arg was not tested due to the poor sample quality. The dashed vertical line indicates the unfolding temperature of wild-type GLT25D1/COLGALT1.



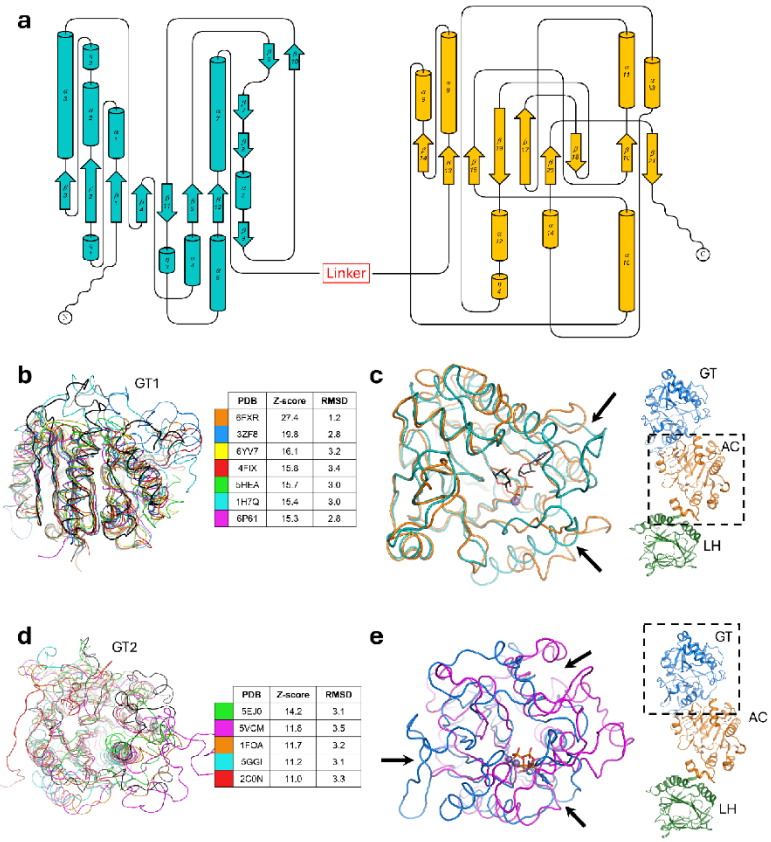
Supplementary Figure 2. Development of a luminescence-based steady-state kinetics assay to evaluate GLT25D1/COLGALT1 enzymatic activity. (a) Individual time points obtained from endpoint measurements using 1 mg/mL GLT25D1/COLGALT1 and 4 mg/mL gelatin substrate displayed exponential profiles, as confirmed by the nonlinear fit performed using *Graphpad Prism*[1], enabling identification of the initial linearity region calculation of initial velocities. (b) Linear fitting of time points collected within the first five minutes of the enzymatic reactions performed in presence of different concentrations of gelatin (represented by different symbols as reported in the figure legend) at constant enzyme concentration (i.e., 1 mg/mL) allows determination of initial velocities. (c) double-reciprocal plot showing the results obtained for three different GLT25D1/COLGALT1 variants (constant enzyme concentration 1 mg/mL) shows linearity, enabling quantitation of K_M and v_{max} values. (d) Results of the steady-state kinetics analysis of GLT25D1/COLGALT1 wild-type and mutants subject to this study. Error bars represent standard deviations from average of triplicate independent experiments. Initial velocity values, expressed as apparent turnover numbers (v_0 / [enzyme]) were subject to nonlinear fit using the Michaelis-Menten equation in *Graphpad Prism*[1]. Results are shown in Supplementary Table 5.



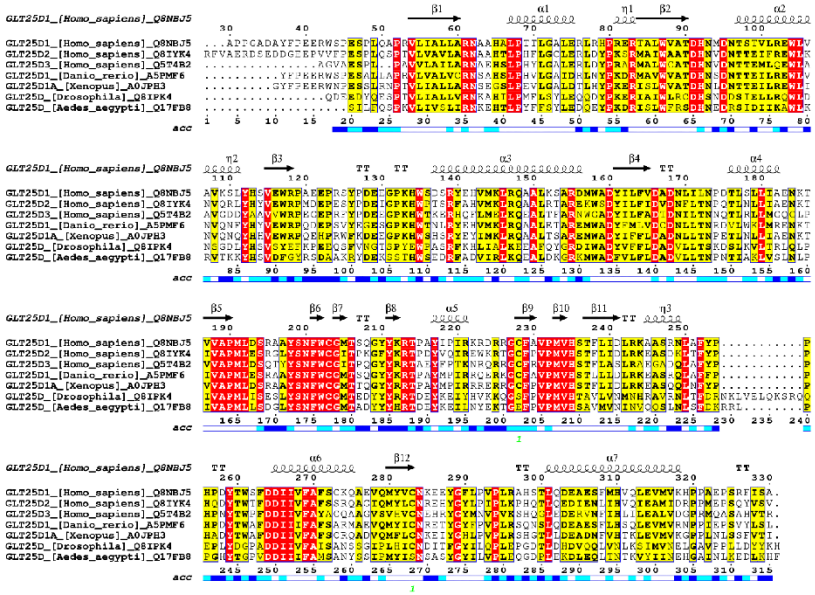
Supplementary Figure 3. GLT25D1/COLGALT1 is a Mn^{2+}-dependent galactosyltransferase. (a) Endpoint luminescence-based assays were used to evaluate the galactosyltransferase enzymatic activity in presence of different metal ions. Control experiments were performed without adding GLT25D1/COLGALT1. (b) Steady-state enzyme kinetics analysis was performed using Mn^{2+} or Mg^{2+} and UDP-Gal donor substrate, highlighting a strong preference for Mn^{2+} or Mg^{2+} as catalytic metal ion cofactor. Quantitation of K_M and k_{cat} was carried out using nonlinear fit through Michaelis-Menten equation in *Graphpad Prism*[1]. Results are shown in Supplementary Table 5. (c) Endpoint luminescence-based assays were used to evaluate the galactosyltransferase enzymatic activity in presence of different donor substrates and their analogs. Control experiments were performed without adding GLT25D1/COLGALT1. Uncoupled measurements were carried out in absence of acceptor (i.e., gelatin) substrate. In all figure panels, error bars represent standard deviations from average of triplicate independent experiments.



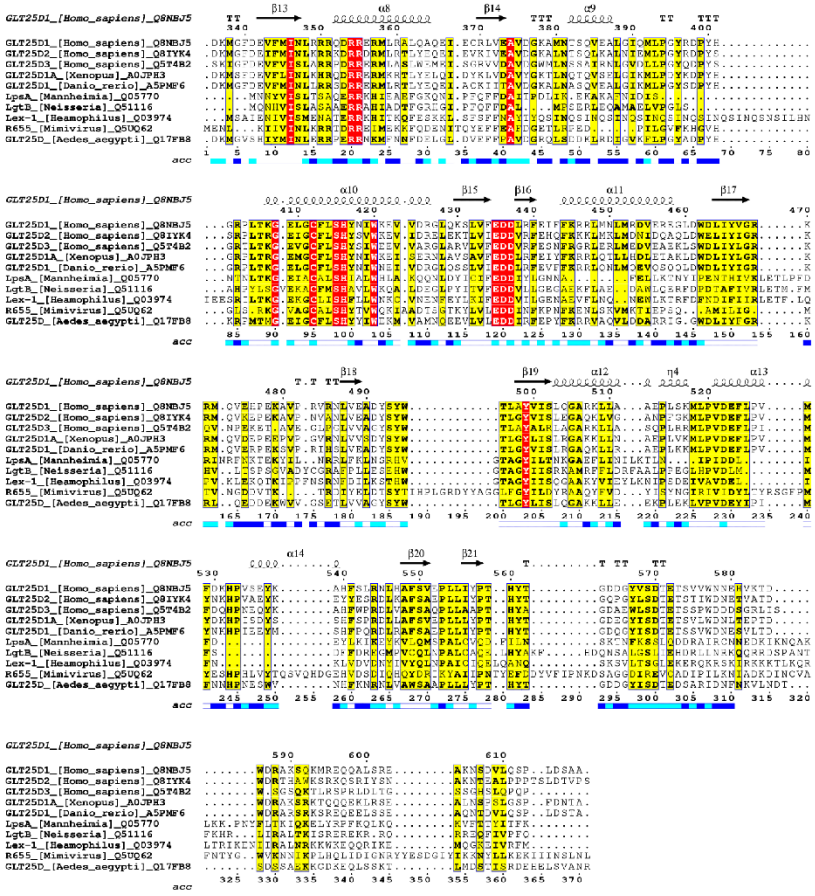
Supplementary Figure 4. Quality of the experimental electron density for the crystal structure of full-length human GLT25D1/COLGALT1. Shown are snapshots of the refined experimental electron density ($2F_o$-F_c, contour level 1 σ) used to generate the GLT25D1/COLGALT1 structural model using the Mn^{2+}, UDP-α-Gal co-crystallized model, covering the GT1 domains (a,d), the GT2 domains (b,e), and the linker regions (c,f) for each of the two copies of the enzyme found in the asymmetric unit.



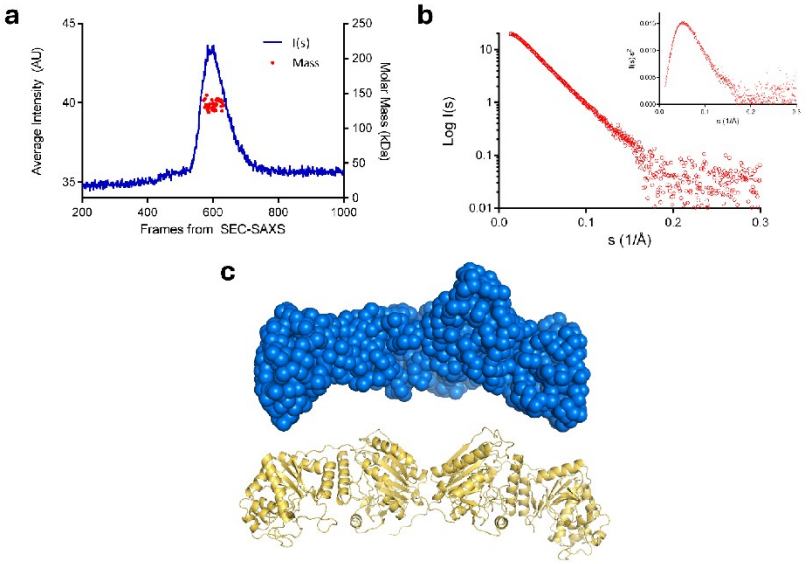
Supplementary Figure 5. Topological details of GLT25D1/COLGALT1 GT1 and GT2 domains. (a) Topology diagrams for the GT1 (teal) and GT2 (gold) domains, respectively, drawn using TOPDRAW [2]; (b) Superposition of structural homologs to the GT1 domain of GLT25D1/COLGALT1 (shown as thick black ribbon), as identified using DALI with Z-score higher than 15. The table inset indicates the PDB ID of each entry, as well as the associated DALI Z-scores and RMSD values. (c) The closest structural homolog of GLT25D1/COLGALT1 GT1 domain is the AC domain of human LH3/PLOD3. For reference, the image shown on the right highlights the position of the AC domain within the LH3/PLOD3 polypeptide (PDB ID 6FXR). The superposition shows the excellent match between the two molecular architectures (shown in teal for GLT25D1/COLGALT1 and orange for LH3/PLOD3, respectively) and the key differences (shown with a black arrow) in the region proximate to the cofactor and donor substrate binding site identified in GLT25D1/COLGALT1. (d) Superposition of structural homologs to the GT2 domain of GLT25D1/COLGALT1, as identified using DALI with Z-score higher than 10. The table inset indicates the PDB ID of each entry, as well as the associated DALI Z-scores and RMSD values. (e) Superposition of the GT2 domain of GLT25D1/COLGALT1 (magenta) with LH3/PLOD3 GT domain (blue) shows limited similarity with the N-terminal GT domain of the multifunctional enzyme, with DALI Z-score lower than 10 and RMSD of 4.7 Å.



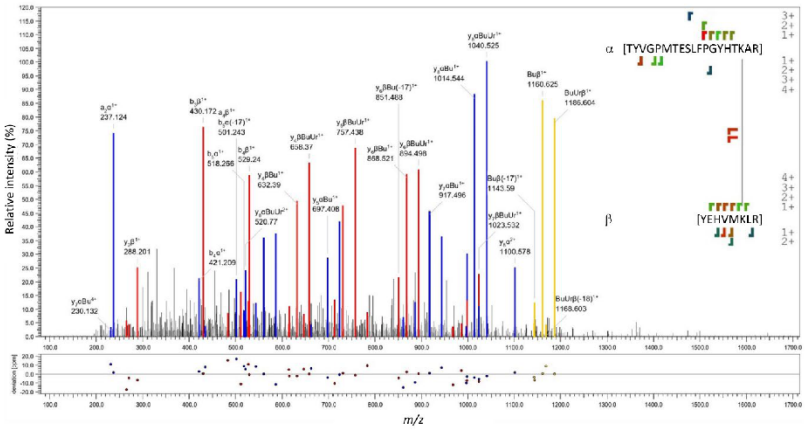
Supplementary Figure 6. Alignment of selected amino acid sequences from different species obtained from a BLAST search against the GLT25D1/COLGALT1 GT1 domain. The alignment has been generated using EBI MUSCLE[3] and drawn using ESPRIPT[4]. The "*consensus*" line computes the sequence consensus according to the result of the multiple sequence alignment (global score 0.7, differential score 0.5). The "*acc*" line indicates the solvent accessibility of each residue based on evaluation of the experimental structure of human GLT25D1/COLGALT1 (white, non-accessible residues; blue, accessible residues).



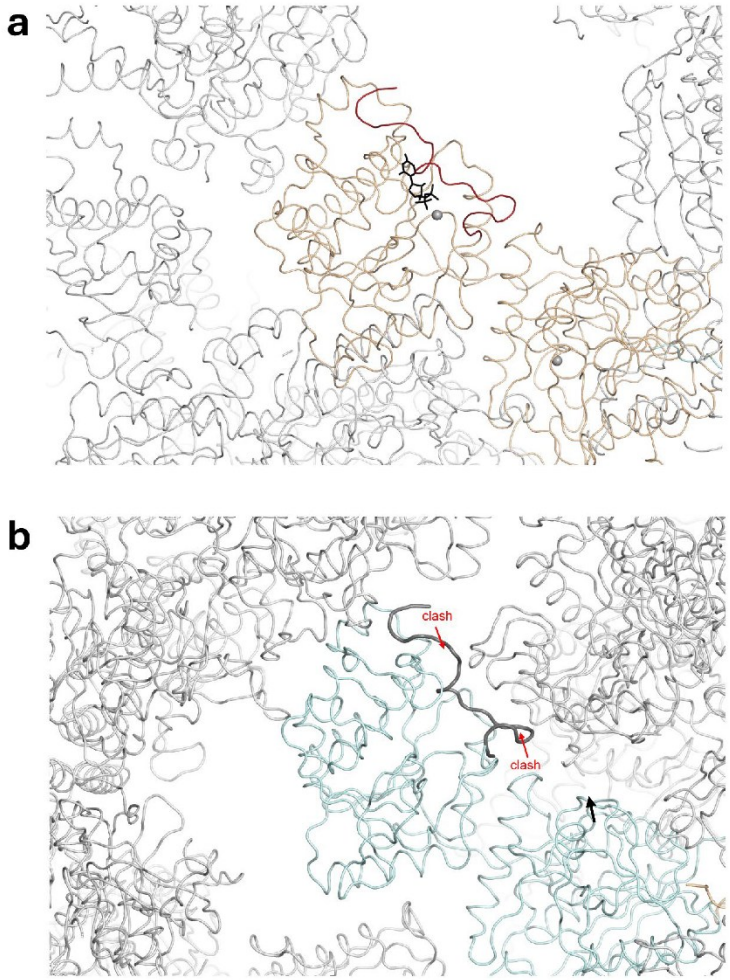
Supplementary Figure 7. Alignment of selected amino acid sequences from different species obtained from a BLAST search against the GLT25D1/COLGALT1 GT2 domain. The alignment has been generated using EBI MUSCLE[3] and drawn using ESPRIPT[4]. The "*consensus*" line computes the sequence consensus according to the result of the multiple sequence alignment (global score 0.7, differential score 0.5). The "*acc*" line indicates the solvent accessibility of each residue based on evaluation of the experimental structure of human GLT25D1/COLGALT1 (white, non-accessible residues; blue, accessible residues).



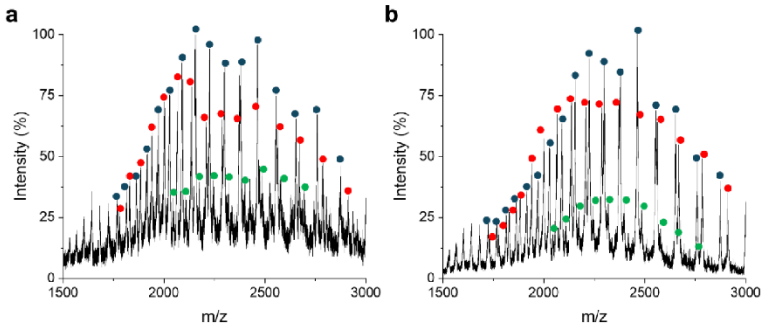
Supplementary Figure 8. SEC-SAXS analysis of GLT25D1/COLGALT1 in solution. (a) SEC-SAXS chromatogram and molecular weight analysis of peak fractions derived from CHROMIXS analysis[5]. (b) SAXS data obtained from the analysis of the peak frames of the SAXS chromatogram shown in (a) (red circles, with Kratky transformation in the inset). (c) *Ab-initio* sphere models based on solution SEC-SAXS data match the observed dimeric arrangement found in the experimental crystal structure of GLT25D1/COLGALT1. Shown is the comparison of the size and shape of an *ab initio* envelope computed from SEC-SAXS data (blue spheres) using *GASBOR*[6] with the crystal structure of GLT25D1/COLGALT1 (gold cartoon).



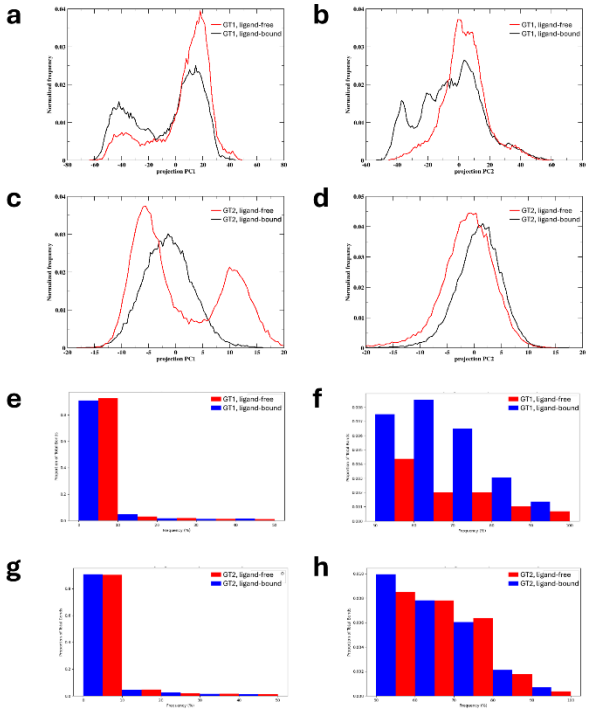
Supplementary Figure 9. Results of the XL-MS analysis of GLT25D1/COLGALT1 and LH3/PLOD3. Product ion mass spectrum of the cross-linked peptides (α) [TYVGPMTESLFPGYHT^{645}KAR] of LH3/PLOD3 and (β) [YEHVM^{145}KLR] of GLT25D1/COLGALT2. In the spectrum, the characteristic DSBU linker-derived fragment ions are highlighted in yellow, while fragment ions originating from the α peptide and β peptide are shown in blue and red, respectively. The precursor ion was detected with a mass accuracy of 1.0 ppm. The bottom panel reports the mass accuracy of fragment ions. The cross-linked species $[M+5H]^{5+}$ at *m/z* 685.9492 was subjected to CID-MS/MS, and fragment ions were annotated using the MeroX software[7]. The cross-link fragmentation profile displayed on the right shows the identified fragment ions with their respective positions in the peptide, as well as the charge. The relative intensity of the identified ion is color-coded from 0% (blue) to 100% (red). Data were obtained from file LH3_GLT_8_GB4_1_2907.d of the PRIDE dataset accompanying this paper, scan ID 25363, with a retention time of 5593.19 s (93.2 min ± 0.4 min) and an inverse ion mobility value of $1/K_0$ = 0.978.



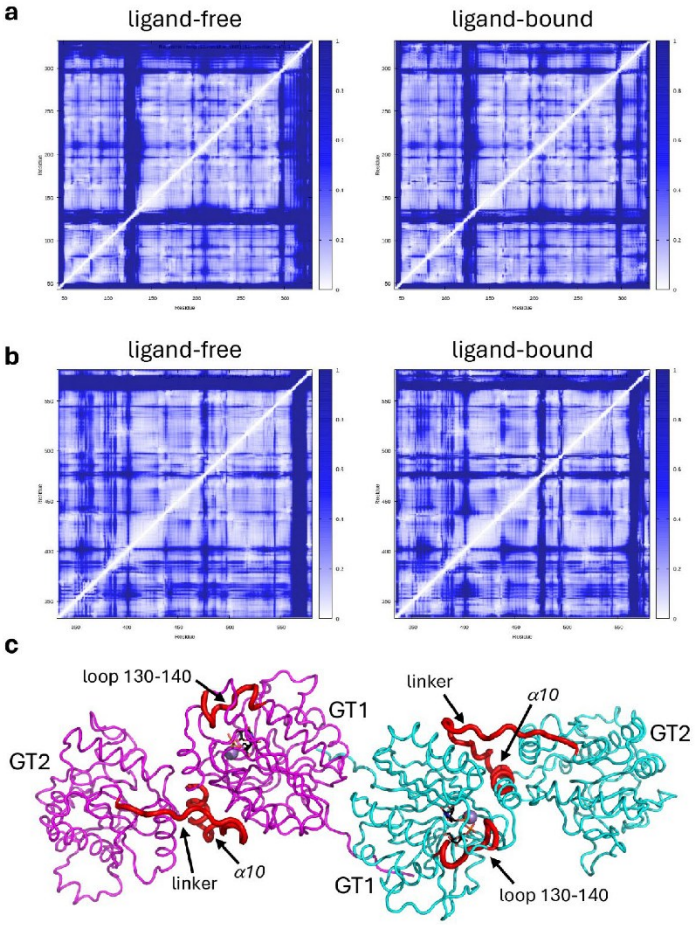
Supplementary Figure 10. Overview of the crystal packing observed in the crystal structures of GLT25D1/COLGALT1. The representations highlight the different contacts surrounding the GT2 domains in the two copies of the enzyme found in the asymmetric unit of the enzyme structure obtained with excess Mn^{2+} and UDP-α-Gal. The conformation adopted by the 560-570 loop in the substrate-bound structure, compatible with the packing of monomer A shown is panel (a), is not compatible with the crystal packing of the second monomer (b). In B, the loop from superposed monomer A is shown as thick grey ribbon to highlight the packing clashes.



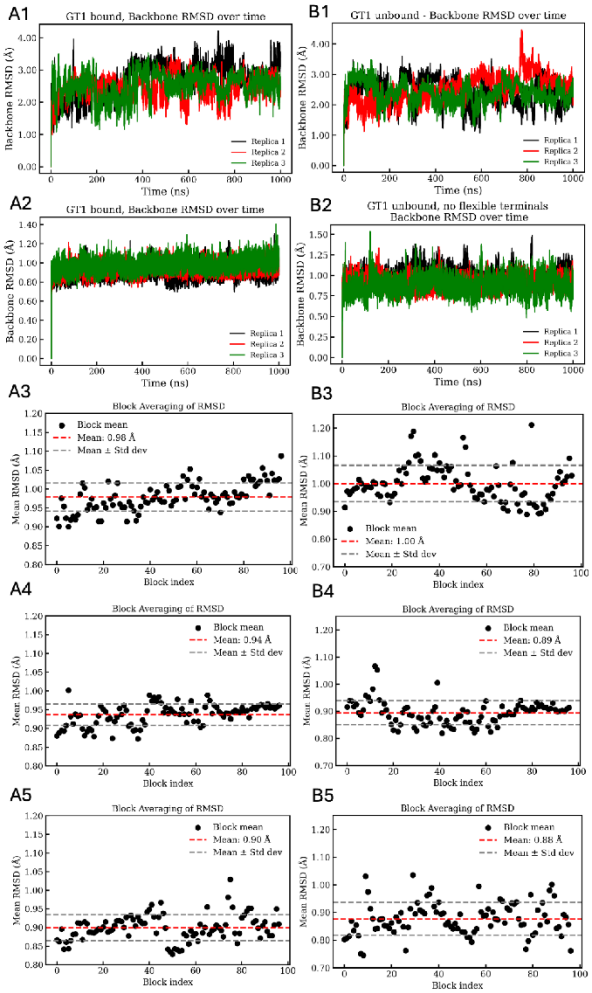
Supplementary Figure 11: Native-MS analysis. Native-MS spectra of wild-type (a) and Trp158Arg (b) GLT25D1/COLGALT1 under partially denaturing conditions (20 µM protein concentration in 50 mM ammonium bicarbonate, 40% acetonitrile). A charge-state distribution of monomeric, partially-unfolded conformation was detected (blue dots), with experimental molecular weights (68,969 ± 8 Da in panel a, 68,942 ± 2 Da in panel b) in close agreement with theoretical ones (68,972.7 Da and 68,942.6 Da, respectively). Putative GLT25D1/COLGALT1 deleted forms (red dots) of 64,049 ± 5 Da (a) and 64,019 ± 10 Da (b) were also present, confirming the doublet bands detected in SDS-PAGE analysis (Supplementary Figure 1). A minor molecular population, ascribable to protein-ligand complex, was also identified (greed dots), with molecular weigths of 69,575 ± 7 Da (a) and 69,550 ± 4 Da (b), compatible with protein interaction with UDP-α-Gal and Ca^{2+}.



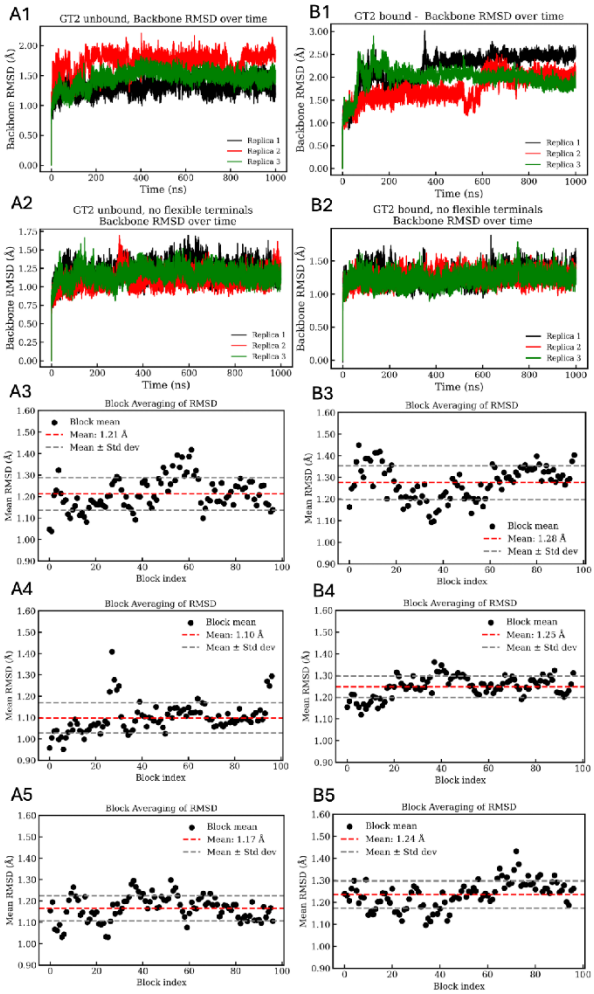
Supplementary Figure 12. Analysis of the MD simulations data. (a,b) Principal Component Analysis (PCA) of the GT1 domain. Normalized frequency distributions of the projections of the molecular dynamics trajectories of the GT1 domain in the unbound (black) and ligand-bound (red) states onto the first two principal components PC1 (a) and PC2 (b), which together capture 79% of the total variance in protein motion. (c,d) Principal Component Analysis (PCA) of the GT2 domain. Normalized frequency distributions of the projections of the molecular dynamics trajectories of the GT2 domain in the unbound (black) and ligand-bound (red) states onto the first two principal components PC1 (c) and PC2 (d), which together capture 84% of the total variance in protein motion. (e,f) Normalized histograms of hydrogen bond persistence in the GT1 domain for the ligand-bound (blue) and unbound (red) states. The normalized frequency persistences were calculated and grouped into 10% intervals; panel (e) represents the normalized frequency of hydrogen bonds persisting for 0–50% of the simulation time, while panel (f) represents those persisting for 50–100%. Frequencies were normalized to account for the total number of hydrogen bonds detected, allowing for a direct comparison between the two states. (g,h) Normalized histograms of hydrogen bond persistence in the GT2 domain for the ligand-bound (blue) and unbound (red) states. The normalized frequency persistences were calculated and grouped into 10% intervals; panel (g) represents the normalized frequency of hydrogen bonds persisting for 0–50% of the simulation time, while panel (h) represents those persisting for 50–100%. Frequencies were normalized to account for the total number of hydrogen bonds detected, allowing for a direct comparison between the two states.



Supplementary Figure 13. MD simulations highlight enhanced flexibility in specific areas of the GT1 domain. (a-b) Characterization of the internal dynamics and flexibility of the simulated domains of GLT25D1/COLGALT1 in terms of residue-pair fluctuations. (a) Matrix of residue-pair distance fluctuations (DF) for the GT1 domain in the ligand-free (left) and ligand-bound (right) states. (b) Matrices of residue-pair DFs for the GT2 domain in the ligand-free (left) and ligand-bound (right) states. The matrices are color-coded from white to dark blue, with darker shades representing greater distance fluctuations, less coordinated motion between residues and higher flexibility. . (c) Localization of the most prominent protein segments showing increased flexibility and less restricted to a well-defined conformation in molecular dynamics simulations upon removal of metal ions and donor substrates from the GLT25D1/COLGALT1 GT1 domain.



Supplementary Figure 14 - Equilibration analysis of MD simulations of GT1 in unbound and bound states. (A1) Backbone RMSD of GT1 in the unbound state across the three independent simulation replicates, calculated over the production phase of the simulations. (A2) Backbone RMSD of GT1 after excluding the flexible terminal regions (first and last 10 residues), also computed over the production phase. (A3–A5) Block-averaging analysis of the backbone RMSD for the unbound GT1 domain, in replicate 1, replicate 2, and replicate 3, respectively, assessing statistical convergence. (B1–B5) Equivalent analyses for GT1 in the bound state, with (B1) showing the backbone RMSD of the full domain, (B2) the backbone RMSD excluding the flexible terminal regions, and (B3–B5) the block-averaging analysis.



Supplementary Figure 15 MD equilibration - Equilibration analysis of MD simulations of GT2 in unbound and bound states. (A1) Backbone RMSD of GT2 in the unbound state across the three independent simulation replicates, calculated over the production phase of the simulations. (A2) Backbone RMSD of GT2 after excluding the flexible terminal regions (first 10 and last 18 residues), also calculated over the production phase. (A3–A5) Block-averaging analysis of the RMSD for the unbound GT2 domain in replicate 1, replicate 2, and replicate 3, respectively, assessing statistical convergence. (B1–B5) Equivalent analyses for GT2 in the bound state, with (B1) showing the backbone RMSD of the full domain, (B2) the backbone RMSD excluding the flexible terminal regions, and (B3–B5) the block-averaging analysis for replicate 1, replicate 2, and replicate 3, respectively.

SUPPLEMENTARY TABLES

Supplementary Table 1. List of oligonucleotides used to generate GLT25D1/COLGALT1 mutants

Site	Mutation	Direction	Sequence (5'→3')
GT1, dimer interface	Trp158Arg	Forward	ggctgattacatcctgtttgtagat
		Reverse	CTaccgagacctcagtcgtatggaacaat
GT1, cofactor binding site	Trp135Arg	Forward	cgtctgactcacgctacgagcatgtcat
		Reverse	Ggtgtttcgggccttcctcgtccgggt
	Asp166Ala	Forward	CAgcggacaacctgatcctcaaccc
		Reverse	ctacaaacaggatgtaatcagcccacatgt
	Asp265Ala	Forward	catcatcgtctttgccttctcctgcaag
		Reverse	Gcgtcaaaggaccaggtgtagtcagggt
	Ile266Gln	Forward	CAAatcgtctttgccttctcctgcaag
		Reverse	gtcgtcaaaggaccaggtgtagtcaggg
	Ile267Gln	Forward	CAAgtctttgccttctcctgcaagcagg
		Reverse	gatgtcgtcaaaggaccaggtgtagt
GT2, cofactor binding site	Arg351Ala	Forward	GCTcaggaccggcgggagcgcatgctgc
		Reverse	ccgcctcaggttaatcatgaagacctcgt
	Cys412Ser	Forward	Gcttcctgagccactacaacatctggaag
		Reverse	cttccagatgttgtagtggctcaggaagc
	Glu435Ala	Forward	ggatgacctgcgttttgagatcttcttc
		Reverse	Gcaaacacaagcgatttctgcagcccc
	Asp437Ala	Forward	Ccctgcgttttgagatcttcttcaagagac
		Reverse	catcctcaaacacaagcgatttctgcag
	Trp495Ala	Forward	gaccctggcctacgtgatctcccctgcaag
		Reverse	GCgtaggaatagtcagcttcgactaggtt
	Asp522Ala	Forward	cgagttcctgcccgtcatgttcgacaaac
		Reverse	Gctactggaagcattttggagagcggc
	Asp570Ala	Forward	CTaccgagacctcagtcgtatggaacaat
		Reverse	cactcacatagccatcgtctcctgtgta

Supplementary Table 2. X-ray diffraction data collection parameters and statistics

[a] Values in parentheses are for reflections in the highest resolution shell.

	GLT25D1/COLGALT1 (Native)	GLT25D1/COLGALT1 (co-crystallized with Mn^{2+}, UDP-α-Gal)	GLT25D1/COLGALT1 (Hg^{2+} Soak)
Data Collection[a]			
X-ray source	ESRF ID23-1	SLS X06SA	ESRF ID23-1
Processing programs	*XDS*, *AIMLESS*	*XDS*, *AIMLESS*	*XDS*, *AIMLESS*, *SHELX*, *HKL2MAP*
Space group	I23	I23	I23
Cell parameters	*a* = 221.9 Å *a* = 90.0° *b* = 221.9 Å *b* = 90.0° *c* = 221.9 Å *g* = 90.0°	*a* = 220.8 Å *a* = 90.0° *b* = 220.8 Å *b* = 90.0° *c* = 220.8 Å *g* = 90.0°	*a* = 220.2 Å *a* = 90.0° *b* = 220.2 Å *b* = 90.0° *c* = 220.2 Å *g* = 90.0°
Wavelength (Å)	0.8856	0.9999	0.7749
Resolution range (Å)	49.62-3.00 (3.13-3.00)	49.38-2.70 (2.79-2.70)	49.24-2.80 (2.91-2.80)
Total reflections	123628 (15938)	425956 (23149)	1013021 (105836)
Unique reflections	35735 (4385)	49018 (4444)	43652 (4551)
CC1/2[b]	0.992 (0.402)	0.998 (0.395)	0.998 (0.599)
Redundancy	3.5 (3.6)	8.7 (5.2)	23.2 (23.3)
Mean I/σ(I)	6.2 (1.0)	13.5 (1.1)	13.2 (0.9)
Completeness (%)	98.4 (99.9)	99.7 (97.9)	100.0 (100.0)
R_{sym}[b]	0.130 (1.047)	0.110 (1.251)	0.169 (3.609)
R_{pim}[c]	0.101 (0.811)	0.058 (0.920)	0.051 (1.093)
Anomalous completeness (%)			100.0 (100.0
Anomalous multiplicity			11.7 (11.7)
Anomalous CC1/2			0.488
SAD resolution limit (Å)[d]			4.4

[b] $R_{sym} = [\ \Sigma_{hkl}\Sigma_j\ |\ I_{hkl,j} - \langle I_{hkl}\rangle\ |\ ]\ /\ [\ \Sigma_{hkl}\Sigma_j\ I_{hkl,j}\ ]$, where I is the observed intensity for a reflection and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[c] $R_{pim} = [\ \Sigma_{hkl}\{1/(n-1)\}^{1/2}\ \Sigma_j\ |\ I_{hkl,j} - \langle I_{hkl}\rangle\ |\ ]\ /\ [\ \Sigma_{hkl}\Sigma_j\ I_{hkl,j}\ ]$ where I is the observed intensity for a reflection and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[d] the resolution limit for SAD phasing has been selected by evaluating the resolution where *d''/s* falls below 1.2 using the *SHELXC* module of *HKL2MAP*.

Supplementary Table 3. X-ray crystallographic refinement statistics

	GLT25D1/COLGALT1 (Native)	GLT25D1/COLGALT1 (co-crystallized with Mn^{2+}, UDP-a-Gal)	GLT25D1/COLGALT1 (Hg^{2+} Soak)
Refinement			
R_{work}/R_{free}[c]	0.1970/0.2417	0.1846/0.2216	0.2024/0.2390
Number of atoms:	8633	8923	8621
Protein	8519	8681	8503
Ligands	105	106	114
Solvent	9	136	4
Average B-factor $(Å)^2$	92.07	74.93	107.55
Protein	92.22	75.27	107.58
Ligands	81.40	64.46	105.82
Solvent	64.47	54.66	85.53
Structure quality			
RMS bond lengths (Å)	0.003	0.002	0.003
RMS bond angles (°)	0.63	0.52	0.55
Ramachandran stats			
Favored (%)	96.03	96.76	95.74
allowed (%)	3.97	3.14	4.26
outliers (%)	0.00	0.10	0.00
PDB ID	9EVK	9EVJ	9EVL

Supplementary Table 4. Details of mass photometry analysis

GLT25D1/COLGALT1 Sample	Shown in	Mass (kDa)	Intensity Counts	Peak mass fraction
wild type	Figure 2B	142.0 ± 8.8	2991	85%
wild type	Figure 3B	137.0 ± 10.4	4597	83%
Trp158Arg	Figure 3B	61.0 ± 7.6	5213	88%
wild type	Figure 6c	131.0 ± 11.4	3236	70%
wild-type, EDTA treated	Figure 6c	124.0 ± 17.2	6703	84%

Supplementary Table 5. Summary of the outcome of the luminescence-based indirect activity assays, and the HRMS direct assays to evaluate GLT25D1/COLGALT1 enzymatic activity. Values are reported as percentage of enzymatic activity observed for each variant compared to that observed using the wild-type enzyme. Values are expressed as average of minimum triplicate independent measurements, with associated standard deviations.

GLT25D1/COLGALT1 variant	Gal-T activity			
	Luminescence-based indirect assay			HRMS direct assay (% activity)
	K_M (mg mL^{-1})	k_{cat} (min^{-1})	k_{cat}/K_M (mL min^{-1} mg^{-1})	
wild-type	5.61 ± 1.71	67272 ± 6060	11991 ± 3811	100.0 ± 22.1
Trp135Arg	n.d.	n.d.	n.d.	n.d.
Trp158Arg	13.37 ± 4.45	73849 ± 10321	5523 ± 1994	97.2 ± 4.3
Ile267Gln	15.21 ± 7.29	82113 ± 17457	5398 ± 2831	93.3 ± 8.9
Arg315Ala	n.d.	n.d.	n.d.	n.d.
Cys412Ser	16.35 ± 9.20	1748 ± 450	107 ± 66	n.d.
Glu435Ala	n.d.	n.d.	n.d.	n.d.
Asp437Ala	n.d.	n.d.	n.d.	n.d.
Trp495Ala	n.d.	n.d.	n.d.	n.d.
Asp522Ala	n.d.	n.d.	n.d.	n.d.
Asp570Ala	n.d.	n.d.	n.d.	n.d.
wild-type using Mg^{2+}	7.23 ± 1.23	16829 ± 933	2327 ± 416	n.d.

Supplementary Table 6. SEC-SAXS data collection parameters and processing statistics

Data Collection	
Beamline	ESRF BM29
Beam energy (keV)	12.5
Sample-detector distance (m)	2.867
Exposure time (s)	1
Sample cell thickness (mm)	1
Temperature (°)	20
Final q range (nm^{-1})	0.01 – 4.00
Data Analysis	
Points used for Guinier analysis	10 - 33
Guinier qR_g limits	0.77 - 1.28
Guinier R_g (nm)	4.33 ± 0.02
I(0) (mm^{-1})	22.91 ± 0.08
D_{max} (nm)	13.97
MW estimation (Bayesian Inference) (kDa)	131 ± 10

Supplementary Table 7: MD systems setup

	GT1 unbound	GT1 bound	GT2 unbound	GT2 bound
Simulation box dimensions	85.39 Å x 85.39 Å x 85.39 Å 109.47° x 109.47° x 109.47°	83.44 Å x 83.44 Å x 83.44 Å 109.47° x 109.47° x 109.47°	71.80 Å x 71.80 Å x 71.80 Å 109.47° x 109.47° x 109.47°	71.92 Å x 71.92 Å x 71.92 Å 109.47° x 109.47° x 109.47°
Total number of atoms	42292	39396	24407	24542
Total number of water molecules	12524	11541	6756	6789
Salt concentration	150 mM	150 mM	150 mM	150 mM

Supplementary Table 8. Molecular dynamics simulations checklist

Reliability and reproducibility checklist for molecular dynamics simulations *All boxes must be marked YES by acceptance unless "Response not needed if No".	Yes	No	Response (Please state where this information can be found in the text)
1. Convergence of simulations and analysis			
1a. Is an evaluation presented in the text to show that the property being measured has equilibrated in the simulations (*e.g.* time-course analysis)?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods; Main text, *"MD simulations supports distinct roles for GT1 and GT2 domains".*
1b. Then, is it described in the text how simulations are split into equilibration and production runs and how much data were analyzed from production runs?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.
1c. Are there at least 3 simulations per simulation condition with statistical analysis?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.
1d. Is evidence provided in the text that the simulation results presented are independent of initial configuration?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods; Main text, *"MD simulations supports distinct roles for GT1 and GT2 domains".*
2. Connection to experiments			
2a. Are calculations provided that can connect to experiments (*e.g.* loss or gain in function from mutagenesis, binding assays, NMR chemical shifts, J-couplings, SAXS curves, interaction distances or FRET distances, structure factors, diffusion coefficients, bulk modulus and other mechanical properties, *etc.*)?	☒	☐	Main text, *"MD simulations supports distinct roles for GT1 and GT2 domains".* All analyses aim to assess whether the biochemical data are consistent with the simulations, providing a perspective to support experimental findings.
3. Method choice			
3a. Do simulations contain membranes, membrane proteins, intrinsically disordered proteins, glycans, nucleic acids, polymers, or cryptic ligand binding?	☐	☒	Response not needed if **No**
3b. Is it described in the text whether the accuracy of the chosen model(s) is sufficient to address the question(s) under investigation (e.g. all-atom vs. coarse-grained models, fixed charge vs. polarizable force fields, implicit vs. explicit solvent or membrane, force field and water model, etc.)?	☐	☐	Main text, *"MD simulations supports distinct roles for GT1 and GT2 domains".*
3c. Is the timescale of the event(s) under investigation beyond the brute-force MD simulation timescale in this study that enhanced sampling methods are needed?	☐	☒	Main text, *"MD simulations supports distinct roles for GT1 and GT2 domains".*
If **YES**, are the parameters and convergence criteria for the enhanced sampling method clearly stated?	☐	☐	
If **NO,** is the evidence provided in the text?	☒	☐	Main text, *"MD simulations supports distinct roles for GT1 and GT2 domains".*
4. Code and reproducibility			
4a. Is a table provided describing the system setup that includes simulation box dimensions, total number of atoms, total number of water molecules, salt concentration, lipid composition (number of molecules and type)?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.

4b. Is it described in the text what simulation and analysis software and which versions are used?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.
4c. Are other parameters for the system setup described in the text, such as protonation state, type of structural restraints if applied, nonbonded cutoff, thermostat and barostat, etc.?	☒	☐	Methods, *"Molecular dynamics simulations";* Supplementary Methods.
4d. Are initial coordinate and simulation input files and a coordinate file of the final output provided as supplementary files or in a public repository?	☒	☐	Data availability statement. The initial coordinate files, simulation input files, and final output coordinate files are available in the public repository Zenodo at the following DOI: 10.5281/zenodo.15020231
4e. Is there custom code or custom force field parameters?	☐	☒	Response not needed if **No**
If **YES**, are they provided as supplementary files or in a public repository?	☐	☐	

Supplementary Methods

System preparation for molecular dynamics simulations

The molecular systems were prepared based on the crystal structures of the GT1 and GT2 domains of the enzyme. Both domains contain Mn^{2+} and UDP-Gal as cofactors in their crystallographic structures. For the C-terminal domain, only UDP was retained, as electron density for the Gal moiety was insufficient to accurately model it.

Five different systems were prepared for simulation: (1) GT1 domain with cofactors (Mn^{2+} and UDP-Gal); (2) GT1 domain without cofactors (cofactors removed from the structure); (3) GT2 domain with cofactors (Mn^{2+} and UDP); (4) GT2 domain without cofactors (cofactors removed from the structure). For simulations in the absence of cofactors, the Mn^{2+} and UDP-Gal (or UDP in the case of the GT2 domain) were removed from the crystallographic complex structures to mimic the apo form. The parameters and charges for UDP and UDP-Gal were generated using *Antechambe*r tool within the *AmberTools* suite (version 24)[8,9], applying the General AMBER Force Field (GAFF)[10], which provides broad compatibility with organic molecules. Atom types and charges were assigned using the the AM1-BCC2 charge method[11], ensuring compatibility with the AMBER ff19SB[12] force field used for the protein. For Mn^{2+}, the parameters developed by Bradbook and co-workers were applied[13]. These parameters were then incorporated into topology file to allow complete system simulations. The *tleap* utility in the *AmberTools* suite (version 24)[9] was used to add hydrogens to all systems, with the protonation states of titratable side chains of Asp, Glu, Arg, Lys, Tyr, His and Cys residues predicted using the H++[14] program at pH 7.0. In the GT1 domain, cysteine residues were evaluated, and a disulfide bond was added between CYS 228 ans CYS 283. Histidine residues 92, 142 and 319 were protonated at both Nε2 and Nδ1, while remaining histidine residues were protonated at Nε2. In the GT2 domain, no disulfide bonds were present. Histidine residues 401, 476 and 547 were protonated at both Nε2 and Nδ1, histidine 540 was protonated at Nδ1, while all other histidine residues were protonated at Nε2. The glutamate residue 523 was predicted in its neutral form. The systems were solvated in a truncated octahedral box with *TIP3P*[15] water molecules ensuring a minimum distance of 10 Å between the protein and the box edges and neutralized by adding Na^+ or Cl^- counter ions modeled using the parameters developed the by Joung and Cheatham[16] for the use with *TIP3P*[15] water. Additional Na^+ Cl^- ions were added to achieve an ionic concentration of 150 mM, distributed randomly throughout the solvent box to mimic physiological ionic conditions. A detailed summary of the system setup is provided in Supplementary Table 7.

Molecular dynamics simulations

All MD simulations in this study were performed using the *Amber* software package (version 20)[8,9], with the ff19SB force field[12]. The *sander* utility was used for the initial minimization and solvent equilibration phases, while GPU-accelerated *pmemd.cuda* utility was employed for all other stages of the simulations. Each system was simulated in three independent MD replicates (different random seeds to assign initial velocities) (Supplementary Fig. 14-15). Each replicate consisted of a 600-step minimization, a 2.069 ns preproduction time, and a production time of 1 μs. Before molecular dynamics production, all solvated systems underwent two sequential energy minimization steps to relieve steric clashes and unfavorable interactions. In the first minimization step, positional restraints (force constant of 5 kcal/mol/Å²) were applied to all heavy atoms, allowing only hydrogen atoms to move freely. The minimization was performed for a total of 300 steps, with 150 steps of steepest descent followed by 150 steps of conjugate gradient minimization. In the second minimization stage, the entire system was minimized without any positional restraints, also for 300 steps (150 steepest descent steps followed by 150 conjugate gradient steps). Both minimizations were conducted with a non-bonded interaction cutoff of 8.0 Å. Following energy minimization, the systems were equilibrated in the constant Number of particles (N), Volume (V), and Temperature (T), NVT, ensemble using a 9 ps simulated annealing protocol to optimize solvent packing around the solute. To this end, all solute atoms were restrained with a force constant of 10 kcal/mol/Å², allowing only the solvent molecules to equilibrate. The simulated annealing protocol consisted of three phases at different temperatires: heating phase (0–3 ps), the system was gradually heated from 25 K to 400 K; stable phase (3–6 ps), the

temperature was maintained at 400 K to allow for solvent relaxation; cooling phase (6–9 ps), the system was slowly cooled from 400 K to 25 K, facilitating optimal solvent packing around the solute. Temperature control was managed using the Berendsen thermostat[17], by adjusting the temperature coupling parameter: during heating and equilibration phases, a tight coupling with τ = 0.2 ps was applied to ensure stable temperature control, while during the cooling phase, a looser coupling was used, with τ = 2.0 ps initially, decreasing to τ = 1.0 ps for a smooth transition to lower temperatures. All equilibration steps were performed with a non-bonded interaction cutoff of 8.0 Å in periodic boundary conditions, and a 1 fs time step. The systems then were gradually heated from 25 to 300 K over a 20 ps molecular dynamics heating step in the NVT ensemble with periodic boundary conditions and a non-bonded interaction cutoff of 8.0 Å . Positional restraints (5 kcal/mol/Å²) were applied to the Cα atoms of the solute to maintain the overall structure. Temperature control was achieved with the Langevin thermostat, applying a collision frequency of 0.75 ps^{-1}. The SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms, allowing for a 2 fs time step[18].

Following initial heating, the system underwent 2.04 ns of equilibration in the the constant Number of particles (N), Pressure (P), and Temperature (T) NPT ensemble (p = 1 bar) with periodic boundary conditions. The Berendsen barostat[17] was used for pressure control, and the temperature was maintained at 300 K with a Langevin thermostat[19] (collision frequency = 1.0 ps^{-1}). Non-bonded interactions were calculated with a cutoff of 8.0 Å, and SHAKE[18] constraints were applied to hydrogen bonds to allow a 2 fs time step. During equilibration, positional restraints on the Cα atoms were progressively reduced and then removed as follows: during the first 20 ps, restraints on Cα atoms were applied with a force constant of 3.75 kcal/mol/Å²; for the next 20 ps the restraint constant was reduced to 1.75 kcal/mol/Å² to allow for further structural relaxation; the final 2 ns, positional restraints were removed entirely, allowing all solute atoms to move freely. The production stage of MD simulations were conducted in the NPT ensemble with a 2 fs time step. Temperature was maintained at 300 K using a Langevin thermostat[19] (collision frequency of 1.0 ps^{-1}), while the pressure was held constant at 1 bar with Berendsen barostat[17] (relaxation time of 1.0 ps). Electrostatic interactions were computed directly within an 8.0 Å cutoff. Beyond this cutoff, Coulomb interactions were calculated with the particle mesh Ewald method[20]. Bond constraints for hydrogen-containing bonds were applied using the SETTLE algorithm for water molecules and SHAKE[18] for the solute.

To assess the equilibration of the generated trajectories, the time evolution of the RMSD of the protein backbone was analyzed for three independent simulation replicates of each system. The RMSD reached a plateau at different times depending on the system and replicate but did not exceed 25 ns in any case when considering the structured core of the domains, excluding the most flexible terminal regions (the first and last 10 residues for GT1, and the first 10 and last 18 residues for GT2). Therefore, the first 25 ns of the production phase were excluded from further analysis, ensuring that only the equilibrated portion of the trajectories was used.

To confirm equilibration, the decorrelation time of the RMSD in the remaining portion of each trajectory was determined by computing its autocorrelation function. The decorrelation time was determined as the first point where the autocorrelation function is below 1/e, yielding a values ranging from 6 to 8 ns across diferent systems. A value of 10 ns was used to define statistically independent blocks for a block averaging analysis, which confirmed that the block-averaged RMSD values remained stable.

To verify that the simulation results were independent of the initial configuration, RMSD distributions and error estimates were compared across the three replicas of each simulated system. The similarity in RMSD distributions and block-averaged means indicates that all three replicates sampled statistically consistent conformational ensembles. These analyses confirm that the simulations were not biased by their starting structures, and that the observed dynamics are representative of an equilibrated system. Analyses on the conformational ensembles were performed on MD metatrajectories constructed by concatenating the equilibrated part of three independent replicates fo each simulated syste, to enhance sampling of the conformational space. Analyses were carried out using the *cpptraj* program from the *AmberTools* suite (version 24)[8,9].

SUPPLEMENTARY REFERENCES

1. Graphpad Software, L.J., California, USA. Graphpad Prism 7. (Graphpad Software, La Jolla, California, USA). doi:
2. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994). doi: 10.1107/S0907444994003112
3. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7 (2004). doi: 10.1093/nar/gkh340
4. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-4 (2014). doi: 10.1093/nar/gku316
5. Panjkovich, A. & Svergun, D.I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small angle X-ray scattering data. *Bioinformatics* (2017). doi: 10.1093/bioinformatics/btx846
6. Petoukhov, M.V. et al. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **45**, 342-350 (2012). doi: 10.1107/S0021889812007662
7. Iacobucci, C. et al. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nat Protoc* **13**, 2864-2889 (2018). doi: 10.1038/s41596-018-0068-8
8. Case, D.A. et al. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-88 (2005). doi: 10.1002/jcc.20290
9. Case, D.A. et al. AmberTools. *J Chem Inf Model* **63**, 6183-6191 (2023). doi: 10.1021/acs.jcim.3c01153
10. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. & Case, D.A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-74 (2004). doi: 10.1002/jcc.20035
11. Jakalian, A., Jack, D.B. & Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **23**, 1623-41 (2002). doi: 10.1002/jcc.10128
12. Tian, C. et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **16**, 528-552 (2020). doi: 10.1021/acs.jctc.9b00591
13. M. Bradbrook, G. et al. X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding. *Journal of the Chemical Society, Faraday Transactions* **94**, 1603-1611 (1998). doi: 10.1039/A800429C
14. Anandakrishnan, R., Aguilar, B. & Onufriev, A.V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **40**, W537-41 (2012). doi: 10.1093/nar/gks375
15. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983). doi: 10.1063/1.445869
16. Joung, I.S. & Cheatham, T.E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B* **112**, 9020-9041 (2008). doi: 10.1021/jp8001614
17. Berendsen, H.J.C., Postma, J.P.M., Gunsteren, W.F.v., DiNola, A. & Haak, J.R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690 (1984). doi: 10.1063/1.448118
18. Miyamoto, S. & Kollman, P.A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952-962 (1992). doi: 10.1002/jcc.540130805
19. Loncharich, R.J., Brooks, B.R. & Pastor, R.W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **32**, 523-35 (1992). doi: 10.1002/bip.360320508
20. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092 (1993). doi: 10.1063/1.464397