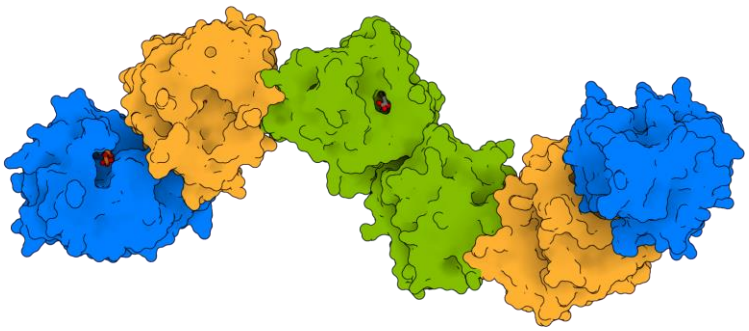
Scuola Universitaria Superiore IUSS Pavia

STRUCTURAL AND BIOCHEMICAL INVESTIGATION OF HUMAN COLLAGEN LYSYL HYDROXYLASES



A Thesis Submitted in Partial Fulfilment of the Requirements

for the Degree of Doctor of Philosophy in

BIOMOLECULAR SCIENCES AND BIOTECHNOLOGY

by

Francesca De Giorgi

November 2020

Scuola Universitaria Superiore IUSS Pavia

STRUCTURAL AND BIOCHEMICAL INVESTIGATION OF HUMAN COLLAGEN LYSYL HYDROXYLASES

A Thesis Submitted in Partial Fulfilment of the Requirements
for the Degree of Doctor of Philosophy in

BIOLOGICAL SCIENCES AND BIOTECHNOLOGY

by

Francesca De Giorgi

Supervisor: Professor Federico Forneris

November 2020

ABSTRACT

Collagen biosynthesis is an intricate pathway which requires multiple post-translational modifications (PTMs) essential for the generation of mature, triple-helical molecules. Among the PTMs, the most important are specific collagen lysine modifications which occur in the endoplasmic reticulum at very initial stages of collagen biogenesis. Lysines are subsequently hydroxylated and glycosylated, which serves as site for the formation of extracellular cross-links, leading to fibrillary or meshwork superstructures in the extracellular matrix (ECM). Enzymes belonging to the family of collagen lysyl hydroxylases (LH or PLOD) catalyze lysine hydroxylation of collagens using Fe^{2+}, 2-oxoglutarate (2-OG), ascorbate and molecular oxygen. In humans, PLOD genes encode for three LH enzyme isoforms: LH1, LH2a/b, and LH3, respectively. Differently from LH1 and LH2a/b, LH3 also displays glycosyltransferase activity, used to further modify collagen performing a specific O-linked conjugation of galactose to hydroxylysines (GalT activity), followed by conjugation of glucose to galactosyl-5-hydroxylysines (GlcT activity), using Mn^{2+} as cofactor. Here, we presented the crystal 3D structures of full-length human LH3 in complex with cofactors and donor substrates. The elongated homodimeric LH3 architecture shows two distinct catalytic sites: the GT domain at the N-terminus and the LH domain at the C-terminus of each monomer, separated by an accessory domain. The GT domain displays both galactosyl- and glucosyl transferase activities, and it also bears distinguishing features compared to other known glycosyltransferases. Various disease-related mutations map in close proximity to the catalytic sites altering the stability of LH3 and/or its normal enzymatic activity. Further, we used structure-based mutagenesis to investigate the broad cooperative network of amino acids in the GT domain. We identified critical "hot spots" leading to selective loss of the GalT activity without affecting the GlcT activity, providing insights into the enzymatic reaction mechanism.We also presented molecular structures of LH3 in complex with UDP-sugar analogs, where the sugar moiety is visible for the first time, thus providing the first structural templates for LH3 glycosyltransferase inhibitors development.

The main aim of this PhD work centered around the structural and biochemical characterization of human LH3 isoform. The results collected were instrumental also for elucidating molecular features associated to other related human LH isoforms. My primary activity included the elucidation of the LH3 biochemical features, and the inference of their significance at the structural level. Also, I developed and standardized the protocols for enzymatic characterization of the enzyme. These methods allowed me

to investigate the multiple activities of LH3 *in vitro*. Collectively, site-directed mutagenesis coupled with enzymology studies allowed defining a comprehensive framework of the complex LH and GT features of LH3, therefore stimulating further studies aiming at elucidating the comprehensive biological significance of this complex multifunctional molecule.

ACKNOWLEDGEMENTS

This work, including my PhD fellowship, was supported by a "My First AIRC Grant" grant (Grant id. 20075) from the Italian Association for Cancer Research (AIRC) to Prof. Forneris. Furhter support obtained by the lab on this project included grants from the Giovanni Armenise-Harvard Career Development Award, the "Programma Rita Levi-Montalcini" from the Italian Ministry of University and Research (MIUR), Cariplo Foundation Grant "COME TRUE" (id. 2015-0768), the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022)—Dept. of Biology and Biotechnology "L. Spallanzani", University of Pavia. We also thank the European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS) for the provision of synchrotron radiation facilities.

I thank Professor Federico Forneris for the supervision and the opportunity to growth professionally and personally in his lab.

My gratitude goes to all the member, present and past, of the AH lab: thank you for the support, the supervision and the friendship. We've been, and still are, a close-knit group, the good environment we established was important to me and allowed me to grow a lot, not only as scientist but also as person.

Dulcis in fundo, the biggest thank goes to my family, who supported me in every moment and made everything possible.

TABLE OF CONTENTS

ABSTRACT v
ACKNOWLEDGEMENTS ix
TABLE OF CONTENTS xi
LIST OF FIGURES xv
LIST OF TABLES xvii
1. Chapter 19
1.1 INTRODUCTION 19
1.2 THE MOST ABUNDANT PROTEIN IN HUMAN BODY: COLLAGEN 19
1.2.1 Collagen structure 20
1.2.2 Collagen biosynthesis 21
1.3 COLLAGEN PROLYL HYDROXYLATION 22
1.3.1 Collagen prolyl 4-hydroxylases 23
1.3.2 Collagen prolyl 3-hydroxylases 25
1.4 COLLAGEN LYSYL HYDROXYLATION 25
1.4.1 Lysyl hydroxylase 1 27
1.4.2 Lysyl hydroxylase 2 27
1.4.3 Lysyl hydroxylase 3 29
1.5 THE FORMATION OF COLLAGEN CROSS-LINKS 30
1.5.1 Lysyl oxidases 30
1.5.2 The LOX-mediated collagen cross-linking 31
1.6 ECM MODULATION AS A HALLMARK OF CANCER 33
1.6.1 Therapeutic strategies for cancer treatment 34
1.7 REFERENCES 35
2. Chapter 45
ABSTRACT 47
2.1 INTRODUCTION 47

2.2 A SIMPLE, ESSENTIAL AND CONSERVED PTM 48
2.3 THE PATH TO GLC-GAL-HYL 48
2.4 THE BIOCHEMISTRY OF UDP-SUGAR TRANSFER 49
2.5 HUMAN COLLAGEN GALACTOSYLTRANSFERASES 50
2.6 COLLAGEN GLUCOSYLTRANSFERASES 51
2.7 THE NEED FOR A COLLAGEN GLUCOSIDASE 53
2.8 GLYCOSYLTRANSFERASE ACTIVITY ASSAYS 54
2.9 CONCLUSIONS 56
2.10 PERSPECTIVES 57
2.11 ACKNOWLEDGEMENTS 59
2.12 CONFLICT OF INTERESTS 59
2.13 AUTHOR CONTRIBUTIONS 59
2.14 REFERENCES 59
3. Chapter 68
ABSTRACT 70
3.1 INTRODUCTION 70
3.2 METHODS 71
3.2.1 Chemicals 71
3.2.2 DNA constructs 72
3.2.3 Recombinant LH3 expression from stable HeLa cell lines 72
3.2.4 Recombinant LH3 expression from transient HEK293 cells 72
3.2.5 Purification of LH3 enzymes 73
3.2.6 LH3 deglycosylation 73
3.2.7 ICP-MS measurements 73
3.2.8 Crystallization of LH3 74
3.2.9 Diffraction data collection and structure refinement 74
3.2.10 SAXS data collection and analysis 75
3.2.11 Determination of LH activity using mass spectrometry 75
3.2.12 Biochemical evaluation of LH activity 76
3.2.13 Biochemical evaluation of GT and GGT enzymatic activities 76
3.2.14 Surface-plasmon resonance 76

3.2.15 Data availability 77
3.3 RESULTS 77
3.3.1 LH3 has three domains encompassing multiple catalytic sites 77
3.3.2 LH3 forms elongated tail-to-tail dimers 80
3.3.3 Structural insights into LH3 glycosyltransferase activity 81
3.3.4 Structural insights into LH3 lysyl hydroxylase activity 85
3.3.5 Excess Fe^{2+} induces a state showing substrate mimicry 87
3.4 DISCUSSION 88
3.5 ACKNOWLEDGEMENTS 91
3.6 REFERENCES 91
4. Chapter 99
ABSTRACT 101
4.1 INTRODUCTION 102
4.2 EXPERIMENTAL PROCEDURES 103
4.2.1 Chemicals 103
4.2.2 Site-directed mutagenesis 104
4.2.3 Recombinant protein expression and purification 104
4.2.4 Crystallization, data collection, structure determination and refinement 104
4.2.5 Evaluation of LH3 GalT/GlcT enzymatic activity 105
4.2.6 Differential Scanning Fluorimetry (DSF) 105
4.3 RESULTS 105
4.3.1 Features and roles of the non-conserved LH3 "glycoloop" 105
4.3.2 LH3 shares features with both retaining and inverting glycosyltransferases 111
4.3.3 Pathogenic LH3 mutations in the LH3 GT domain affect protein folding 114
4.3.4 Molecular structures of LH3 in complex with UDP-sugar analogs provide insights on how glycan moieties are processed inside the LH3 catalytic cavity 114
4.4 DISCUSSION 117
4.5 ACKNOWLEDGEMENTS 120
4.6 REFERENCES 120
CONCLUSIONS 127

LIST OF FIGURES

Figure 1-1 Collagen prolyl-4-hydroxylases.. 24
Figure 1-2 Procollagen Lysine Hydroxylation 26
Figure 1-3 Scheme of LOX reaction mechanism. 32
Figure 2-1 Biochemistry of collagen hydroxylysine glycosylation. 49
Figure 2-2 Detection methods for the measurement of collagen lysine post-translational modifications. 56
Figure 3-1 Molecular architecture of human LH3 79
Figure 3-2 Insights into LH3 glycosyltransferase activity 83
Figure 3-3 Insights into LH3 lysyl hydroxylase activity 87
Figure 3-4 Mapping of disease-related mutations identified in LH enzymes on the LH3 crystal structure 90
Figure 4-1 Features of the LH3 glycosyltransferase (GT) domain. 107
Figure 4-2 Evaluation of the effect of LH3 GT domain mutations in the GT site on glycosyltransferase activities. 108
Figure 4-3 Structural characterization of the LH3 Val80Lys mutant. 109
Figure 4-4 Characterization of UDP-sugar analogs. 115

LIST OF TABLES

Table 4-1: List of glycosyltransferase enzymes used for comparisons with human LH3. 112

1.Chapter

THE COLLAGEN POST-TRANSLATIONAL MODIFICATIONS

1.1 INTRODUCTION

The extracellular matrix (ECM) is a cell-secreted dynamic network of several proteins distinct into a "core matrisome" and a "matrisome-associated" ensemble (Naba, 2012). The "core matrisome" comprises glycoproteins, fibronectins and laminins, proteoglycans, elastin, and collagens. Instead, the matrisome-associated ensemble includes growth factors and cytokines, and ECM-remodelling enzymes, such as lysyl oxidases (LOXs) and matrix metalloproteinases (MMPs) (Hynes, 2012; McKee, 2019). Depending on the organization of the protein ensemble, the ECM can adopt two forms, (i) the interstitial matrix, a hydrated porous 3D lattice surrounding cells, or (ii) the basement membrane, a thin layer of specialized ECM at the interface between epithelial cells and connective tissue. Generally, the collagen fibers and elastin are submerged in a complex network of proteoglycans, while fibronectin anchors the matrix to cell by interacting with integrins, transmembrane proteins keeping the extracellular space in contact with the cytoplasm of the cell. This connection between the extracellular and the intracellular compartments is essential for the regulation of cell behaviour. Indeed, the ECM is a reservoir of factors, such as epidermal growth factor (EGF), fibroblast growth factor (FGF) and transforming growth factor-β (TGFβ) that are released during degradation and remodelling, thus inducing the onset of signal transduction cascades. The ECM does not only provide structural support defining cell form and distribution, but is also capable of influencing cell survival, proliferation, and migration (Engler, 2006; Hadjipanayi, 2009; Hynes, 2009). Numerous severe diseases indeed result from alteration of the ECM; therefore, the various components are targets for development of therapeutic molecules.

1.2 THE MOST ABUNDANT PROTEIN IN HUMAN BODY: COLLAGEN

The major component of the ECM are collagens, representing approximately 30% of the total human body dry weight. Collagen are found mostly in connective tissues, such as skin, cartilage, tendon, and bone (Bielajew, 2020). So far, 28 different collagen types have been described and grouped into subfamilies: (i) fibrillar collagens (I, II, III, V, XI, XXIV, XXVII), representing 90% of the total collagen content; (ii) fibril-associated

collagens with interrupted triple helices (IX, XII, XIV, XVI, XIX, XX, XXI, XXII); (iii) network-forming collagens (IV, VIII, X); (iv) membrane collagens (XIII, XXIII, XXV); (v) multiplexins (XV, XVIII) and collagens VI, VII and XXVIII (Ricard-Blum, 2011). Several molecular isoforms exist as result of splicing events, specific for each tissue and developmental stage, increasing the wide structural and functional diversity across the collagen superfamily.

1.2.1 Collagen structure

Despite the remarkable diversity in molecular and supramolecular organization, tissue distribution and function, collagens show common features. In general, collagens are synthesized as procollagen precursors constituted by a central large continuous triple helix bordered by N- and C-terminal extensions called the N- and C-telopeptides, respectively. Also, two additional sequences typically constituting globular elements flank the telopeptides. The N- and the C-propeptides are enriched in cysteine residues. In collagen I and III, these propeptides are usually removed during the process of maturation into tropocollagen. In other collagen types, namely collagen V and XI, the N-propeptide is instead maintained even in the mature molecule where it sterically limits lateral molecule addition in the extracellular space, influencing heterotypic fibril growth (Ricard-Blum, 2005).

The mature collagen consists in the assembly of left-handed α-chains into homo- or heterotrimers (Sharma, 2017), which are in turn supercoiled with a right-handed twist (Beck, 1998). Chain trimerization is favoured by close packing and hydrogen bonding, resulting in a triple-helical structure which is a unique feature proper of collagens that can range from most of their structure (96% for collagen I) to less than 10% (collagen XII). The α-chains are constituted by $(Z\text{-}X\text{-}Y)_n$ repeating pattern, where glycine residues (Gly) are always in the Z position, buried inside the triple-helix, and any variation leads to severe forms of osteogenesis imperfecta (OI), with the exception of nonfibrillar collagens where the Gly-X-Y pattern is often interrupted (Brodsky, 2008). Conversely, any amino acid can be found in positions X and Y, however proline (Pro) and hydroxyproline (HyP) occupies quite often these two solvent exposed positions respectively, stabilizing the triple-helix. Out of 400 combinations, only 25 triplets are found at a frequency greater than 1%, the reason is that the presence of certain residue either in X or Y position is energetically favoured, for example, phenylalanine (Phe), leucine (Leu), and glutamate (Glu) are preferentially in the X position, while arginine (Arg) and lysine (Lys) are commonly in the Y position. Indeed, the stability of the triple-helix depends on the triplets and the resulting network of intra- and interchain interactions (Ramshaw, 1998; Persikov, 2000).

1.2.2 Collagen biosynthesis

Collagen biosynthesis is an intricate and tightly regulated process. It starts in the rough endoplasmic reticulum (rER), where synthesizing ribosomes translocate single procollagen α-chains into the luminal region. Here, the unfolded procollagen chains are subjected to several post-translational modifications. As first, proline and lysine residues are hydroxylated to hydroxyproline (HyP) and hydroxylysine (HyK). Most of these modifications take place while the procollagen chains are growing on the ribosomes and continue after the release of the complete procollagen chains, until formation of the triple-helical folding prevents any further hydroxylation. HyK residues serve as sites for additional modifications: specific collagen glycosyltransferases, in the ER and the Golgi apparatus, perform O-linked galactosylation of HyK and subsequent glucosylation, producing galactosyl-hydroxylysines (Gal-HyK) and glucosyl-galactosyl-hydroxylysines (Glc-Gal-Hyk), the most peculiar and conserved collagen mono- and disaccharide (Gjaltema, 2017; Salo & Myllyharju 2020).

Once modified, the pro-alpha chains have to assemble. The recognition and association of the three pro-α-chains occur at variable regions called chain recognition sequences (CRS) (Lees, 1997) present in the C-propeptide (COLF1). CRS are the nucleus for triple-helix folding which proceed in a "zipper-like" fashion, along the helical region, toward the N-propeptide (Engel, 1991; Bourish, 2012). This process requires the interplay of several molecular chaperones. Heath shock protein 47, for example, binds to Gly–X–Arg repeats within triple-helical procollagen in the ER, preventing its local unfolding or aggregation, and accelerating triple-helix formation (Ishida, 2011; Oecal, 2016; Ito, 2019). Protein disulfide isomerase (PDI) is another chaperone taking part to this process. PDI associates independently with the C-propeptide of monomeric procollagen chains and catalyses the formation of intra- and interchain disulfide bonds, thus prompting the assembly of trimeric molecules (Wilson, 1998). One of the rate-limiting steps of collagen folding is the cis-trans isomerization of peptide bonds, which is catalysed by peptidyl-prolyl cis-trans isomerases (PPIase). Among collagen PPIases there are FK506-binding protein 22 (FKBP22, encoded by FKBP14), FK506 binding protein 65 (FKBP65, encoded by FKBP10) and cyclophilin B (CypB, encoded by PPIB). FKBP22 acts after proline-4-hydroxylation, while FKBP65 interacts both with unfolded α-chains and triple-helical collagen (Ishikawa, 2008 and 2014), preventing premature association of procollagen chains, as well as aggregation of triple-helical collagen molecules. Instead, CypB might accelerate the initiation of triple-helix assembly by assisting the folding of the C-propeptides (Pyott, 2011).

The triple-helical procollagen molecules are semiflexible and can be incorporated into COP-II vesicles, then can traffic from the ER to the Golgi apparatus, for final modifications. The misfolded molecules are retro-translocated back to the ER inside

COP-I vesicles, whereas the properly folded molecules are incorporated into secretory vesicles and directed to the plasma membrane (McCaughey, 2019; Malhotra, 2015).

Upon secretion in the extracellular space, the propeptides are cleaved by procollagen N-proteinases, belonging to the A Disintegrin And Metalloproteinase with Thrombospondin motifs (ADAMTs) (Bekhouche, 2015), and the procollagen C-proteinase, also termed Bone Morphogenetic Protein-1 (BMP-1) (Vadon-Le Goff, 2015). The proteolytically processed collagen molecules become the substrate of lysyl oxidases (LOXs), a family of enzymes that perform oxidative deamination of ε-group of Lys and HyK residues present in the telopeptides. The resulting aldehydes react with residues from nearby collagen molecules establishing interchain cross-links (Vallet, 2018; Yamauchi, 2019). Finally, the cross-linked collagen molecules spontaneously assemble and accumulate into supramolecular structures in the ECM where can accomplish to their functional roles (for review, Sorushanova, 2019).

In summary, collagen biosynthesis requires a large number of PTMs, many of them being unique to these complex multimeric macromolecules. Collagen PTMs occur primarily inside the cells, where the characteristic triple-helix folds, but continue also outside the cell, allowing the formation of supramolecular structures such as fibrils and mesh networks. The intracellular modifications comprise hydroxylation of proline and lysine residues, and subsequent glycosylation of the latter. The extracellular modifications result in the conversion of procollagen into tropocollagen, pivotal for the maintenance of ECM properties and homeostasis are the oxidation of lysyl and hydroxylysyl residues and the subsequent formation of interchain cross-linking.

1.3 COLLAGEN PROLYL HYDROXYLATION

Collagen consists of repeating $(Gly\text{-}Xaa\text{-}Yaa)_n$ sequences, the most frequent is the Gly-Pro-HyP, where HyP corresponds to hydroxyproline. Hydroxylation of proline residues occurs exclusively on individual unfolded procollagen α-chains and it is essential for the proper assembly and stabilization of the collagen triple helical structure. Proline residues can be hydroxylated either on their C3 or C4 by collagen prolyl-3-hydroxylases (C-P3Hs) and collagen prolyl-4-hydroxylases, respectively, producing 3-hydroxyproline (3-HyP) and 4-hydroxyproline (4-HyP) (Gjaltema, 2016). The 4-hydroxylation of proline is the most common modification in collagen, and it enhances the thermal stability of the triple-helix through the formation of water-mediated hydrogen bonds. Instead, 3-HyP is a rare amino acid, with an occurrence of two residues per α-chain in collagen types I and II, between three to six residues per α-chain of collagen types V and XI, and over 10 residues per α-chain of collagen type IV (Pokidysheva, 2014). The role of 3-Hyp is still unclear. Based on its position in the collagen chain, 3-Hyp might be involved in the formation of short-range hydrogen bonding between triple-helices, having a role in supramolecular

assemblies. (Hudson & Eyre, 2013). Furthermore, 4- and 3-HyP may constitute the binding sites for several proteins, such as integrins, important for the maintenance of collagen and ECM homeostasis. Absence of hydroxylated prolines results in the disruption of interaction interfaces and subsequently in diverse pathological conditions, thus highlighting the importance of such collagen modifications (Rappu, 2019).

1.3.1 Collagen prolyl 4-hydroxylases

Collagen prolyl 4-hydroxylases (C-P4Hs) are multi-domain ER-resident proteins. In humans, they are α2β2-tetramers of about 240 kDa and are distinguished by their α subunit into type I [α(I)2β2], type II [α(II)2β2] and type III [α(III)2β2]. The β subunit is identical in all the isoenzymes, it is a protein disulfide isomerase (PDI) domain necessary to keep soluble the α subunits, preventing their aggregation, and is responsible for the retention of the entire tetramer in the rER through its KDEL retention sequence (Myllyharju, 2003). C-P4H-I is the main form in most cell types and tissues, while C-P4H-II is the major form in osteoblasts, chondrocytes, endothelial and epithelial cells. The critical functions are present on the α subunits, each divided into three domains: the N domain, whose function is still unknown; the middle peptide-substrate-binding (PSB) domain; the C-terminal catalytic domain (CAT) responsible for proline hydroxylation. According to crystallographic structures, the PSB domain is constituted by 100 amino acids organized into 5 helices, folded in two tetratricopeptide (TRP) repeats and an extra solvating helix (Pekkala, 2004). Small variations in the amino acidic sequence can impact on the responsiveness to inhibitors, for example the Poly-(L)-proline has a more powerful inhibitory effect on C-P4H-I but C-P4H-II, due to the replacement of Ile82 and Tyr223, in C-P4H-I, by a Glu and Gln, in C-P4H-II. Also, sequence variations are responsible of differences in substrate affinity. Indeed, type I P4H has higher affinity for (GPP)n sequences than type II, which prefers GPxP sequences. By the way, it has been shown that P4Hs have higher affinity for longer peptides, affinity that decreases after hydroxylation occurs, contributing to the release of the product and preventing sequestration on the enzymes. The CAT domain presents conserved features that classified P4Hs into the superfamily of 2-oxoglutarate and iron dependent dioxygenases (2OGDD). Indeed, the CAT domain is characterized by a double-stranded β-helix core fold termed "jelly-roll" and by a His-Asp-His triad coordinating an Fe^{2+} atom (Hieta, 2003; Anantharajan, 2013; Murthy, 2018). The enzyme requires Fe^{2+}, 2-oxoglutarate (2-OG), molecular oxygen and a reducing agent (preferably ascorbate). During catalysis, the 2-OG is stoichiometrically decarboxylated producing carbon dioxide (CO_2) and succinate, that incorporates one atom belonging to oxygen, as side products. The other oxygen atom is incorporated into the hydroxyl group transferred on proline C-4. The Fe^{2+} undergoes oxidation in Fe^{3+}, thus ascorbate is needed to return the Fe^{3+} into the initial reduced state (Tuderman, 1997; Myllylä, 1997).

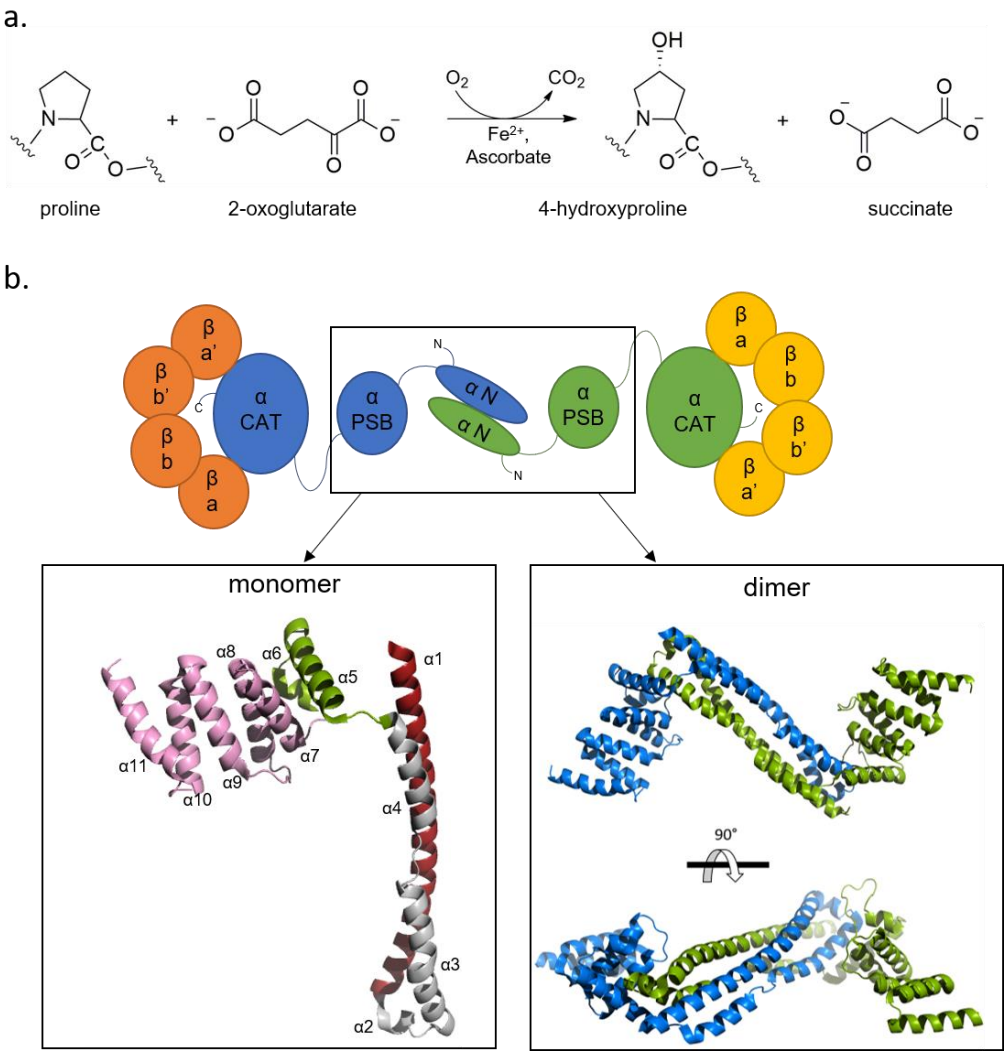


Figure 1-1 Collagen prolyl-4-hydroxylases. a) Schematic representation of the reaction catalyzed by P4Hs. b) Rapresentation of the tetrameric $\alpha_2\beta_2$ organization of P4H enzymes (adapted from Anantharajan et al. (2013)). The α subunit is divided into the N-terminal domain, the peptide binding domain (PSB) and the catalytic domain (CAT) which interacts with the β subunit. According to the crystallographic structure (PDB 4BT8), the N-terminal domain is constituted by four α helices (α1-4) and the α1 (red) is involved in the homodimerization between two α subunits. The N-terminal domain is connected by a linker (green) to the PSB domain, which is formed by five α helices (pink) organized into two TRP sequences. The $(Pro)_9$ peptide binds to the α9, α10 and α11 of the PSB domain (not shown in figure, see PDB 4BTB and Koski et al. 2017).

1.3.2 Collagen prolyl 3-hydroxylases

P3Hs catalyse transfer of the hydroxyl group on C-3 of Pro residue in position Xaa of the Gly-X-HyP triplet, where 4-hydroxylation of the Y proline seems to be a compulsory requirement. P3Hs are members of the leprecan family, in human there are three isoforms, namely P3H1, P3H2 and P3H3 (encoded by LEPRE1, LEPREL1 and LEPREL2, respectively). P3Hs belong to the 2-OGDDs, maintaining the conserved features of the catalytic domain, and also have a P4Hα domain at the C-terminal region (Jarnum, 2004). The fourth member of the leprecan family is P3H4, also known as Synaptonemal complex protein SC65, which possess the PBS domain but lacks the catalytic residues, being an inactive prolyl-3-hydroxylase (Gruenwald, 2014). The first isolated P3H is P3H1, initially identified as leprecan (leucine and proline-enriched proteoglycan) or Gros1 (growth suppressor 1). P3H1 is a prolyl-3-hydroxylase involved in modification of fibrillar collagens, precisely collagen I, and localizes in tendon, cartilage, and skin (Vranka, 2004). Furthermore, P3H1 is part of an ER-complex together with cyclophilin B (CypB) and cartilage-associated protein (CRTAP), the fifth member of the leprecan family sharing 55% sequence identity with P3H4. *In vivo*, the trimeric complex is essential for the hydroxylation of collagen I Pro986 and Pro707 residues (Pokidysheva, 2013), and mutations affecting the expression of these three proteins result in severe or lethal autosomal recessive osteogenesis imperfecta (OI) (Cabral, 2007). P3H1 shares 40% sequence identity with P3H2. Both isoforms possess four TRP domains involved in protein-protein interactions, a central leucine-zipper domain, and a C-terminal KDEL sequence needed for retention in the ER. Despite the similarity with P3H1, P3H2 hydroxylates efficiently collagen IV and is highly expressed in tissues rich of basement membranes (Tiainen, 2008). Mutations in LEPREL1 are associated with ocular abnormalities due to the disruption of basement membranes, enriched in collagen IV, present in eye structures such as the lens capsule and the retina. A single point mutant produces a completely inactive P3H2 and is associated with high myopia (Mordechai, 2011). Recent studies on P3H3 and P3H4 revealed that the two proteins work in complex with LH1 and potentially CypB in the ER, regulating LH1 activity on collagen (Heard and Besio, 2016). Indeed, it has been seen that P3H3-null mice have under-hydroxylated collagen lysines and cross-linking deficiency (Hudson, 2017).

1.4 COLLAGEN LYSYL HYDROXYLATION

Collagen-specific lysine modifications occur in the endoplasmic reticulum at the very initial stages of biosynthesis when the triple helix is still not folded. Alterations of this process generate a sort of "domino effect" resulting in untreatable developmental conditions. Most of the alterations cause overexpression or misfunctioning of enzymes belonging to lysyl hydroxylase family.

As prolyl hydroxylases (PHs), lysyl hydroxylases (LHs) are ER-resident Fe^{2+}, 2-OG-dependent dioxygenases catalysing the conversion of procollagen lysine into 5-hydroxylysine (HyK) residues (LH activity). The newly generated HyK residues are the main sites for collagen glycosylation and cross-linking.

The LH reaction, reported in Figure 1-2, requires 2-OG, Fe^{2+}, molecular oxygen (O_2) and ascorbate. 2-OG is needed both as cofactor and cosubstrate, being decarboxylated into succinate and carbon dioxide (CO_2). The Fe^{2+} is the metallic cofactor, that reacts with the O_2. One molecule of oxygen is incorporated into succinate, the other forms a Fe^{3+}-superoxo intermediate which reacts with the acceptor substrate, lysine, and is transferred as a hydroxyl group, finally producing the HyK. The ascorbate is the reducing cofactor needed to regenerate the Fe^{2+} and allow another reaction cycle. LHs can also catalyse the uncoupled reaction, meaning the decarboxylation of 2-OG in absence of the acceptor

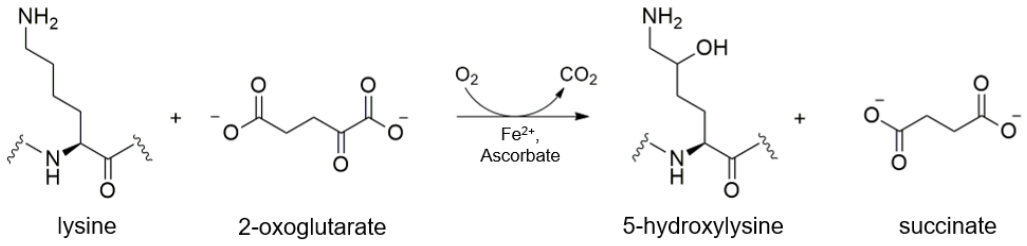


Figure 1-2 Procollagen Lysine Hydroxylation

substrate (Yamauchi, 2012; Gjaltema, 2017; Scietti and Forneris, 2020).

Hydroxylation occurs on lysine at Y position in collagenous Gly-X-Lys triplets; however, not all lysines are modified, and the extent of hydroxylation varies among collagen types, and even across the same collagen molecule. The origin of such variability is still unclear, although an explanation might be found in differences among LH enzymes. In humans, three different genes *PLOD1*, *PLOD2* and *PLOD3* encode for distinct isoenzymes, LH1, LH2 and LH3 respectively, sharing about 70% amino acid sequence identity (Yamauchi, 2012; Gjaltema, 2017; Scietti and Forneris, 2020). Besides the high similarity, the three enzymes act on different portions of the procollagen molecule: LH1 and LH3 modifies lysines in the triple-helical region, whereas LH2 is the only one capable of hydroxylating the telopeptidyl lysines (Takaluoma, 2007). There is no strict requirement for the amino acids flanking the Gly-X-Lys triplet, however *in vitro* studies on collagen-like peptides show a more efficient hydroxylation for certain peptides compared to others, for example, a positively-charged amino acid in proximity of the lysine increases the binding affinity of the LHs for the peptide, promoting enzyme catalysis (Risteli, 2004).

1.4.1 Lysyl hydroxylase 1

LH1 performs preferentially lysyl hydroxylation of the triple helical domains of both collagens I and III. Over 30 mutations in *PLOD1* gene (mainly duplications) result in Ehlers-Danlos syndrome type VIA (EDS VIA), a rare inherited connective tissue disorder characterized by skin hyperextensibility and fragility, joint hypermobility, severe muscle hypotonia and progressive kyphoscoliosis (Rohrbach, 2011). At molecular level, the mechanical instability of tissues found in EDS VIA patients is due to deficiency of LH1, causing collagen under hydroxylation and glycosylation, and thus affecting cross-linking patterns (Walker, 2000; Giunta, 2005). The LH1 enzymatic activity is modulated by a "local molecular ensemble" composed of P3H3, SC65 (P3H4) and CypB (Ishikawa, 2019). P3H3 interacts with SC65, forming a stable complex which can bind to LH1 (Heard and Besio, 2016). P3h3- and Sc65-null mice, showing EDS-like phenotypes, displayed under hydroxylation of specific lysine residues, such as Lys87 and Lys930 in the triple helical region of α1 and α2 chains of collagen I, involved in intermolecular cross-links. The absence of such modifications enables the formation of intrachain cross-links which alters the physiological pattern, resulting in diseased condition. Indeed, P3H3/SC65/LH1 complex is pivotal for the hydroxylation at cross-linking sites (Hudson, 2017). Instead, the LH1 activity at the remaining lysine residues is most likely influenced by CypB. The complex is not well characterized; however, mutations disrupting the interaction between CypB and LH1 lead to under modification of collagen I, again resulting in EDS phenotypes (Ishikawa, 2012). At the moment, structural information is unfortunately incomplete, although essential for better understanding on how and why these proteins associate together. Finally, it is important to underline also that LH enzymes reside in the ER although lacking any retention sequence, thus the interaction of LH1 with P3H3, SC65 and CypB, which possess such feature (KDEL or REEL sequences), might have important implications in LH1 ER-retention.

1.4.2 Lysyl hydroxylase 2

PLOD2 gene, encoding for LH2, is subjected to alternative splicing events, accordingly the exclusion or the inclusion of exon 13 A produces respectively a short variant LH2a (737 amino acids) and a longer variant LH2b (758 amino acids), both lacking ER-retention sequences. LH2b is ubiquitously expressed in several tissues and display hydroxylase activity within telopeptide sequences Gly-X-Lys-, –Ala-X-Lys- and –Ser-X-Lys-, being the only known telopeptidyl lysine hydroxylase (t-LH). LH2 activity generates tissue-specific hydroxylation sites involved in formation of stable collagen cross-links, which are necessary for correct collagen deposition and fibrillogenesis in the ECM (Guo, 2017).

LH1 activity is modulated by formation of ER-chaperone complexes, the same happens for LH2. FKBP65, encoded by FKBP10, is a peptidyl prolyl cis-trans isomerase (PPIase)

reported to prevent collagen aggregation during synthesis and to assist LH2 homodimerization in the ER. The binding outcome differs depending on the involved isoform: the FKBP65-LH2a complex does form but fails in binding and hydroxylation of the collagen telopeptides, whereas the FKBP65-LH2b result in normal t-LH activity. Homozygous mutations in FKBP10 and PLOD2, resulting in depletion of FKBP65 and LH2, cause respectively Bruck syndrome type I and II, rare autosomal recessive disorders both characterized by fractures and congenital contractures with pterygia, severe limb deformity and progressive scoliosis. Such clinical outcome is the result of abnormal bone collagen I, which present under hydroxylation in the telopeptides and a subsequent reduction of pyridinoline cross-links. To note is that triple-helix hydroxylation pattern is not altered, meaning that FKBP65 does not bind to other LH isoforms and is a specific positive regulator of LH2b (Gjaltema, 2016; Chen, 2017). In presence of procollagen type I, actually the FKBP65-LH2 binomial forms a bigger complex together with Hsp47 and BiP (immunoglobulin heavy-chain-binding protein). BiP directly binds to Hsp47 and increases the affinity of the complex for collagen. Hsp47 instead negatively regulates LH2 activity possibly inducing its monomerization. Thus, the competition between FKBP65 and Hsp47 modulates LH2 and ensures the correct hydroxylation of telopeptidyl lysines. Further, defective Hsp47 results in abnormal trafficking of the complex into collagen I containing-vesicles, suggesting a possible mechanism for retention of LH2 into the ER, as occurs for the homologous LH1 (Duran, 2017).

Overexpression of *PLOD2* correlates with tissue fibrosis and metastatic cancer progression, and can be induced by TGF-β1, hypoxia-inducible factor 1α (HIF-1α), profibrotic cytokines and microRNA (van der Slot, 2005; Eisinger-Mathason, 2013; Gilkes, 2013; Kurozumi, 2016). In several solid tumours, high levels of LH2 function as a regulatory switch, primarily controlling the relative abundance of different collagen cross-links and pushing it toward a unique type. As result, altered collagen accumulates and subsequent stiffening of the matrix promotes fibrosis and tumour cell invasion (van der Slot, 2004; Chen, 2015). Moreover, LH2 overexpression leads to unusual active enzyme secretion in the extracellular space: in lung cancer tissues, LH2 colocalizes in the tumour stroma with collagen I which is accessible and can be further modified by LH2, without interfering with lysyl oxidases activity (Chen, 2016). In head-and-neck squamous cell carcinomas (SCCs), LH2 exerts its pro-metastatic role specifically interacting with and modifying integrin β1. The hydroxylation of three lysine residues in the sequence AFNKGEKK has a stabilizing effect on integrin β1, which can be efficiently recruited at the plasma membrane. Here, it dimerizes with integrins α1, α2 and α11 and mediates cell adhesion to collagen molecules (Popov, 2011). The LH2-integrin β1 interaction has been observed in SCCs, but also in lung adenocarcinoma, breast adenocarcinoma and in HeLa cells. The result of such interaction is the accelerated cell motility, promoting cells invasion and metastatic spreading of the tumour (Ueki, 2020).

1.4.3 Lysyl hydroxylase 3

Considered as the ancestor of the LH family, LH3 has unique features and differs from LH1 and LH2 for its enzymatic activity, substrate specificity and cellular localization. LH3 is a multifunctional enzyme displaying LH activity and two additional collagen glycosyltransferase (GT) activities: starting from UDP-sugar in presence of Mn^{2+} cofactor, LH3 catalyses the O-linked conjugation of galactose to hydroxylysines (GalT activity), and the subsequent transfer of glucose to galactosyl-5-hydroxylysines (GlcT activity). The three different activities have been detected *in vitro* (Heikkinen, 2000), and LH3 was found to preferentially modify collagen types II, IV and V. The *in vivo* activities of LH3 have been investigated using genetically modified mouse lines. Mutations affecting only the LH activity result in normally developed mice, bearing defective basement membranes, and altered collagen fibrils in skin and lung (Ruotsalainen, 2006). Instead, reduction of LH3 GlcT activity leads to abnormal collagen IV deposition, causing developmental disorders, ECM abnormalities and defects in collagen and cytoskeleton arrangement (Risteli, 2009). Likewise, total depletion of LH3 is embryonically lethal in mice, due to prevented collagen IV secretion and subsequent intracellular accumulation (Ruotsalainen, 2006). Interestingly, the GT activity of LH3 enables secretion of the ER-retained enzyme. Indeed, LH3 has been found correctly folded and active also in the Golgi apparatus and in the extracellular space (Salo, 2006; Banushi, 2016). Treatment with Brefeldin A (inhibitor of COP-I coated trans-Golgi vesicles) identified two pathways for LH3 secretion in cells: the first one is a non-conventional route which bypasses the Golgi-apparatus, resulting in extracellular LH3 bearing non-processed high-mannose N-glycosylation; the second route instead requires the formation of a Golgi complex (Wang, 2012). The LH3 interacts with VIPAR and VPS33B that, in association with RAB10 and RAB25, drive LH3 trafficking trough the trans-Golgi network and delivery into collagen IV carriers (CIVC), also regulating the post-Golgi sorting process. VPS33B and VIPAR deficiency reduces LH3-dependent post-translational modification of collagen IV, followed by abnormal deposition of the ECM (Banushi, 2016). Overall, these findings open questions about possible compartmentalisation of the LH3 activities, suggesting different and relevant roles in the Golgi-apparatus and in extracellular space that must be further investigated.

Our group determined the first three-dimensional structure of full-length human LH3 in complex with cofactors (PDB 6fxr; the paper by Scietti et al. 2018 is presented in Chapter 3 of this thesis). This result constitutes a major milestone: combining the structural knowledge with biochemical information, it was possible to generate homology models for LH1 and LH2 and map several disease-causing mutations in all isoforms. Most of the mutations have been manually annotated and collected on SiMPLOD (the Structurally-integrated database for Mutations of *PLOD* genes) and can be visualized directly on the available molecular models (Scietti and Campioni, 2019). Thus, mutagenesis studies

became more accurate in the identification of mutation sites and in the interpretation of resulting phenotypes (an example is reported in Chiapparino et al. 2019, Chapter 4 of this thesis). Further, the structural information can advance the research of efficient therapeutic drugs for the treatment of severe connective tissue disorders, fibrosis and solid tumors, speeding up the rational design of structure-based candidate molecules able to bind LH enzymes and to modulate their activity.

1.5 THE FORMATION OF COLLAGEN CROSS-LINKS

Once in the extracellular space, the secreted collagen molecules have to assemble in supramolecular structures, as example collagen type I molecules are packed in parallel and staggered with one another at a distance of 67 nm, forming fibrils. As already mentioned, the fibrillogenesis requires the formation of stable collagen cross-links in a process tightly regulated in terms of initiation and maturation (Yamauchi, 2012).

1.5.1 Lysyl oxidases

Collagen cross-linking is initiated by the cleavage of collagen N- and C- propeptides that makes collagen molecules accessible to enzyme of the lysyl oxidase family. This family is composed by five members, lysyl oxidase (LOX) and four LOX-like enzymes (LOXL1,2,3 and 4). Evolutionary studies grouped LOX and LOXL1, sharing a common ancestor, into a different subgroup compared to LOXL2, LOXL3 and LOXL4 enzymes (Grau-Bovè, 2015). LOX and LOXL1 are secreted as quiescent enzymes that become active upon proteolytic cleavage. The pro-LOX is a 50 kDa protein and is cleaved by BMP1 proteinase, the same involved in processing of procollagen C-propeptide (Trackman, 2016). The pro-LOX cleavage releases a 20 kDa propeptide (LOX-PP) identified as "matricryptin", a bioactive ECM fragment that accomplishes several functions different from the parental molecule. Indeed, LOX-PP is essential for the secretion of pro-LOX and favours the cleavage by interacting with fibronectin on the cell surface. Also, LOX-PP exhibits several functions in several compartments, and it has been found in the nucleus of different cell types where it modulates cell signalling pathways and cell proliferation (Vallet, 2018). Mature LOX is a 30 kDa enzyme whose activity is located in the C-terminal domain, conserved among all the family members. The catalytic domain presents a histidine-rich Cu^{2+} binding site and a lysine tyrosyl quinone (LTQ) cofactor. The latter is the result of an autocatalytic Cu^{2+}-assisted oxidative deamination of a covalent bound Lys of the active site, and on turn LTQ provides the carbonyl moiety needed for catalysis (Lucero, 2006; Yamauchi, 2012). Lysyl oxidases are characterized by highly variable N-terminal regions and also LOXL2-3-4 bear scavenger receptor cysteine-rich (SRCR) sequences. This variability contributes to the establishment of substrate specificity and a diversification of lysyl oxidase roles. Indeed, LOX oxidizes *in vivo* Lys residues embedded in Asp-Glu-Lys-Ser sequences in the N-telopeptide region of collagen I, whereas LOXL2 prefers collagen IV, and LOXL1 works on elastin

(Trackman, 2016). LOX and LOXLs are copper-dependent enzymes and catalyse the oxidative deamination of the primary amine in Lys and HyK residues, forming respectively allysine and hydroxyallysine, and ammonia and hydrogen peroxide as side products. Lysyl oxidases do not directly form cross-linking, however the introduction of the carbonyl group changes the chemical microenvironment around the modified residues, as the newly formed aldehydes are highly reactive, they can spontaneously condense with Lys and HyK residues.

Cancer cells overexpress LOX and LOXLs via HIF-1α or TGF-β induction, as for LH2, resulting in increased proliferative rates and invasiveness. Inhibition of LOX reduces the metastatic potential of tumour cells, preventing pro-inflammatory signal transduction pathways leading to fibrosis, fibronectin expression and formation of altered collagen cross-links. For these reasons, LOXs are target for the development of drugs effective in cancer therapy (Nishioka,2012).

1.5.2 The LOX-mediated collagen cross-linking

Primary interchain cross-links form between telopeptidyl aldehydic residues and Lys/HyK in the helical region of two different neighbouring collagen molecules. Further maturation leads to formation of di-, tri- and tetravalent cross-links, tying together two or three different collagen molecules. An intrachain cross-link can occur as result of condensation of two aldehydic residues, forming an aldol condensation product (ACP), that can mature into an intrachain cross-link. Two different cross-linking pathways can be identified: lysine aldehyde collagen cross-linking (LCC) and the hydroxylysine aldehyde collagen cross-linking (HLCC) pathways. LCCs are prominent in soft tissues, whereas HLCCs are more stable and provide tensile strength in skeletal tissues, in which are predominantly found (Yamauchi & Sricolpech, 2012; Yamauchi, 2017).The fine balance between LCCs and HLCCs is essential for proper ECM deposition, ensuring tissue integrity and homeostasis, thus their formation is tightly controlled at level of quantity, quality, and maturation. Quality and quantity of cross-links mostly depend on the activity of LH2b, generating the telopeptidyl HyK, and lysyl oxidase, converting the HyK into the respective reactive aldehyde. As previously discussed, alterations of LH2b and LOXs activities cause perturbation of this equilibrium inducing reduction or increase in HLCCs content. HLCCs firstly form from the condensation of telopeptidyl hydroxyallysine with the ε-amino group of juxtaposed triple-helical Lys or HyK forming divalent cross-link (deH-HLNL and deH-DHLNL) that can mature into a trivalent one: pyridinoline (Pyr), deoxypyridinoline (d-Pyr), pyrrole (Prl) and deoxy-pyrrole (d-Prl) cross-links. The maturation process is mostly regulated by glycosylation of HyK residues (it will be discussed much in detail in the following dedicated chapters). Divalent cross-links are mono or diglycosylated, whereas in trivalent cross-links such modification is nearly absent, with the exception of few Gal-HyK residues. Probably the bulky structure and the

hydrophilic character of such modification would impair the correct formation of such cross-links impeding proper packing of collagen molecules (Terajima, 2014). Dramatic rearrangements of the ECM are the consequence of LCCs to HLCCs switch, leading to severe fibrosis, a proper characteristic of malignant and highly metastatic solid tumours.

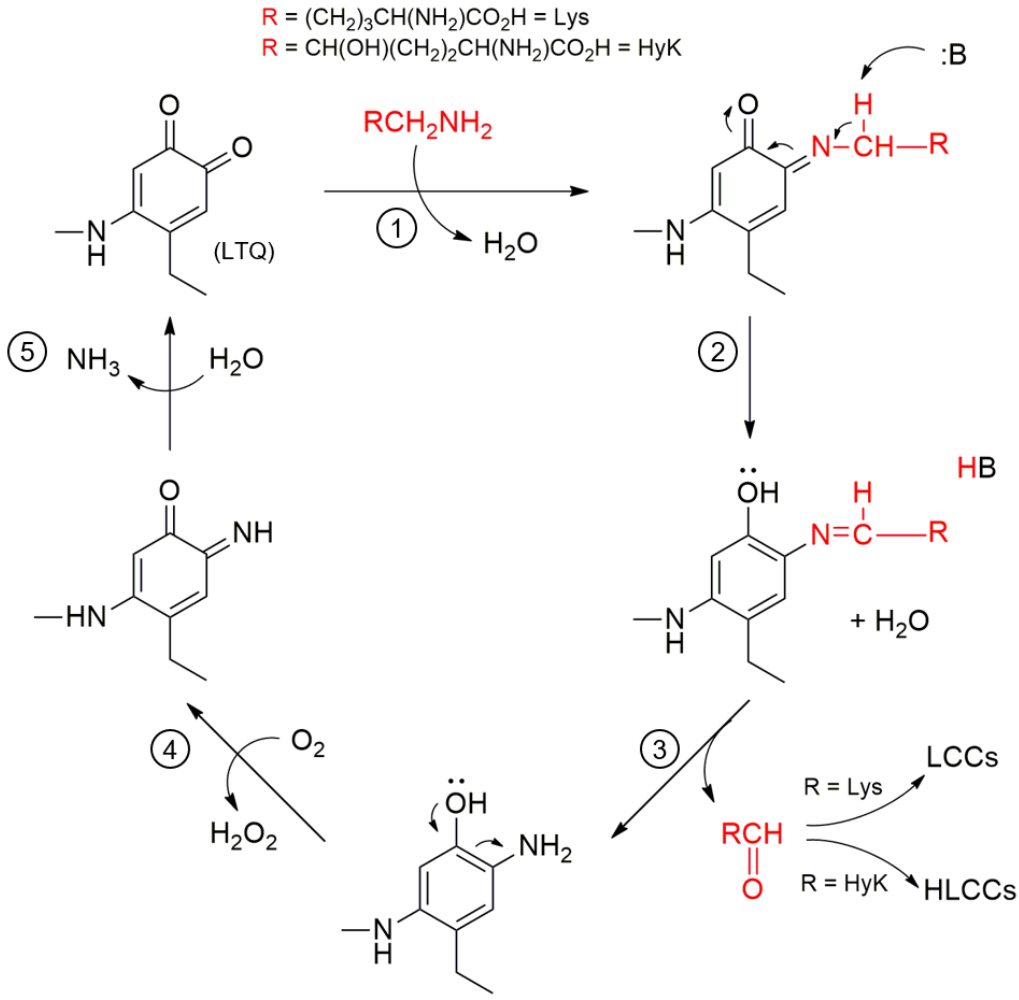


Figure 1-3 Scheme of LOX reaction mechanism. 1) The ε-amino group of Lys or HyK (RCH_2NH_2) condenses with a carbonyl of the enzyme bound LTQ cofactor. 2) A general base (His 303 in LOX) abstracts an electron from C-6 of Lys or HyK, and the transfer of 2 electrons on LTQ generates a reduced peptidyl lysyl tyrosyl aminoquinol. 3) The hydrolysis of the imine linkage releases the (hydroxy)allysine that will be involved in cross-linking, LCC when R = Lys or HLCC when R = HyK. 4) The molecular oxygen abstracts 2 electrons from the aminoquinol, releasing hydroxgen peroxide. 5) Then, hydrolysis of the quinoneimine regenerates the LTQ cofactor with production of ammonia. (Adapted from Oldfield, 2018)

1.6 ECM MODULATION AS A HALLMARK OF CANCER

A tumour is any abnormal proliferation of cells, which starts with a single cell accumulating genetic mutations altering its normal behaviour. Thus, the genetic instability of the progenitor cell triggers the abnormal proliferation leading to primary tumour growth. Tumour aggressiveness depends on its ability to migrate: a tumour is defined "benign" as long as it remains confined to its original location. When the primary tumour becomes capable of invading the surrounding normal tissues and spreads into the body, exploiting the circulatory and lymphatic system, it becomes "malignant". A tumour is a complex assembly composed of a microenvironment surrounding a stroma. The tumour microenvironment (TME) comprises the tumour stroma itself, a network of blood vessels, secreted factors, and extracellular matrix proteins. The tumour stroma consists of a heterogeneous population of cancer cells interacting with a variety of resident and infiltrating host cells, such as immune inflammatory cells, pericytes, endothelial cells and cancer-associated fibroblasts (CAFs).

The acquisition of specific capabilities, defined as the hallmarks of cancer, drives development, growth and spreading of malignancy. Cancer cells must sustain highly proliferative states, evading from growth suppression and cell death mechanisms and acquire replicative immortality. Further, the hallmarks comprise also the capabilities of induce angiogenesis and initiate invasion into the neighbouring tissues, the dysregulation of normal cellular metabolic circuits, the avoidance of immune destruction and the development of chronic inflammation (Hanahan, 2011). Establishment of such hallmarks can be reached through different mechanisms depending on cancer type and developmental stages; however, it always implies the alteration of normal cell metabolic and signalling pathways. In several cancers, such as pancreatic, gastric, and colorectal cancers, the normal TGF-β signalling pathway is redirected from suppressing cell division toward the induction of epithelial-to-mesenchymal transition (EMT), promoting cell proliferation and migration. The EMT is a feature proper of highly malignant tumour cells, in such transition the epithelial cells lose their polarity and cell-to-cell adhesion capability and acquire motility proper of mesenchymal cells. TGF-β is frequently produced in high amounts by stromal cells and can stimulate angiogenesis and suppress the immune system (Colak, 2017). Growing tumors need the continuous formation of new vasculature in order to receive oxygen and nutrients necessary for their growth and survival. The angiogenesis inducers are the vascular endothelial growth factor (VEGF) and fibroblast growing factor (FGF), and their expression is upregulated by tumour cells in presence of hypoxia and fibrosis (Hori, 2017).

The hallmarks of cancer are directly influenced by the deposition and remodelling of a 'provisional matrix', which contributes to define tumour histopathology and behaviour, mostly affecting its progression, promoting cellular transformation and metastasis. The cancer-associated ECM also deregulates behaviour of stromal cells, facilitates angiogenesis induction, and stimulates a persistent state of chronic inflammation (Pickup, 2014).

Tumor cells secrete their own ECM components, in particular CAFs which promote intratumoral fibrosis. CAFs are tissue resident fibroblasts and/or stellate cells (pancreas and liver) undergoing an activation process upon stimulation by tumour-derived stimuli, namely signalling molecules such as TGF-β, lysophosphatidic acid, FGF and interleukin-1 (IL-1) and 6 (IL-6). CAFs are divided into (i) secretory and immune modulatory CAFs, producing cytokines essential for the recruitment of macrophages and promoting an immune suppressive effect, (ii) contractile and matrix producing (Park, 2020). CAFs are able to synthesise and remodel the ECM modulating the expression of matrix remodelling enzymes, in particular by overexpressing LH2b and/or LOX enzymes CAFs secrete a heavily cross-linked collagen, presenting a higher amount of HLCCs, which deposits on a thick scaffold of fibronectin molecules. The resulting microenvironment is characterized by hypoxia, and releases of factors such as HIF-1α in turn stimulates the overexpression of LH2b and LOX, thus generating a self-sustaining positive feedback loop perpetuating a fibrotic state (Yamauchi, 2018). Progressive stiffening of the matrix creates compressive remodelling of the TME that alters the fluid transport, thus CAFs necessarily produce VEGF and FGF to induce angiogenesis and promote irroration of the thick ECM, but also the migration of invasive cancer cells (Ferruzzi, 2019). Cancer cells acquire motility capabilities through the EMT, then the formation of invadopodia directs the invasion of the surrounding tissues through the release of metalloproteases (MMPs), which perform the remodelling of matrix scaffolding proteins. In lung and breast cancers, through a continuous process of synthesis and degradation, collagen type I accumulates and forms "highways" along which invasive cancer cells can adhere and migrate (Eisinger-Mathason, 2013; Gilkes, 2013; Pankova, 2016; Park, 2020). Such collagen highways prompt the dissemination of cancer cells inside the body, and favour formation of tumour metastasis at secondary sites.

1.6.1 Therapeutic strategies for cancer treatment

Desmoplasia characterizes aggressive solid tumors and is predictive of poor prognosis. Overexpression of LH2b and LOX correlates with tumour progression and metastatisation. Inhibition of collagen cross-linking reduces ECM stiffening and prevents tumour metastasis, therefore a potential strategy to arrest malignancy is the development of therapeutic molecules targeting matrix remodelling enzymes. Minoxidil has been used to inhibit LH2 in preclinical models. Minoxidil regulates *PLOD2* transcription, thus

lowering lysyl hydroxylation in telopeptides also results in normal HLCC/LCC ratio and deposition of normally cross-linked collagen molecules (Zuurmond, 2005). LH2 might also be targeted indirectly with tacrolimus, an inhibitor of FKBP65. In this context, the 3D structure of LH3 can also serve to generate a reliable model of LH2b, which can be used for developing selective inhibitors targeting the LH catalytic domain or its dimerization interface. The pro-metastatic effect of collagen modifying enzymes could be counteracted also by targeting HIF-1α and TGF-β1, which promotes migratory, invasive, and adhesive capabilities of cancer cells via induction of EMT. Indeed, inhibitors of hypoxia effectively block metastatisation of breast and lung cancers (Wong, 2012), whereas neutralizing antibodies and ligand traps have been developed in order to revert the effect of altered TGF-β1 signalling and suppress tumour cell proliferation (Akhurst, 2017).

1.7 REFERENCES

1. Akhurst RJ. Targeting TGF-β Signaling for Therapeutic Gain. Cold Spring Harb Perspect Biol. 2017;9(10):a022301. Published 2017 Oct 3. doi:10.1101/cshperspect.a022301

2. Anantharajan J, Koski MK, Kursula P, Hieta R, Bergmann U, Myllyharju J, Wierenga RK. The structural motifs for substrate binding and dimerization of the α subunit of collagen prolyl 4-hydroxylase. Structure. 2013 Dec 3;21(12):2107-18. doi: 10.1016/j.str.2013.09.005. Epub 2013 Oct 24. PMID: 24207127.

3. Banushi B, Forneris F, Straatman-Iwanowska A, Strange A, Lyne AM, Rogerson C, Burden JJ, Heywood WE, Hanley J, Doykov I, Straatman KR, Smith H, Bem D, Kriston-Vizi J, Ariceta G, Risteli M, Wang C, Ardill RE, Zaniew M, Latka-Grot J, Waddington SN, Howe SJ, Ferraro F, Gjinovci A, Lawrence S, Marsh M, Girolami M, Bozec L, Mills K, Gissen P. Regulation of post-Golgi LH3 trafficking is essential for collagen homeostasis. Nat Commun. 2016 Jul 20;7:12111. doi: 10.1038/ncomms12111. PMID: 27435297; PMCID: PMC4961739

4. Beck K, Brodsky B. Supercoiled protein motifs: the collagen triple-helix and the alpha-helical coiled coil. J Struct Biol. 1998;122(1-2):17-29. doi: 10.1006/jsbi.1998.3965. PMID: 9724603

5. Bekhouche M, Colige A. The procollagen N-proteinases ADAMTS2, 3 and 14 in pathophysiology. Matrix Biol. 2015 May-Jul;44-46:46-53. doi: 10.1016/j.matbio.2015.04.001. Epub 2015 Apr 8. PMID: 25863161.

6. Bielajew, B. J., Hu, J. C., & Athanasiou, K. A. (2020). Collagen: quantification, biomechanics and role of minor subtypes in cartilage. Nature Reviews Materials. doi:10.1038/s41578-020-0213-1

7. Bourhis, J. M. et al. Structural basis of fibrillar collagen trimerization and related genetic disorders. Nat. Struct. Mol. Biol. 19, 1031–1036 (2012)

8. Brodsky B, Thiagarajan G, Madhan B, Kar K. Triple-helical peptides: an approach to collagen conformation, stability, and self-association. Biopolymers. 2008 May;89(5):345-53. doi: 10.1002/bip.20958. PMID: 18275087.

9. Colak, S., & ten Dijke, P. (2017).Targeting TGF-β Signaling in Cancer. Trends in Cancer, 3(1), 56–71.doi:10.1016/j.trecan.2016.11.008

10. Cabral WA, Chang W, Barnes AM, Weis M, Scott MA, Leikin S, Makareeva E, Kuznetsova NV, Rosenbaum KN, Tifft CJ, Bulas DI, Kozma C, Smith PA, Eyre DR, Marini JC. Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nat Genet. 2007 Mar;39(3):359-65. doi: 10.1038/ng1968. Epub 2007 Feb 4. Erratum in Nat Genet. 2008 Jul;40(7):927. PMID: 17277775; PMCID: PMC7510175.

11. Daniel P. Keeley, Eric Hastie, Ranjay Jayadev, Laura C. Kelley, Qiuyi Chi, Sara G. Payne, Jonathan L. Jeger, Brenton D. Hoffman, David R. Sherwood, Comprehensive Endogenous Tagging of Basement Membrane Components Reveals Dynamic Movement within the Matrix Scaffolding, Developmental Cell, Volume 54, Issue 1, 2020, Pages 60-74.e7, ISSN 1534-5807, https://doi.org/10.1016/j.devcel.2020.05.022.

12. Engel, J. & Prockop, D. J. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. Ann. Rev. Biophys. Biophys. Chem. 20, 137–152 (1991)

13. Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix elasticity directs stem cell lineage specification. Cell. 2006 Aug 25;126(4):677-89. doi: 10.1016/j.cell.2006.06.044. PMID: 16923388.

14. Ferruzzi, J et al. "Compressive Remodeling Alters Fluid Transport Properties of Collagen Networks - Implications for Tumor Growth." Scientific reports vol. 9,1 17151. 20 Nov. 2019, doi:10.1038/s41598-019-50268

15. Gjaltema RA, Bank RA. Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease. Crit Rev Biochem Mol Biol. 2017 Feb;52(1):74-95. doi: 10.1080/10409238.2016.1269716. Epub 2016 Dec 23. PMID: 28006962.

16. Grau-Bové, X., Ruiz-Trillo, I. & Rodriguez-Pascual, F. Origin and evolution of lysyl oxidases. Sci Rep 5, 10568 (2015). https://doi.org/10.1038/srep10568

17. Gruenwald K, Castagnola P, Besio R, et al. Sc65 is a novel endoplasmic reticulum protein that regulates bone mass homeostasis. J Bone Miner Res. 2014;29(3):666-675. doi:10.1002/jbmr.2075

18. Guo, H., Tsai, C., Terajima, M. et al. Pro-metastatic collagen lysyl hydroxylase dimer assemblies stabilized by Fe2+-binding. Nat Commun 9, 512 (2018). https://doi.org/10.1038/s41467-018-02859-z

19. Hadjipanayi, E., Mudera, V., & Brown, R. A. (2009). Close dependence of fibroblast proliferation on collagen scaffold matrix stiffness. Journal of Tissue Engineering and Regenerative Medicine, 3(2), 77–84. doi:10.1002/term.136

20. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar 4;144(5):646-74. doi: 10.1016/j.cell.2011.02.013. PMID: 21376230.

21. Hieta R, Kukkola L, Permi P, Pirilä P, Kivirikko KI, Kilpeläinen I, Myllyharju J. The peptide-substrate-binding domain of human collagen prolyl 4-hydroxylases. Backbone assignments, secondary structure, and binding of proline-rich peptides. J Biol Chem. 2003 Sep 12;278(37):34966-74. doi: 10.1074/jbc.M303624200. Epub 2003 Jun 24. PMID: 12824157.

22. Heard ME, Besio R, Weis M, Rai J, Hudson DM, Dimori M, Zimmerman SM, Kamykowski JA, Hogue WR, Swain FL, Burdine MS, Mackintosh SG, Tackett AJ, Suva LJ, Eyre DR, Morello R. Sc65-Null Mice Provide Evidence for a Novel Endoplasmic Reticulum Complex Regulating Collagen Lysyl Hydroxylation. PLoS Genet. 2016 Apr 27;12(4):e1006002. doi: 10.1371/journal.pgen.1006002. PMID: 27119146; PMCID: PMC4847768.

23. Hori Y, Ito K, Hamamichi S, Ozawa Y, Matsui J, Umeda IO, Fujii H. Functional Characterization of VEGF- and FGF-induced Tumor Blood Vessel Models in Human Cancer Xenografts. Anticancer Res. 2017 Dec;37(12):6629-6638. doi: 10.21873/anticanres.12120. PMID: 29187438.

24. Hudson DM, Eyre DR. Collagen prolyl 3-hydroxylation: a major role for a minor post-translational modification? Connect Tissue Res. 2013;54(4-5):245-251. doi:10.3109/03008207.2013.800867

25. Hudson DM, Weis M, Rai J, Joeng KS, Dimori M, Lee BH, Morello R, Eyre DR. P3h3-null and Sc65-null Mice Phenocopy the Collagen Lysine Under-hydroxylation and Cross-linking Abnormality of Ehlers-Danlos Syndrome Type VIA. J Biol Chem. 2017 Mar 3;292(9):3877-3887. doi: 10.1074/jbc.M116.762245. Epub 2017 Jan 23. PMID: 28115524; PMCID: PMC5339768.

26. Huynh LK, Hipolito CJ, Ten Dijke P. A Perspective on the Development of TGF-β Inhibitors for Cancer Treatment. Biomolecules. 2019;9(11):743. Published 2019 Nov 17. doi:10.3390/biom9110743

27. Hynes RO, Naba A. Overview of the matrisome- an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol. 2012;4(1):a004903. Published 2012 Jan 1. doi:10.1101/cshperspect.a004903

28. Hynes RO. The extracellular matrix: not just pretty fibrils. Science. 2009;326:1216–1219.

29. Ishida Y, Nagata K. Hsp47 as a collagen-specific molecular chaperone. Methods Enzymol. 2011;499:167-82. doi: 10.1016/B978-0-12-386471-0.00009-2. PMID: 21683254.

30. Ishikawa Y, Bächinger HP. A substrate preference for the rough endoplasmic reticulum resident protein FKBP22 during collagen biosynthesis. J Biol Chem. 2014;289(26):18189-18201. doi:10.1074/jbc.M114.561944

31. Ishikawa Y, Vranka J, Wirz J, Nagata K, Bächinger HP. The rough endoplasmic reticulum-resident FK506-binding protein FKBP65 is a molecular chaperone that interacts with collagens. J Biol Chem. 2008 Nov 14;283(46):31584-90. doi: 10.1074/jbc.M802535200. Epub 2008 Sep 10. PMID: 18786928.

32. Ishikawa Y., Vranka J. A., Boudko S. P., Pokidysheva E., Mizuno K., Zientek K., Keene D. R., Rashmir-Raven A. M., Nagata K., Winand N. J., Bächinger H. P. (2012) Mutation in cyclophilin B that causes hyperelastosis cutis in American Quarter Horse does not affect peptidylprolyl cis-trans isomerase activity but shows altered cyclophilin B-protein interactions and affects collagen folding. J. Biol. Chem. 287, 22253–22265

33. Ito S, Nagata K. Roles of the endoplasmic reticulum-resident, collagen-specific molecular chaperone Hsp47 in vertebrate cells and human disease. J Biol Chem. 2019;294(6):2133-2141. doi:10.1074/jbc.TM118.002812

34. Järnum S, Kjellman C, Darabi A, Nilsson I, Edvardsen K, Aman P. LEPREL1, a novel ER and Golgi resident member of the Leprecan family. Biochem Biophys Res Commun. 2004 Apr 30;317(2):342-51. doi: 10.1016/j.bbrc.2004.03.060. PMID: 15063763.

35. Koski M. K., Anantharajan J., Kursula P., Dhavala P., Murthy A.V. et al. Assembly of the elongated collagen prolyl 4-hydroxylase α2β2 heterotetramer around a central α2 dimer. Biochem J 1 March 2017; 474 (5): 751–769. doi: https://doi.org/10.1042/BCJ20161000

36. Lees, J. F., Tasab, M. & Bulleid, N. J. Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. EMBO J. 16, 908–916 (1997)

37. Lucero HA, Kagan HM. Lysyl oxidase: an oxidative enzyme and effector of cell function. Cell Mol Life Sci. 2006 Oct;63(19-20):2304-16. doi: 10.1007/s00018-006-6149-9. PMID: 16909208.

38. Malhotra V, Erlmann P, Nogueira C. Procollagen export from the endoplasmic reticulum. Biochem Soc Trans. 2015 Feb;43(1):104-7. doi: 10.1042/BST20140286. PMID: 25619253.

39. McCaughey J, Stevenson NL, Cross S, Stephens DJ. ER-to-Golgi trafficking of procollagen in the absence of large carriers. J Cell Biol. 2019;218(3):929-948. doi:10.1083/jcb.201806035

40. McKee, T.J., Perlman, G., Morris, M. et al. Extracellular matrix composition of connective tissues: a systematic review and meta-analysis. Sci Rep 9, 10542 (2019). https://doi.org/10.1038/s41598-019-46896-0

41. Mordechai S, Gradstein L, Pasanen A, et al. High myopia caused by a mutation in LEPREL1, encoding prolyl 3-hydroxylase 2. Am J Hum Genet. 2011;89(3):438-445. doi:10.1016/j.ajhg.2011.08.003

42. Murthy AV, Sulu R, Koski MK, et al. Structural enzymology binding studies of the peptide-substrate-binding domain of human collagen prolyl 4-hydroxylase (type-II): High affinity peptides have a PxGP sequence motif. Protein Sci. 2018;27(9):1692-1703. doi:10.1002/pro.3450

43. Myllyharju J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix Biol. 2003 Mar;22(1):15-24. doi: 10.1016/s0945-053x(03)00006-4. PMID: 12714038.

44. Myllylä R, Tuderman L, Kivirikko KI. Mechanism of the prolyl hydroxylase reaction. 2. Kinetic analysis of the reaction sequence. Eur J Biochem. 1977 Nov 1;80(2):349-57. doi: 10.1111/j.1432-1033.1977.tb11889.x. PMID: 200425.

45. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and *in vivo* characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics. 2012;11(4):M111.014647. doi:10.1074/mcp.M111.014647

46. Oldfield RN, Johnston KA, Limones J, Ghilarducci C, Lopez KM. Identification of Histidine 303 as the Catalytic Base of Lysyl Oxidase via Site-Directed Mutagenesis. Protein J. 2018 Feb;37(1):47-57. doi: 10.1007/s10930-017-9749-3. PMID: 29127553.

47. Oecal S, Socher E, Uthoff M, Ernst C, Zaucke F, Sticht H, Baumann U, Gebauer JM. The pH-dependent Client Release from the Collagen-specific Chaperone HSP47 Is Triggered by a Tandem Histidine Pair. J Biol Chem. 2016 Jun 10;291(24):12612-26. doi: 10.1074/jbc.M115.706069. Epub 2016 Apr 19. PMID: 27129216; PMCID: PMC4933464.

48. Pankova D, Chen Y, Terajima M, Schliekelman MJ, Baird BN, Fahrenholtz M, Sun L, Gill BJ, Vadakkan TJ, Kim MP, Ahn YH, Roybal JD, Liu X, Parra Cuentas ER, Rodriguez J, Wistuba II, Creighton CJ, Gibbons DL, Hicks JM, Dickinson ME, West JL, Grande-Allen KJ, Hanash SM, Yamauchi M, Kurie JM. Cancer-Associated Fibroblasts Induce a Collagen Cross-link Switch in Tumor Stroma. Mol Cancer Res. 2016 Mar;14(3):287-95. doi: 10.1158/1541-7786.MCR-15-0307. Epub 2015 Dec 2. PMID: 26631572; PMCID: PMC4794404.

49. Park D, Sahai E, Rullan A. SnapShot: Cancer-Associated Fibroblasts. Cell. 2020 Apr 16;181(2):486-486.e1. doi: 10.1016/j.cell.2020.03.013. PMID: 32302576.

50. Pekkala M, Hieta R, Bergmann U, Kivirikko KI, Wierenga RK, Myllyharju J. The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetratricopeptide repeat domain with functional aromatic residues. J Biol Chem. 2004 Dec 10;279(50):52255-61. doi: 10.1074/jbc.M410007200. Epub 2004 Sep 28. PMID: 15456751.

51. Persikov AV, Ramshaw JA, Brodsky B. Collagen model peptides: Sequence dependence of triple-helix stability. Biopolymers. 2000;55(6):436-50. doi: 10.1002/1097-0282(2000)55:6<436::AID-BIP1019>3.0.CO;2-D. PMID: 11304671.

52. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. 2014 Dec;15(12):1243-53. doi: 10.15252/embr.201439246. Epub 2014 Nov 8. PMID: 25381661; PMCID: PMC4264927.

53. Pokidysheva E, Boudko S, Vranka J, Zientek K, Maddox K, Moser M, Fässler R, Ware J, Bächinger HP. Biological role of prolyl 3-hydroxylation in type IV collagen. Proc Natl Acad Sci U S A. 2014 Jan 7;111(1):161-6. doi: 10.1073/pnas.1307597111.

54. Pokidysheva E, Zientek KD, Ishikawa Y, et al. Posttranslational modifications in type I collagen from different tissues extracted from wild type and prolyl 3-hydroxylase 1 null mice. J Biol Chem. 2013;288(34):24742-24752. doi:10.1074/jbc.M113.464156

55. Pyott SM, Schwarze U, Christiansen HE, Pepin MG, Leistritz DF, Dineen R, Harris C, Burton BK, Angle B, Kim K, Sussman MD, Weis M, Eyre DR, Russell DW, McCarthy KJ, Steiner RD, Byers PH. Mutations in PPIB (cyclophilin B) delay type I procollagen chain association and result in perinatal lethal to moderate osteogenesis imperfecta phenotypes. Hum Mol Genet. 2011 Apr 15;20(8):1595-609. doi: 10.1093/hmg/ddr037. Epub 2011 Jan 31. PMID: 21282188; PMCID: PMC3063987.

56. Ramshaw JA, Shah NK, Brodsky B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. J Struct Biol. 1998;122(1-2):86-91. doi: 10.1006/jsbi.1998.3977. PMID: 9724608.

57. Rappu P, Salo AM, Myllyharju J, Heino J. Role of prolyl hydroxylation in the molecular interactions of collagens. Essays Biochem. 2019;63(3):325-335. Published 2019 Sep 13. doi:10.1042/EBC20180053

58. Ricard-Blum S, Ruggiero F. (2005). The collagen superfamily: from the extracellular matrix to the cell membrane. Pathologie biologie, 53, 430-442.

59. Ricard-Blum, S. The collagen family. Cold Spring Harb. Perspect. Biol. 3, a004978 (2011).

60. Salo AM, Cox H, Farndon P, Moss C, Grindulis H, Risteli M, Robins SP, Myllylä R. A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. Am J

Hum Genet. 2008 Oct;83(4):495-503. doi: 10.1016/j.ajhg.2008.09.004. Epub 2008 Oct 2. PMID: 18834968; PMCID: PMC2561927.

61. Salo AM, Myllyharju J. Prolyl and lysyl hydroxylases in collagen synthesis. Exp Dermatol. 2020 Sep 23. doi: 10.1111/exd.14197. Epub ahead of print. PMID: 32969070.

62. Scietti L, Campioni M, Forneris F. SiMPLOD, a Structure-Integrated Database of Collagen Lysyl Hydroxylase (LH/PLOD) Enzyme Variants. J Bone Miner Res. 2019 Jul;34(7):1376-1382. doi: 10.1002/jbmr.3692. Epub 2019 Mar 12. PMID: 30721533.

63. Scietti, L. and Forneris, F. (2020). Full-Length Human Collagen Lysyl Hydroxylases. In Encyclopedia of Inorganic and Bioinorganic Chemistry, R.A. Scott (Ed.). doi:10.1002/9781119951438.eibc2739

64. Scietti, L., Chiapparino, A., De Giorgi, F. et al. Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. Nat Commun 9, 3163 (2018). https://doi.org/10.1038/s41467-018-05631-5

65. Sharma, U., Carrique, L., Vadon-Le Goff, S. et al. Structural basis of homo- and heterotrimerization of collagen I. Nat Commun 8, 14671 (2017). https://doi.org/10.1038/ncomms14671

66. Sorushanova A, Delgado LM, Wu Z, Shologu N, Kshirsagar A, Raghunath R, Mullen AM, Bayon Y, Pandit A, Raghunath M, Zeugolis DI. The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development. Adv Mater. 2019 Jan;31(1):e1801651. doi: 10.1002/adma.201801651.

67. Terajima M, Perdivara I, Sricholpech M, Deguchi Y, Pleshko N, Tomer KB, Yamauchi M. Glycosylation and cross-linking in bone type I collagen. J Biol Chem. 2014 Aug 15;289(33):22636-47. doi: 10.1074/jbc.M113.528513. Epub 2014 Jun 23. PMID: 24958722; PMCID: PMC4132771.

68. 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. J Biol Chem. 2008 Jul 11;283(28):19432-9. doi: 10.1074/jbc.M802973200. Epub 2008 May 15. PMID: 18487197.

69. Tuderman L, Myllylä R, Kivirikko KI. Mechanism of the prolyl hydroxylase reaction. 1. Role of co-substrates. Eur J Biochem. 1977 Nov 1;80(2):341-8. doi: 10.1111/j.1432-1033.1977.tb11888.x. PMID: 200424.

70. Vadon-Le Goff S, Hulmes DJ, Moali C. BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling. Matrix Biol. 2015 May-Jul;44-46:14-23. doi: 10.1016/j.matbio.2015.02.006. Epub 2015 Feb 18. PMID: 25701650.

71. Vallet, S.D., Miele, A.E., Uciechowska-Kaczmarzyk, U. et al. Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners. Sci Rep 8, 11768 (2018). https://doi.org/10.1038/s41598-018-30190-6

72. Vranka JA, Sakai LY, Bächinger HP. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. J Biol Chem. 2004 May 28;279(22):23615-21. doi: 10.1074/jbc.M312807200. Epub 2004 Mar 24. PMID: 15044469.

73. Wilson R, Lees JF, Bulleid NJ. Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. J Biol Chem. 1998 Apr 17;273(16):9637-43. doi: 10.1074/jbc.273.16.9637. PMID: 9545296.

74. Wong CC, Zhang H, Gilkes DM, Chen J, Wei H, Chaturvedi P, Hubbi ME, Semenza GL. Inhibitors of hypoxia-inducible factor 1 block breast cancer metastatic niche formation and lung metastasis. J Mol Med (Berl). 2012 Jul;90(7):803-15. doi: 10.1007/s00109-011-0855-y. Epub 2012 Jan 10. PMID: 22231744; PMCID: PMC3437551.

75. Yamauchi M, Barker TH, Gibbons DL, Kurie JM. The fibrotic tumor stroma. J Clin Invest. 2018 Jan 2;128(1):16-25. doi: 10.1172/JCI93554. Epub 2018 Jan 2. PMID: 29293090; PMCID: PMC5749516.

76. Yamauchi M, Sricholpech M. Lysine post-translational modifications of collagen. Essays Biochem. 2012;52:113-133. doi:10.1042/bse0520113

77. Yamauchi M, Taga Y, Hattori S, Shiiba M, Terajima M. Analysis of collagen and elastin cross-links. Methods Cell Biol. 2018;143:115-132. doi: 10.1016/bs.mcb.2017.08.006. Epub 2017 Nov 22. PMID: 29310773.

78. Yamauchi M, Terajima M, Shiiba M. Lysine Hydroxylation and Cross-Linking of Collagen. Methods Mol Biol. 2019;1934:309-324. doi: 10.1007/978-1-4939-9055-9_19. PMID: 31256387.

79. Zuurmond AM, van der Slot-Verhoeven AJ, van Dura EA, De Groot J, Bank RA. Minoxidil exerts different inhibitory effects on gene expression of lysyl hydroxylase 1, 2, and 3: implications for collagen cross-linking and treatment of fibrosis. Matrix Biol. 2005 Jun;24(4):261-70. doi: 10.1016/j.matbio.2005.04.002. PMID: 15908192.

2.Chapter

COLLAGEN HYDROXYLASE GLYCOSYLATION: NON-CONVENTIONAL SUBSTRATES FOR ATYPICAL GLYCOSYLTRANSFERASE ENZYMES

This is a mini-review submitted to the Biochemical Society Transactions (Portland Press). I wrote the paper together with all the other authors, and I made the figures using ChemDraw 12 and BioRender.com.

Authors:
Francesca De Giorgi, Marco Fumagalli, Luigi Scietti* and Federico Forneris*

Affiliation(s):
The Armenise-Harvard Laboratory of Structural Biology, Department of Biology and Biotechnology, University of Pavia, via Ferrata 9/A, 27100 Pavia, Italy, http://fornerislab.unipv.it

+ Corresponding authors: LS luigi.scietti@unipv.it, FF federico.forneris@unipv.it

ABSTRACT

Collagen is a major constituent of the extracellular matrix (ECM) that confers fundamental mechanical properties to tissues. To allow proper folding in triple-helices and organization in quaternary super-structures, collagen molecules require essential post-translational modifications (PTMs), including hydroxylation of proline and lysine residues, and subsequent attachment of glycan moieties (galactose and glucose) to specific hydroxylysine residues on procollagen alpha chains, catalyzed by specialized glycosyltransferase enzymes. The resulting glucosyl-galactosyl-hydroxylysine (Glc-Gal-Hyl) is one of the simplest glycosylation patterns found in nature and is essential for collagen and ECM homeostasis. Although the biochemical reactions leading to formation of Glc-Gal-Hyl have been widely studied, several key biological questions about the possible functions of this essential PTM are still missing. In addition, the lack of three-dimensional structures of the enzymes catalyzing these reactions hinders our understanding of the catalytic mechanisms producing this modification, as well as the impact of genetic mutations causing severe connective tissue pathologies. In this mini-review, we summarize the current knowledge on the biochemical features of the enzymes involved in the production of collagen glycosylations and the current state-of-the-art methods for the identification of this key PTM.

2.1 INTRODUCTION

Collagen O-linked glycosylations are unique post-translational modifications (PTM) occurring at early stages of the complex biosynthesis sequence leading to mature extracellular collagen molecules. Starting from hydroxylation of lysine residues in the endoplasmic reticulum (ER), the subsequent 5-hydroxylysine (Hyl) glycosylations generate a very simple and extremely conserved PTM (1-3). Its distribution and abundance on collagen molecules strongly depend on collagen amino acid sequence, tissue type and organism developmental stage (4-8). Micro- (i.e. the extent of glycosylation at a specific collagen site) and macro-heterogeneities (i.e. the glycosylation occupancy at different collagen sites) characterize different collagen types and seem to be responsible for the diverse structural organizations and functions in the extracellular matrix (ECM). Fibrillar collagens such as type I and II display only few glycosylated residues (5,9,10), whereas mesh-type collagens (such as type IV) are glycosylated much more extensively (11). However, the molecular paths leading to Hyl PTMs are still missing fundamental pieces, thus preventing a comprehensive understanding of the mechanisms and the exact biological impact of collagen glycosylation. Furthermore, how, where and why specific lysines are differentially modified in the diverse collagen types are far to be understood and rationalized, although increasing efforts have been made to efficiently map these PTMs (11). Novel mass-spectrometry based approaches started to shed light on such heterogeneity (see the "glycosyltransferase activity assays" paragraph),

but a complete description of the requirements to obtain different glycosylation types on specific collagen sequences is still missing.

2.2 A SIMPLE, ESSENTIAL AND CONSERVED PTM

Firstly described by Spiro late in the 60's, the α-(1,2)-glucosyl-β-(1,*O*)-galactosyl-5-hydroxylysine (Glc-Gal-Hyl) is one of the simplest existing glycosylation type in the animal protein world, and is uniquely occurring on collagen and collagen-like proteins (12,13). This PTM is extremely conserved from sponges to humans (14-18). The biological functions of collagen glycans remained unclear for long time, but critical roles in collagen triple-helix stabilization and organization of collagen superstructure in the ECM are emerging (1,2,19,20). Although not completely understood, Hyl glycosylation was reported to be involved in several step of collagen biosynthesis, such as control of secretion, cross-linking, and fibrillogenesis. These processes are crucial to maintain the ECM homeostasis by modulating cell-ECM adhesion events and integrin-mediated signaling (1,19,21). Accumulating evidence supporting the fundamental role of Hyl glycosylation in ECM structural organization are provided by the diverse disease conditions arising from alterations in the specific collagen glycosylation patterns (*see also box i*). In this respect, genetic mutations affecting the enzymatic machineries associated to collagen glycosylations lead to severe connective tissue disorders (22-27). Similarly, overexpression and mislocalization of these enzymes in the tumor microenvironment have been described to enhance the metastatic progression of many solid tumors (28-30).

2.3 THE PATH TO GLC-GAL-HYL

Complete collagen glycosylation requires three distinct and consecutive enzymatic activities on collagen lysine residues: after generation of Hyl in the ER by means of procollagen lysyl hydroxylases (LH enzymes, encoded by the procollagen-lysine 2-oxoglutarate 5-dioxygenase (*PLOD*) genes), 5-hydroxylysyl galactosyltransferases (Gal-T, EC 2.4.1.50) form a β-glycosidic bond between a galactose molecule and the 5-hydroxyl group of Hyl, thus generating β-(1,*O*)-galactosyl-5-hydroxylysine (Gal-Hyl); subsequent galactosyl-5-hydroxylysyl glucosyltransferases (Glc-T, EC 2.4.1.66) form a α-glycosidic bond between a glucose molecule and Gal-Hyl, yielding the final Glc-Gal-Hyl PTM pattern (*see also box ii*). The two enzymatic glycosyltransferase reactions rely on the use of activated nucleotide sugars in the form of UDP-conjugates (UDP-α-D-galactose and UDP-α-D-glucose) as donor substrates. Considering that UDP-glycan are almost exclusively alpha anomers, the catalytic mechanisms and therefore the glycosyltrasferases involved in the process can be defined as "inverting" when the anomeric carbon configuration changes from α to β and "retaining" when the configuration of the anomeric carbon is maintained (2, 31-33) (Figure 1). These enzymes use a bivalent metal ion (typically Mn^{2+}) to allow proper coordination of the phosphate moiety of the UDP-glycan within the active site. The metal ion is in turn coordinated by two aspartate

residues, which are usually part of extremely conserved Asp-X-Asp (DXD) motifs. In humans, two main glycosyltransferase enzyme families are responsible for the biosynthesis of Glc-Gal-Hyl: the galactosyltransferases belonging to the *COLGALT* gene family and the multifunctional lysyl hydroxylase-glucosyltransferases from the *PLOD* gene family.

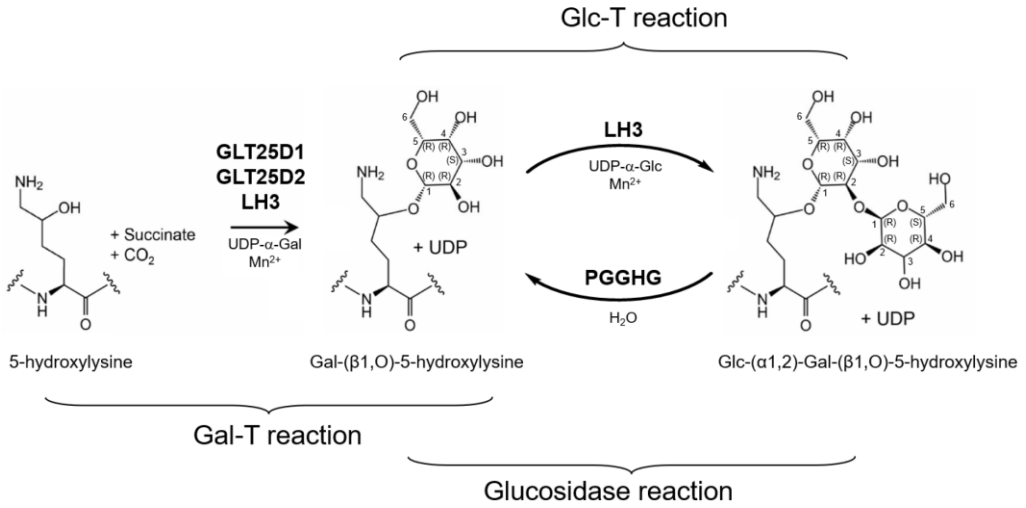


Figure 2-1 Biochemistry of collagen hydroxylysine glycosylation. Collagen hydroxylysine residues are modified through the subsequent addition or removal of sugar moieties. The first reaction is a galactosyltransferase (Gal-T) reaction, performed by GLT25D1/2 and, possibly, also by multifunctional LH3. Using UDP-galactose as donor substrate, the galactose is transferred by the enzyme through an inverting mechanism to the hydroxylysine 5-OH group, forming a β-(1,O)-glycosidic bond. A subsequent glycosyltransferase (Glc-T) reaction is carried out by LH3 through a retaining reaction. The glucose is transferred from the donor UDP-glucose to the β-(1,O)-galactosyl-5-hydroxylysine residue, forming a α-(1,2)-glycosidic bond. The resulting α-(1,2)-glucosyl-β-(1,O)-galactosyl-5-hydroxylysine may also undergo further processing: an α-glucosidase, PGGHG, can use a water molecule to hydrolyze the glucose moiety, thus yielding back the β-(1,O)-galactosyl-5-hydroxylysine residue.

2.4 THE BIOCHEMISTRY OF UDP-SUGAR TRANSFER

Glycosyltransferase enzymes using nucleoside conjugated sugars are termed Leloir glycosyltransferases, and are structurally characterized either by the presence of a GT-A or a GT-B fold (32,34,35). The GT-A glycosyltransferases comprises two close β/α/β Rossmann domains and the conserved DXD motif, coordinating a divalent metal ion cofactor, usually Mg^{2+} or Mn^{2+} (32,35). GT-B glycosyltransferases are membrane associated. They possess two β/α/β Rossmann domains facing each other and are independent from metallic cofactors (36). Depending on the reaction mechanism, both

GT-A and GT-B glycosyltransferases can be inverting or retaining GTs. The inverting reaction proceeds via a single displacement S_N2 mechanism mediated by a catalytic base, with the concomitant nucleophilic attack by the acceptor substrate and leaving group departure (32,37). The retaining reaction implicates the maintenance of the same configuration at the anomeric carbon. In this context, two mechanisms are possible: the first one implies a double displacement involving the formation of a covalent glycan-enzyme intermediate, whereas the second one consists in a S_Ni mechanism (termed 'internal return') in which the attacking catalytic nucleophile and the leaving group departure occur from the same face (38,39).

2.5 HUMAN COLLAGEN GALACTOSYLTRANSFERASES

The transfer of a galactose moiety on the hydroxyl group of Hyl is catalyzed through an inverting reaction by two specialized procollagen hydroxylysine galactosyltransferases (Figure 1): GLT25D1 (Procollagen galactosyltransferase 1, UniProt Q8NBJ5) and GLT25D2 (Procollagen galactosyltransferase 2, UniProt Q8IYK4), respectively encoded by *COLGALT1* and *COLGALT2* genes. Their Gal-T-associated activity is essential: *COLGALT1* knock-out is embryonically lethal in mouse (40), whereas knock-down experiments consistently resulted in pronounced reduction of collagen galactosylation in cell-based experiments (20,41) and in murine models (40), underlying the importance of these enzymes in ECM homeostasis and suggesting that the *in vivo* LH3 Gal-T activity (see next paragraph) may not be sufficient for normal collagen glycosylation. Genetic mutations on *COLGALT* genes in mouse and humans cause cerebral small vessel disease, a hereditary syndrome characterized by small arteries and capillaries in the brain. The pathogenic mutations Leu151Arg, Ala154Pro, and Glu366Arg produce a highly unstable and inactive GLT25D1, causing a strong decrease in Gal-T activity and leading to abnormal intracellular accumulation of collagen type IV (42). A loss of function mutation (Trp130Arg) of GLT25D1 was also reported in mice and zebrafish models with musculoskeletal defects associated with impaired secretion of collagen IV (43).

GLT25D1 is ubiquitously found in human tissues especially in placenta, heart, lung, and spleen, whereas GLT25D2 is present at low levels in brain and skeletal muscle tissues (44). The two enzymes share 55% sequence identity with the cell-adhesion protein CerCAM (or CEECAM1, UniProt Q5T4B2), which is expressed in secretory tissues (salivary gland, pancreas, placenta) and in the nervous system, and is involved in leukocyte transmigration across the blood-brain barrier (44,45). CerCAM does not display galactosyltransferase activity either *in vitro* or *in vivo* and is considered as the inactive GLT25D3 member (44). Human GLT25D1 and GLT25D2 are not active on isolated Hyl residues and require collagenous peptides of at least 500-600 Da as acceptor substrates. However, no preference for specific collagen sequences has been reported, resulting in enzymatic activity on several Hyl residues found in different collagen types

(44). In addition, GLT25D1 can exert its Gal-T activity also on collagenous domain of mannan-binding lectin (MBL) and high molecular weight (HMW) adiponectin. Introduction of glycan moieties stabilizes MBL and HMW adiponectin and facilitates their secretion into the extracellular space. Intracellular accumulation of these two proteins has been observed as consequence of GLT25D1 depletion (44,46,47).

Biochemical analyses coupled with site-directed mutagenesis and domain-swapping experiments enabled the identification of GLT25D1/2 residues essential for catalysis. GLT25D1 presents three DXD motifs: one at the N-terminus (Asp166-Asp168, using GLT25D1 numbering) fully conserved among the three human GLT25D isoforms, and two at the C-terminus (Asp461-Asp463 and Asp585-Asp587), not conserved in GLT25D2 and CerCAM. The first two motifs appear critical for enzymatic activity, whereas the third one is dispensable (48). However, no unambiguous identification of the metal ion binding site is currently available, as well as accurate structural description of the GLT25D molecular architecture. Based on amino acid sequence comparisons, GLT25D1 and GLT25D2 appear to share features with other glycosyltransferases. The N-terminal domain is similar to the GT2 domain of heparin synthase, chondroitin synthase and hyaluronan synthase (46,48). Interestingly, GLT25D1 also shares about 30% sequence identity with the N-terminal glycosyltransferase domain of the multifunctional procollagen lysyl hydroxylase and glycosyltransferase LH3 (46,49), which displays also collagen Gal-T activity *in vitro* (see next paragraphs). The C-terminal domain of GLT25D enzymes is instead closely related to the inverting-type Leloir GT25 family (2,41,44).

An intriguing aspect of GLT25D biology is the possible association of these enzymes with multifunctional lysyl hydroxylases-glycosyltransferases. Colocalization studies indeed identified GLT25D1 and LH3 in the ER, suggesting the possibility of a complex multiprotein enzyme system capable of LH, Gal-T and Glc-T activity (46,47). However, numerous reports detected LH3 (and also related isoforms (28)) in the Golgi apparatus and in the extracellular space (50-52), whereas GLT25D enzymes were always observed as resident in the ER. This is consistent with the presence of specific ER retention sequences for GLT25D enzymes but not for LH, supporting the possibility of transient LH-GLT25D multiprotein systems in the ER during specific stages of collagen biosynthesis and subsequent additional enzymatic modifications by LH within other cellular compartments.

2.6 COLLAGEN GLUCOSYLTRANSFERASES

Gal-Hyl undergo subsequent Glc-T reaction, yielding Glc-Gal-Hyl. In this case, differently from the previous Gal-T reaction, the acceptor substrate is a glycan moiety (Gal) instead of a modified amino acid (Hyl). Opposite to the Gal-T reaction, in the Glc-T step the glucose moiety is transferred from the UDP-α-Glc donor to the Gal-Hyl

acceptor in alpha configuration, thus through a retaining reaction mechanism. Since the initial identification of the Glc-Gal-Hyl pattern, several enzymes residing in the ER have been proposed as putative collagen glucosyltransferases (2). However, accumulating evidence assigned this role exclusively to the multifunctional procollagen lysil hydroxylase and glycosyltransferase LH3 (UniProt O60568) (49). Encoded by the *PLOD3* gene, LH3 is a multifunctional enzyme displaying procollagen glycosyltransferase activities in its N-terminal (GT) domain, and lysyl hydroxylase activity in the C-terminal (LH) domain (53,54). The function(s) of LH3 are essential for life, as total knock-out of the *PLOD3* gene is embryonically lethal in mice, due to prevented collagen IV secretion and subsequent intracellular accumulation (55,56). Published reports suggest the ability of LH3 to carry out the complete conversion of collagen Lys residues to Glc-Gal-Hyl *in vitro* (54,57,58), but the measured catalytic efficiency of the two reactions is strongly different, with the Glc-T activity being at least one order of magnitude more efficient than the Gal-T activity in the same experimental conditions (54). The observed coexistence of both Gal-T and Glc-T activities within the same LH3 catalytic site is controversial and so far not supported by clear evidence of biological significance: whilst the functional implications of LH3 as collagen glucosyltransferase were extensively verified (51,59-62), no LH3 Gal-T activity has ever been observed *in vivo*. Mouse studies showed that mutations exclusively affecting the lysyl hydroxylase activity domains of the enzyme result in normally developed animals without dramatic alterations in tissue morphology, but with reduced thickness of the epidermal basal lamina (55,56). On the other hand, several reports indicate that the LH3 Glc-T activity is essential for correct secretion and assembly and extracellular deposition of several collagen types, ranging from the fibrillar type I (62) to the highly glycosylated type IV and VI (63), with strong implications in correct formation of cytoskeletal structures, basement membranes and ECM architectures (55,64,65).

The recent determination of the three-dimensional structures of full-length human LH3 enabled precise identification of the amino acids involved in Mn^{2+} binding and stabilization of the UDP moiety of the donor substrate (54). Nevertheless, due to the lack of electron density for the sugar moiety of UDP-glycan in the catalytic site, none of the obtained structures allowed to uniquely characterize the residues responsible for catalysis. A surprising, unexpected structural feature is the presence of a second well defined glycosyltransferase domain in the quaternary structure of the enzyme that is not competent for any enzyme's catalytic activity (54). The striking sequence similarity between this noncatalytic "accessory" domain and classic glycosyltransferases constituted one of the major difficulties associated to the understanding of the localization and mechanisms of LH3 glycosyltransferase activities (53,57,66). Conversely, the structure of the only functional LH3 glycosyltransferase domain only partially resembles Leloir-type GT-A. The classical Rossmann fold architecture is decorated with long, flexible loops characterized by unique amino acid sequences and incorporates a non-canonical DXXD

motif (Asp112-Asp115) coordinating the Mn^{2+} ion (54). Structure-guided site-directed mutagenesis encompassing the unique features of the glycosyltransferase domain confirmed the critical features associated to the two aspartate residues responsible for Mn^{2+} coordination, and for two aromatic residues (Trp75 and Tyr 114) critical for trapping the uridine moiety of the UDP-glycan donor substrate within the catalytic site (54). Additional mutants probing the precise significance of the numerous "hot spots" localized within the LH3 glycosyltransferase catalytic site are currently subject of thorough investigation (67).

Combined with molecular structure knowledge, mouse and zebrafish models helped in the rationalization of the increasing reports describing mutations occurring on the human *PLOD3* gene affecting its glycosyltransferase domain, causing severe connective disorders such as recessive Dystrophic Epidermolysis Bullosa and Stickler syndrome (26,27,64,68,69).

2.7 THE NEED FOR A COLLAGEN GLUCOSIDASE

In late 1970s, a glucose hydrolase, different from the well-known α-glucosidase, has been characterized in rat kidney, spleen, and chick embryos (70-72). Encoded by *PGGHG/ATHL1* gene, the protein-glucosylgalactosyl-hydroxylysine glucosidase (PGGHG) is also conserved in humans (73). Intriguingly, PGGHG activity was not observed in jawless vertebrates (73), possibly tracing a distinctive role for this enzyme in jaw and joint formation.

PGGHG is specific for the Glc-Gal-Hyl of collagens and catalyzes the hydrolytic removal of the glucose moiety, yielding Gal-Hyl residues. The exact PGGHG contribution to collagen homeostasis is still unclear: the critical roles of this enzyme in Glc-Gal-Hyl catabolism are well established, however the mechanisms underlying the balance between glucosyltransferase and glucosidase activities *in vivo*, as well as the selectivity of the enzyme for different collagen types and glycosylated Hyl residues within protein sequences demand an in-depth investigation. Although in absence of structural data, biochemical characterizations provided insight into its substrate recognition properties and catalytic mechanism: the positive charge of the ε-amino group in Hyl seems to guide substrate recognition through a conserved aspartate residue (Asp301 in human PGGHG), directly interacting with the glucose moiety. Judged by the strong product inhibition observed *in vitro* and *in vivo* (74,75), the interaction of the enzyme with glucose plays fundamental roles in catalysis. Two additional glutamate residues (Glu430 and Glu574) were identified respectively as the catalytic acid and base, judged by their critical impact on glucose hydrolysis (73).

2.8 GLYCOSYLTRANSFERASE ACTIVITY ASSAYS

Assaying glycosyltransferase activity has always been challenging (76). This biochemical complexity extends much further in presence of non-conventional acceptor substrates such as the large multimeric collagen molecules (Figure 2). Historically, the most common method to probe these enzymatic functions focused on monitoring the transfer of radiolabeled sugars from donor to acceptor molecules (77,78). However, these experimental approaches are becoming less and less frequently used, as they rely on cumbersome extraction processes to recover and measure glycosylated products, and require management of radioactive materials and waste. In this respect, various assays not relying on radiolabeled reagents have been developed, but most of these have been tailored to specific glycosyltransferase reactions, requiring modified sugar-nucleotide donors and additional separation steps (76). Highly sensitive immunological methods, based on very specific antibodies or lectins, offer the advantage of a precise identification of the glycosyltransferase reaction products. Immunostaining has long been used for product detection with glycolipid or glycoprotein acceptors (79). The downside of immunological assays is the lack of readiness for quantitative enzyme kinetics experiments. Furthermore, the availability of specific antibodies or acceptor conjugates can also be problematic. In this respect, it should be noted that these assays were not broadly applied to collagen glycosyltransferases, due to the intrinsic difficulties in detecting Hyl glycosylations at specific collagen sites within the highly repeated amino acid sequences (16). With the advent of cheap and efficient fluorescent labeling techniques, HPLC and capillary electrophoresis (CE) coupled to detection of fluorescence increasingly became the standard approach to separate and quantitate glycosyltransferase reaction products (80,81). These methods couple high resolution and sensitivity to reliable product identification; the main limitation is the requirement of a fluorescent label in the acceptor substrate at sites that do not interfere with enzyme catalysis.

Mass spectrometry (MS) has always been a powerful analytical method to analyze collagen PTMs, and the advances in this field enabled the characterization of collagen glycosylations and their micro- and macro-heterogeneity. Classical MS analytical procedures require enzymatic degradation of collagens followed by separation of the resulting peptides by liquid chromatography (LC) and MS analysis of fractions by matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) coupled to peptide sequencing through automated Edman degradation and tandem mass spectrometry (MS/MS) (4). A novel method for glycopeptide preparation requiring hydrazide chemistry and galactose oxidase was developed by Taga et al. (82). The hydrazide-based method coupled with LC-electrospray ionization (ESI)-MS/MS can allow the simultaneous detection of substrates and reaction products even in presence of complex, highly repeated amino acid sequences. The ESI technique enabled the identification of several

and unreported O-glycosylation sites in bovine type I and type II collagens, discriminating between Gal-Hyl and Glc-Gal-Hyl residues (83). A major challenge in collagen analysis is the characterization of post-translational heterogeneities in collagen peptide mixtures. In this respect, the development of high-resolution mass spectrometry (HRMS) instrumentation offered swiftness, ease of use, and reliable enabling the characterization of the glycosylation patterns. HRMS key to success in recent investigations describing the features of distinct glycosylation sites on specific collagen sequences (84-86). Among the others, Basak et al. implemented HRMS with fragmentation methods and bioinformatics workflows providing a comprehensive characterization of glycosylation and hydroxylation heterogeneity of collagen IV from mouse and human basement membranes with about 80% sequence coverage (11). Similarly, Merl-Pham et al. performed a quantitative proteomic analysis of collagen PTMs in crude ECM preparation, yielding a comprehensive map of lysine modifications for 15 collagen types and quantifying the micro-heterogeneity of the O-glycosylation sites on human collagen I alpha-1 chain (8).

A viable alternative to probe glycosyltransferase activity is the usage of a universal indirect assay based on detection of reaction byproducts (e.g., free UDP) using either fluorescence (87,88) or luminescence (54). Similar methods were developed and successfully applied for the characterization of the lysyl hydroxylase enzymatic activity (54,89,90). Application of these methods to collagen glycosyltransferases requires enzyme coupling to convert the UDP to a detectable product, such as ATP or phosphate, or coupling with UDP-glucose dehydrogenase to produce NADH before conversion to a fluorescent resazurin by diaphorase (76). However, these coupled assays are inherently prone to false positives from compound interference with the coupling enzymes, and thus require many additional tests for counter-screening. In addition, numerous glycosyltransferases (including LH3) systematically display processing of donor substrates uncoupled to transfer of the glycan moiety to the acceptor molecule (54). In this respect, a general recommendation in using indirect assays involves combining the quantitation of enzymatic activity with direct evidence of substrate formation, e.g., through radio- or immune- labeling, or using mass spectrometry.

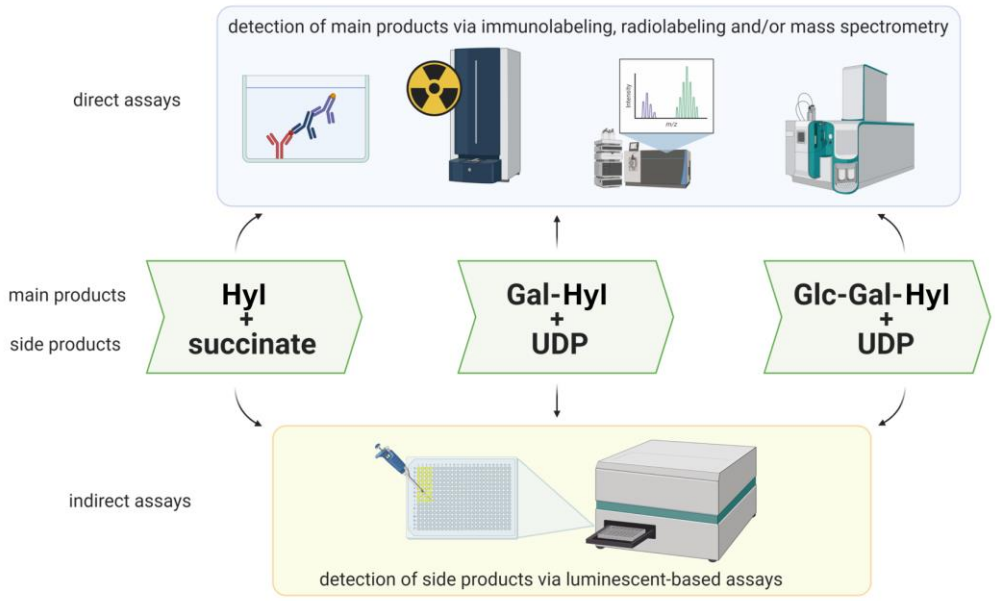


Figure 2-2 Detection methods for the measurement of collagen lysine post-translational modifications. Collagen lysines are modified by hydroxylation and subsequent glycosylations. The first hydroxylase reaction produces hydroxylysine (Hyl) and succinate as side product. Hyl is then glycosylated by consecutive addition of galactose and glucose moieties, yielding galactosyl-hydroxylysine (Gal-Hyl) and glucosyl-galactosyl-hydroxylysine (Glc-Gal-Hyl). Both glycosylation reactions produce UDP as side product. Detection of collagen lysine post-translational modifications may require either direct or indirect assays, exploiting the formation of main or side products, respectively. Direct assays (upper blue box) include well-established classical biochemical methods, such as immunolabelling and radiolabelling of main products, but also advanced methods based on coupling of liquid chromatography with hybrid ion trap and/or quadrupole time-of-flight mass spectrometry (LC/MS, LC/QTOF MS). Indirect assays (lower yellow box) are based on the enzymatic conversion of side products, succinate and UDP, into detectable reagents such as ATP (suitable for luminescence-based assays), resazurin or phosphates (suitable for chromogenic and fluorescence-based assays). Created with BioRender.com.

2.9 CONCLUSIONS

Collagen lysine modifications are highly conserved PTMs essential for the correct biogenesis of collagen and deposition of ECM. To date, some pieces of this intricate picture are missing. Although collagen lysine hydroxylations and glycosylations have been extensively studied and characterized, it is still unclear how and why certain types of collagen display most of their lysine residues in the glycosylated form (e.g. collagen IV and VI), whereas in other collagen types only a small percentage of lysines is subject to hydroxylation. Macro-heterogeneities depend on the specificity of the glycosyltransferase machineries operating only on certain Hyl residues and thus arise as a consequence of a temporally- and spatially-controlled process, as observed in zebrafish (91). Likewise, they can also correlate with differential expression of collagen glycosyltransferase enzymes depending on tissue type and developmental stage. It is important to underline that not

all Hyl residues are found fully glycosylated, as numerous Gal-Hyl residues can be found (4) contributing to increase the overall micro-heterogeneity of collagen PTM. This in turn poses the question about the mechanism by which Gal-Hyl is produced, and the possible biological roles exerted by different mono- and disaccharide glycosylations. For example, the presence of O-linked monosaccharide or disaccharide units covalently attached to Hyl impacts on the formation and the extent of extracellular collagen cross-links. Fully glycosylated Glc-Gal-Hyl are often found at the bridge of divalent cross-links, whereas the less bulky Gal-Hyl modifications are more frequently present at mature trivalent cross-links (1). Furthermore, reduction of the Gal-T activity accelerates collagen cross-link maturation, resulting in an increased number of trivalent cross-links (20). Alterations in glycosylation and cross-linking patterns affect the final shape and diameter of collagen fibrils. Ultimately, the distribution of mono- and disaccharide units on collagen molecules impacts on the supramolecular architecture of the ECM and its associated protein-protein interaction networks, modulating tissue homeostasis and, in case of malfunctions, leading to severe connective tissue disorders (22-27,83,92). Taken together, the available data on distribution and abundance of different collagen glycosylations support the intriguing hypothesis of a precise biological significance for each Lys, Hyl, Gal-Hyl and Glc-Gal-Hyl pattern found across different collagen molecules during different stages of tissue development (*see also box iii*). In this respect, the enzymatic activities of lysyl hydroxylases, hydroxylysine glycosyltransferases and PGGHG may contribute as a whole to the establishment and maintenance of a possible "collagen code", indispensable to prevent the dramatic impact displayed by altered macro- and micro-heterogeneities on tissue homeostasis (93). The accurate description of the fine details underlying catalytic selectivity and specificity mechanisms of these enzymes will provide crucial insights to fully unravel the outstanding complexity associated to this relatively simple, yet unique and essential PTM pattern.

2.10 PERSPECTIVES

i) Collagen hydroxylysine glycosylations are essential PTMs that allow proper collagen weaving and correct ECM deposition. Alterations in the biosynthesis of this relatively simple pattern lead to collagen aberrations resulting in severe connective tissue disorders.
ii) Collagen glycosylation is established by the synergistic activity of lysyl hydroxylases, GLT25D1/2, LH3 and PGGHG. The combined action of these enzymes results in the production of a specific micro- and macro-heterogeneity which are peculiar for each collagen type.
iii) Synergistic approaches combining comprehensive MS-based glycosylation analyses of collagen molecules, molecular structure investigations and biochemical characterization of the enzymatic players involved in the process are key to understanding the biological significance of such heterogeneity.

i. Altered collagen glycosylation patterns cause a variety of developmental connective tissue disorders. Combined with the recent identification of alterations in Hyl patterns in metastatic progression of solid tumors, the need for accurate investigations on collagen glycosyltransferase mechanisms becomes evident, for fundamental knowledge and for possible development of advanced therapeutic strategies. In this respect, a fascinating observation relates to the identification of multifunctional LH3 in the ER, in the Golgi apparatus. and the extracellular space, where this enzyme can further process collagen molecules. The precise functional role associated to LH3 trafficking has still to be unveiled, but its involvement in developmental diseases suggests critical importance for this process, likely associated to LH3 Glc-T activity and its possible regulation due to collagen processing in different cell compartments.

ii. Glc-Gal-Hyl is a simple, yet unique glycosylation pattern essential for proper collagen weaving in the ECM. The Glc-Gal-Hyl biosynthesis is catalyzed by two families of glycosyltransferases which make use of activated UDP-glycan to transfer sugars: the GLT25D Gal-T and the LH3 multifunctional Glc-T. These enzymes use different donor substrates and catalytic mechanisms: in the inverting Gal-T reaction the substrate is the hydroxyl group of specific collagen lysines generating Gal-Hyl, whereas in the retaining Glc-T reaction the glucose is transferred to the acceptor substrate (galactose), yielding Glc-Gal-Hyl. The enzymes involved in the Glc-Gal-Hyl biosynthesis have peculiar structural features that resemble only partially other known glycosyltransferases. Multifunctional LH3 has also been reported capable of Gal-T activity *in vitro*, however the lack of *in vivo* evidence, the possibility of synergies and colocalization with GLT25D1, and the requirement of a different biochemical mechanism for catalysis renders this observation quite controversial.

iii. New proteomics-based analyses are progressively unveiling details about the specificity of collagen Hyl glycosylation, however how and why only specific lysines are processed is still unclear. The simultaneous identification of Lys, Hyl, Gal-Hyl and Glc-Gal-Hyl residues in collagen sequences in different regions of the molecule may represent a hallmark of PTM specificity. Alternatively, the possible lack of PTM specificity might suggest the possibility of a fine balance between the action of hydroxylases, glycosyltrasferases (GLT25D1/2, LH3) and glucosidases (PGGHG), adding further complexity to the system. Furthermore, the precise structural organization of the Glc-Gal moiety with respect to the main collagen triple-helix chain is unknown, as well as its precise molecular impact on formation of complex quaternary structures. Our understanding of the enzymatic properties of collagen glycosyltransferases and their impact on collagen assembly demands a more comprehensive biochemical and structural characterization of LH, GLT and PGGHG enzymes and their complexes with collagen substrates and products, to shed light on the differences and similarities driving to their diverse role in Glc-Gal-Hyl and Gal-Hyl biosynthesis.

2.11 ACKNOWLEDGEMENTS

We thank Dr. A Chiapparino and Prof. S. Vidal for useful discussions. 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), by Fondazione Giovanni Armenise-Harvard (CDA2013 to FF), and by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022, to the Dept. of Biology and Biotechnology "L. Spallanzani", University of Pavia). None of the funding sources had roles in study design, collection, analysis and interpretation of data, in the writing of the report and in the decision to submit this article for publication.

2.12 CONFLICT OF INTERESTS

The authors declare no conflict of interests.

2.13 AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

2.14 REFERENCES

1. Terajima M, Perdivara I, Sricholpech M, Deguchi Y, Pleshko N, Tomer KB, et al. Glycosylation and cross-linking in bone type I collagen. J Biol Chem. 2014 Aug 15;289(33):22636-47.

2. Hennet T. Collagen glycosylation. Curr Opin Struct Biol. 2019 Jun;56:131-8.

3. Cummings RD. The repertoire of glycan determinants in the human glycome. Mol Biosyst. 2009 Oct;5(10):1087-104.

4. Perdivara I, Yamauchi M, Tomer KB. Molecular Characterization of Collagen Hydroxylysine O-Glycosylation by Mass Spectrometry: Current Status. Aust J Chem. 2013 Jul 18;66(7):760-9.

5. Yamauchi M, Terajima M, Shiiba M. Lysine Hydroxylation and Cross-Linking of Collagen. Methods Mol Biol. 2019;1934:309-24.

6. Sorushanova A, Delgado LM, Wu Z, Shologu N, Kshirsagar A, Raghunath R, et al. The Collagen Suprafamily: From Biosynthesis to Advanced Biomaterial Development. Adv Mater. 2019 Jan;31(1):e1801651.

7. Stammers M, Ivanova IM, Niewczas IS, Segonds-Pichon A, Streeter M, Spiegel DA, et al. Age-related changes in the physical properties, cross-linking, and glycation of collagen from mouse tail tendon. J Biol Chem. 2020 Jul 31;295(31):10562-71.

8. Merl-Pham J, Basak T, Knüppel L, Ramanujam D, Athanason M, Behr J, et al. Quantitative proteomic profiling of extracellular matrix and site-specific collagen post-translational

modifications in an in vitro model of lung fibrosis. Matrix Biology Plus. 2019 2019/02/01/;1:100005.

9. Song E, Mechref Y. LC–MS/MS Identification of the O-Glycosylation and Hydroxylation of Amino Acid Residues of Collagen α-1 (II) chain from Bovine Cartilage. Journal of Proteome Research. 2013 2013/08/02;12(8):3599-609.

10. 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;30(10):830-43.

11. Basak T, Vega-Montoto L, Zimmerman LJ, Tabb DL, Hudson BG, Vanacore RM. Comprehensive Characterization of Glycosylation and Hydroxylation of Basement Membrane Collagen IV by High-Resolution Mass Spectrometry. J Proteome Res. 2016 Jan 4;15(1):245-58.

12. Spiro RG. Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. J Biol Chem. 1969 Feb 25;244(4):602-12.

13. Spiro RG. The structure of the disaccharide unit of the renal glomerular basement membrane. J Biol Chem. 1967 Oct 25;242(20):4813-23.

14. Hill RC, Wither MJ, Nemkov T, Barrett A, D'Alessandro A, Dzieciatkowska M, et al. Preserved Proteins from Extinct Bison latifrons Identified by Tandem Mass Spectrometry; Hydroxylysine Glycosides are a Common Feature of Ancient Collagen. Molecular & Cellular Proteomics. 2015;14(7):1946-58.

15. Tenni R, Rimoldi D, Zanaboni G, Cetta G, Castellani AA. Hydroxylysine glycosides: preparation and analysis by reverse phase high performance liquid chromatography. Ital J Biochem. 1984 Mar-Apr;33(2):117-27.

16. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennet T. Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J Biol Chem. 2011 Dec 23;286(51):43701-9.

17. Myllyharju J, Kivirikko KI. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet. 2004 Jan;20(1):33-43.

18. Junqua S, Robert L, Garrone R, Pavans de Ceccatty M, Vacelet J. Biochemical and morphological studies on collagens of horny sponges. Ircinia filaments compared to spongines. Connect Tissue Res. 1974;2(3):193-203.

19. Stawikowski MJ, Aukszi B, Stawikowska R, Cudic M, Fields GB. Glycosylation modulates melanoma cell alpha2beta1 and alpha3beta1 integrin interactions with type IV collagen. J Biol Chem. 2014 Aug 1;289(31):21591-604.

20. 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 Dec 17;58(50):5040-51.

21. Terajima M, Taga Y, Cabral WA, Liu Y, Nagasawa M, Sumida N, et al. Cyclophilin B control of lysine post-translational modifications of skin type I collagen. PLoS Genet. 2019 Jun;15(6):e1008196.

22. 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 Jun;10(6):e1004465.

23. 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 Jan 15;45(2):252-6.

24. Dominguez LJ, Barbagallo M, Moro L. Collagen overglycosylation: a biochemical feature that may contribute to bone quality. Biochem Biophys Res Commun. 2005 Apr 29;330(1):1-4.

25. Tajima S, Takehana M, Azuma N. Production of overmodified type I procollagen in a case of osteogenesis imperfecta. J Dermatol. 1994 Apr;21(4):219-22.

26. Salo AM, Cox H, Farndon P, Moss C, Grindulis H, Risteli M, et al. A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. Am J Hum Genet. 2008 Oct;83(4):495-503.

27. Ewans LJ, Colley A, Gaston-Massuet C, Gualtieri A, Cowley MJ, McCabe MJ, et al. Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. J Med Genet. 2019 Sep;56(9):629-38.

28. Chen YL, Guo HF, Terajima M, Banerjee P, Liu X, Yu J, et al. Lysyl Hydroxylase 2 Is Secreted by Tumor Cells and Can Modify Collagen in the Extracellular Space. Journal of Biological Chemistry. 2016 Dec 9;291(50):25799-808.

29. Dayer C, Stamenkovic I. Recruitment of Matrix Metalloproteinase-9 (MMP-9) to the Fibroblast Cell Surface by Lysyl Hydroxylase 3 (LH3) Triggers Transforming Growth Factor-beta (TGF-beta) Activation and Fibroblast Differentiation. J Biol Chem. 2015 May 29;290(22):13763-78.

30. Qi Y, Xu R. Roles of PLODs in Collagen Synthesis and Cancer Progression. Front Cell Dev Biol. 2018;6:66.

31. Salo AM, Myllyharju J. Prolyl and lysyl hydroxylases in collagen synthesis. Exp Dermatol. 2020 Sep 23.

32. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem. 2008;77:521-55.

33. Ikeda Y, Takahashi M. 3.07 - Glycosyltransferases and Glycosidases: Enzyme Mechanisms. In: Kamerling H, editor. Comprehensive Glycoscience. Oxford: Elsevier; 2007. p. 115-28.

34. Mestrom L, Przypis M, Kowalczykiewicz D, Pollender A, Kumpf A, Marsden SR, et al. Leloir Glycosyltransferases in Applied Biocatalysis: A Multidisciplinary Approach. Int J Mol Sci. 2019 Oct 23;20(21).

35. Taujale R, Venkat A, Huang LC, Zhou Z, Yeung W, Rasheed KM, et al. Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. elife. 2020 Apr 1;9.

36. Albesa-Jove D, Giganti D, Jackson M, Alzari PM, Guerin ME. Structure-function relationships of membrane-associated GT-B glycosyltransferases. Glycobiology. 2014 Feb;24(2):108-24.

37. Ardevol A, Rovira C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. J Am Chem Soc. 2015 Jun 24;137(24):7528-47.

38. Ardevol A, Iglesias-Fernandez J, Rojas-Cervellera V, Rovira C. The reaction mechanism of retaining glycosyltransferases. Biochem Soc Trans. 2016 Feb;44(1):51-60.

39. Albesa-Jove D, Sainz-Polo MA, Marina A, Guerin ME. Structural Snapshots of alpha-1,3-Galactosyltransferase with Native Substrates: Insight into the Catalytic Mechanism of Retaining Glycosyltransferases. Angew Chem Int Ed Engl. 2017 Nov 20;56(47):14853-7.

40. He L, Ye X, Gao M, Yang J, Ma J, Xiao F, et al. Down-regulation of GLT25D1 inhibited collagen secretion and involved in liver fibrogenesis. Gene. 2020 Mar 1;729:144233.

41. Baumann S, Hennet T. Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase. J Biol Chem. 2016 Aug 26;291(35):18514-24.

42. Miyatake S, Schneeberger S, Koyama N, Yokochi K, Ohmura K, Shiina M, et al. Biallelic COLGALT1 variants are associated with cerebral small vessel disease. Ann Neurol. 2018 Dec;84(6):843-53.

43. Geister KA, Lopez-Jimenez AJ, Houghtaling S, Ho TH, Vanacore R, Beier DR. Loss of function of Colgalt1 disrupts collagen post-translational modification and causes musculoskeletal defects. Dis Model Mech. 2019 Jun 17;12(6).

44. Schegg B, Hulsmeier AJ, Rutschmann C, Maag C, Hennet T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. Mol Cell Biol. 2009 Feb;29(4):943-52.

45. Starzyk RM, Rosenow C, Frye J, Leismann M, Rodzinski E, Putney S, et al. Cerebral cell adhesion molecule: a novel leukocyte adhesion determinant on blood-brain barrier capillary endothelium. J Infect Dis. 2000 Jan;181(1):181-7.

46. Liefhebber JM, Punt S, Spaan WJ, van Leeuwen HC. The human collagen beta(1-O)galactosyltransferase, GLT25D1, is a soluble endoplasmic reticulum localized protein. BMC Cell Biol. 2010;11:33.

47. Webster JA, Yang Z, Kim YH, Loo D, Mosa RM, Li H, 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. 2017 Jun 30;37(3):BSR20170105.

48. 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(12):e29390.

49. Scietti L, Forneris F. Full-Length Human Collagen Lysyl Hydroxylases. In: Scott RA, editor. Encyclopedia of Inorganic and Bioinorganic Chemistry: Wiley; 2020. p. 1-12.

50. Salo AM, Wang C, Sipila L, Sormunen R, Vapola M, Kervinen P, et al. Lysyl hydroxylase 3 (LH3) modifies proteins in the extracellular space, a novel mechanism for matrix remodeling. J Cell Physiol. 2006 Jun;207(3):644-53.

51. Banushi B, Forneris F, Straatman-Iwanowska A, Strange A, Lyne AM, Rogerson C, et al. Regulation of post-Golgi LH3 trafficking is essential for collagen homeostasis. Nat Commun. 2016 Jul 20;7:12111.

52. Wang C, Ristiluoma MM, Salo AM, Eskelinen S, Myllyla R. Lysyl hydroxylase 3 is secreted from cells by two pathways. J Cell Physiol. 2012 Feb;227(2):668-75.

53. Rautavuoma K, Takaluoma K, Passoja K, Pirskanen A, Kvist AP, Kivirikko KI, et al. Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1-3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity. J Biol Chem. 2002 Jun 21;277(25):23084-91.

54. 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. Nat Commun. 2018 Aug 8;9(1):3163.

55. Ruotsalainen H, Sipila L, Vapola M, Sormunen R, Salo AM, Uitto L, et al. Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. J Cell Sci. 2006 Feb 15;119(Pt 4):625-35.

56. Risteli M, Ruotsalainen H, Salo AM, Sormunen R, Sipila L, Baker NL, et al. Reduction of lysyl hydroxylase 3 causes deleterious changes in the deposition and organization of extracellular matrix. J Biol Chem. 2009 Oct 9;284(41):28204-11.

57. Heikkinen J, Risteli M, Wang C, Latvala J, Rossi M, Valtavaara M, et al. Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. J Biol Chem. 2000 Nov 17;275(46):36158-63.

58. 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 Nov;21(7):559-66.

59. Sricholpech M, Perdivara I, Nagaoka H, Yokoyama M, Tomer KB, Yamauchi M. Lysyl hydroxylase 3 glucosylates galactosylhydroxylysine residues in type I collagen in osteoblast culture. J Biol Chem. 2011 Mar 18;286(11):8846-56.

60. Wang C, Kovanen V, Raudasoja P, Eskelinen S, Pospiech H, Myllyla R. The glycosyltransferase activities of lysyl hydroxylase 3 (LH3) in the extracellular space are important for cell growth and viability. J Cell Mol Med. 2009 Mar;13(3):508-21.

61. Risteli M, Ruotsalainen H, Bergmann U, Venkatraman Girija U, Wallis R, Myllyla R. Lysyl hydroxylase 3 modifies lysine residues to facilitate oligomerization of mannan-binding lectin. PLoS One. 2014;9(11):e113498.

62. Sricholpech M, Perdivara I, Yokoyama M, Nagaoka H, Terajima M, Tomer KB, et al. Lysyl hydroxylase 3-mediated glucosylation in type I collagen: molecular loci and biological significance. J Biol Chem. 2012 Jun 29;287(27):22998-3009.

63. Sipila L, Ruotsalainen H, Sormunen R, Baker NL, Lamande SR, Vapola M, et al. Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. J Biol Chem. 2007 Nov 16;282(46):33381-8.

64. Watt SA, Dayal JH, Wright S, Riddle M, Pourreyron C, McMillan JR, et al. Lysyl Hydroxylase 3 Localizes to Epidermal Basement Membrane and Is Reduced in Patients with Recessive Dystrophic Epidermolysis Bullosa. PLoS One. 2015;10(9):e0137639.

65. Taler K, Weiss O, Rotem-Bamberger S, Rubinstein AM, Seritrakul P, Gross JM, et al. Lysyl hydroxylase 3 is required for normal lens capsule formation and maintenance of lens epithelium integrity and fate. Dev Biol. 2019 Oct 24:in press.

66. Wang C, Risteli M, Heikkinen J, Hussa AK, 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 May 24;277(21):18568-73.

67. Chiapparino A, De Giorgi F, Scietti L, Faravelli S, Roscioli T, Forneris F. A cooperative network of molecular "hot spots" highlights the complexity of LH3 collagen glycosyltransferase activities. bioRxiv. 2020:841486.

68. 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 Biol. 2018 Nov 18.

69. Scietti L, Campioni M, Forneris F. SiMPLOD, a structure-integrated database of collagen lysyl hydroxylase (LH/PLOD) enzyme variants. J Bone Miner Res. 2019 Feb 5.

70. Sternberg M, Spiro RG. Studies on the catabolism of the hydroxylysine-linked disaccharide units of basement membranes and collagens. Isolation and characterization of a rat kidney alpha-glucosidase of high specificity. J Biol Chem. 1979 Oct 25;254(20):10329-36.

71. Hamazaki H, Hotta K. Enzymatic hydrolysis of disaccharide unit of collagen. Isolation of 2-O-alpha-D-glucopyranosyl-O-beta-D-galactopyranosyl-hydroxylysine glucohydrolase from rat spleens. Eur J Biochem. 1980 Oct;111(2):587-91.

72. Hamazaki H, Hotta K. Purification and characterization of an alpha-glucosidase specific for hydroxylysine-linked disaccharide of collagen. J Biol Chem. 1979 Oct 10;254(19):9682-7.

73. Hamazaki H, Hamazaki MH. Catalytic site of human protein-glucosylgalactosylhydroxylysine glucosidase: Three crucial carboxyl residues were determined by cloning and site-directed mutagenesis. Biochem Biophys Res Commun. 2016 Jan 15;469(3):357-62.

74. Sternberg M, Andre J. Glucose inhibition of the alpha-glucosidase specific for basement membrane and collagen disaccharide units. FEBS Lett. 1982 Mar 8;139(1):53-6.

75. Sternberg M, Andre J, Peyroux J. Inhibition of the alpha-glucosidase specific for collagen disaccharide units in diabetic rat kidney by in vivo glucose levels: possible contribution to basement membrane thickening. Diabetologia. 1983 Apr;24(4):286-9.

76. Palcic M, Sujino K. Assays for Glycosyltransferases. Trends in Glycoscience and Glycotechnology. 2001;13:361-70.

77. Kivirikko KI, Myllyla R. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods Enzymol. 1982;82 Pt A:245-304.

78. Kivirikko KI, Myllyla R. Collagen glycosyltransferases. Int Rev Connect Tissue Res. 1979;8:23-72.

79. Yan L, Smith DF, Cummings RD. Determination of GDP-Fuc:Gal beta 1-4GlcNAc-R (Fuc to GlcNAc) alpha 1,3 fucosyltransferase activity by a solid-phase method. Anal Biochem. 1994 Nov 15;223(1):111-8.

80. Schachter H, Brockhausen I, Hull E. High-performance liquid chromatography assays for N-acetylglucosaminyltransferases involved in N- and O-glycan synthesis. Methods Enzymol. 1989;179:351-97.

81. Wolff MW, Bazin HG, Lindhardt RJ. Analysis of fluorescently labeled oligosaccharides by capillary electrophoresis and electrospray ionization mass spectrometry. Biotechnology Techniques. 1999 1999/11/01;13(11):797-801.

82. Taga Y, Kusubata M, Ogawa-Goto K, Hattori S. Development of a Novel Method for Analyzing Collagen O-glycosylations by Hydrazide Chemistry. Molecular & Cellular Proteomics. 2012;11(6):M111.010397.

83. Taga Y, Kusubata M, Ogawa-Goto K, Hattori S. Site-specific Quantitative Analysis of Overglycosylation of Collagen in Osteogenesis Imperfecta Using Hydrazide Chemistry and SILAC. Journal of Proteome Research. 2013 2013/05/03;12(5):2225-32.

84. Yamauchi M, Taga Y, Hattori S, Shiiba M, Terajima M. Chapter 6 - Analysis of collagen and elastin cross-links. In: Mecham RP, editor. Methods in Cell Biology: Academic Press; 2018. p. 115-32.

85. Yamauchi M, Sricholpech M, Terajima M, Tomer KB, Perdivara I. Glycosylation of Type I Collagen. In: Kannicht C, editor. Post-Translational Modification of Proteins: Tools for Functional Proteomics. New York, NY: Springer New York; 2019. p. 127-44.

86. Zhang Y, Yu C-Y, Song E, Li SC, Mechref Y, Tang H, et al. Identification of Glycopeptides with Multiple Hydroxylysine O-Glycosylation Sites by Tandem Mass Spectrometry. Journal of Proteome Research. 2015 2015/12/04;14(12):5099-108.

87. Trubetskoy OV, Shaw PM. A fluorescent assay amenable to measuring production of beta-D-glucuronides produced from recombinant UDP-glycosyl transferase enzymes. Drug Metab Dispos. 1999 May;27(5):555-7.

88. Kumagai K, Kojima H, Okabe T, Nagano T. Development of a highly sensitive, high-throughput assay for glycosyltransferases using enzyme-coupled fluorescence detection. Anal Biochem. 2014 Feb 15;447:146-55.

89. Guo HF, Cho EJ, Devkota AK, Chen Y, Russell W, Phillips GN, Jr., et al. A scalable lysyl hydroxylase 2 expression system and luciferase-based enzymatic activity assay. Arch Biochem Biophys. 2017 Mar 15;618:45-51.

90. Guo HF, Tsai CL, Terajima M, Tan X, Banerjee P, Miller MD, et al. Pro-metastatic collagen lysyl hydroxylase dimer assemblies stabilized by Fe(2+)-binding. Nat Commun. 2018 Feb 6;9(1):512.

91. Schneider VA, Granato M. The myotomal diwanka (lh3) glycosyltransferase and type XVIII collagen are critical for motor growth cone migration. Neuron. 2006 Jun 1;50(5):683-95.

92. Gruber R, Rogerson C, Windpassinger C, Banushi B, Straatman-Iwanowska A, Hanley J, et al. Autosomal recessive Keratoderma-Ichthyosis-Deafness (ARKID) syndrome is caused by VPS33B mutations affecting Rab protein interaction and collagen modification. J Invest Dermatol. 2016 Dec 22;S0022-202X(16):32800-7.

93. De Luca G, Tenni R, Lauria A, Cetta G, Salvini R, Zanaboni G, et al. Hypothesis on the role of hydroxylysyl glycosides in collagen fibre organization. Ital J Biochem. 1983 Nov-Dec;32(6):418-30.

3.Chapter

MOLECULAR ARCHITECTURE OF THE MULTIFUNCTIONAL COLLAGEN LYSYL HYDROXYLASE AND GLYCOSYLTRANSFERASE LH3

This chapter is a research article published the 8 August 2018 in Nature Communication (doi: 10.1038/s41467-018-05631-5), it includes the work I did during my first year as PhD student. In details, I optimized and carried out the enzymatic assays for the biochemical evaluation of the activities of LH3 wild-type and mutanted variants. Also, I helped in the expression and purification of LH3 mutants.

Authors:
Luigi Scietti[1], Antonella Chiapparino[1], Francesca De Giorgi[1], Marco Fumagalli[2], Lela Khoriauli[3], Solomon Nergadze[3], Shibom Basu[4], Vincent Olieric[4], Lucia Cucca[5], Blerida Banushi[6,8], Antonella Profumo[5], Elena Giulotto[3], Paul Gissen[6,7] & Federico Forneris[1]

Affiliations:
1 The Armenise-Harvard Laboratory of Structural Biology, Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9/A, 27100 Pavia, Italy. 2 Laboratory of Biochemistry, Department of Biology and Biotechnology, University of Pavia, Via Taramelli 3/B, 27100 Pavia, Italy. 3 Laboratory of Molecular Biology, Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9/A, 27100 Pavia, Italy. 4 Swiss Light Source, Paul Scherrer Institut, Villigen 5232, Switzerland. 5 Laboratory of Analytical Chemistry, Department of Chemistry, University of Pavia, Via Taramelli 12, 27100 Pavia, Italy. 6 MRC Laboratory for Molecular Cell Biology, University College London, London WC1E 6BT, UK. 7 UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK. 8 Present address: Translational Research Institute, The University of Queensland Diamantina Institute, Princess Alexandra Hospital, 37 Kent Street, Brisbane, Australia.

Corresponding author: F.F. (email: federico.forneris@unipv.it)

Supplementary information available online at the following link: https://www.nature.com/articles/s41467-018-05631-5#additional-information
They include supplementary figures from 1 to 17 and supplementary tables 1-2-3 and 4.

ABSTRACT

Lysyl hydroxylases catalyze hydroxylation of collagen lysines, and sustain essential roles in extracellular matrix (ECM) maturation and remodeling. Malfunctions in these enzymes cause severe connective tissue disorders. Human lysyl hydroxylase 3 (LH3/PLOD3) bears multiple enzymatic activities, as it catalyzes collagen lysine hydroxylation and also their subsequent glycosylation. Our understanding of LH3 functions is currently hampered by lack of molecular structure information. Here, we present high resolution crystal structures of full-length human LH3 in complex with cofactors and donor substrates. The elongated homodimeric LH3 architecture shows two distinct catalytic sites at the N- and C-terminal boundaries of each monomer, separated by an accessory domain. The glycosyltransferase domain displays distinguishing features compared to other known glycosyltransferases. Known disease-related mutations map in close proximity to the catalytic sites. Collectively, our results provide a structural framework characterizing the multiple functions of LH3, and the molecular mechanisms of collagen-related diseases involving human lysyl hydroxylases.

3.1 INTRODUCTION

Collagen biosynthesis requires multiple post-translational modifications essential for the generation of mature, triple-helical molecules[1]. Modification of collagen lysines enables subsequent glycosylation and formation of extracellular cross-links, leading to fibrillary or meshwork superstructures[2]. Enzymes belonging to the family of collagen lysyl hydroxylases (LH or PLOD) catalyze lysine hydroxylation of collagens using Fe^{2+}, 2-oxoglutarate (2-OG), ascorbate and molecular oxygen[3,4]. In humans, PLOD genes encode for three LH enzyme isoforms sharing >60% amino acid sequence identity: LH1, LH2a/b, and LH3, respectively[5]. Mutations in PLOD genes that reduce or abolish LH activity are associated with severe connective tissue diseases including Ehlers-Danlos[6] and Bruck syndromes[7,8]. In mouse models, LH3 knock-outs are embrionically lethal[9,10]. Mutations in the PLOD3 gene also result in impaired collagen glycosylation, secretion, and basement membrane formation, yielding phenotypes resembling osteogenesis imperfecta[11]. Conversely, PLOD overexpression and upregulated enzymatic activity have been linked to fibrosis[12], and recently also to hypoxia-induced metastatic spreading of solid tumors with poor prognosis[13,14,15].

LH3 is considered the evolutionary ancestor of the LH family: this isoform is the only one capable of further processing of hydroxylysines through glycosylation, whereas other isoforms might have lost such capability during evolution[16]. LH3 is therefore a multifunctional enzyme capable of converting collagen lysines into 1,2-glucosylgalactosyl-5-hydroxylysines through three consecutive reactions: hydroxylation of collagen lysines

(LH activity), N-linked conjugation of galactose to hydroxylysines (GT activity), and conjugation of glucose to galactosyl-5-hydroxylysines (GGT activity)[17,18]. Biochemical data suggest that these different enzymatic activities are localized in distinct compartments of the enzyme[19], but despite the extensive evidence available, the current knowledge of LH enzymes is far from exhaustive. These enzymes are known to act together with prolyl hydroxylases, respectively introducing hydroxylations of lysine and proline residues on procollagens in the endoplasmic reticulum (ER), prior to the formation of triple-helical assemblies[20]. In line with this, LH enzymes are found as ER-resident proteins albeit they do not possess specific ER-retention sequences[21,22.] Reports suggest that ER retention could be mediated via interaction with distinct ER-resident proteins: LH1 is described to be part of a macromolecular complex with SC65, P3H3 and CYPB[23]; while LH2 forms a complex with HSP47, FKBP65 and BiP[24,25]; LH3 was found colocalizing with collagen galactosyltransferases GLT25D1/2[26]. Multiple reports identify LH3 also in the extracellular space and suggest dedicated trafficking mechanisms for its secretion[27,28,29,30]. Abnormalities in LH3 post-Golgi trafficking are associated with devastating developmental diseases with phenotypes characterized by immature collagen accumulation and lack of its secretion, very similar to those observed in case of enzyme malfunctions caused by knock-down or inactivation[9,10,11,30,31]. Very recently, LH2 secretion has been reported associated with hypoxia-induced PLOD2 overexpression in metastatic tumor microenvironments[13,15]. Extracellular LHs were reported to be active, suggesting implications for ECM stability and remodeling[27]. These data indicate that although lysine modifications are known to occur in the ER prior to collagen triple helical formation, secreted variants of LH3 and LH2 can modify collagens in different compartments and, possibly, in different folding states[32].

The accumulated knowledge about the precise molecular roles and mechanisms associated with LH enzymes has suffered from the lack of molecular structure models fundamental to shed light on the complexity and the diversity of this important enzyme family. Here, we present the crystal structures of multifunctional full-length human LH3 in complex with various cofactors and donor substrates. The structures reveal a multidomain architecture characterized by two independent catalytic sites devoted to the different enzymatic activities and provide a molecular understanding that has implications for various disease-related mutations found in LH enzymes. Altogether, our results offer new insights into the complex mechanisms of collagen biosynthesis and homeostasis, and provide structural templates for the development of targeted therapies for LH-related diseases and cancer.

3.2 METHODS

3.2.1 Chemicals

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

3.2.2 DNA constructs

Human LH3 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 LH3 sequence devoid of the N-terminal signal peptide into the pUPE.106.08 expression vector, kindly provided by U-protein Express, BV (U-PE, The Netherlands) and into the pPuro-DHFR, containing the Streptomyces alboniger puromycin resistance gene (isolated from the pPUR plasmid, Clontech) and the mouse dihydrofolate reductase cDNA47. Both expression vectors bear N-terminal signal peptide followed by a N-terminal 6xHis-tag and a recognition site for Tobacco Etch Virus (TEV) protease prior to an in-frame BamHI restriction site, as well as an in-frame stop codon after the NotI restriction site. LH3 mutants were generated using the Phusion Site Directed Mutagenesis Kit (ThermoFisher Scientific) following manufacturer's instructions. The entire plasmid was amplified using phosphorylated primers. For all mutants, the forward primer introduced the mutation of interest (Supplementary table 4). All expression plasmids were checked by Sanger sequencing prior to usage.

3.2.3 Recombinant LH3 expression from stable HeLa cell lines

The pPuro-DHFR-LH3 construct was transfected into human cervical carcinoma cells (HeLa S3, provided by ATCC and further selected for high transfection efficiency by Dr.F. Peverali, Consiglio Nazionale delle Ricerche, Pavia) using the Lipofectamine LTX reagent (Invitrogen). Cells were not authenticated and not tested for mycoplasma contamination. Stably transfected clones, isolated with 1 mg mL^{-1} puromycin (Invivogen), were subjected to step-wise selection with increasing methotrexate concentrations to select for cells containing multiple copies of the plasmid. Cells were routinely cultured at 37 °C in 5% CO2 in high-glucose DMEM supplemented with 10% foetal calf serum (Biowest), 1× non-essential amino acids, 2 mM L-glutamine and 1× penicillin-streptomycin. Clones expressing high yields of PLOD3 were identified by SDS-PAGE analysis after imidazole elution from small-scale immobilized metal ion affinity purification using Nickel sepharose beads (GE Healthcare).

3.2.4 Recombinant LH3 expression from transient HEK293 cells

Recombinant tagged LH3 mutants were produced using suspension cultures of HEK293F (Invitrogen) cells. Cells were not authenticated and not tested for mycoplasma contamination. Cells were transfected at cell densities of 1 million mL−1 using 3 µg of polyethyleneimine (PEI; Polysciences, Germany) for 1 µg of pUPE.106.08-LH3 plasmid DNA per mL of cells. Cultures were supplemented with 0.6% Primatone RL 4 h after transfection. The cell medium containing secreted LH3 was collected 6 days after transfection by centrifugation at 1000 × g for 15 min.

3.2.5 Purification of LH3 enzymes

The LH3-containing medium from either HeLa or HEK293 cell cultures was filtered through a syringe 0.2 mm filter. The pH and ionic strength of the filtrated medium were adjusted using a 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. LH3 was purified using affinity and size-exclusion chromatography on Äkta systems (GE Healthcare). 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 HiTrap desalting FF column (GE Healthcare) equilibrated in 25 mM HEPES, NaOH, 500 mM NaCl, pH 8.0. The N-terminal histidine-tag was cleaved using overnight TEV protease digestion at 4 °C followed by affinity-based removal of TEV protease and His-tag using a 5 mL HisTrap FF (GE Healthcare). The protein was concentrated to 5 mg mL^{-1} 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-containing fractions as assessed from SDS-PAGE analysis were pooled, concentrated and stored at −80 °C until further usage.

3.2.6 LH3 deglycosylation

Wild-type LH3 and mutants T672N, L714N, and W148N-L150T were subjected to deglycosylation to validate the introduction of an additional glycosylation site through mutagenesis. 20 µL protein at 0.15 mg mL^{-1} were first incubated with 1X Glycoprotein Denaturing Buffer (New England BioLabs) and denatured at 95 °C for10 min. 1X glycobuffer 2 (New England BioLabs), 1% NP-40, and 0.2 µL (100 Units) of PNGase F (New England BioLabs) were added to the reaction mix, which was further incubated for 2 h at 37 °C. PNGase F-treated and untreated samples were then analyzed using western blotting with rabbit anti-LH3 antibody (Proteintech 11027–1-AP) in a 1:1000 ratio followed by anti-rabbit HRP-conjugate antibody (Sigma-Aldrich A0545) in a 1:3000 ratio.

3.2.7 ICP-MS measurements

To measure the number of tightly bound divalent metal ions 3 mg of LH3 were diluted in 5 mL 25 mM HEPES, 500 mM NaCl, pH 8.0. Three aliquots of 1 ml each were kept overnight with 0.5 mL 65% ultra-pure HNO3 and 0.1 mL 30% w/w H2O2, then diluted to 5 mL with Milli-Q water and analyzed by ICP-MS. The measurements of Fe and Mn were performed on a Perkin Elmer Mod ELAN DRC-e instrument, following the standard procedures suggested by the manufacturer. Quantitative determinations were obtained by the external standard calibration with five standards (0, 5, 10, 50, 100, and 300 µg L^{-1}) daily prepared in the same buffer used for samples preparation, at the same dilution and HNO3 concentration. Only Fe was quantified in each replicate solution, with

standard deviation of the mean value of 8%, obtaining a molar ratio Fe/LH3 = 1, while Mn was present as impurity in the blank and in the samples.

3.2.8 Crystallization of LH3

LH3 spherulites were found in nanoliter-dispensed droplets (0.1 µL protein at 4 mg mL^{-1} + 0.1 µL reservoirs) using commercial crystallization screens in sitting vapor diffusion drop plates. These spherulites were initially optimized by mixing 0.5 µL of protein concentrated at 3.5 mg mL^{-1} and 0.5 µL of reservoir solution composed of 600 mM sodium formate, 12% poly-glutamic-acid (PGA-LM, Molecular Dimensions), 100 mM HEPES/NaOH, pH 7.8. These crystals diffracted to a maximum of 5 Å. Crystal quality was improved through sequential runs of macro-seeding of LH3 crystals using the same crystallization mixture with slight variations in protein concentration. Co-crystallization experiments were performed by setting up the same seeding conditions, and supplementing the protein solution with mixtures of 500 µM FeCl2, 500 µM MnCl2, 1 mM UDP-galactose, 1 mM UDP-glucose (Supplementary Table 1). Crystals were cryo-protected with the mother liquor supplemented with 20% glycerol, harvested using MicroMounts Loops (Mitegen), flash-cooled and stored in liquid nitrogen prior to data acquisition. Heavy atom derivatives were prepared by soaking the LH3 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 cryo protection, harvesting and flash-cooling in liquid nitrogen.

3.2.9 Diffraction data collection and structure refinement

Diffraction data from LH3 crystals were collected at various beamlines of the European Synchrotron Radiation Facility, Grenoble, France and at the Swiss Light Source, Villigen, Switzerland (details in Supplementary Table 1). Single wavelength Anomalous Dispersion (SAD) experiments at the Hg edge were performed at the ESRF ID30B beamline, whereas high multiplicity long-wavelength native SAD data (57 data sets of 360° from 6 different crystals) were collected at the SLS X06DA beamline as described elsewhere[48]. The data, which showed strong anisotropy (Supplementary Fig. 3), were processed with autoPROC[49] including STARANISO[50]. Statistics are summarized in Suppl. table 1. Hg heavy atom sites were identified using SHELXC/D[51] with the HKL2MAP GUI[52]. Experimental phasing with the Hg SAD data using SHELXE[51] and SHARP[53] yielded a partial model. Completion of the model could only be achieved by combining the latter with high multiplicity native SAD data using the CRANK2 pipeline[54] followed by iterations of automatic and manual model building using BUCCANEER[55] and COOT[56]. Subsequent LH3 structures were determined using the initial LH3 structural model in molecular replacement runs with PHASER[57]. Final 3D models were generated using iterations of automatic refinement using PHENIX[58] alternated with manual adjustments using COOT[56]. Validation of structure quality was carried out with Molprobity[59], the

RCSB PDB Validation Server[60], and PDB-CARE[61]. Final refinement statistics are listed in Suppl. Table 1. Structural figures were generated using PyMol[62].

3.2.10 SAXS data collection and analysis

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 0.03–4.5 nm with a wavelength of 0.01 nm. SEC-SAXS experiments were carried out using Nexera High Pressure Liquid/Chromatography (HPLC; Shimadzu) system connected online to SAXS sample capillary[63]. For these experiments, 50 µL of LH3 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 offline batch sample analysis, 50 µL of LH3 at concentrations ranging from 0.5 to 9 mg mL^{-1} were injected using the dedicated automatic sample changer available at the BM29 beamline[64]. For SEC-SAXS data, frames corresponding to LH3 protein peak were identified, blank subtracted and averaged using CHROMIXS[65], whereas batch concentration series were analyzed using PRIMUS[66]. Radii of gyration (Rg), molar mass estimates and distance distribution functions P(r) were computed using the ATSAS package[67] in PRIMUS[66]. Comparison of experimental SAXS data and 3D models from crystal structures was performed using CRYSOL[68]. A summary of SAXS data collection and analysis results is shown in Suppl. Table 3.

3.2.11 Determination of LH activity using mass spectrometry

Synthetic collagen peptides were purchased from China peptides. Peptides tested were ARGIKGIRGFS, GIKGIKGIKGIK, and IKGIKGIKG sequences. LH3 5 µM was incubated with 500 µM FeCl2, 1 mM 2-OG, 2 mM ascorbate and 1 mM peptide substrate. Reactions were allowed to proceed for 1 h at 37 °C. In total 20 µL of each sample were previously acidified by addition of 1 µL of formic acid (FA) and then analyzed on an LC–MS system (Thermo Finnigan, USA) consisting of a thermostated column oven Surveyor autosampler controlled at 25 °C; a quaternary gradient Surveyor MS pump equipped with an UV/vis detector and an Ion Trap (LCQ Advantage Max) mass spectrometer with electrospray ionization ion source controlled by Xcalibur software 2.0.7. Peptides were separated by RP-HPLC on a Jupiter (Phenomenex, USA) C18 column (150 × 2 mm, 4 µm, 90 Å particle size) using a linear gradient (2–60% solvent B in 60 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 ion mode under constant instrumental conditions: source voltage 5.0 kV, capillary voltage 46 V, sheath gas flow 20 (arbitrary units), auxiliary gas flow 10 (arbitrary units), sweep gas flow 1 (arbitrary units), capillary temperature 200 °C, tube lens voltage −105 V. Spectra analyses were performed using Xcalibur software 2.0.7.

3.2.12 Biochemical evaluation of LH activity

Reaction mixtures (5 µL total volume) containing wild-type or mutant LH3 samples at 0.2 mg mL^{-1} were prepared by sequentially adding 0–1 mM peptide substrate or 4 mg mL^{-1} gelatin in water, (solubilized through heating denaturation at 95 °C for 10 min), 500 µM ascorbate, 100 µM 2-OG, and 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 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, 2-OG or peptide substrates. Data were analyzed and plotted using the GraphPad Prism 7 software[69].

3.2.13 Biochemical evaluation of GT and GGT enzymatic activities

Reaction mixtures (5 µL total volume) containing wild-type or mutant LH3 samples at 0.2 mg mL^{-1} were prepared by sequentially adding 0–1 mM peptide substrate, 500 µM ascorbate, 100 µM 2-OG, 50 µM FeCl2, and let incubate for 1 h at 37 °C. Reactions mixtures were then supplemented with 50 µM MnCl2, and 50µM UDP-Gal or UDP-Glc, and let incubate for 1 h at 37 °C. Experiments using gelatin as substrate were performed by sequentially adding 4 mg mL^{-1} gelatin in water, (solubilized through heating denaturation at 95 °C for 10 min), 50 µM MnCl2, and 100 µM UDP-Gal or UDP-Glc to the LH3 samples at 0.2 mg mL^{-1}, and let incubate for 1 h at 37 °C. All reactions were stopped by heating at 95 °C for 2 min, prior to transfer into Proxiplate white 384-well plates (Perkin-Elmer), then 5 µL of the UDP-Glo luminescence detection reagent (Promega) were added and let incubate 1 h at 25 °C. Detection was carried out as described for the LH enzymatic activity. All experiments were performed in triplicates. Control experiments were performed using identical conditions by selectively removing LH3, donor or acceptor substrates. Data were analyzed and plotted using the GraphPad Prism 7 software[69].

3.2.14 Surface-plasmon resonance

Wild-type and mutant LH3 preparations were immobilized onto a carboxymethylated dextran (CM5) sensor chip (GE Healthcare) using a mixed solution of 200 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) in 50 mM N-hydroxysuccinimide (NHS) in a Biacore T200 SPR instrument (GE Healthcare). Excess reactive groups were blocked with 1 M ethanolamine. Efficient immobilization of LH3 samples was judged based on the SPR signal collected. For each of the tested samples,

3000 RU were reached. A control flow cell 1 was pre-activated and blocked using the same protocol as above but without protein samples, and used as reference cell during measurements. Collagen peptides were dissolved in running buffer (PBS-P 0.01%) and injected as two-fold dilution concentration series of eight points each using a flow of 5 µl min^{-1}. Two replicates of each concentration were injected. Data analysis was performed using the Biacore T200 evaluation software (GE Healthcare) using a 1:1 steady-state affinity model.

3.2.15 Data availability

Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession codes 6FXK, 6FXM, 6FXR, 6FXT, 6FXX, 6FXY. SEC-SAXS experimental data and ab-initio model have been deposited in Small Angle Scattering Biological Data Bank (SASBDB) with accession code SASDDW4. Other data are available from the corresponding author upon reasonable request.

3.3 RESULTS

3.3.1 LH3 has three domains encompassing multiple catalytic sites

We have generated human stable cell lines for large-scale production of full length, glycosylated human LH3, and established methods for its purification and evaluation of its LH and GT enzymatic activities (Supplementary Fig. 1). ICP-MS analyses indicated that all enzyme preparations contained Fe^{2+} with a 1:1 stoichiometry (see Methods). We observed significant uncoupling (up to 25%) of donor substrate activation, with substrate-independent generation of the succinate or UDP reaction products, respectively. Nevertheless, we could detect significantly increased enzymatic activity in the presence of synthetic peptides (Supplementary Fig. 1C) or gelatin (Supplementary Fig. 1D) as acceptor substrates. We could confirm such reactivity by detection of concentration-dependent binding (Supplementary Fig. 2A) and appearance of post-translationally modified lysine residues on synthetic peptides upon LH3 treatment (Supplementary Fig. 2B). In our peptide binding measurements, all surface plasmon resonance (SPR) profiles were characterized by very fast association and dissociation events and very weak (millimolar) binding affinities (Supplementary Fig. 2A).

Crystal structures of LH3 were determined in complex with various substrates and cofactors at resolutions ranging between 2.1 and 3.0 Å (Supplementary Table 1). Diffraction data were systematically affected by strong anisotropy (Supplementary Fig. 3), thus structure determination required anisotropy correction followed by a combination of experimental phasing with heavy atom and highly redundant native single wavelength anomalous dispersion (SAD), eventually yielding electron density maps of superb quality (Supplementary fig. 4A, B). Residues Asn63 and Asn548 showed extended electron densities protruding from their side chains (Supplementary fig. 4C), indicating the

expected N-linked glycosylations. The LH3 monomer encompasses three domains aligned along one direction (Fig. 3-1a). The first two N-terminal domains show Rossmann-fold architectures reminiscent of glycosyltransferases[33,34], whereas the C-terminal domain is characterized by a double-stranded β-helix (DSBH) fold, highly conserved among the 2-OG, Fe^{2+}-dependent dioxygenases[35,36]. Overall, the three-dimensional structure of the LH3 full-length enzyme provides a molecular blueprint to elucidate previous suggestions on the enzyme architecture based on biochemical data: GT and GGT activities localize at the N-terminus of the enzyme, whereas the LH activity is segregated at the LH3 C-terminus[37,38].

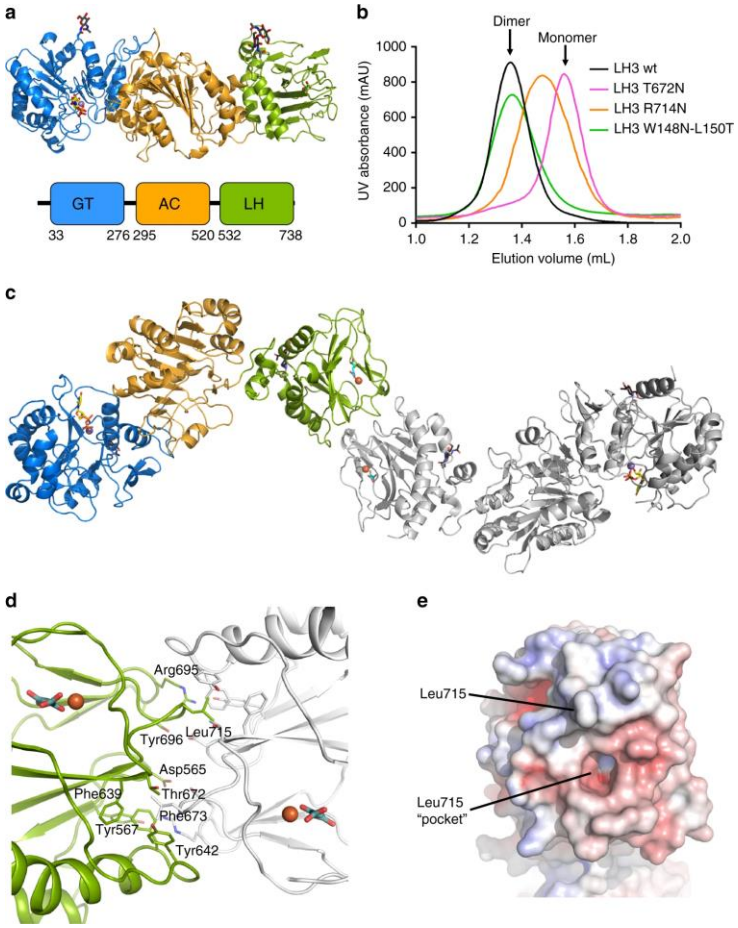


Figure 3-1 Molecular architecture of human LH3

a. Cartoon representation of the LH3 enzyme, showing its organization with three domains aligned from the N- to the C- terminus. Based on results of functional assays, the first two glycosyltransferase domains have been named GT (catalytic glycosyltransferase, blue) and AC (accessory, orange), respectively. The C-terminal domain hosts the Fe^{2+} and 2-OG cofactors necessary for lysyl hydroxylase activity, and therefore has been named LH (green). Metal ions are shown as spheres, cofactors and glycans as sticks. b. Introduction of additional glycosylation sites identifies the LH3 dimer interface in solution. Glycosylated mutants R714N and T672N induce disruption of the dimeric assembly, as observed in analytical size exclusion chromatography experiments. Control glycosylated mutant W148N-L150T, located at a crystal contact interface, is a dimer as wild-type LH3. c. Overview of the LH3 dimer as observed in the crystal structures. The quaternary arrangement highlights an elongated tail-to-tail dimer extending for over 20 nm in one direction, connected through strong electrostatic and hydrophobic interactions near the LH catalytic site. Fe^{2+} and 2-OG in the neighboring catalytic site are shown with spheres and sticks, respectively. For clarity, one LH3 monomer is colored as in a, whereas the other is shown in white. d. The molecular interface connecting two C-terminal LH domains in the LH3 crystal structure is characterized by strong hydrophobic interactions involving Leu715 of one monomer (green) and various aromatic residues shaping a cavity on the opposite monomer (white). This hydrophobic contact is surrounded by electrostatic interactions. For clarity, only the amino acids of one monomer are labeled. e. Details of the LH3 dimer interface. Shown is the electrostatic potential computed using APBS[70]colored from $-10\ k_bTe_c^{-1}$ (red) to $+10\ k_bTe_c^{-1}$ (blue)

3.3.2 LH3 forms elongated tail-to-tail dimers

Although the asymmetric units of our LH3 crystal structures contain a single LH3 monomer, previous biochemical studies[20,38] and our size exclusion chromatography coupled to small-angle X-ray scattering (SEC-SAXS) analyses consistently showed 200 kDa dimers in solution (Supplementary Fig. 1A). The crystal packing indeed suggests two homodimeric arrangements with physiologically plausible assemblies (Supplementary Fig. 5). A first, elongated, tail-to-tail quaternary structure shares nearly identical dimer interface with that observed recently in a C-terminal fragment of a viral LH homolog (L230-LH)[39]. This interface interconnects the C-terminal domains of LH3 and exposes individual glycosyltransferase domains at the two sides of the dimer (Supplementary Fig. 5B). A second, more compact antiparallel conformation is characterized by contacts between the glycosyltransferase domains and exposes individual lysyl hydroxylase domains at the two sides of the dimer (Supplementary Fig. 5C). Both interfaces are characterized by a large, buried surface area and numerous hydrogen bonds and hydrophobic interactions. Previous biochemical characterizations based on C-terminal LH3 deletions indicated residues Lys541-Glu547 as essential for dimerization[38]. In our structures, both observed dimeric assemblies fully support this statement, as this region is located in a linking platform connecting the central glycosyltransferase domain with the C-terminal domain (Fig. 3-1a). Initial attempts using SAXS and computational methods to discriminate between crystallographic contacts and stable dimers in solution (Supplementary Fig. 5–6) were not conclusive. A recent report suggested that in both homologous viral L230-LH domain and in human LH2, replacement of the fully conserved, surface-exposed C-terminal residue Leu715 (LH3 numbering, located in the middle of the "elongated" dimer interface found in the LH3 crystal structure) with a charged Asp could disrupt the enzyme's dimeric assembly and generate inactive, monomeric species in solution[39]. We took advantage of this information and introduced the corresponding L715D mutation in LH3; we also opted for generating a second mutant, bearing a positive Arg side chain replacing Leu715. Surprisingly, the L715D mutant was comparable to wild-type LH3 in activity assays and in analytical size exclusion chromatography experiments, while we observed slightly increased retention volumes and abolished LH activity for the L715R mutant (Supplementary Fig. 7). Similarly, removal of Fe^{2+} using chelating agents and acidification, successfully exploited to destabilize both viral LH and human LH2 dimer interfaces[39] did not seem to affect LH3 stability (Supplementary Fig. 7C). We therefore decided to introduce more pronounced steric hindrance at the observed crystallographic dimer interfaces, through mutations carrying additional glycosylation sites. Mutants T672N and R714N, both adding a glycosylation near the C-terminal dimer interface as verified by SDS-PAGE and enzymatic deglycosylation experiments (Supplementary Fig. 7A, B, Supplementary Fig. 17), shifted the size exclusion retention volumes towards monomeric species (Fig. 3-1b) and abolished LH activity (Supplementary Fig. 7D). On the contrary, the LH3 mutant

W148N-L150T, introducing an additional glycosylation at the "compact" N-terminal interface did not affect the dimeric quaternary structure of LH3 nor the LH enzymatic activity (Fig. 3-1b, Supplementary Fig. 7D). We therefore concluded that the physiological LH3 dimeric assembly corresponds to the elongated, tail-to-tail arrangement shown in Fig. 3-1c. This interface is characterized by two-fold symmetric interactions involving a rather limited set of aminoacid side chains engaged in electrostatic contacts, plus a deep hydrophobic cavity shaped by Phe673, Phe639, Tyr642 and Thr672 hosting the side chain of Leu715 from the opposite monomer (Fig. 3-1d, e).

3.3.3 Structural insights into LH3 glycosyltransferase activity

The N-terminal LH3 glycosyltransferase domain partially shares tertiary structure topology with divalent metal ion-dependent class-A glycosyltransferase folds (GT-A)[33], but with distinguishing structural features as expected given the very low sequence identity conservation, lower than 12%. Indeed, although numerous three-dimensional structures of GT-A glycosyltransferases are available in the protein data bank, superpositions with even the closest structural homolog yielded root mean square deviations (r.m.s.d.) higher than 3 Å (Supplementary Fig. 8A). Notably, this LH3 domain lacks a highly conserved α-β hairpin at its N-terminus near residue Gly70, and includes other structural elements surprisingly well conserved at the primary sequence level within the LH enzyme family (Supplementary Fig. 9), but distinct from other glycosyltransferases. In particular, the conformations of four loops differ from GT-A structures and shape the substrate-binding face of the N-terminal domain of LH3 (Supplementary Fig. 8A). Among these, a very flexible surface loop comprising residues Gly72 to Gly87 is not visible in the electron density of ligand-free LH3 structures. This loop contains several residues highly conserved among LH isoforms (Supplementary Fig. 9), which are not found in other glycosyltransferases. A cavity characterized by aspartate residues 112 and 115 and His253 shapes the metal ion binding site. Co-crystallizations with Mn^{2+} resulted in appearance of strong electron density proximate to these residues for the metal ion and two coordinating water molecules, without observable conformational changes compared to ligand-free LH3 (Supplementary Fig. 10A). Co-crystallizations with Mn^{2+} and donor substrates UDP-galactose or UDP-glucose yielded additional clear electron densities for UDP, but not for the glycan moieties (Fig. 2a). We observed weak electron density near the UDP pyrophosphate group partially compatible with glycan donor substrates, but we refrain from modeling anything inside this weak density, likely representative of multiple conformations simultaneously trapped in the substrate binding cavity (Supplementary Fig. 10B). We could not detect significant differences when comparing co-crystal structures obtained using UDP-Gal or UDP-Glc donor substrates. Nonetheless, binding of these donor substrates induced dramatic conformational changes in the enzyme's catalytic site, with full stabilization of the flexible Gly72-Gly87 loop in a "closed" conformation (Fig. 3-2b). The UDP pyrophosphate

group is stabilized by interactions with Mn^{2+} and hydrogen bonding with Lys259 and backbone nitrogen of Gly256; both residues are positioned in a uniquely shaped α-helix located at the C-terminus of the domain. The hydroxyl groups of the ribose form a network of hydrogen bonds with backbone atoms of Ser113 and Tyr114. The uracil moiety is sandwiched through π-π stacking interactions between Trp75 and Tyr114, and is stabilized by hydrogen bonding with Thr46 (Fig. 3-2a, b, Supplementary Fig. 10). Of note, these two residues highlight an unprecedented arrangement of UDP binding residues in glycosyltransferases: Trp75 belongs to the distinctive LH3 flexible loop covering residues Gly72-Gly87, that becomes fully stabilized upon substrate binding (Fig. 3-2b); Tyr114 is part of a non-canonical DxxD motif (Supplementary fig. 9), where Asp112, and Asp115 are responsible for Mn^{2+} coordination. Structurally related glycosyltransferases also often bear tyrosine residues stacking with the UDP moiety, but these residues are located far in sequence from the canonical DxD motif[40,41] responsible for metal ion coordination. Site-directed mutagenesis on Trp75 or Tyr114 into alanine residues yielded folded, but almost completely inactive LH3 variants (Fig. 3-2c, Supplementary fig. 11). Binding data using SPR on synthetic collagen peptides showed very limited differences between wild-type and mutant LH3 (Supplementary Fig. 11D). Together, these results highlight the distinguishing roles of Trp75 and Tyr114 in donor substrate binding and stabilization.

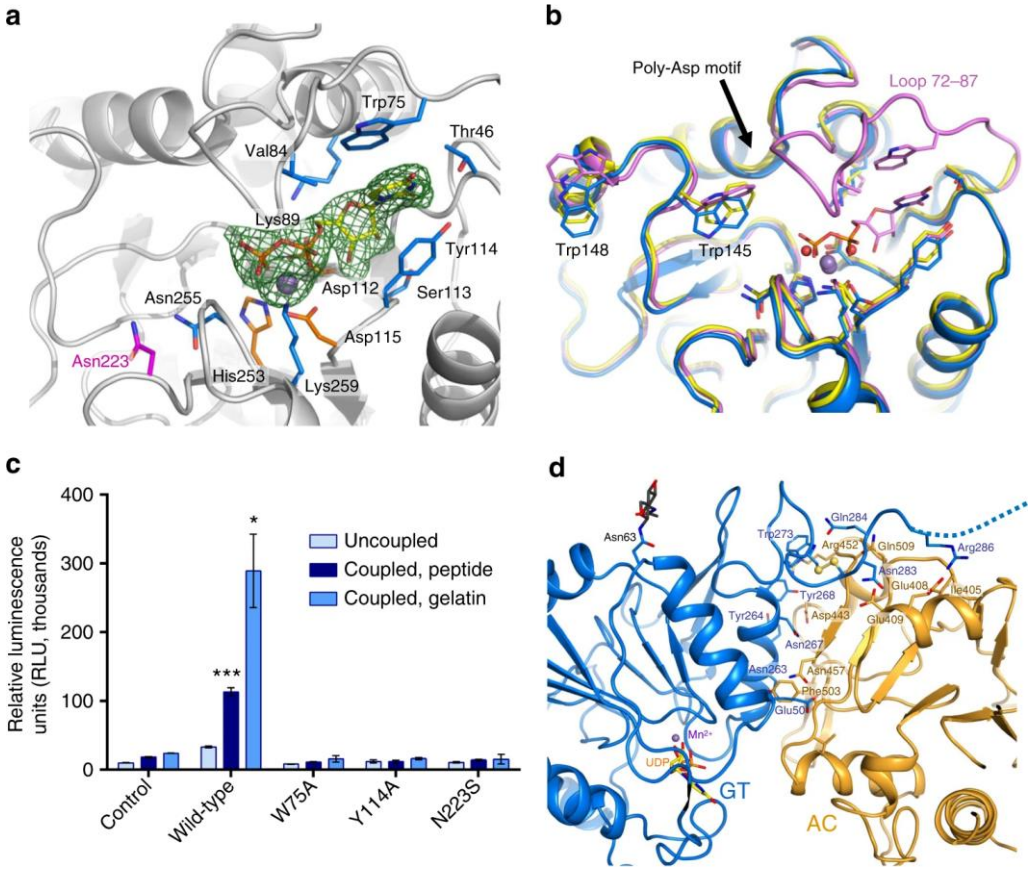


Figure 3-2 Insights into LH3 glycosyltransferase activity

a Co-crystallizations with Mn^{2+} and donor substrates revealed clear electron density ($2F_o$-F_c omit electron density maps, green mesh, contour level 1.2 σ) for the metal ion and for UDP in the catalytic site of the N-terminal GT domain. Residues involved in coordination of the metal ion are shown with orange sticks, while residues interacting and stabilizing UDP binding are shown in blue. Residue Asn223, found mutated into a Ser and causing pathogenic phenotypes similar to osteogenesis imperfecta, is shown with magenta sticks. b Binding of donor substrates induces conformational changes in the GT domain. Shown is a superposition of ligand-free (yellow), Mn^{2+}-bound (blue), and Mn^{2+}-donor substrate-bound (violet) structures of LH3 GT domain. Only the UDP-Gal-bound structure is shown; the conformation observed in UDP-Glc-bound structure is identical. Flexible loop 72–87 becomes well defined only in donor substrate-bound structures; in these structures, residues Trp145 and Trp148 adopt different conformations. c Luminescence-based assays for the evaluation of LH3 GT enzymatic activity show that mutants W75A, Y114A, and N223S are inactive. Control experiments were performed without adding enzyme. Error bars represent standard deviations from average of triplicate independent experiments. Statistical evaluations based on pair sample comparisons between uncoupled and coupled assay values using Student's t-test. *P-value <0.05; ***P-value <0.001. d Details of the interface between the N-terminal (GT, blue) and the central (AC, orange) LH3 glycosyltransferase domain. Residues found at the interface are shown as sticks (side chain view only). The disulfide bond found in the linker region between the two domains is shown with yellow spheres on the sulfur atoms.

Two tryptophan residues Trp145 and Trp148 proximate to the UDP-binding cavity consistently change their side chain conformations in structures with bound donor substrates (Fig. 3-2b). In UDP-glycan-bound structures, residue Trp145 adopts a conformation that can easily accommodate the sugar moiety during catalysis (Fig. 3-2b). In the absence of donor substrates this residue partially obstructs the cavity, thus acting as a gating residue to host the UDP-glycan (Fig. 3-2b). Trp148 localizes on the LH3 surface, in a region distant from the glycosyltransferase catalytic site located at the interface between two crystallographically related LH3 molecules. Interestingly, both Trp145 and Trp148 residues are located in one of the loops that are not conserved neither in glycosyltransferases, nor in other LH enzymes (Supplementary Fig. 9): this region may therefore have a role in recognition and binding of donor and/or acceptor substrates, highlighting the unusual mechanisms of LH3 GT/GGT activity. Of note, mutagenesis studies aimed at characterizing LH3 GT/GGT activities showed that residues in this loop are indeed critical for LH3-mediated collagen glycosylation[19] (Supplementary Table 2). The neighboring region comprising residues 187–191 is characterized by a non-conserved poly-Asp repeat (Supplementary Fig. 10B), and mutations on Asp190 and Asp191 were reported to abolish the glycosyltransferase activity[19] (Supplementary Table 2). In UDP-bound structures, this short loop forms contacts with the stabilized Gly72-Gly87 loop directly involved in UDP interaction (Fig. 3-2b). Although Asp190 and Asp191 are not directly involved in binding of enzyme cofactors and substrates, both residues point towards the active site near Trp145, possibly playing roles as nucleophiles during the glycosyltransferase reaction, or supporting solvent-bridged interactions with the glycan co-substrates.

Interestingly, the pathogenic LH3 mutation N223S, responsible for an LH3-dependent developmental connective tissue disorder with phenotype resembling osteogenesis imperfecta11 localizes in close proximity to the identified LH3 GT/GGT catalytic site (Fig. 3-2a). This mutation was reported to introduce a new N-linked glycosylation site on residue Asn221, resulting in strongly reduced LH3 GT/GGT activity (Supplementary Table 2). We produced this enzyme variant obtaining a folded, dimeric enzyme comparable with wild type LH3 (Supplementary Fig. 11A, B). However, due to pronounced instability and high propensity to degradation for this mutant, we could not unambiguously confirm that this disease-linked mutant bears an additional glycosylation (Supplementary Fig. 11A). LH3 N223S showed severely reduced lysyl hydroxylase and fully abolished glycosyltransferase activities (Fig. 3-2c, Supplementary Fig. 11C). As we did not observe changes in LH3 oligomeric assembly in the presence of this pathogenic mutation (Supplementary Fig. 11B), we concluded that the lack of enzymatic activity caused by this variant is likely due to the alterations in enzyme stability, possibly introduced by the novel glycosylation on Asn221, which would interfere with recognition of acceptor substrate molecules.

The C-terminal part of the first glycosyltransferase domain of LH3 incorporates other uniquely structured regions: a long, non-conserved beta hairpin constituted by residues Val229-Ala244 points towards the domain face opposite to the GT/GGT catalytic site. This segment is stabilized by numerous hydrophobic contacts including Phe233, Trp273 and with a non-conserved α-helix formed by residues Pro257-Glu276, which is sandwiched between the first and the second domain of LH3 (Fig. 3-2d). Residues linking these two domains extend to a solvent-exposed region of the enzyme comprising the Cys279-Cys282 disulfide bridge and a flexible loop extending from Asp285 to Gly292, and not visible in the electron density. Antibodies targeting these residues were reported to reduce GT activity[17], suggesting a role for this surface-exposed region in modulating accessibility of acceptor substrates to the identified glycosyltransferase catalytic site (Supplementary Table 2).

In the central LH3 domain, two previously suggested candidate metal ion binding sites are present, identified by a DxD (Asp392, Ala393, Asp394) and a DxDxD (Asp486, Thr487, Asp488, Pro489, Asp490) motif, respectively[18,19] (Supplementary Fig. 12). Co-crystallization experiments with metal ions and glycosyltransferase substrates and products did not highlight any appreciable differences within this domain compared to ligand-free structures (Supplementary Fig 13A). On the contrary, superposition of the two glycosyltransferase domains of LH3 showed remarkable differences, emphasized by the unique features found in the first domain (Supplementary Fig. 8, Supplementary Fig. 13B). Although the overall fold of the second domain shares higher similarity than the first with known GT-A type glycosyltransferases (Supplementary Fig. 13C, D), extensive mutagenesis experiments did not allow clear identification of residues implicated in metal ion or substrate binding for LH3 GT catalytic activity[18,19] (Supplementary Table 2). Our structures highlight an unusual conformation for the Val304-Phe310 loop: this region overlaps with the donor substrate binding site observed in structural homologs, strongly interfering with donor substrate binding (Supplementary Fig. 13E). Furthermore, recombinant N-terminal LH3 constructs devoid of this domain were still capable of glycosylating collagen peptides similar to the wild-type enzyme[19] (Supplementary Table 2). Collectively, these data indicate that despite the overall conservation of the glycosyltransferase fold, this domain may have lost its enzymatic capabilities during evolution. It may therefore constitute a non-catalytic accessory element within the LH3 architecture, possibly involved in collagen substrate recognition or interactions with LH binding partners. Thus, we named the LH3 N-terminal domain catalytic glycosyltransferase (GT), and the central domain accessory (AC).

3.3.4 Structural insights into LH3 lysyl hydroxylase activity

The C-terminal domain of LH3 hosts the lysyl hydroxylase catalytic site of the enzyme. This domain shows the typical DSBH fold, characterized by two β-sheets with antiparallel

β-strands flanked by three α-helices (Fig. 3-3a). This fold is preceded by two additional short helices covering residues Thr523 to Asp554, serving as a buffer platform between the AC and the LH domain (Fig. 3-1a). Residues 590–610 constitute a flexible loop capping the LH catalytic site. Superposition with identified structural homologs highlighted the high conservation of the overall fold and the consistent presence of highly flexible residues near the Fe^{2+}, 2-OG binding site (Supplementary Fig. 14A). The recently determined structure of a viral L230-LH domain[39] shows very high overall similarity (r.m.s.d. = 1.3 Å), with the exception of surface-exposed residues Ser550-Ile558, likely due to the non-conserved N-linked glycosylation at residue Asn548 (Supplementary Fig. 14B); also the structure of the viral LH fragment shows a completely flexible capping loop. All our LH3 structures systematically showed Fe^{2+} and 2-OG in the electron density near the core of this domain (Fig. 3-3a), confirming that these cofactors are tightly bound to the enzyme and providing a possible explanation for the observed substrate uncoupled LH enzymatic activity (Supplementary Fig. 1C, D). Given the high structural similarity of our LH3 structure with the LH domain of the viral variant, we were surprised not to find the 2-OG cofactor in this structure. We interpreted this difference as due to usage of a prokaryotic expression system for the recombinant production of the viral homolog[39]. In human LH3, the Fe^{2+} ion is stabilized by interactions with residue His719 and with residues His667 and Asp669, which constitute part of a HxD motif fully conserved in the Fe^{2+}, 2-OG dioxygenase enzyme family (Supplementary Fig. 15). Mutations in these residues[17], including a recently identified pathogenic variant of LH2 causing Bruck syndrome[8], were found to completely abolish LH activity (Supplementary Table 2). Pathogenic LH3 mutations causing premature C-terminal enzyme truncations lacking His719 were found incapable of hydroxylating collagen lysines, although retaining GT/GGT activities11 (Supplementary Table 2). Residues Tyr656, Cys691 and Arg729 are fully conserved among LH isoforms and delimit the pocket hosting the 2-OG cofactor (Supplementary Fig. 14C). In particular, the side chain of free Cys691 contributes to the positioning of the carboxyl moiety of the 2-OG cofactor in the direction of Arg729, forming a salt bridge with the guanidinium group of this residue. In addition, biochemical mutations of Arg729 in homologous LH1 (Arg719 in LH3) were reported to cause complete loss of LH activity due to decreased binding affinity for 2-OG[42]; these data are also in agreement with a recent report analyzing these mutations in the viral L230-LH homolog[39] (Supplementary Table 2).

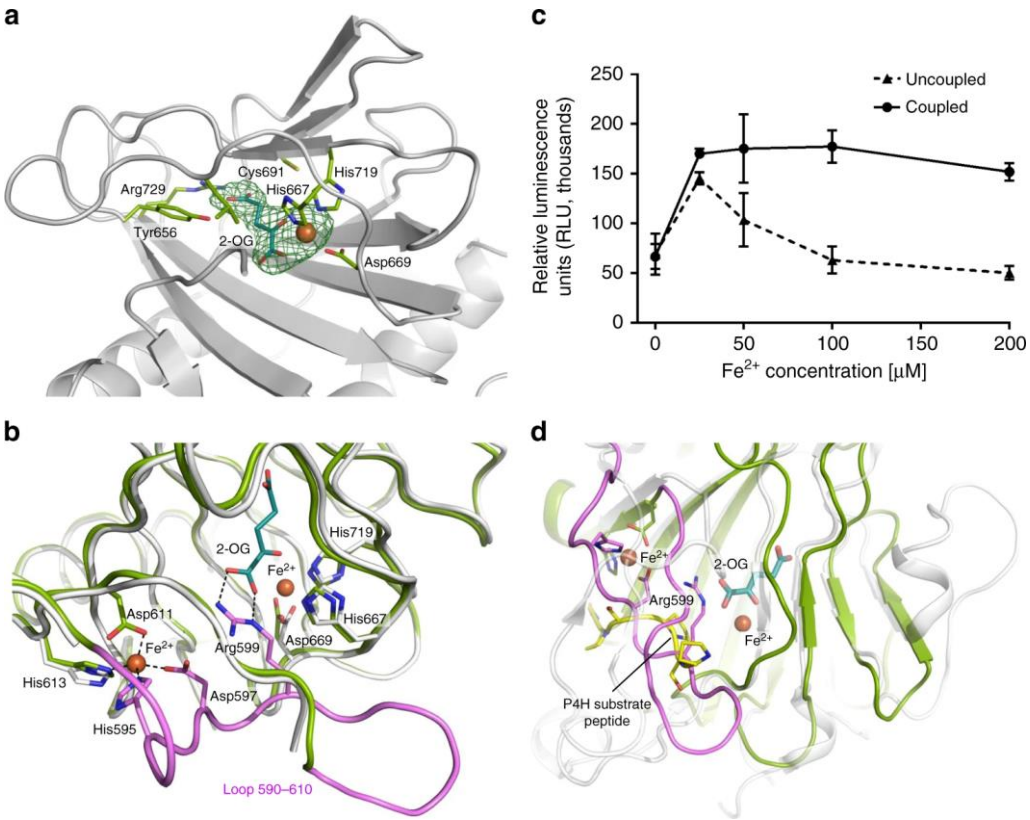


Figure 3-3 Insights into LH3 lysyl hydroxylase activity
a All our LH3 structures consistently show clear electron density for bound Fe^{2+} and 2-OG in the LH catalytic cavity (2Fo-Fc omit electron density maps, green mesh, contour level 1.2 σ). Residues involved in interactions with Fe2+ and 2-OG are shown with green sticks. b Co-crystallizations with Fe^{2+} allow identification of a second metal ion bound near the LH catalytic site that stabilizes the flexible capping loop 590–610 (shown in pink). c Evaluation of collagen substrate coupled and uncoupled LH3 LH activities as a function of Fe^{2+} concentration. Error bars represent standard deviations from average of triplicate independent experiments. d In Fe^{2+} co-crystal structures, the conformation observed for residue Arg599 mimics the collagen lysine substrate in front of the 2-OG cofactor, yielding a non-productive ternary complex. Shown is the superposition of this conformation found in the LH3 LH domain (green) with a homologous algal prolyl-4-hydroxylase (white) in complex with a short poly-PS peptide (yellow)[43]. The LH3 loop 590–610 is shown in pink.

3.3.5 Excess Fe^{2+} induces a state showing substrate mimicry

By increasing the Fe^{2+} concentration in crystallization experiments, we serendipitously found that this metal ion contributes to the overall enzyme stabilization, systematically enhancing the quality and the resolution of X-ray diffraction data. Analysis of electron

density maps allowed the identification of a second Fe^{2+} in the LH domain, coordinated by residues His595, Asp597, Asp611 and His613 (Fig. 3-3b). Metal ion coordination stabilizes the flexible capping loop 590–610 into a conformation that completely plugs the LH catalytic site, in proximity to the dimer interface (Supplementary Fig. 16A). We could observe a very similar arrangement, although slightly more flexible, by replacing Fe^{2+} with Mn^{2+} in crystallization experiments (Supplementary Fig. 16B). Early reports indicated that LH enzymes may be inhibited by high concentrations of metal ions20. We probed LH3 enzymatic activity in the presence of increasing concentrations of Fe^{2+}, and found that the enzymatic activity peaks at 25 μM Fe^{2+} concentration. Higher metal ion concentrations yield significant reduction of LH uncoupling, but do not seem to affect the enzymatic reactivity in the presence of synthetic peptide substrates (Fig. 3-3c). In the metal-ion stabilized conformations we also found that residue Arg599 forms a salt bridge with the 2-OG co-substrate, yielding a conformation that may mimic the collagen lysine substrate. The positioning of Arg599 indeed superimposes to that of a collagen proline residue as observed in a homologous prolyl hydroxylase structure in complex with a short peptide[43] (Fig. 3-3d)

3.4 DISCUSSION

Despite being known for over [40] years[44], several molecular aspects underlying collagen lysyl hydroxylases function are still obscure. Our crystal structures of full-length human LH3 rationalize the accumulated biochemical knowledge, offering a template to better understand the molecular mechanisms of LH3-dependent collagen hydroxylation and glycosylation.

The two different enzymatic activities are segregated in two distinct catalytic sites 80 Å apart within the same enzyme molecule. The overall dimeric arrangement, experimentally validated through site-directed mutagenesis, shows an elongated tail-to-tail quaternary assembly (Fig. 3-1c), characterized by a relatively small hydrophobic contact platform surrounded by electrostatic contacts that stabilize the dimer interface near the LH catalytic site (Fig. 3-1c–e) and closely resemble those recently reported for homologous human LH2 and viral L230[39] (Supplementary Fig. 14B). As previously suggested, dimerization is essential for LH activity, whereas disruption of physiological dimers does not significantly perturb the N-terminal glycosyltransferase activities of LH3[38,39].

We identified two glycosyltransferase domains at the N-terminus of the enzyme characterized by Rossmann fold-like domain architecture and partial similarity with known GT-A glycosyltransferases. However, only the first domain is active: the GT domain (residues 33–277) is the solely responsible for glycosyltransferase activities (both GT and GGT), albeit showing unique features in multiple regions of its fold strongly divergent from known glycosyltransferases (Supplementary Fig. 8). Sequence alignments

comparing human LH isoforms (Supplementary Fig. 9) and analysis of donor substrate-free and bound structures do not provide a direct explanation for the lack of GT/GGT activities in homologous LH1 and LH2. Nearly all residues surrounding the UDP moiety and shaping the LH3 glycosyltransferase catalytic site are conserved, including those involved in the unprecedented mode of UDP-substrate stabilization characterized by dual π-π stacking with Trp75 and Tyr114. Nevertheless, sequence conservation at the rim of the catalytic site is much lower (Supplementary Fig. 9), suggesting that lack of collagen substrate recognition may be the reason for the absence of glycosyltransferase activity in homologous LH isoforms.

The adjacent AC domain does not appear to be directly involved in enzyme function. Surprisingly, this domain shares the highest structural similarities with known glycosyltransferases, and yet it may have lost its function during evolution through mutations that introduced steric hindrance in the donor and acceptor substrate binding regions (Supplementary Fig. 13E). Despite the lack of direct evidence of substrate recognition capabilities, the elongated quaternary structure of the LH3 enzyme allows to speculate about possible long-range collagen substrate recognition mechanisms involving this domain. Alternatively, it may serve as a docking module for interactions with binding partners, such as the proposed collagen glycosyltransferases GLT25D1/2[26] or chaperones like FKBP65[24,25].

In the C-terminal LH domain both Fe^{2+} and 2-OG are natively retained and consistently found in the electron densities of all our structures (Fig. 3-3a). Enzymatic assays show that LH activity is present without the need for Fe^{2+} supplementation, demonstrating that the metal ion remains tightly bound into the active site during catalysis, and supporting recent reports also indicating essential roles for the metal ion in folding maintenance of LH domain[39]. Concentrations higher than 25 μM affect the LH3 uncoupled catalytic activity, but do not seem to affect substrate processing. Crystals grown with excess Fe^{2+} or Mn^{2+} show stabilization of a substrate gating loop, with formation of a self-inactivated ternary complex between Fe^{2+}, 2-OG and the side chain of Arg599 mimicking the collagen lysine substrate and obstructing the enzyme's active site (Fig. 3-3b, c). Despite gating loops have been described for prolyl-hydroxylases[43,45,46], the presence of a metal ion-induced gating seems a prerogative of LH enzymes. Given the full conservation in the LH enzyme family of amino acid residues involved in coordination of the second Fe^{2+} (Supplementary Fig. 15), we postulate that a general mechanism for metal ion-dependent regulation of LH activity may rely on generation of a self-inactivated resting state with interlocking of the 591–610 gating loop in front of the catalytic site. Remarkably, mutation of LH2 conserved homologous Arg599 into a histidine, as well as other mutations localized on this gating loop, perturb the enzymatic activity and cause Bruck syndrome[7] (Supplementary Table 2). This corroborates the proposed role of the

590–610 gating loop in modulating substrate accessibility to the LH active site, and in particular the role of conserved Arg599 as non-productive substrate mimicry.

The LH3 structures allow understanding the molecular phenotypes associated to most of the pathogenic mutations involving LH enzymes (Fig. 3-4). The LH3 N223S variant causing connective tissue disorders similar to osteogenesis imperfecta was reported to introduce a novel N-linked glycosylation near the cavity hosting the glycosyltransferase donor substrates. Our experiments show that this alteration strongly affects the enzyme's stability, causing degradation, and completely abolishes GT and GGT activities. Pathogenic mutations of LH1 responsible for the kyphoscoliotic variant of the Ehlers-Danlos syndrome were reported to reduce, but not abolish, mechanisms of collagen recognition and catalysis (Supplementary Table 2). This is fully consistent with their localization on the enzyme's surface, distant from the identified catalytic sites, with the exception of pathogenic mutation H706R, which maps to a conserved residue in close proximity to the homodimer interface. Conversely, LH2 mutations causing severe Bruck syndrome are located within the LH active site, either proximate to the 2-OG binding site, directly involved in Fe^{2+} coordination, or in the self-inactivated capping loop, also including the lysine mimicry residue Arg599 (Supplementary Table 2). Given the recent reports describing implications of human LH2 and its upregulated enzymatic activity in metastatic spreading of numerous solid tumors[13], our full-length structures of self-inhibited human LH3 constitute a valid template for structure-based drug discovery campaigns aiming at blocking unwanted collagen lysine hydroxylation in tumor microenvironments.

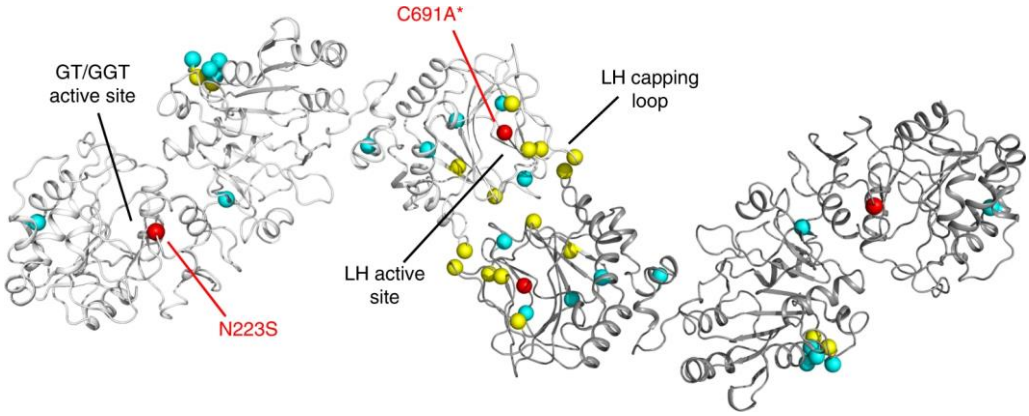


Figure 3-4 Mapping of disease-related mutations identified in LH enzymes on the LH3 crystal structure. The positions of pathogenic mutations causing Ehlers-Danlos type VI syndrome in LH1 (cyan), Bruck syndrome type II in LH2 (yellow), and connective tissue diseases sharing phenotype features with osteogenesis imperfecta (red) in LH3 are shown as spheres on the LH3 dimer structure. Mutations N223S and C691A*, found on LH3, are labeled in red in one of the two monomers. A complete list of mutations reported in the literature and their significance based on evaluation of their mapping on the LH3 crystal structure is shown in Supplementary Table 2

In conclusion, the crystal structures of full-length human LH3 in complex with various cofactors and donor substrates provide a molecular understanding of the biochemical knowledge underlying the multiple functions of this enzyme. Our data shed light on the unique molecular architecture of the LH3 glycosyltransferase domains, and allow understanding of the molecular bases of multiple genetic diseases involving LH3 and homologous human lysyl hydroxylases.

3.5 ACKNOWLEDGEMENTS

We thank scientists at ARDIS S.R.L. Pavia for technical support during the SPR experiments. We thank the European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS) for the provision of synchrotron radiation facilities and beamline scientists of the SLS, ESRF and the European Molecular Biology Laboratory for assistance. In particular, we would like to thank M. Brennich (EMBL Grenoble) and P. Pernot (ESRF) for support with SAXS data collection and analysis. We thank M. Miao for support in crystallization experiments and Prof. A. Mattevi for useful discussions. This work was supported by the Giovanni Armenise-Harvard Career Development Award, the "Programma Rita Levi-Montalcini" from the Italian Ministry of University and Research (MIUR), Cariplo Foundation Grant "COME TRUE" (id. 2015-0768), a "My First AIRC Grant" grant (Grant id. 20075) from the Italian Association for Cancer Research (AIRC), and by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022)—Dept. of Biology and Biotechnology "L. Spallanzani", University of Pavia. For X-ray diffraction experiments we were supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (Grant agreements 7551 and 10205). A.C. is supported by a Marie Curie Individual Fellowship from the Horizon 2020 EU Program (Grant agreement no. 745934 – COTETHERS). P.G. is supported by the NIHR HS&DR Programme (14/21/45) and supported by the NIHR GOSH BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

3.6 REFERENCES

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

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

3. Kivirikko, K. I., Ryhanen, L., Anttinen, H., Bornstein, P. & Prockop, D. J. Further hydroxylation of lysyl residues in collagen by protocollagen lysyl hydroxylase *in vitro*. Biochemistry 12, 4966–4971 (1973).

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

5. Hautala, T. et al. Cloning of human lysyl hydroxylase: complete cDNAderived amino acid sequence and assignment of the gene (PLOD) to chromosome 1p36.3→p36.2. Genomics 13, 62–69 (1992).

6. Rohrbach, M. et al. Phenotypic variability of the kyphoscoliotic type of Ehlers-Danlos syndrome (EDS VIA): clinical, molecular and biochemical delineation. Orphanet J. Rare Dis. 6, 46 (2011).

7. Hyry, M., Lantto, J. & Myllyharju, J. Missense mutations that cause Bruck syndrome affect enzymatic activity, folding, and oligomerization of lysyl hydroxylase 2. J. Biol. Chem. 284, 30917–30924 (2009).

8. Leal, G. F. et al. Expanding the clinical spectrum of phenotypes caused by pathogenic variants in PLOD2. J. Bone Miner. Res. 33, 753–760 (2018).

9. Ruotsalainen, H. et al. Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. J. Cell Sci. 119, 625–635 (2006).

10. Risteli, M. et al. Reduction of lysyl hydroxylase 3 causes deleterious changes in the deposition and organization of extracellular matrix. J. Biol. Chem. 284, 28204–28211 (2009).

11. Salo, A. M. et al. A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. Am. J. Hum. Genet. 83, 495–503 (2008).

12. van der Slot, A. J. et al. Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. J. Biol. Chem. 278, 40967–40972 (2003).

13. Du, H., Pang, M., Hou, X., Yuan, S. & Sun, L. PLOD2 in cancer research. Biomed. Pharmacother. 90, 670–676 (2017).

14. Gilkes, D. M. et al. Procollagen lysyl hydroxylase 2 is essential for hypoxia induced breast cancer metastasis. Mol. Cancer Res. 11, 456–466 (2013).

15. 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).

16. Ruotsalainen, H., Sipila, L., Kerkela, E., Pospiech, H. & Myllyla, R. Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3, their phylogenetic analysis and tissue-specific expression in the mouse. Matrix Biol. 18, 325–329 (1999).

17. Heikkinen, J. et al. Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. J. Biol. Chem. 275, 36158–36163 (2000).

18. 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).

19. Wang, C. et al. Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). J. Biol. Chem. 277, 18568–18573 (2002).

20. Kivirikko, K. I. & Myllyla, R. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods Enzymol. 82, 245–304 (1982). Pt A.

21. Salo, A. M. et al. The lysyl hydroxylase isoforms are widely expressed during mouse embryogenesis, but obtain tissue- and cell-specific patterns in the adult. Matrix Biol. 25, 475–483 (2006).

22. Kellokumpu, S., Sormunen, R., Heikkinen, J. & Myllyla, R. Lysyl hydroxylase, a collagen processing enzyme, exemplifies a novel class of luminally-oriented peripheral membrane proteins in the endoplasmic reticulum. J. Biol. Chem. 269, 30524–30529 (1994).

23. Heard, M. E. et al. Sc65-null mice provide evidence for a novel endoplasmic reticulum complex regulating collagen lysyl hydroxylation. PLoS Genet. 12, e1006002 (2016).

24. Duran, I. et al. A chaperone complex formed by HSP47, FKBP65, and BiP modulates telopeptide lysyl hydroxylation of Type I procollagen. J. Bone Miner. Res. 32, 1309–1319 (2017).

25. Gjaltema, R. A., van der Stoel, M. M., Boersema, M. & Bank, R. A. Disentangling mechanisms involved in collagen pyridinoline cross-linking: The immunophilin FKBP65 is critical for dimerization of lysyl hydroxylase 2. Proc. Natl Acad. Sci. USA 113, 7142–7147 (2016).

26. 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).

27. 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).

28. 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).

29. Watt, S. A. et al. Lysyl hydroxylase 3 localizes to epidermal basement membrane and is reduced in patients with recessive dystrophic epidermolysis bullosa. PLoS ONE 10, e0137639 (2015).

30. Banushi, B. et al. Regulation of post-Golgi LH3 trafficking is essential for collagen homeostasis. Nat. Commun. 7, 12111 (2016).

31. Gruber, R. et al. Autosomal recessive Keratoderma-Ichthyosis-Deafness (ARKID) syndrome is caused by VPS33B mutations affecting Rab protein interaction and collagen modification. J. Invest. Dermatol. S0022–202X, 32800–32807 (2016).

32. Myllyla, R. et al. Expanding the lysyl hydroxylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3). J. Cell. Physiol. 212, 323–329 (2007).

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

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

35. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. Curr. Opin. Struct. Biol. 22, 691–700 (2012).

36. McDonough, M. A., Loenarz, C., Chowdhury, R., Clifton, I. J. & Schofield, C. J. Structural studies on human 2-oxoglutarate dependent oxygenases. Curr. Opin. Struct. Biol. 20, 659–672 (2010).

37. Pirskanen, A., Kaimio, A. M., Myllyla, R. & Kivirikko, K. I. Site-directed mutagenesis of human lysyl hydroxylase expressed in insect cells. Identification of histidine residues and an aspartic acid residue critical for catalytic activity. J. Biol. Chem. 271, 9398–9402 (1996).

38. Heikkinen, J. et al. Dimerization of human lysyl hydroxylase 3 (LH3) is mediated by the amino acids 541-547. Matrix Biol. 30, 27–33 (2011).

39. Guo, H. F. et al. Pro-metastatic collagen lysyl hydroxylase dimer assemblies stabilized by Fe(2+)-binding. Nat. Commun. 9, 512 (2018).

40. Gastinel, L. N. et al. Bovine alpha1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. EMBO J. 20, 638–649 (2001).

41. Wiggins, C. A. & Munro, S. Activity of the yeast MNN1 alpha-1, 3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proc. Natl Acad. Sci. USA 95, 7945–7950 (1998).

42. Passoja, K., Myllyharju, J., Pirskanen, A. & Kivirikko, K. I. Identification of arginine-700 as the residue that binds the C-5 carboxyl group of 2-oxoglutarate in human lysyl hydroxylase 1. FEBS Lett. 434, 145–148 (1998).

43. Koski, M. K. et al. The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. J. Biol. Chem. 284, 25290–25301 (2009).

44. Ryhanen, L. & Kivirikko, K. I. Hydroxylation of lysyl residues in native and denatured protocollagen by protocollagen lysyl hydroxylase *in vitro*. Biochim. Biophys. Acta 343, 129–137 (1974).

45. Horita, S. et al. Structure of the ribosomal oxygenase OGFOD1 provides insights into the regio- and stereoselectivity of prolyl hydroxylases. Structure 23, 639–652 (2015).

46. Longbotham, J. E. et al. Structure and mechanism of a viral collagen prolyl hydroxylase. Biochemistry 54, 6093–6105 (2015).

47. Subramani, S., Mulligan, R. & Berg, P. Expression of the mouse dihydrofolate reductase complementary deoxyribonucleic acid in simian virus 40 vectors. Mol. Cell. Biol. 1, 854–864 (1981).

48. Olieric, V. et al. Data-collection strategy for challenging native SAD phasing. Acta Crystallogr. D Struct. Biol. 72, 421–429 (2016).

49. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. Acta Crystallogr. Sect. D 67, 293–302 (2011).

50. Tickle, I. J. et al. STARANISO. (Global Phasing Ltd., Cambridge, United Kingdom, 2018).

51. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. 66, 479–485 (2010).

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. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. 59, 2023–2030 (2003).

54. Skubak, P. & Pannu, N. S. Automatic protein structure solution from weak Xray data. Nat. Commun. 4, 2777 (2013).

55. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. 62, 1002–1011 (2006).

56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. Sect. D 66, 486–501 (2010).

57. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).

58. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221 (2010).

59. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21 (2010).

60. Gore, S. et al. Validation of structures in the protein data bank. Structure 25, 1916–1927 (2017).

61. Lutteke, T. & von der Lieth, C. W. pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files. BMC Bioinf. 5, 69 (2004).

62. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. (2010).

63. Brennich, M. E., Round, A. R. & Hutin, S. Online size-exclusion and ionexchange chromatography on a SAXS beamline. J. Vis. Exp. 119, e54861 (2017).

64. Round, A. et al. BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. Acta Crystallogr. D Biol. Crystallogr. 71, 67–75 (2015).

65. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small angle X-ray scattering data. Bioinformatics 34, 1944–1946 (2017).

66. 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).

67. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for smallangle scattering from macromolecular solutions. J. Appl. Crystallogr. 50, 1212–1225 (2017).

68. 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).

69. Graphpad Prism 7. (Graphpad Software, La Jolla, California, USA).

70. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 27, 112–128 (2018).

4.Chapter

A COOPERATIVE NETWORK OF MOLECULAR "HOT SPOTS" HIGHLIGHTS THE COMPLEXITY OF LH3 COLLAGEN GLYCOSYLTRANSFERASE ACTIVITIES

This chapter is a research manuscript, available since 4 April 2020 on the preprint server bioRxiv (doi:/10.1101/841486), and it includes most of the work I did during my PhD program. I carried out the biochemical evaluation of LH3 wild-type and mutated variants by performing luminescent-based enzymatic assays. I helped in the purification of some LH3 mutants. I helped in the design of the study, in particular what regards the usage of UDP-sugar analogs, indeed I performed DSF experiments and LH3 wild-type co-crystallization wih UDP-sugar analogs. I went to syncrothron facilities to perform data collection and I helped in the structural refinement and analysis pipeline.

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

*This authors contributed equally

Affiliations:
a The Armenise-Harvard Laboratory of Structural Biology, Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9/A, 27100 Pavia, Italy.
b NSW Health Pathology East laboratory, Prince of Wales Hospital, Randwick, NSW, Australia
c Centre for Clinical Genetics, Sydney Children's Hospital, Randwick, Australia
d Neuroscience Research Australia (NeuRA), Prince of Wales Clinical School, University of New South Wales, Sydney, Australia

Corresponding author: Federico Forneris (email: federico.forneris@unipv.it)

Supplementary information are available online at the following link: https://www.biorxiv.org/content/10.1101/841486v2.supplementary-material?versioned=true
They include supplementay tables S1A, S1B, S2, S3 and S4; and supplementary figures S1, S2, S3, S4

ABSTRACT

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

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

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

4.1 INTRODUCTION

Collagens are the most abundant proteins in the human body and are highly conserved from sponges to mammals (Luther et al, 2011; Myllyharju & Kivirikko, 2004). The different oligomeric architectures and roles of collagen molecules strongly depend on a variety of post-translational modifications, including proline and lysine hydroxylations, as well highly specific glycosylations of hydroxylated lysines (HyK) (Cummings, 2009; Myllyharju & Kivirikko, 2004). The disaccharide present on HyK contains a highly conserved glucosyl(α−1,2)-galactosyl(β−1,O) glycan moiety, whose identity was discovered in the late sixties (Spiro & Spiro, 1971; Spiro, 1967; Spiro, 1969). Monosaccaridic galactosyl-(β−1,O) HyK have been identified as well, as a result of catabolic reactions carried out by the collagen α-glucosidase, an enzyme highly specific for the disaccharide present on collagenous domains. The role for this enzyme is to localize collagen in the glomerular basement membrane (Hamazaki & Hamazaki, 2016; Sternberg et al, 1982; Sternberg & Spiro, 1980). The spread of glycosylation largely depends on collagen type (Bornstein & Sage, 1980; Spiro, 1969; Terajima et al, 2014), on the functional area inside tissues (Moro et al, 2000; Schofield et al, 1971; Toole et al, 1972), on the developmental stage (Rautavuoma et al, 2004; Sipila et al, 2007) and on disease states (Lehmann et al, 1995; Salo et al, 2008; Tenni et al, 1993). Although extensively studied, the precise mechanisms of collagen glycosylation and their biological relevance in collagen homeostasis have remained poorly understood.

The identity and exquisite stereochemistry of the Glc(α−1,2)-Gal(β−1,O)-HyK-linked carbohydrate supports the idea that at least two distinct enzyme types are needed to fully incorporate these complex post-translational modifications on collagen molecules (Hennet, 2019). The first reaction indeed requires an inverting-type galactosyltransferase (GalT) acting on HyK, whereas the subsequent glucosylation is catalyzed by a retaining-type glucosylgalactosyltransferase (GlcT). Multifunctional lysyl hydroxylase 3 (LH3) was the only enzyme identified as possessing lysyl hydroxylase activity as well as GalT and GlcT activities *in vitro* (Wang et al, 2002a). The glycosyltransferase activities are specific of LH3, as highly homologous LH1 and LH2a/b are not capable of catalyzing these reactions (Heikkinen et al, 2000). Conversely, *in vivo* studies have demonstrated that decreased LH3 protein levels and/or pathogenic mutations in LH3 GT domain, exclusively impair the GlcT activity (Ewans et al, 2019; Salo et al, 2008; Savolainen et al, 1981). This occurs secondarily to the LH3 p.Asn223Ser, which introduces an additional glycosylation site within the enzyme's GT domain leading to an osteogenesis imperfecta-like phenotype (Salo et al, 2008); and in the recently identified LH3 p.Pro270Leu, which results in a Stickler-like syndrome with vascular complications and variable features typical of Ehlers-Danlos syndrome and Epidermolysis Bullosa (Ewans et al, 2019). Mouse studies have also shown that only the LH3 GlcT activity is indispensable for the biosynthesis of collagen IV and formation of the basement membrane during embryonic

development (Rautavuoma et al, 2004; Ruotsalainen et al, 2006), consistent with the presence of additional collagen galactosyltransferases. Two genes encoding for O-galactosyltransferases (GLT25D1 and GLT25D2) were recently identified (Perrin-Tricaud et al, 2011; Schegg et al, 2009). It is of interest that GLT25D1 and LH3 were proposed to act in concert on collagen molecules (Schegg et al, 2009; Sricholpech et al, 2011; Yamauchi & Sricholpech, 2012). Studies on osteosarcoma cell lines which produce large amounts of fibrillar collagens, showed that the simultaneous deletion of GLT25D1 and GLT25D2 resulted in growth arrest due to lack of glycosylation, further indicating that the GalT activity of LH3 might not be as essential as its GlcT activity (Baumann & Hennet, 2016).

These data support the existence of a specific and highly conserved machinery for collagen O-glycosylation, and sustain the hypothesis that *in vivo* the entire collagen glycosylation machinery may involve distinct proteins and protein complexes for GalT and GlcT reactions. This raises the intriguing question of how this highly conserved process is spatiotemporally regulated at the molecular level. Our current understanding of collagen glycosyltransferases is however restricted to three-dimensional structures of human LH3 in complex with UDP-sugar donor substrates (Scietti et al, 2018) and to few mutagenesis studies focusing on the main hallmarks of Mn^{2+}-dependent glycosyltransferase catalysis (Wang et al, 2002a; Wang et al, 2002b).

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

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

4.2 EXPERIMENTAL PROCEDURES

4.2.1 Chemicals

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

4.2.2 Site-directed mutagenesis

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

4.2.3 Recombinant protein expression and purification

Wild-type and mutant LH3 coding sequences were cloned into the pUPE.106.08 expression vector (U-protein Express BV, The Netherlands) in frame with a 6xHis-tag followed by a Tobacco Etch Virus (TEV) protease cleavage site. Suspension growing HEK293F cells (Life Technologies, UK) were transfected at a confluence of 10^6 cells/ml, using 1 μg of plasmid DNA and 3 μg of linear polyethyleneimine (PEI; Polysciences, Germany). Cells were harvested 6 days after transfection by centrifuging the medium for 15 minutes at 1000 x g. The clarified medium was filtered using a 0.2 mm syringe filter and the pH was adjusted to 8.0 prior to affinity purification as previously described (Scietti et al, 2018). All proteins were isolated from the medium exploiting the affinity of the 6xHis tag for the HisTrap Excel (GE Healthcare, USA) affinity column. The purified protein was further polished using a Superdex 200 10/300 GL (GE Healthcare) equilibrated in 25 mM HEPES/NaOH, 200 mM NaCl, pH 8.0, to obtain a homogenous protein sample; peak fractions containing the protein of interest were pooled and concentrated to 1 mg mL^{-1}.

4.2.4 Crystallization, data collection, structure determination and refinement

Wild-type and mutant LH3 co-crystallization experiments were performed using the hanging-drop vapor-diffusion method protocols as described in (Scietti et al, 2018), by mixing 0.5 μL of enzyme concentrated at 3.5 mg mL^{-1} with 0.5 μL of reservoir solutions composed of 600 mM sodium formate, 12% poly-glutamic-acid (PGA-LM, Molecular Dimensions), 100 mM HEPES/NaOH, pH 7.8, 500 μM FeCl2, 500 μM MnCl2, supplemented with 1 mM of the appropriate UDP-sugar analogs (UDP, UDP-glucose, UDP-glucuronic acid, UDP-xylose). Crystals were cryo-protected with the mother liquor supplemented with 20% ethylene glycol, harvested using MicroMounts Loops (Mitegen), flash-cooled and stored in liquid nitrogen prior to data acquisition. X-ray diffraction data were collected at various beamlines of the European Synchrotron Radiation Facility,

Grenoble, France and at the Swiss Light Source, Villigen, Switzerland. Data were indexed and integrated using XDS (Kabsch, 2010) and scaled using Aimless (Evans & Murshudov, 2013). Data collection statistics are summarized in Suppl. Table 2. The data showed strong anisotropy and therefore underwent anisotropic cut-off using STARANISO (Tickle et al, 2018) prior to structure determination and refinement. The structures were solved by molecular replacement using the structure of wild-type LH3 in complex with Fe^{2+}, 2-OG and Mn^{2+} (PDB ID: 6FXM) (Scietti et al, 2018) as search model using PHASER (McCoy et al, 2007). The structure was refined with successive steps of manual building in COOT (Emsley et al, 2010) and automated refinement with phenix.refine (Adams et al, 2010). Model validation was performed with MolProbity (Chen et al, 2010). Refinement statistics for the final models are reported in supplementary Table 2.

4.2.5 Evaluation of LH3 GalT/GlcT enzymatic activity

LH, GalT and GlcT activities were tested using luminescence-based enzymatic assays with a GloMax Discovery (Promega, USA) as described in Scietti et al. (Scietti et al, 2018). Minor modifications have been done for GalT/GlcT competitive inhibition assays, where 1 µL of a mixture of 250 µM MnCl2, 500 µM UDP-galactose (GalT assay) or UDP-glucose (GlcT assay) and increasing concentrations of either UDP-GlcA or UDP-Xyl were initially added to the enzyme and gelatin substrate to start the reactions. All experiments were performed in triplicates. Control experiments were performed in the same conditions by selectively removing LH3. Data were analyzed and plotted using the GraphPad Prism 7 (Graphpad Software, USA).

4.2.6 Differential Scanning Fluorimetry (DSF)

DSF assays were performed on LH3 wild-type and mutants using a Tycho NT.6 instrument (NanoTemper Technologies GmbH, Germany). LH3 samples at a concentration of 1 mg/mL in a buffer composed of 25 mM HEPES, 500 mM NaCl, pH 8. Binding assays were performed by incubating LH3 with 50 µM MnCl2 and 5 mM free UDP or UDP-sugar donor substrates or their analogs. Data were analyzed and plotted using the GraphPad Prism 7 (Graphpad Software, USA).

4.3 RESULTS

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

The LH3 N-terminal GT domain shares its fold with Mn2+-dependent GT-A glycosyltransferases, encompassing a UDP-donor substrate binding cavity stretched towards a GalT/GlcT catalytic pocket (Scietti et al, 2018). We firstly inspected the structural organization of the UDP binding cavity to identify signatures for LH3 glycosyltransferase activity by comparing the amino acid sequences for the residues

surrounding the UDP-sugar donor substrate with those of GalT/GlcT-inactive LH1 and LH2a/b isoforms. We found that nearly all residues involved in Mn^{2+} and UDP-sugar binding are conserved in orthologs, with the exception of Val80, becoming Lys68 in LH1 and Gly80 in LH2a/b (Fig 4-1). The presence of a different amino acid side chain surrounding the donor substrate cavity led us to consider whether this could be a discriminating functional feature among the GT domains in LH enzymes. In LH3, Val80 is located in the middle of a flexible "glycoloop" (Gly72-Gly87), not visible in the electron density of the ligand-free LH3 structure and stabilized upon UDP-substrate binding (Scietti et al, 2018). Within this glycoloop, residue Val80 is in close proximity to the ribose ring of the UDP-sugar donor substrates. We hypothesized that introduction of a large, positively charged residue such as Lys in LH1, or alterations due to complete removal of side chain steric hindrance such as Gly in LH2 could lead to inability of binding donor substrates. We therefore generated the LH3 Val80Lys and Val80Gly mutants. Both 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 (supplementary Fig S1). Conversely, the mutation resulted in significant reduction of both GalT and GlcT activities when the reaction was carried out in presence of both donor and acceptor (gelatin) substrates (Fig 4-2, supplementary Table S1). Considering that wild-type LH3 is also capable of activating donor UDP-sugar substrates and release UDP in absence of the acceptor collagen substrate ("uncoupled" activity, as defined in (Scietti et al, 2018)), we investigated the impact of the Val80Lys and the Val80Gly mutations also in absence of acceptor substrates. In this case, the experiments yielded minor, but significant reduction in the enzymatic activities between wild-type and mutant LH3 (Fig 4-2, supplementary Table S1), indicating that the Val80 residue might be involved in the productive positioning of the donor substrate during transfer of the glycan moiety to the acceptor substrate, rather than stabilizing the UDP moiety in the catalytic pocket.

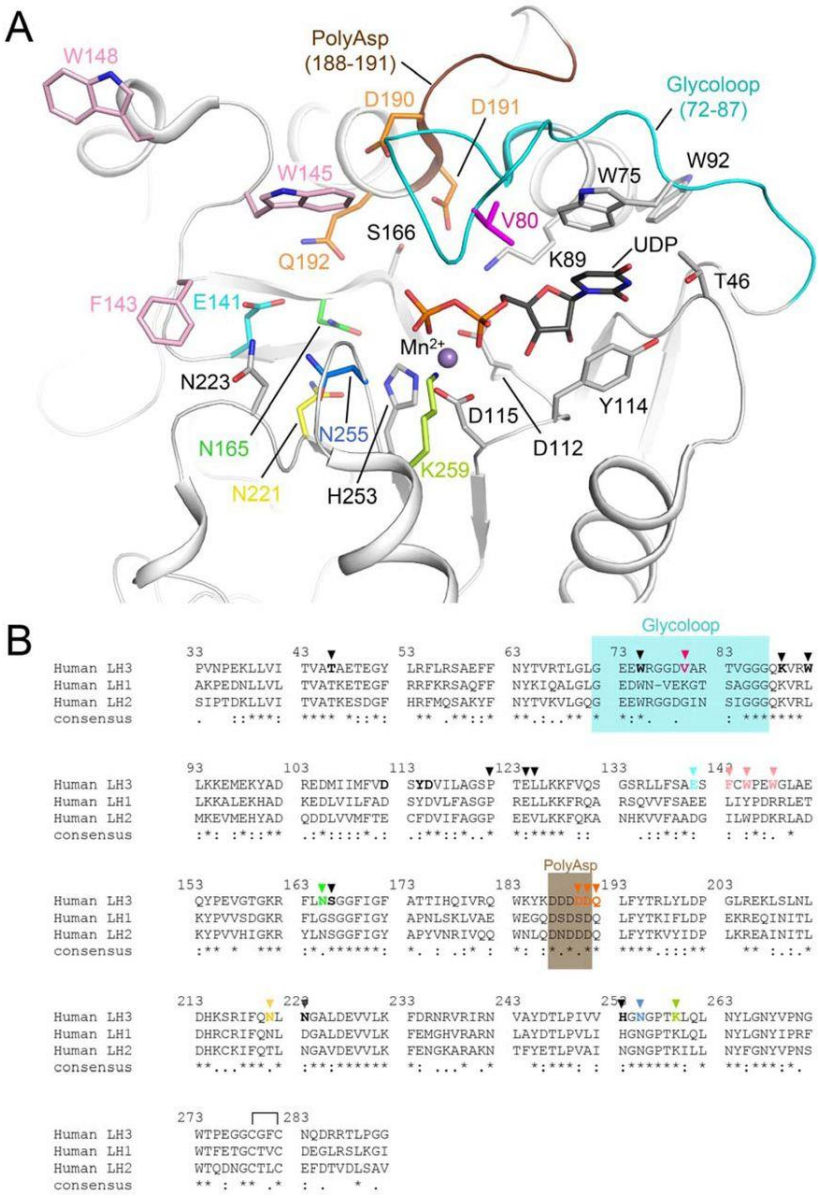


Figure 4-1 Features of the LH3 glycosyltransferase (GT) domain.
(A) Cartoon representation of the LH3 GT domain (PDB ID: 6FXR) showing the key residues shaping the catalytic site as sticks. The PolyAsp motif (brown) and the Glycoloop (cyan) involved in binding of UDP-sugar donor substrates are shown. The residues implicated in the catalytic activity and investigated in this works are colored, while the residues depicted in grey have already been shown to be essential in Mn^{2+} (purple sphere) and UDP (black sticks) coordination. (B) Sequence alignment of human LH1, LH2 and LH3 GT domains, highlighting similarities and differences in the amino acid residues within the active site. Residues shown in Fig 4-1A are indicated with a triangle. Colored boxes indicate the PolyAsp motif (brown) and the Glycoloop (cyan).

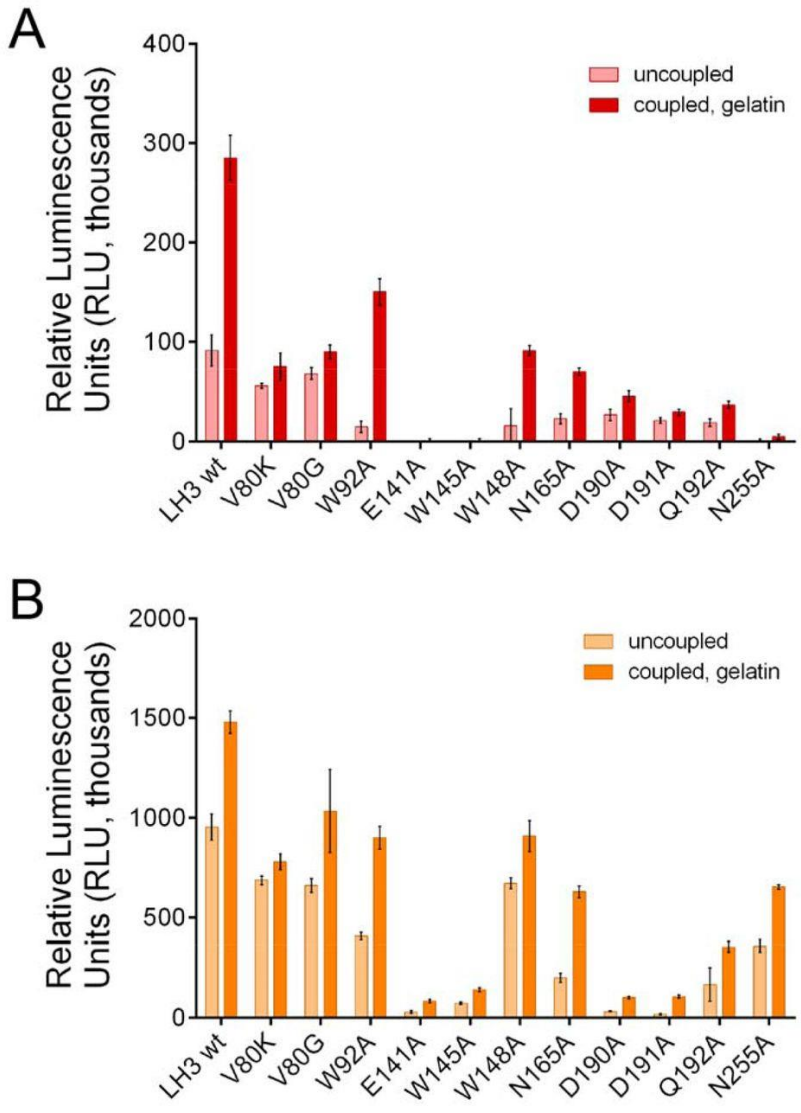


Figure 4-2 Evaluation of the effect of LH3 GT domain mutations in the GT site on glycosyltransferase activities.
Evaluation of the GalT activity (A) and GlcT activity (B) of LH3 mutants compared to the wild-type. Each graph shows the enzymatic activity detected in absence (i.e., "uncoupled") or in presence of gelatin, used as acceptor substrate. The plotted data are baseline-corrected, where the baseline was the background control. Error bars represent standard deviations from average of triplicate independent experiments.

To further rationalize the implications of LH3 Val80 in GalT and GlcT activities, we crystallized and solved the 3.0-Å resolution structure of the Val80Lys mutant in complex with Mn^{2+}, and also obtained its 2.3-Å resolution structure in presence of both Mn^{2+} and the UDP-Glc donor substrate (supplementary Table S2). Overall, both structures superimpose almost perfectly with wild-type LH3 for all domains (supplementary Fig S2). The structure of the LH3 Val80Lys mutant bound to Mn^{2+} appeared identical to that of wild-type LH3. In both structures, the glycoloop containing Val/Lys80 could not be modelled in the electron density due to its high flexibility (supplementary Fig S2B). On the other hand, the side chain of the Val80Lys residue could be modelled unambiguously in the experimental electron density of the UDP-donor substrate bound structure (Fig 4-3A). Despite the increased steric hindrance, the mutated Lys80 residue adopted a conformation compatible with the simultaneous presence of the UDP-Glc in the catalytic cavity. However, similar to what observed for wild-type LH3, the glycan moiety of UDP-Glc was completely flexible and therefore not visible in the electron density (Fig 4-3A). Collectively, these data are consistent with the alteration introduced by the Val80Lys mutation, impacting partially on the LH3 glycosyltransferase catalytic activities.

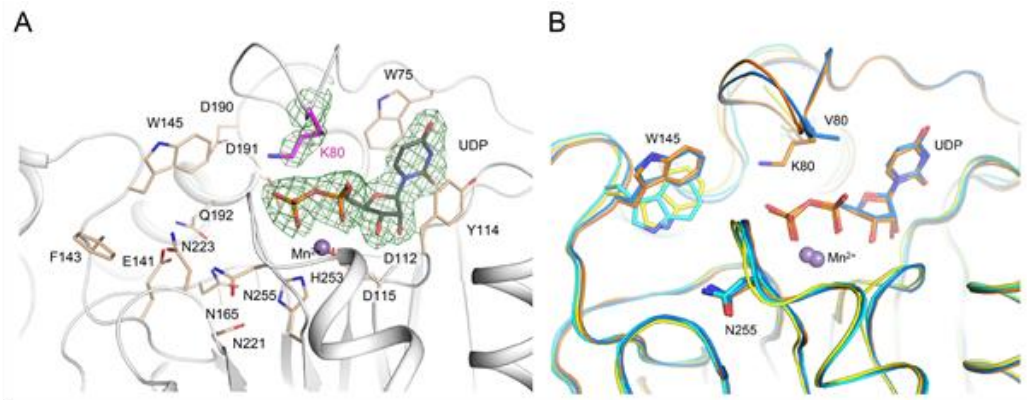


Figure 4-3 Structural characterization of the LH3 Val80Lys mutant.
(A) Crystal structure of the LH3 Val80Lys mutant in complex with UDP-glucose and Mn^{2+}. Electron density is visible for the mutated lysine and the UDP donor substrate (green mesh, 2Fo-Fc omit electron density map, contoured at 1.3 σ). Catalytic residues shaping the enzyme cavity are shown as sticks, Mn^{2+} is shown as purple sphere. Consistent with what observed in the crystal structure of wild-type LH3, the glucose moiety of the donor substrate is not visible in the experimental electron density. (B) Superposition of wild-type and Val80Lys LH3 crystal structures in substrate-free (cyan and yellow, respectively) with UDP-glucose bound (marine and orange, respectively) states. Notably, the conformations adopted by the side chain of Trp145 upon ligand binding are consistent in the wild-type and in the mutant enzyme. As the glycoloop is flexible in substrate-free structures, the side chains of Val/Lys80 are visible only in the in UDP-sugar bound structures.

The glycoloop is a structural feature found exclusively in the GT domains of LH enzymes. It incorporates Trp75, a residue whose aromatic side chain stabilizes the uridine moiety of the donor substrate and, together with residue Tyr114 of the DxxD motif (a distinguishing feature of LH3 GT domain, shared among LH enzymes, (Scietti et al, 2018)) "sandwiches" the donor substrate in an aromatic stacking environment (Fig 4-1A). Both residues are critical for the LH3 GalT and GlcT enzymatic activities (Scietti et al, 2018). The conformation adopted by the LH3 glycoloop in the presence of UDP-donor substrates is however not accompanied by other significant structural changes in surrounding amino acids, with the exception of minor rearrangements of distant residue Trp92 (not conserved in other LH isoenzymes (Fig 4-1)), 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 (supplementary Fig S1A-B) nor its LH enzymatic activity (supplementary Fig S1C). Conversely, the mutant showed 40% decrease for both GalT and GlcT activities in presence of donor and acceptor substrate compared to LH3 wild-type (Fig 4-2, supplementary Table S1); the impact of the Trp92Ala mutation on reactions in absence of acceptor substrate seemed to affect both activities at similar levels, with lower residual GalT (30%) compared to GlcT (40%) (Fig 4-2, supplementary Table S1). These findings suggest that LH3-specific long-range interactions in the GT domain may contribute to the productive conformations of the glycoloop in donor substrate-bound states.

In UDP-sugar bound structures, the glycoloop contacts a poly-Asp sequence (Asp188-Asp191, Fig 4-1A); this sequence is partially conserved in LH isoforms lacking glycosyltransferase activities (Fig 4-1B). Mutations of Asp190 and Asp191 were reported to affect the glycosyltransferase enzymatic activities of LH3 (Wang et al, 2002b). Based on LH3 crystal structures, such behaviour is expected, since residues Asp190 and Asp191 point towards the GalT/GlcT catalytic cavity. Interestingly, superposition of LH3 molecular structures with GT-A fold glycosyltransferases show that Asp191 caps the N-terminal end of an α-helix in a highly conserved position, with functional relevance in both retaining (Flint et al, 2005; Persson et al, 2001) and inverting enzymes (Charnock & Davies, 1999; Pedersen et al, 2000) (supplementary Fig S3, Table 4-1). We designed and generated individual alanine mutants for both LH3 Asp190 and Asp191, and found that both variants were compatible with folded and functional LH enzymes (supplementary Fig S1). When tested for GalT and GlcT activity, both these mutants caused severe impairment in activation of donor UDP-sugar (as shown by the strong reduction in uncoupled activity to less than 5% using UDP-Glc as substrate) as well as in transfer of the sugar moiety to the acceptor substrate (Fig 4-2, supplementary Table S1). However, none of these mutations resulted in a complete inactive LH3 GalT nor GlcT glycosyltransferase. Collectively, these data point towards an extended 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.

4.3.2 LH3 shares features with both retaining and inverting glycosyltransferases

After investigating the amino acid residues involved in stabilization of the UDP moiety of donor substrates, we focused on another group of residues within the GT catalytic pocket, opposite to the putative position of the flexible sugar rings of the same substrates (Fig 4-1A). Many LH3 residues shaping this part of the glycosyltransferase catalytic cavity matched catalytic amino acids found in other GT-A type glycosyltransferases (Fig 4-1A, supplementary Fig S3) (Ardevol et al, 2016; Lairson et al, 2008). In particular, LH3 Trp145, a residue located in one of the loops of the GT domain uniquely found in LH3, was previously suggested as a possible candidate for modulation of LH3 GalT and GlcT activities. This residue was found to adopt different side chain conformations in substrate-free and substrate-bound structures, affecting the shape and steric hindrance of the enzyme's catalytic cavity (Scietti et al, 2018); interestingly, nearly identical conformational changes were observed when comparing substrate-free and substrate bound LH3 Val80Lys structures (Fig 4-3B). Mutating the LH3 Trp145 residue into alanine strongly reduced both GalT (6% residual) and GlcT (10% residual) enzymatic activities (Fig 4-2, supplementary Table S1), without affecting other enzyme's properties (supplementary Fig S1). This supports previous hypotheses of a "gating" role for Trp145 in the GT catalytic cavity, assisting the productive positioning of sugar moieties of donor substrates for effective transfer during catalysis. Comparison with molecular structures of other glycosyltransferases (including distant homologs) highlighted that most structurally-related enzymes manage to position aromatic side chains from different structural elements of their fold in their catalytic cavities. Such structural arrangement is reminiscent to that of Trp145 in LH3, but relies on completely different structural features of the glycosyltransferase domain. 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 4-1, supplementary Fig S3). This supports glycosyltransferases being highly versatile enzymes, displaying an impressive structural plasticity to carry out reactions characterized by a very similar mechanism on a large variety of specific donor and acceptor substrates.

Protein Name	Type	PDB ID	catalytic base residue	nucleophile acceptor residue	corresp. LH3 residue	reference paper
LgtC - GALACTOSYL TRANSFERASE LGTC *(N. meningitidis)*	retaining	1GA8				Persson et al., 2001 10.1038/84168
GYG1 - Glycogenin *(O. cuniculus)*	retaining	1LL2		Asp163	Asp191	Gibbons et al., 2002 10.1016/S0022-2836(02)00305-4
mgs - Mannosylglycerate synthase (*R. marinus)*	retaining	2BO8				Flint et al., 2005 10.1038/nsmb950
GALNT10 - Polypeptide N-acetylgalactosaminyl transferase 10 (*H. sapiens)*	retaining	2D7R		Gln346	Asp190	Kubota et al., 2006 10.1016/j.jmb.2006.03.061
GGTA1 - N-Acetyllactosaminide □-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 4-1: List of glycosyltransferase enzymes used for comparisons with human LH3. The list includes the indication of the catalytic bases and nucleophile residues as proposed in the original papers describing the various glycosyltransferases, with the corresponding residue number in human LH3 based on structural superpositions.

Our previous structural comparisons of ligand-free and substrate-bound LH3 highlighted the additional possibility of a concerted mechanism involving conformational changes of a non-conserved aromatic residue located on the LH3 surface (Trp148, Fig 4-1), together with Trp145 (Scietti et al, 2018). To investigate such possibility, we mutated Trp148 into alanine. The mutant enzyme was also found to behave like wild-type LH3 in folding and LH activity (supplementary Fig S1). Glycosyltransferase assays showed that this variant had reduced GalT (30% residual) and GlcT (50% residual) activities compared to the wild-type, in presence of both donor and acceptor substrates (Fig 4-2, supplementary Table S1). Despite less pronounced alterations compared to those observed when mutating Trp145, these data support possible synergistic mechanisms between long-range acceptor substrate recognition on the enzyme's surface and conformational rearrangements in the enzyme's catalytic site.

Molecular structures of LH3 in complex with UDP-Glc and Mn^{2+}, showed weak electron density near the UDP pyrophosphate group partially compatible with glycan donor substrates, likely representative of multiple conformations simultaneously trapped in the substrate binding cavity (Scietti et al, 2018). We explored the LH3 catalytic cavity in its proximity, looking for additional amino acids potentially critical for catalysis. In particular, we searched for residues carrying carboxylic or amide side chains, thus capable of acting as candidate catalytic nucleophiles for the formation of a (covalent) glycosyl-enzyme intermediate prior to glycosylation of the acceptor substrate (Gloster, 2014).

In retaining-type glycosyltransferases belonging to the GT-6 family, a conserved glutamate has been found to act as a nucleophile (supplementary Fig S3) (Albesa-Jove et al, 2017; Coutinho et al, 2003; Gomez et al, 2012; Lombard et al, 2014; Patenaude et al, 2002). In LH3 structures, we noticed that residues Gln192, Asn165 and Glu141 point towards the cavity that accommodates the glycan moiety (Fig 4-1A). We generated Ala mutations of all these residues, obtaining in all cases folded functional LH enzymes (supplementary Fig S1). When probed for GalT and GlcT activity, we found that both the Asn165Ala and the Gln192Ala mutants were less efficient, but still capable, in activating UDP-donor substrates and performing sugar transfer to acceptor substrates. Conversely, the Glu141Ala mutant was completely deprived of both GalT and GlcT activities (Fig 4-2, supplementary Table S1). These data suggest Glu141 as essential for catalysis and the surrounding negatively charged pocket composed of Asp190, Asp191, Gln192, Asn165 comprising a broad network of amino acids which may concertedly assist the LH3 glycosyltransferase activity.

Proximate to Glu141 in the GalT/GlcT cavity, residue Asn255 is the closest amino acid to the UDP phosphate-sugar bond. Despite being fully conserved in LH isoforms (Fig 4-1B), this residue is not found in any GT-A-type glycosyltransferases with known structures to date. The side chain of Asn255 consistently points to a direction opposite to

the donor substrate in all LH3 structures (Fig 4-3B, supplementary Fig S2B). We wondered whether the side chain amide group might be involved in catalysis, possibly through recognition of acceptor substrates given the conformation displayed by this side chain. Surprisingly, LH3 Asn255Ala mutants showed that their GalT activity was completely abolished, whereas the GlcT enzymatic activity was reduced to 50% (Fig 4-2, supplementary Table S1); the protein was properly folded and showed LH activity comparable to LH3 wild-type (supplementary Fig S1). These results identify Asn255 as a possible critical discriminating residue for the two glycosyltransferase activities of LH3, and rule out possible functions for this residue as catalytic nucleophile for retaining-type glycosyltransferase mechanisms, given its major impact restricted to the GalT catalytic activity.

4.3.3 Pathogenic LH3 mutations in the LH3 GT domain affect protein folding

Recently a new pathogenic LH3 mutation, Pro270Leu, has been identified and mapped at the interface of the AC and GT domain (Ewans et al, 2019). This residue localizes in a loop which is critical for shaping the GT cavity, although given its position it is unlikely to play direct roles in catalysis. To better understand the impact of this pathogenic mutation on LH3 enzymatic activity, we generated a Pro270Leu mutant. In this case, due to very low expression levels, we could not reliably carry out any *in vitro* investigations. Considering the high reproducibility associated to recombinant production of a large variety of LH3 point mutants, this result may indicate that this mutation is likely to severely impact the overall enzyme stability rather than its enzymatic activity, resulting in extremely low protein expression levels *in vitro* and likely also *in vivo*.

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

A frequent limitation associated to molecular characterizations of glycosyltransferases is the high flexibility of the donor substrate glycan moiety within the catalytic cavity. Such limitation becomes even more relevant when the enzyme is capable of processing UDP-sugar molecules in absence of acceptor substrates, such as in the case of LH3. Considering our previous (Scietti et al, 2018) and current co-crystallization results, we wondered whether free UDP, the product of the enzymatic reaction, could remain bound in the LH3 GT domain with the same efficiency as physiological donor substrates even after processing. We therefore compared the binding of free UDP and donor UDP-sugars using DSF and detected a thermal shift of 3.5 °C for free UDP, compared to a 2-2.5 °C shift using UDP-sugar substrates (Fig 4-4A). These results suggested that free UDP may bind to LH3, likely with even higher affinity than UDP-glycan substrates, and that the GalT and GlcT reactions may therefore be affected by product inhibition. To our surprise, the increase in thermal stability did not correlate with efficient trapping of the reaction product in LH3 molecular structures. Independently from the UDP

concentration used in co-crystallization and soaking experiments, we never observed any electron density for free UDP, yielding LH3 structures completely identical to ligand-free enzyme (supplementary Fig S4A).

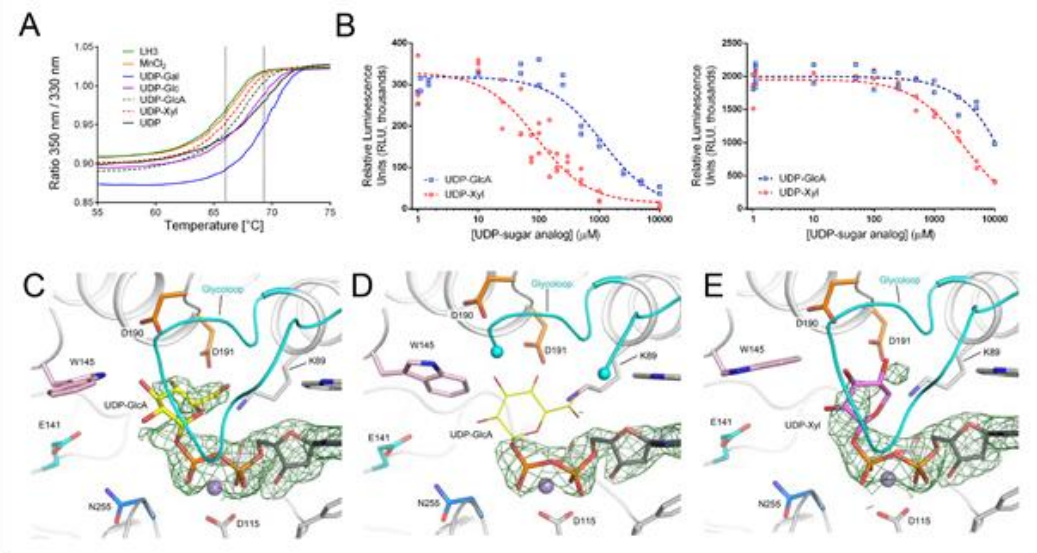


Figure 4-4 Characterization of UDP-sugar analogs.
(A) Thermal stability of LH3 wild-type (solid green) using differential scanning fluorimetry (DSF) in presence of various Mn^{2+} and several UDP-sugars. A prominent stabilization effect is achieved in presence of the biological donor substrates UDP-galactose (solid blue), UDP-Glucose (solid purple) and free UDP (solid black). A milder stabilization effect is also obtained with UDP-xylose (red dash) and UDP-glucuronic acid (green dash). (B) Evaluation of GalT and GlcT enzymatic activities of LH3 in the presence of increasing concentrations of UDP-GlcA or UDP-Xyl. (C) Crystal structure of LH3 wild-type in complex with Mn^{2+} and UDP-glucuronic acid shows clear electron density for UDP (2Fo-Fc omit electron density maps, green mesh, contour level 1.2 σ). The glucuronic acid (shown in yellow) can be modelled even if with partial electron density. (D) Crystal structure of the LH3 Val80Lys mutant in complex with Mn^{2+} and UDP-glucuronic acid. Whereas the UDP backbone can be modelled in the electron density (black sticks) (2Fo-Fc 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 (2Fo-Fc 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 from Kivirikko and colleagues (Kivirikko & Myllyla, 1979) suggested that UDP-glucuronic acid (UDP-GlcA) could act as competitive inhibitor of collagen glycosyltransferases. Based on that, UDP-GlcA was used to isolate LH3 from chicken embryos preparation (Myllyla et al, 1977; Wang et al, 2002a). However, no follow-up biochemical studies could be found in the literature. We used DSF and luminescence-

based GalT and GlcT activity assays to investigate whether and how UDP-GlcA could affect LH3 enzymatic activity. DSF showed that UDP-GlcA indeed binds weakly to LH3, resulting in a thermal shift of 1-1.5 °C (Fig 4-4A), highlighting limited stabilization compared to UDP-glycan substrates and free UDP. Enzymatic assays also confirmed the competitive inhibition displayed by this molecule (Fig 4-4B), with IC50 values in the millimolar range (supplementary Table S1). We also successfully co-crystallized and determined the 2.2-Å resolution crystal structure of wild-type LH3 in complex with Mn^{2+} and UDP-GlcA (supplementary Table 2), and found that the inhibitor could efficiently replace UDP-sugar donor substrates in the substrate cavity (Fig 4-4C). We observed additional electron density for the glucuronic acid moiety of the inhibitor in the enzyme's catalytic cavity, however this density could not be interpreted with a single inhibitor conformation. Nevertheless, analysis of the experimental electron density for the glucuronic acid moiety unambiguously showed that the inhibitor adopts a "bent" conformation: the glycan moiety is deeply buried in the enzyme's catalytic cavity proximate to residues Lys89, Asp190, Asp191, but distant from the residues found critical for catalysis, including Trp145, Asn255 and Glu141 (Fig 4-4C), thereby leaving the remaining space in the cavity for accommodating acceptor substrates.

Considering the possible conformations adopted by the glucuronic acid moiety based on analysis of the electron density and the close proximity of the glucuronic acid moiety to LH3 Val80, we wondered whether the LH3 Val80Lys mutation could interfere with inhibitor binding. We therefore co-crystallized and solved the 2.7-Å resolution structure of LH3 Val80Lys mutant in complex with UDP-GlcA (supplementary Table S2), and surprisingly observed partial displacement of the glycoloop, for which we could not observe the typical well defined electron density present in UDP-sugar-bound wild-type LH3 structures (Fig 4-4D). At the same time, we could not observe improvements in the quality of the electron density for the glucuronic acid moiety, resulting even poorer than what observed in wild-type LH3 (Fig 4-4D). This suggests that the intrinsic flexibility of the sugar-like moiety is not influenced by specific conformations of the glycoloop, but rather by lack of specific protein-ligand interactions that could provide stabilization of the sugar ring in a unique structural arrangement.

The lack of a precise conformation for the glucuronic acid moiety observed crystal structures prompted us for a further investigation of another UDP-sugar substrate analog, characterized by lack of the carboxylic moiety of UDP-GlcA: UDP-Xylose (UDP-Xyl). Similar to UDP-GlcA, UDP-Xyl resulted to be a weak inhibitor of LH3 GalT and GlcT activities (Fig 4-4A-B), with IC50 in the high micromolar range (supplementary Table S1). The 2.0-Å resolution structure of LH3 in complex with Mn^{2+} and UDP-Xyl also showed the inhibitor bound inside the enzyme's catalytic cavity, with weak electron density associated to the sugar moiety suggesting multiple conformations of the xylose moiety attached to UDP, similar to what observed for UDP-GlcA (Fig 4-4E). Taken

together, these results suggest that the LH3 GT binding cavity is capable of hosting a variety of UDP-sugar substrates, and that inhibition likely depends on the reduced flexibility (and therefore increased stabilization) of the ligand within the cavity. In this respect, we could expect that UDP-sugar analogs strongly interacting with side chains proximate to the glycan moieties of UDP-GlcA and UDP-Xyl, may have the potential to become powerful inhibitors of LH3 glycosyltransferase activities.

4.4 DISCUSSION

Glycosyltransferases are highly versatile, yet very specific enzymes. If carefully inspected, they reveal a series of recurrent features that allow their comparative characterization even in presence of very low sequence/structure conservation. LH3 has been known for long time as a promiscuous enzyme able to exploit both an inverting and a retaining catalytic mechanism *in vitro* for the specific transfer of different sugars to at least two different acceptor substrates: the HyK and the GalHyK of collagens (Myllyharju & Kivirikko, 2004). Our *in vitro* investigations highlight multiple areas surrounding the glycosyltransferase catalytic site that can be considered as critical "hot spots" for the LH3 GalT and GlcT activities: the glycoloop, the poly-Asp helix, the acceptor substrate cavity, and the region proximate to the UDP-sugar donor substrate (Fig 4-1A).

LH3 is the only isoform of its family found capable of glycosyltransferase activities, however the strong sequence conservation among human LH isoforms cannot be used to elucidate the key determinants for this additional function. The GT domain has features (such as the DxxD motif, the poly-Asp region, the glycoloop) not found in other glycosyltransferases, yet highly conserved within the LH family (Fig 4-1B). Computational homology models of homologous LH1 and LH2a/b (Scietti et al, 2019) (supplementary Fig S5) support the possibility of UDP-sugar donor binding and processing. To further investigate LH enzymes' GT domains, we focused on the only non-matching residue present within the whole amino acid sequence directly surrounding the UDP-donor substrate. This residue (Val80 in LH3, corresponding to Lys68 in LH1 and Gly80 in LH2a/b) is located in the middle of the glycoloop, in close proximity to the ribose ring of the UDP-sugar donor substrate(s). Our data were consistent with this residue being important for catalysis, as it assists the positioning of the bound donor substrate and in particular the glycan moiety. However, the results also emphasize how sequence alterations at this site are not sufficient to justify the lack of GalT/GlcT enzymatic activities in homologous LH1/2. We therefore expanded our investigation to the second-shell environment surrounding the donor substrate, and found that non-conserved LH3 Trp92 (Leu80 in LH1, Leu92 in LH2a/b) positions its aromatic side chain in a conformation that stabilizes the entire glycoloop to facilitate the enzymatic reactions. Again, mutating this residue did not lead to full loss of the glycosyltransferase activities. This is consistent with the loss of GalT/GlcT functions in LH1 and LH2a/b being

associated to a broad set of subtle alterations, possibly involving residues distant from the actual enzyme's catalytic site, likely essential for recognition of collagen acceptor substrates.

Prompted by these observations, we expanded our investigation to the LH3 catalytic cavity expected to host the glycan moieties of the UDP-sugar donor substrates and the acceptor molecules. Residues Asp190 and Asp191 of the characteristic LH3 poly-Asp helix lie in a conserved position compared to other retaining and inverting type GT-A (supplementary Fig S3, Table 4-1) and our data indicate that both these carboxylate moieties are critical for efficient donor substrate activation and sugar transfer. Although residues matching these positions have been proposed to act as catalytic nucleophiles in retaining type GT-A (Flint et al, 2005; Persson et al, 2001; Wang et al, 2002b), and as catalytic bases in inverting type GT-A (Charnock & Davies, 1999; Pedersen et al, 2000), their distances and the relative orientations with respect to UDP-sugar donor substrates in the LH3 GT domain (Fig 4-1A) do not support this hypothesis. Nevertheless, these residues play critical roles in recognizing and assisting the proper positioning of donor UDP-sugar substrates, as shown by their proximity to glycan moieties in LH3 co-crystal structures with UDP-sugar analogs (Fig 4-4C-E), a feature observed also in mutagenesis studies on other GT-A retaining type glycosyltransferases (Lairson et al, 2004).

On the opposite site of the UDP-binding pocket, LH3 exhibits a non-conserved loop shaping the GT catalytic cavity bearing two aromatic residues, the Trp145 and the Trp148 that seem to act in a concerted way during catalysis, as suggested by comparisons between substrate-free and substrate-bound LH3 molecular structures (Fig 4-1A). Trp145 is indispensable for both GalT/GlcT activities: its conformational changes seem to respond to the presence and conformational positioning of the donor substrate inside the catalytic cavity. Although located in a loop that is uniquely found in LH3, the Trp145 side chain matches a site frequently occupied by bulky aromatic residues in other GT-A glycosyltransferases that shape a portion of the GT cavity to facilitate donor substrate processing and catalysis. This further highlights the versatility of glycosyltransferases, in which many different structural features have evolved to specifically recognize distinct donor and acceptor substrates, while preserving the ability to carry out the same catalytic reaction. The implication in catalysis of the less conserved Trp148 on the surface of the GT domain is even more intriguing: this residue seems to coordinately respond to the rearrangements of its counterpart Trp145 in the catalytic site, suggesting involvement in recognition of the acceptor substrate prior to its access into the GT cavity. Alternatively, Trp148 may contribute to long-range stabilizing interactions with collagen molecules, while they dock their HyK or GalHyK residues in the acceptor substrate site during the GalT or GlcT reactions, respectively.

Catalytic nucleophiles have been clearly identified so far only in the retaining-type glycosyltransferases belonging to the GT-6 family (Coutinho et al, 2003; Lombard et al, 2014), such as the α−1,3 galactosyltransferase (GGTA1) where a conserved glutamate is found positioned on the β−face of the donor sugar (Albesa-Jove et al, 2017; Gomez et al, 2012; Patenaude et al, 2002) (supplementary Fig S3). Conversely, extensive structural comparisons and mutagenesis experiments have been performed in the O-galactosyltransferase LgtC from Neisseria menengitidis, focusing on matching residue, Gln189 (Lairson et al, 2004). However, the role of this residue as catalytic nucleophile was ruled out. This site is occupied by Gln192 in LH3. This residue is next to the poly-Asp helix, distant from the sites occupied by donor substrates and in an arrangement that is not compatible with a direct role in catalysis. However, our mutagenesis data indicate that removal of the Gln192 side chain has a strong impact on LH3 glycosyltransferase activity. In close proximity, we identified two other amino acid residues potentially involved in donor substrate activation or transfer of sugar moieties to the acceptor molecule. Both Asn165 and Glu141 point directly towards the glycan moiety of the donor substrate (Fig 4-1A). Whilst the Asn165Ala mutation only reduced the glycosyltransferase activity by a factor of two (Fig 2, supplementary Table 1), we found that Glu141 is essential for both GalT and GlcT activities, as the Glu141Ala mutation yields results in a completely inactive LH3 glycosyltransferase (Fig 4-2, supplementary Table 1). In LH3, 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 (supplementary Fig 3, Table 4-1). In LH3, residue Asn255 is the closest amino acid to the UDP phosphate-sugar bond, but in crystal structures its side chain consistently points to a direction opposite to the donor substrate (Fig 4-1A). When inspecting the molecular structures of other GT-A glycosyltransferases, we noticed that this residue is not conserved (supplementary Fig 3). On the contrary, this residue is fully conserved among human LH isoforms (Fig 4-1B). Strikingly, the LH3 Asn255Ala mutant showed a complete loss of GalT activity, but partially preserved the GlcT activity. Although the significance of *in vivo* LH3 GalT activity is uncertain, such activity is clearly detectable *in vitro* (Scietti et al, 2018; Wang et al, 2002a). The ability of Asn255 to selectively abolish only LH3 GalT activity highlights how LH3 can promiscuously accept and recognize very different acceptor substrates (i.e., collagen HyK versus GalHyK) within the same catalytic site and carry out two glycosyltransferase reactions that rely on different mechanisms (i.e., inverting GalT and retaining GlcT). Presently, LH3 is the only known glycosyltransferase capable of such promiscuity, and our data show that it shares numerous features with both retaining and inverting glycosyltransferases. Collectively, these results suggest the intriguing possibility that LH3 is not just the ancestor of the whole LH family, but may preserve in its sequence features belonging to the evolutionary precursors of both retaining and inverting glycosyltransferases.

In addition, the present work provides a set of 3D structures of LH3 in complex with UDP-sugar analogs, which work as mild inhibitors (Fig 4-4). Despite the high flexibility observed for the glycan moieties of the bound molecules, the new molecular structures presented provide valuable insights for structure-based drug development of inhibitors of LH3 GalT/GlcT enzymatic activities. These molecules may give the spark to innovative therapeutic strategies against pathological conditions characterized by excess collagen glycosylations, such as osteogenesis imperfecta (Raghunath et al, 1994).

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

4.5 ACKNOWLEDGEMENTS

We thank the European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS) for the provision of synchrotron radiation facilities. We thank Dr. M. Campioni for useful discussions and M. Miao for support in crystallization experiments.

4.6 REFERENCES

1. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213–221

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

3. Ardevol A, Iglesias-Fernandez J, Rojas-Cervellera V, Rovira C (2016) The reaction mechanism of retaining glycosyltransferases. Biochem Soc Trans 44: 51–60

4. Baumann S, Hennet T (2016) Collagen Accumulation in Osteosarcoma Cells lacking GLT25D1 Collagen Galactosyltransferase. J Biol Chem 291: 18514–18524

5. Bornstein P, Sage H (1980) Structurally distinct collagen types. Annu Rev Biochem 49: 957–1003

6. Charnock SJ, Davies GJ (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. Biochemistry 38: 6380–6385

7. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66: 12–21

8. Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. J Mol Biol 328: 307–317

9. Cummings RD (2009) The repertoire of glycan determinants in the human glycome. Mol Biosyst 5: 1087–1104

10. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallographica Section D 66: 486–501

11. Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69: 1204–1214

12. Ewans LJ, Colley A, Gaston-Massuet C, Gualtieri A, Cowley MJ, McCabe MJ, Anand D, Lachke SA, Scietti L, Forneris F, Zhu Y, Ying K, Walsh C, Kirk EP, Miller D, Giunta C, Sillence D, Dinger M, Buckley M, Roscioli T (2019) Pathogenic variants in PLOD3 result in a Stickler syndrome-like connective tissue disorder with vascular complications. J Med Genet

13. Flint J, Taylor E, Yang M, Bolam DN, Tailford LE, Martinez-Fleites C, Dodson EJ, Davis BG, Gilbert HJ, Davies GJ (2005) Structural dissection and high-throughput screening of mannosylglycerate synthase. Nat Struct Mol Biol 12: 608–614

14. Gloster TM (2014) Advances in understanding glycosyltransferases from a structural perspective. Curr Opin Struct Biol 28: 131–141

15. Gomez H, Lluch JM, Masgrau L (2012) Essential role of glutamate 317 in galactosyl transfer by alpha3GalT: a computational study. Carbohydr Res 356: 204–208

16. Hamazaki H, Hamazaki MH (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

17. Heikkinen J, Risteli M, Wang C, Latvala J, Rossi M, Valtavaara M, Myllyla R (2000) Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. J Biol Chem 275: 36158–36163

18. Hennet T (2019) Collagen glycosylation. Curr Opin Struct Biol 56: 131–138

19. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66: 125–132

20. Kivirikko KI, Myllyla R (1979) Collagen glycosyltransferases. Int Rev Connect Tissue Res 8: 23–72

21. Lairson LL, Chiu CP, Ly HD, He S, Wakarchuk WW, Strynadka NC, Withers SG (2004) Intermediate trapping on a mutant retaining alpha-galactosyltransferase identifies an unexpected aspartate residue. J Biol Chem 279: 28339–28344

22. Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem 77: 521–555

23. Lehmann HW, Wolf E, Roser K, Bodo M, Delling G, Muller PK (1995) Composition and posttranslational modification of individual collagen chains from osteosarcomas and osteofibrous dysplasias. J Cancer Res Clin Oncol 121: 413–418

24. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42: D490–495

25. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennet T (2011) Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J Biol Chem 286: 43701–43709

26. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674

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

28. Myllyharju J, Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20: 33–43

29. Myllyla R, Anttinen H, Risteli L, Kivirikko KI (1977) Isolation of collagen glucosyltransferase as a homogeneous protein from chick embryos. Biochim Biophys Acta 480: 113–121

30. Patenaude SI, Seto NO, Borisova SN, Szpacenko A, Marcus SL, Palcic MM, Evans SV (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis. Nat Struct Biol 9: 685–690

31. Pedersen LC, Tsuchida K, Kitagawa H, Sugahara K, Darden TA, Negishi M (2000) Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. J Biol Chem 275: 34580–34585

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

33. Persson K, Ly HD, Dieckelmann M, Wakarchuk WW, Withers SG, Strynadka NC (2001) Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. Nat Struct Biol 8: 166–175

34. Raghunath M, Bruckner P, Steinmann B (1994) Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. J Mol Biol 236: 940–949

35. Rautavuoma K, Takaluoma K, Sormunen R, Myllyharju J, Kivirikko KI, 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

36. Ruotsalainen H, Sipila L, Vapola M, Sormunen R, Salo AM, Uitto L, Mercer DK, Robins SP, Risteli M, Aszodi A, Fassler R, Myllyla R (2006) Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. J Cell Sci 119: 625–635

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

38. Savolainen ER, Kero M, Pihlajaniemi T, Kivirikko KI (1981) Deficiency of galactosylhydroxylysyl glucosyltransferase, an enzyme of collagen synthesis, in a family with dominant epidermolysis bullosa simplex. N Engl J Med 304: 197–204

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

40. Schofield JD, Freeman IL, Jackson DS (1971) The isolation, and amino acid and carbohydrate composition, of polymeric collagens prepared from various human tissues. Biochem J 124: 467–473

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

42. 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, Forneris F (2018) Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. Nat Commun 9: 3163

43. Sipila L, Ruotsalainen H, Sormunen R, Baker NL, Lamande SR, Vapola M, Wang C, Sado Y, Aszodi A, Myllyla R (2007) Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. J Biol Chem 282: 33381–33388

44. Spiro MJ, Spiro RG (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

45. Spiro RG (1967) The structure of the disaccharide unit of the renal glomerular basement membrane. J Biol Chem 242: 4813–4823

46. Spiro RG (1969) Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. J Biol Chem 244: 602–612

47. Sricholpech M, Perdivara I, Nagaoka H, Yokoyama M, Tomer KB, Yamauchi M (2011) Lysyl hydroxylase 3 glucosylates galactosylhydroxylysine residues in type I collagen in osteoblast culture. J Biol Chem 286: 8846–8856

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

49. Sternberg M, Spiro RG (1980) Studies on the catabolism of the hydroxylysine-linked disaccharide units of basement membranes and collagens: isolation and characterization of a new rat-kidney alpha-glucosidase of high specificity. Ren Physiol 3: 1–3

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

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

52. Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, Bricogne G. (2018) STARANISO. Global Phasing Ltd., Cambridge, United Kingdom.

53. Toole BP, Kang AH, Trelstad RL, Gross J (1972) Collagen heterogeneity within different growth regions of long bones of rachitic and non-rachitic chicks. Biochem J 127: 715–720

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

55. Wang C, Risteli M, Heikkinen J, Hussa AK, Uitto L, 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

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

CONCLUSIONS

Collagen lysine modifications are highly conserved PTMs essential for the correct biosynthesis of collagen and deposition of ECM. The first modification is lysine hydroxylation, catalyzed by LH enzymes. In humans, three isoforms are found: LH1, LH2a/b and LH3, sharing approximately 70% of sequence identity. LH3 is the only multifunctional enzyme, bearing lysyl hydroxylase but also additional glycosyltransferase activities, absent in LH1 and LH2. Malfunctions of LH enzymes leading to abnormal collagen PTMs correlate with severe connective tissue disorders and cancer metastatisation. The biological and medical relevance of this enzyme family arose our interest. Thus, we firstly established innovative strategies of large-scale LH recombinant protein expression using mammalian expression systems, and developed enzymatic assays to monitor lysine hydroxylation activity on different substrates. This thesis reports on the successful results obtained about the comprehensive biochemical and structural characterization of human full-length human LH3, and also about some follow-up studies. We have determined the 3D structure of LH3 isoform, highlighting unprecedented features of this multi-functional enzyme. Using site-directed mutagenesis, we identified the crucial features essential for catalysis in both catalytic sites of LH3. Collectively, our results provided a structural framework to describe the multiple functions of LH3, and shed light on some of the molecular mechanisms causing collagen-related diseases involving human lysyl hydroxylases. The high sequence conservation allowed using the 3D structure of LH3 as template for the entire LH family, and enabled the *in silico* homology modelling prediction of LH1 and LH2 structures. Furthermore, it paved the way to structure-based drug design studies, now in progress in the lab. The results presented in this PhD thesis also provide critical insights into the LH3 glycosyltransferase activities, and expand the available knowledge for the development of collagen GalT/GlcT inhibitors. These insights will be instrumental for the manipulation of LH3 protein functions and donor substrate specificity for biomedical applications, including possible development of antimetastatic agents. Our studies will advance the understanding of the molecular genotype-phenotype correlations associated to several connective tissue disorders, such as Osteogenesis Imperfecta, Ehlers-Danlos syndrome, Bruck syndrome and Dystrophic Epidermolysis Bullosa. However, some pieces of the puzzle are missing, in particular regarding the *in vivo* significance of the molecular insights obtained: it is unclear how and why only a small percentage of lysines in fibrillar collagens

are modified through hydroxylation and glycosylation. The elongated structure of LH3 together with the double activity suggests a potential processive catalytic mechanism, however no clear experimental evidence is currently available and the existence of the LH3 GalT activity must be confirmed in order to validate such hypothesis. Also, we still do not know much about the exact localization of LH3 enzyme, as well as about the functional role of its secretion. Future studies, encouraged by the results presented in this work, should aim at the *in vivo* characterization of LH3 and the other LH isoforms.