nature portfolio

Peer Review File

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

Corresponding Author: Professor Federico Forneris

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Reviewer #2

(Remarks to the Author)
Collagen is a critical extracellular protein, which functions are regulated by lysyl post-translational modification. Certain lysine residues are hydroxylated into hydroxylysine to serve as an attaching site for a unique type of glycosylation initiated collagen galactosyltransferase GLT25D family. However, the structural basis of GLT25D is still not known. Marco et al provide the first structural insights into collagen galactosyltransferase GLT25D1. The structure reveals a head-to-head homodimer assembly, although dimer is not required for the in vitro enzymatic activity. So, the dimer's function is still not clear. Each monomer has two Rossman fold domains linked by a loop, both of which bind cofactor Mn2+ and sugar donor UDP galactose. Site-directed mutagenesis and enzymatic activity assays suggest the second Rossman domain houses the active site. The manuscript is novel and largely technically sound with enough details.

The following suggestions may help further improve the work and advance the findings.
1. These authors showed an interesting dimer assembly of GLT25D1, however, the function of dimerization is still not clear. Is the dimer needed for the collagen glycosylation in cells? Does the dimer help interact with PLODs/LHs which form tail-to-tail homodimers?
2. This work showed that both Rossman fold domains bind cofactor and UDP sugar, which is very interesting. Is there any structure difference that prevents the first Rossman fold domain from becoming an active enzyme? There are about 30 families of collagen subtypes. Is it possible the different Rossman fold domains modify different collagen subtypes? The assays in the manuscript cannot rule out this possibility.
3. The triplet Gly-X-Y motif description is confusing. It sounds like X is hydroxyproline, and Y is hydroxylysine, which is not the case. Please explain in a clear manner.
4. Page 3 line 58 and 59, it may be better to simplify it since we focus on GLT25D1. If we decide to include LHs/PLODs in the manuscript, it may be worth talking about 2-oxoglutarate and succinate.
5. Page 4 line 71, please reference the work suggesting GLT25Ds and LHs form a complex.
6. EDD motif has been found in other glycosyltransferases. It is not unprecedented.
7. The work showing the first Rossman fold domain cofactor binding stabilizes the folding is not conclusive. Do the results still hold if we use the second Rossman fold domain mutant and measure its thermostability with and with EDTA?
8. Fig 3C, the structure does not match the ab-initio SAXS envelop well. Does the protein used for SAXS still have sumo tag?
9. Mutagenesis of Trp130/135 to show its relevance in GLT25D1 function?
10. Fig 2, the contour level at 3.2 σ is high.

Reviewer #3

(Remarks to the Author)
This manuscript deals with the molecular structure of an enzyme, GLT25D1/COLDALT1, to discuss the enzymatic mechanism. Various methods, X-ray diffraction, mass photometry, SEC-MALS and SEC-SAXS have been performed to reveal the dimeric quaternary structure. These are mostly thorough. Although the evaluated higher order structure is not directly related to the enzymatic activity, the current finding may be important to elucidate the mechanism to construct ECM in vivo. The following points can be considered before publication.
1. Both mass photometry and SEC-MALS results show some molar mass (or aggregation number) dispersion for the analyzed samples. Authors can discuss the effect of dispersion on enzymatic and/or physiological activity.
2. The above mentioned dispersity may affect the analysis of the SAXS data. Some comments may be added to the main text. In this context, can the evaluated dimeric quaternary structure be described as rigid at room temperature? If not, this will also affect the scattering profile, and hence the resulting 3-d structure shown in Figure 3d.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Review of revised Nature Communications manuscript NCOMMS-24-37599A
Molecular Structure and Enzymatic Mechanism of the Human Collagen Hydroxylysine Galactosyltransferase GLT25D1/COLGALT1

Firstly, congratulations to the authors on the revised manuscript - they really have made an excellent effort to address all of my points and those of the other reviewers. I think the manuscript is informative and will genuinely useful to the field of collagen biosynthesis. Also, I fully support keeping the SAD workflow in the manuscript, despite the success of Alpha fold. Regarding keeping ‘enzymatic’ in the title, this is up to the authors but it is redundant as a transferase in an enzyme by definition. Anyway, well done!

Reviewer #2

(Remarks to the Author)
The work is solid and innovative. The reviewers have addressed most of the minor concerns; while the following major concern (What is the function of dimierization?) has not been addressed.

The evidence supporting GT2, not GT1, as a active domain is still somewhat weak.

Moreover, nowadays, AlphaFold can predict the structure with good accuracy but can not predict dynamics or conformation changes. Thus the initial ab-initio SAXS data that did not agree well with the crystal structure may still be worth keeping. It is potentially meaningful and may add more to the manuscript if properly explained.

Trp130/130 mutagenesis is surprising. Have we tried an alanine mutation?

Reviewer #3

(Remarks to the Author)
The authors have adequately revised the manuscript. Therefore, the reviewer recommends publication.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
This revised manuscript has addressed my concerns.

POINT BY POINT RESPONSE TO REVIEWER'S COMMENTS

Reviewer #1:

Structural studies on collagen biosynthesis enzymes are both of considerable basic biological interest and are relevant to understanding the molecular basis of multiple diseases. The structural work on human galactosyl transferase described by De Marco et al is therefore timely and, I think, appropriate for publication in Nature Communications after appropriate modifications, including substantial new experiments. The strengths of the work is the crystallographic analyses. The biochemical work is much less strong. At very least basic kinetic characterisation of the enzymes and key mutant needs to be reported, i.e. kcat and KM values. A second weakness is the characterisation of the interactions of the GT's with metal ions. Much more extensive screening with biologically relevant needs to be done; again, at least with basic kinetic characterisation. Does Mg2+ work? Note the electron density maps are not sufficient to define the metal ions, let alone the oxidation states.

Finally, after all the hard work to get the crystals, more extensive structurally informed bioinformatic structures need to be done. Now do these GT's vary across evolution, are they in plants, archaea etc. So overall, I think this is a potentially very nice paper, but it needs much more biochemistry and the bigger biological/evolution picture needs to be included (well done on the structures!)

We would like to thank the reviewer for his/her comments on our work. We appreciate his/her consideration and have addressed his/her requests as detailed in the next paragraphs. In particular, we have added numerous experiments, as requested, to characterize the enzyme kinetics as well as the metal ion binding properties of GLT25D1/COLGALT1, also through molecular dynamics simulations that further supported the experimental results. Thanks to the insightful suggestions, we have now identified new features further strengthening the conclusions of our manuscript regarding the structural role of the GT1 domain and the functional role of the GT2 domain of GLT25D1/COLGALT1. We have also included broader investigations on protein sequences and homologous structures, and provided more extensive reasoning on the possible evolution of this enzyme family.

Title

Delete "Enzymatic" from the title

In consideration of the new experiments added to characterize the catalytic properties of GLT25D1/COLGALT1, we find appropriate to maintain the word "Enzymatic" in the title of our revised manuscript.

Main Text – specific points

Abstract - check grammar

"During collagen biosynthesis," "supramolecular"

L29 replace "integrative molecular" with "biophysical"

Studies"

L28 delete "for……"

L36 "which is essential"

L37 delete "conversely"

Following the reviewer's suggestions, we have thoroughly checked the sentence construction of our abstract and revised where needed.

Introduction

L48 "; it is essential…"

L49 delete "Contrary to proline residues"

L74 "the subject of"

L76 "which are responsible"

L95 "have yielded"

L99 "wildtype" more correct than "native"

L103 "both of which support binding of the …"

Following the reviewer's suggestions, we have thoroughly checked the sentence construction of our introduction and revised according to the suggestions provided.

P6 Refer to/give evidence of protein purity by SDS-PAGE and MS.

We have now included all quality control checks on recombinant samples, performed using SDS-PAGE and DSF, in the new Supplementary Fig. 1. We have included reference to these quality control checks where appropriate in the materials and methods section as well as in the results section as indicated by the reviewer.

Supplementary Information

Fig S3 Please label secondary structures elements

We have modified Figure S3 (now Supplementary Fig. 5) as requested by adding labels on secondary structure elements.

Fig S7 maybe worth adding non-animal eukaryotes?

We think that the reviewer refers to former Fig S4, not S7. We have added sequence and structural alignments focusing on non-animal eukaryotes and provided additional discussions on the evolution of GLT25D1/COLGALT1 glycosyltransferases in the revised main text.

Fig S8 "of the ". Check grammar of this legend – purification doesn't result in expression

We thank the reviewer for this suggestion. We have now replaced the "purification" word with a more appropriate "production".

P7 Crystallisation. Give more details of optimisation procedure on screens used etc.

We thank the reviewer for this suggestion. We have added details of the crystallization optimization procedures in the revised materials and methods section (lines 518-521).

P12 As requested above give evidence of protein purity (SDS-PAGE/MS)

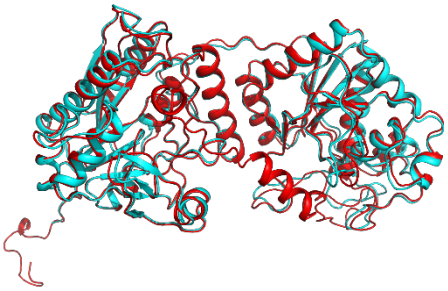
We have now included all quality control checks on recombinant samples, performed using SDS-PAGE and DSF, in the new Supplementary Fig. 1. Where appropriate, we have included reference to these quality control checks in the materials and methods section as well as in the results section as indicated by the reviewer.

At least basic kinetic characterisation should be reported, i.e. Kcat 4, Km values. Please screen other metal ions more extensively, e.g. Fe (II), Co (II), 2nC (II) Ca (II), etc. There does appear to be some activity with Mg (II)? Test for inhibition as well as activity.

Following the reviewer's suggestion, we have developed a new luminescence-based time-course assay to evaluate the enzymatic activity, which enabled measuring K_M and k_{cat} values for wild-type and mutant GLT25D1/COLGALT1 and for Mg^{2+} other than Mn^{2+}. Assays performed using other metals did not show activity. The assay methodology is described in materials and methods (lines 576-591). Associated results are included in Supplementary Fig. 2 and in Supplementary Table 5.

L291 "The experimental" to "A" crystal…."
Corrected.

Out of interest, presumably the authors tried to solve the structure by Alpha fold - and presumably it didn't work? Maybe connect.
We thank the reviewer for pointing out about the possibility of structure determination using Alphafold. The initial crystal structure determination of GLT25D1/COLGALT1 was obtained before the release Alphafold2/3, and given the extremely limited sequence identity conservation between GLT25D1/COLGALT1 and other glycosyltransferases with known structure at that time (including LH3/PLOD3), molecular replacement attempts were not successful. After the introduction of Alphafold2/3, we generated predictions of GLT25D1/COLGALT1 structure and obtained models with RMSD values less than 1 Å from our crystal structure. These predictions likely enable structure determination using molecular replacement.



The figure shows a superposition between the crystal structure of human GLT25D1/COLGALT1 (revised PDB 9EVJ, this work, shown in cyan) and the computational model generated by Alphafold3 using the same protein sequence (shown in red). The superposition RMSD is 0.87 Å.

Nevertheless, given the excellent quality of the experimental data collected, and consistent with the strategy implemented for structure determination, we wish to maintain the heavy atom SAD workflow and the associated structure in this manuscript.

The structure similarity with LH3/PLOD is ??. This needs to be discussed in more detail. More extensive sequence alignments need to be given, including in relation to evolution of the GT domain across the animal kingdom and maybe beyond – what about plants?
We have now included additional superpositions considering the possible similarities also involving GLT25D1/COLGALT1 GT2 domain and LH3/PLOD3 GT domain (Supplementary Fig. 5e). Other cross-comparisons with LH3/PLOD3 (i.e., GLT25D1/COLGALT1 GT2 domain with LH3/PLOD3 AC domain, and GLT25D1/COLGALT1 GT1 domain with LH3/PLOD3 GT domain) could not be performed due to the topological differences between the GT1 and GT2 domain of GLT25D1/COLGALT1. We have also included new alignments, incorporating glycosyltransferases beyond the animal kingdom in Supplementary Fig. 6-7, and associated discussions in the revised main text (lines 140-160).

Nevertheless, we wish to point out that the Rossman fold containing glycosyltransferases typically share limited sequence similarity and that their specificity is defined by hallmark arrangements of the surface loops surrounding the rim of the catalytic cavity. In this respect, we highlight that DALI and PDBeFOLD searches for related glycosyltransferase structures have identified very limited structural homologs (shown in Supplementary Fig. 5 and described at lines 137-140 and 152-156 of the revised text).

L313 change subtitle – populated is not appropriate – "complexed"
We have replaced "can be populated" with "can bind" and revised accordingly.

L316 " unambiguously modelled by ?" this is very ? – firstly the oxidation state of the x-ray exposed crystals is unknown, secondly ? between Mn and (e.g.) as the basis of electron density alone is difficult/not possible? Please be more cautious.
We thank the reviewer for this suggestion. Throughout the manuscript, we have revised the description of the metal ions by replacing "Mn^{2+}" with a general "metal ion" unless Mn^{2+} was added to the co-crystallization and reaction mixtures. In addition, we initially performed ICP-MS and found that GT1 does not bind manganese. Unfortunately, ICP-MS investigations addressing possible metal ions such as Mg^{2+} or Ca^{2+} were not conclusive because of the large amounts of sample required to obtain clear data, hence we decided not include these results in the manuscript. We therefore opted for native mass spectrometry (Native-MS) analysis of recombinant GLT25D1/COLGALT1 under controlled partially-unfolding conditions for the characterization of the metal ion bound in the "structural" GT1 domain, and surprisingly found that the bound metal ion in this domain of the enzyme is not Mn^{2+}, but Ca^{2+}. This observation motivated us towards a more comprehensive investigation of the structural role of this metal ion, which was carried out by combining molecular dynamics simulations and evaluation of protein stability after EDTA/EGTA treatment. We thank the reviewer for this challenging yet very stimulating request that has strongly improved the significance of our biochemical work. The manuscript, display items and PDB files have been revised accordingly.

L329 please justify the 'unusual' ?
We have removed the word "unusual" from this statement.

L333 "cosubstrate"
Modified as suggested

L228 What about comparison width related GT domains, e.g. in PLOD?
We assume the reviewer refers to L338, and not 228 as stated. A structural comparison between GLT25D1/COLGALT1 GT1 domain and the related AC domain of LH3/PLOD3 was already present in our original manuscript and has been kept in Supplementary Fig. 5c. We have also added a comparison between GLT25D1/COLGALT1 GT2 domain with PLOD3 GT domain in Supplementary Fig. 5e, albeit showing very limited structural similarity as suggested by the high RMSD value (4.7 Å) and the very low DALI score (<10). Other cross-comparisons with LH3/PLOD3 (i.e., GLT25D1/COLGALT1 GT2 domain with LH3/PLOD3 AC domain, and GLT25D1/COLGALT1 GT1 domain with LH3/PLOD3 GT domain) could not be performed due to the topological differences between the GT1 and GT2 domain of GLT25D1/COLGALT1.

L354 “was supported”
Modified as suggested

L367 can it be certain that dimers don’t form during catalysis ?
We thank the reviewer for pointing out this possibility. Our data show that in solution, at both high concentrations (as shown by SEC-MALS and SEC-SAXS analysis) as well as at very low concentrations (as shown by mass photometry), as well as in crystal structures the enzyme is in dimeric form regardless of the presence of substrates and/or cofactors needed for catalysis.

L380 ICP MS/atomic absorption spectrometry need sto be done to assign the identity of metal ions. As stated above more careful/extensive studies on metal ions needs to be carried out.
We wish to thank the reviewer for his/her insightful suggestion. We have performed ICP-MS analysis on recombinant GLT25D1/COLGALT1 preparations and found that Mn^{2+} is indeed not present. This has pointed us towards identifying which other metal ion could bind in the “structural” GT1 site. Unfortunately, ICP-MS investigations addressing possible metal ions such as Mg^{2+} or Ca^{2+} were not conclusive because of the large amounts of sample required to obtain clear data, hence we decided not include these results in the manuscript. Using Native-MS analysis of recombinant GLT25D1/COLGALT1 under controlled partially unfolding conditions, we consistently identified peaks compatible with adducts composed of Ca^{2+} and UDP-α-Gal that led us to the suggestion of Ca^{2+} as the possible metal ion bound in the structural GT1 site. This unexpected result motivated us towards validation using DSF on EDTA as well as EGTA-treated enzyme samples. We also consistently observed destabilization using EDTA as well as EGTA. As the latter is highly selective for Ca^{2+}, we concluded that Ca^{2+} is the bound metal ion present in the GT1 domain. To our knowledge, this result, experimentally corroborated by two different techniques, shows the unique presence of Ca^{2+} in a glycosyltransferase domain bound to UDP-α-Gal. We have revised the manuscript, the PDB files and the associated display items accordingly.
Regarding the GT2 site, enzyme activity assays (now carried out with comprehensive measurements of K_M and k_{cat} values) have revealed that Mn^{2+} is the catalytic cofactor essential for galactosyltransferase enzymatic activity. These results have been included in the results and discussion sections of the manuscript.

L408 “unusual” more correct than “unprecedented” – is where the statement needs to be justified.
Modified as suggested.

L423 are there any clinically relevant mutations?
In our revised manuscript, we have included a paragraph about “Molecular significance of GLT25D1/COLGALT1 pathogenic mutations” (lines 277-294) describing clinically relevant mutations as suggested. We are grateful to the reviewer for this suggestion, as this analysis has likely broadened the significance of the data presented in our manuscript.

L435 delete “the”
L442 “reveal an”
L446 replace “populated”
Discussion, please modify appropriately in the light of the requested new experiments/suggested
We have revised according to the suggestions provided by the reviewer. Thank you very much.

Reviewer #2:

Collagen is a critical extracellular protein, which functions are regulated by lysyl post-translational modification. Certain lysine residues are hydroxylated into hydroxylysine to serve as an attaching site for a unique type of glycosylation initiated collagen galactosyltransferase GLT25D family. However, the structural basis of GLT25D is still not known. De Marco et al provide the first structural insights into collagen galactosyltransferase GLT25D1. The structure reveals a head-to-head homodimer assembly, although dimer is not required for the in vitro enzymatic activity. So, the dimer's function is still not clear. Each monomer has two Rossman fold domains linked by a loop, both of which bind cofactor Mn2+ and sugar donor UDP galactose. Site-directed mutagenesis and enzymatic activity assays suggest the second Rossman domain houses the active site. The manuscript is novel and largely technically sound with enough details.

We would like to thank the reviewer for his/her comments on our work. In particular, we appreciated his/her comments regarding novelty and technical soundness of the experiments described.

The following suggestions may help further improve the work and advance the findings.

1. These authors showed an interesting dimer assembly of GLT25D1, however, the function of dimerization is still not clear. Is the dimer needed for the collagen glycosylation in cells?

We thank the reviewer for this question. Based on the available data, we are not in the position to provide a conclusive statement regarding the relevance of GLT25D1/COLGALT1 dimerization in cells. Likewise, we understand the importance of this point, which has been further emphasized in the discussion section. We wish to point out that several enzymes responsible for introduction of post-translational modifications on collagens are dimeric (e.g., multifunctional lysyl hydroyxlases/glycosyltransferases as well as prolyl hydroxylases), and that most eukaryotic glycosyltransferases form homomers, as reviewed in (Kellokumpu et al. 2016). We speculate about roles associated with formation of multi-enzyme assemblies assisted by dimerization interfaces, sure that this will stimulate exciting future investigations.

Does the dimer help interact with PLODs/LHs which form tail-to-tail homodimers?

Related to our response to the previous point, we have expanded our reasoning in the discussion section (lines 447-466) related to dimerization and possible protein-protein interaction contacts. We wish to point out that an in-depth characterization of possible macromolecular complexes involving GLT25D1/COLGALT1 and LH3/PLOD3 is clearly beyond the scope of the current work and would require extensive structure-function studies as well as validation experiments.

2. This work showed that both Rossman fold domains bind cofactor and UDP sugar, which is very interesting. Is there any structure difference that prevents the first Rossman fold domain from becoming an active enzyme? There are about 30 families of collagen subtypes. Is it possible the different Rossman fold domains modify different collagen subtypes? The assays in the manuscript cannot rule out this possibility.

We thank the reviewer for proposing this intriguing possibility. The revised version of our manuscript includes a more extensive comparison of homologous protein sequences and/or structures related to the individual GT domains composing GLT25D1/COLGALT1, including vertebrates, invertebrates, giant viruses and bacteria. Although very similar to many other GT-A glycosyltransferase domains, the amino

acid residues surrounding the metal ion and UDP-α-Gal in the GT1 domain, in particular those proximate to the galactose moiety, are mostly hydrophobic where essential negatively charged amino acid side chains required for the enzymatic activity of inverting GT-A glycosyltransferases are usually present (Lairson et al. 2008). In the GT2 domain, this spot is occupied by Asp522 (a residue that has been found critical for catalysis based on site-directed mutagenesis analysis). Combined with previous results showing that a chimera of CERCAM GT1 and GLT25D1/COLGALT1 GT2 domain retains enzymatic activity (Perrin-Tricaud et al. 2011), we believe that our biochemical assays, the new evidence about structural roles of the GT1 domain, the identification of unusual Ca^{2+} bound in this domain interacting with UDP-α-Gal, and the structural alignments provide sufficient evidence about the GT1 domain of GLT25D1 not being competent for catalysis. To clarify these considerations, we have extensively modified our statements in the results and discussion sections.

3. The triplet Gly-X-Y motif description is confusing. It sounds like X is hydroxyproline, and Y is hydroxylysine, which is not the case. Please explain in a clear manner.
We apologize for the confusion our statement might have created. We have revised the text in lines 41-44 of our revised manuscript: "The repetitive unit consists of the Gly-Xaa-Yaa sequence motif, where positions Xaa and Yaa are frequently occupied by proline and hydroxyproline residues, respectively. The role of proline hydroxylation in the global process of collagen biosynthesis is the most studied and understood to date, and it is essential to the winding process of the collagen right-handed triple helix. Lysine residues are less abundant than proline residues in collagen polypeptides and are typically found at Yaa positions of the Gly-Xaa-Yaa motif".

4. Page 3 line 58 and 59, it may be better to simplify it since we focus on GLT25D1. If we decide to include LHs/PLODs in the manuscript, it may be worth talking about 2-oxoglutarate and succinate.
We thank the reviewer for his/her suggestion. As this manuscript is focusing on GLT25D1/COLGALT1, we have simplified the statement about the LH reaction (lines 52-56).

5. Page 4 line 71, please reference the work suggesting GLT25Ds and LHs form a complex.
We thank the reviewer for this suggestion. The reference has been added.

6. EDD motif has been found in other glycosyltransferases. It is not unprecedented.
We have modified the text and removed the "unprecedented" statement about the EDD motif.

7. The work showing the first Rossman fold domain cofactor binding stabilizes the folding is not conclusive. Do the results still hold if we use the second Rossman fold domain mutant and measure its thermostability with and with EDTA?
In the revised manuscript, we have included several new experiments supporting our description of the structural role for the GT1 domain of GLT25D1/COLGALT1. Using molecular dynamics simulations of GT1 and GT2 domains in the presence or in absence of metal ions and UDP-Gal. We have systematically observed increased flexibility and unfolding propensity in the GT1 structure without cofactors compared to the same structure bound to metal ion and UDP-Gal. Such stability alteration was not observed for GT2, further supporting structural roles for the former and functional roles for the latter. In addition, following the reviewer's insightful suggestion, we carried out thermostability assays using DSF for two GLT25D1/COLGALT1 mutants (i.e., Glu435Ala and Asp437Ala) bearing alterations associated with Mn^{2+} binding. There results are shown in Fig. 6 of our revised manuscript. The DSF

results suggested that the EDTA-induced destabilization does not depend on metal ion binding in the GT2 domain. Furthermore, DSF experiments carried out in the presence of Mn^{2+} showed that neither Glu435Ala nor Asp437Ala undergo stabilization by additional supplementation of metal ion, as it occurs instead for the wild-type enzyme.

8. Fig 3C, the structure does not match the ab-initio SAXS envelop well. Does the protein used for SAXS still have sumo tag?

We have realized that the original figure shown in Fig. 3C contained two superimposed models obtained by two distinct SAXS analyses that, due to the flexibility of the N- and C- termini of GLT25D1/COLGALT1, could result in possible confusion when compared. SAXS models of protein ensembles encompassing flexible segments, and in particular ab initio models such as the one displayed with the envelope in Fig. 3C of our original manuscript, are intrinsically ambiguous. We apologize for the confusion this could have generated. In the revised version of our manuscript, we have revised the graphical presentation of the SAXS results in the new Figure 2 and Supplementary Fig. 8, including consistent results from CORAL and CRYSOL in the revised Fig. 2 and the sphere model computed using the GASBOR/DAMAVER pipeline in the revised Supplementary Fig. 8, together with a size/shape comparison with the crystal structure of GLT25D1/COLGALT1 dimer.

9. Mutagenesis of Trp130/135 to show its relevance in GLT25D1 function?

Following the reviewer's suggestion, we have generated the human GLT25D1/COLGALT1 Trp135Arg mutant matching the mouse "fosse" mutant Trp130Arg but, similar to what we found for the other GT1 mutants involved in metal ion and/or binding of the UDP-α-Gal cofactor, we could not obtain samples suitable for subsequent analysis, likely due to the roles of Trp135 in keeping UDP-α-Gal inside the GT1 domain to preserve folding stability. We have added such description at lines 232-236 of the revised manuscript.

10. Fig 2, the contour level at 3.2 σ is high.

We wish to point out that the maps provided are F_o-F_c omit maps, which are typically represented at contour levels between 3.0 and 4.0 σ. We apologize for not having been sufficiently clear in the presentation of these data in our original version of the manuscript. We have now improved the description of these maps in the legend to Fig. 4 of our revised manuscript to avoid possible confusion.

Reviewer #3:

This manuscript deals with the molecular structure of an enzyme, GLT25D1/COLGALT1, to discuss the enzymatic mechanism. Various methods, X-ray diffraction, mass photometry, SEC-MALS and SEC-SAXS have been performed to reveal the dimeric quaternary structure. These are mostly thorough. Although the evaluated higher order structure is not directly related to the enzymatic activity, the current finding may be important to elucidate the mechanism to construct ECM in vivo. The following points can be considered before publication.

We would like to thank the reviewer for his/her comments on our work. In particular, we appreciated his/her comments regarding the thoroughness of our work and its possible relevance for the in vivo ECM community.

1. Both mass photometry and SEC-MALS results show some molar mass (or aggregation number)

dispersion for the analyzed samples. Authors can discuss the effect of dispersion on enzymatic and/or physiological activity. ”

We thank the reviewer for his/her comment. Results shown in Figure 2b, Figure 3b-c and Figure 6c highlight that the majority of the sample is in its dimeric form, with limited dispersity. In the revised discussion section of our manuscript (lines 452-457), we have commented about how this could suggest the need for further stabilizing homo- or hetero- protein-protein interactions with physiological relevance in vivo.

2. The above-mentioned dispersity may affect the analysis of the SAXS data. Some comments may be added to the main text. In this context, can the evaluated dimeric quaternary structure be described as rigid at room temperature? If not, this will also affect the scattering profile, and hence the resulting 3-d structure shown in Figure 3d. ”

Consistent with the mass photometry and SEC-MALS data shown in Figure 3, the SAXS data shown in Supplementary Fig. 8 show minimal dispersion, consistent with a stable dimeric ensemble. Regarding the question about “rigidity” at room temperature, given the flexibility associated in particular to the C- terminus of the enzyme (as highlighted by the CORAL modelling shown in Figure 2c), we would not consider GLT25D1/COLGALT1 as a rigid protein in solution, but rather as a rod-like dimeric enzyme complex with flexible loops. In agreement also with the suggestions from Reviewer 2, we have modified the presentation of the modelling results from SAXS data focusing more on the CORAL-based rigid body modelling of domains with addition of random loops (Figure 2c), which better represents our results, compared to the ab initio sphere models (now moved to Supplementary Fig. 8c) which can be characterized by ambiguity due to presence of flexible elements. We assume that the Reviewer was referring to the structural model originally shown in Fig. 3c and not 3d, as panel 3d included a visual comparison of projections obtained from crystal structure data and negative stain electron microscopy.

Literature references cited in this rebuttal:

Kellokumpu, S., et al. (2016). "Glycosyltransferase complexes in eukaryotes: long-known, prevalent but still unrecognized." Cell Mol Life Sci 73: 305-325.

Lairson, L. L., et al. (2008). "Glycosyltransferases: structures, functions, and mechanisms." Annu Rev Biochem 77: 521-555.

Perrin-Tricaud, C., et al. (2011). "Identification of domains and amino acids essential to the collagen galactosyltransferase activity of GLT25D1." PLoS One 6: e29390.

POINT BY POINT RESPONSE TO REVIEWER'S COMMENTS

Reviewer #1:

Firstly, congratulations to the authors on the revised manuscript - they really have made an excellent effort to address all of my points and those of the other reviewers. I think the manuscript is informative and will genuinely useful to the field of collagen biosynthesis. Also, I fully support keeping the SAD workflow in the manuscript, despite the success of Alpha fold. Regarding keeping 'enzymatic' in the title, this is up to the authors but it is redundant as a transferase in an enzyme by definition. Anyway, well done!

We would like to thank the reviewer for his/her positive feedback about our revised work. We sincerely appreciated his/her consideration for the additional efforts that we have included in the revised version.

Reviewer #2:

The work is solid and innovative. The reviewers have addressed most of the minor concerns; while the following major concern (What is the function of dimerization?) has not been addressed.

We would like to thank the reviewer for stimulating further efforts addressing the significance of GLT25D1/COLGALT1 dimerization. The newly revised version contains new cross-linking mass spectrometry experiments, showing that the enzyme's dimer interface, which does not seem to have direct implications in catalysis, is instrumental for interactions with LH3/PLOD3, ultimately demonstrating previously hypothesized direct molecular interactions between the two enzymes.

The evidence supporting GT2, not GT1, as a active domain is still somewhat weak.

All mutants designed to interfere with the bound metal ion and/or UDP-Gal binding in the GT1 domain resulted in no expression or in recombinant preparations characterized by folding instability, whereas individual point mutations in the GT2 domain result in well-folded but completely inactive enzyme variants. These data are consistent with previously published evidence highlighting that replacing the GT1 domain of GLT25D1/COLGALT1 with that of CERCAM/GLT25D3/COLGALT3 does not impact on enzymatic activity, whereas the opposite (i.e., replacing the GT2 domain) generates inactive chimeras. MD simulations corroborate the experimental data, showing that without metal ions and UDP-Gal cofactors the GT1 domain is strongly destabilized. In addition, our revised manuscript shows that the GT1 domain unexpectedly binds Ca^{2+} and not Mn^{2+}. This is now further demonstrated also by additional ICP-MS measurements. Furthermore, EDTA/EGTA treatment of the enzyme causes a 5 °C destabilization according to DSF. All experimental data, from our group and from others, converge towards the evidence of the GT2 domain being the active galactosyltransferase domain of GLT25D1/COLGALT1.

Moreover, nowadays, AlphaFold can predict the structure with good accuracy but can not predict dynamics or conformation changes. Thus the initial ab-initio SAXS data that did not agree well with the crystal structure may still be worth keeping. It is potentially meaningful and may add more to the manuscript if properly explained.

We wish to point out that we have not removed any of the SAXS analyses provided in our first submission, and that we fully agree with the reviewer about the potential meaningfulness of the models derived from solution SAXS data analysis, which may provide useful insights about molecular

flexibility. In our revision, we have only corrected what was previously displayed in an unclear manner, i.e. the ab-initio sphere models (supplementary figure 8c in current version) with the model based on the crystal structure (figure 2c). As stated in our previous reply to the same reviewer, "We have realized that the original figure shown in Fig. 3C contained two superimposed models obtained by two distinct SAXS analyses that, due to the flexibility of the N- and C- termini of GLT25D1/COLGALT1, could result in possible confusion when compared. SAXS models of protein ensembles encompassing flexible segments, and in particular ab initio models such as the one displayed with the envelope in Fig. 3C of our original manuscript, are intrinsically ambiguous." Such ambiguity is addressed by showing multiple models that fit the experimental data, thus highlighting the presence of flexible elements (in particular N- and C-terminal loops) that cannot be observed and unambiguously modelled from the crystal structure data.

Trp130/130 mutagenesis is surprising. Have we tried an alanine mutation?

The experimental results obtained on the GLT25D1/COLGALT Trp135Arg mutant are consistent with previously published data about the loss of protein expression associated with the mouse Trp130Arg fosse mutation (Geister et al., 2019, https://doi.org/10.1242/dmm.037176). Such result is also consistent with what we could observe when we attempted production and purification of other GT1 mutants involved in binding of metal ions and/or UDP-Gal, including Asp166Ala, Asp265Ala, Ile266Gln. These experiments were repeated multiple times and they all, consistently, resulted in no recombinant expression or very poor recombinant preparations due to unfolding/degradation. On the contrary, control mutants such as Ile267Gln (proximate to UDP-Gal in GT1, but not directly involved in cofactor binding), Trp158Arg (located at the GT1 homodimer interface but not directly involved in cofactor binding), as well as all other point mutants generated in the GT2 domain resulted in well folded and active enzyme preparations. In this respect, we do not find the Trp135Arg mutagenesis results surprising, but quite consistent with the evidence that metal ions and cofactors in the GT1 domain have structural, rather than functional roles. Considering the amount of experimental work presented in our revised manuscript, we refrain from generating another GT1 mutant that is extremely likely to result in no expression or unfolded enzyme preparations.

Reviewer #3:

The authors have adequately revised the manuscript. Therefore, the reviewer recommends publication.

We would like to thank the reviewer for his/her positive feedback about our revised work. We sincerely appreciated his/her consideration for the additional efforts that we have included in the revised version.

Report on N comms – 24 – 37599

Structural studies on collagen biosynthesis enzymes are both of considerable basic biological interest and are relevant to understanding the molecular basis of multiple diseases. The structural work on human galactosyl transferase described by De Marco et al is therefore timely and, I think, appropriate for publication in Nature Communications after appropriate modifications, including substantial new experiments.

The strengths of the work is the crystallographic analyses. The biochemical work is much less strong. At very least basic kinetic characterisation of the enzymes and key mutant needs to be reported, i.e. k_{cat} and K_M values. A second weakness is the characterisation of the interactions of the GT's with metal ions. Much more extensive screening with biologically relevant needs to be done; again, at least with basic kinetic characterisation. Does Mg^{2+} work? Note the electron density maps are not sufficient to define the metal ions, let alone the oxidation states.

Finally, after all the hard work to get the crystals, more extensive structurally informed bioinformatic structures need to be done. Now do these GT's vary across evolution, are they in plants, archaea etc.

So overall, I think this is a potentially very nice paper, but it needs much more biochemistry and the bigger biological/evolution picture needs to be included (well done on the structures!)

Title

Delete "Enzymatic" from the title

Main Text – specific points

Abstract - check grammar

"During collagen biosynthesis," "supramolecular"

L9 replace "integrative molecular" with "biophysical"

Studies"

L28 delete "for……"

L36 "which is essential"

L37 delete "conversely"

Introductions

L48 "; it is essential…"

L49 delete "Contrary to proline residues"

L74 "the subject of"

L76 "which are responsible"

L95 "have yielded"

L99 "wildtype" more correct than "native"

L103 "both of which support binding of the …"

P6 Refer to/give evidence of protein purity by SDS-PAGE and MS.

Supplementary Information

Fig S3 Please label secondary structures elements

Fig S7 maybe worth adding non-animal eukaryotes?

Fig S8 "of the ". Check grammar of this legend – purification doesn't result in expression

P7 Crystallisation

Give more details of optimisation procedure on screens used etc.

P12 As requested above give evidence of protein purity (SDS-PAGE/MS)

At least basic kinetic characterisation should be reported, i.e. Kcat 4, K_m values. Please screen other metal ions more extensively, e.g. Fe (II), Co (II), 2nC (II) Ca (II), etc. There does appear to be some activity with Mg (II)? Test for inhibition as well as activity

L291 "The experimental" to "A" crystal…."

Out of interest, presumably the authors tried to solve the structure by Alpha fold - and presumably it didn't work? Maybe connect. The structure similarity with LH3/PLOD is ??. This needs to be discussed in more detail. More extensive sequence alignments need to be given, including in relation to evolution of the GT domain across the animal kingdom and maybe beyond – what about plants?

L313 change subtitle – populated is not appropriate – "complexed"

L316 " unambiguously modelled by ?" this is very ? – firstly the oxidation state of the x-ray exposed crystals is unknown, secondly ? between Mn and (e.g.) as the basis of electron density alone is difficult/not possible? Please be more cautious.

L329 please justify the 'unusual' ?

L333 "cosubstrate"

L228 What about comparison width related GT domains, e.g. in PLOD?

L354 "was supported"

L367 can it be certain that dimers don't form during catalysis ?

L380 ICP MS/atomic absorption spectrometry need sto be done to assign the identity of metal ions. As stated above more careful/extensive studies on metal ions needs to be carried out.

L408 "unusual" more correct than "unprecedented" – is where the statement needs to be justified.

L423 are there any clinically relevant mutations?

L435 delete "the"

L442 "reveal an"

L446 replace "populated"

Discussion, please modify appropriately in the light of the requested new experiments/suggested