

An in vitro approach for simulating divergent Golgi O-glycosylation of tumour-associated MUC1 from normal MUC1

Corresponding Author: Professor Kevin Naidoo

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)

"An in vitro one-pot synthetic biology approach to simulating diverging Golgi O-glycosylation of tumor-associated MUC1 from normal tissue MUC1" by Abdullateef Nashed and coauthors is a very interesting paper in the field of protein O-glycosylation. They develop an assay to study the effect of the order in which glycosyltransferases act, mimicking spatial organization of the Golgi. As GALNTs are known to migrate from the cis Golgi to the ER in tumors, they apply GALNTs before other GTs leads to higher site occupancy and shortened glycans, similarly to what is seen in cancer tissues. Their assay also allows for investigation of lectin dependent versus lectin independent mechanisms, exact substrate specificity of GTs. Finally, they model the sialylation reaction by ST6GALNAC1 to rationalize site specificity.

The paper is well-written and easy to follow, altogether an informative and enjoyable read that definitely add to the understanding of the protein O-glycosylation field, and provide an assay that seems quite straightforward and practical to use.

However, there are some issues relative to the peptide design, the study of the lectin-dependent mechanism, and the QM/MM modelling and the ensuing rationale.

Peptide design. To me the biggest limitation is to have limited the analysis to a single repeat unit of MUC-1. While the design of the 23 and 27-mer can a little bit differentiate between the effects of intra and inter repeats, it is limited. The authors just state that the optimal sequence should only have each glycosylation site once, without justifying it. They conclude by stating that the commonly used TAP24MUC1 is unsuitable, using the principle that they propose without justifying them. They also only have this discussion in the methods section. Is it impossible that the glycosylation on one site affects the same site on unit $n+1$ or $n-1$? The kinetics with their UGS assay should be all doable with several repeats without having to adjust it, and it should be doable to express their constructs with several repeats to see if there are any differences. Did the authors attempted that, or why did they not do it? They should at least provide some rationale for why it is fine to study single repeat units rather than the full repeat, and provide it early in the manuscript, when they explain the design of their assay.

Lectin dependent mechanism. To further confirm which glycosylation are lectin-dependent or not, why not express only the catalytic domain? Or at least a single point mutant of the lectin binding site? If it is known it cannot be done, the authors should explain why.

QM/MM simulations. The first two glycosylation reactions show activation energy of 25 kcal/mol, which means that there is basically no reaction. The authors justify an observed difference of less than 2-fold in rate by modelling showing a 7.5 kcal/mol difference, which amount for a 100 000 fold difference in rate. They justify the slow rate by some far-fetched mechanism in which the catalytic base would actually be covalently bond to the sugar donor, making it a suicide inhibitor. Worse, they actually do not show anything of their modelling. There is not a single figure in the SI, and in the main text only one figure that does not show any conformation, distance, etc. The authors should at least provide some of those. As well, they should tune down some interpretation on their simulations given the results they obtain. They could also remove the two simulations on T20 glycosylation and only keep the one on T23, which provides a reasonable energy barrier, and expand its description.

Also, the SI needs to be formatted. There are not even supplementary figure numbers.

Minor, throughout the text.

There are also formatting issues:

L13 O-GalNAc, Galnac "O" is positional and should be italicized, L47 N-acetyl "N" is positional and should be italicized. L94 N-acetylglucosaminylation, L103 O-GalNAc, Other occurrences.

Numbers for stoichiometry are subscript, e.g. MnCl₂ / MgCl₂ L371.

"Core" is written either as "core" or Core", chose one and be consistent. Same thing for "one-pot" and "One-pot",

Serine/serine, Threonine/threonine, etc.

"SN1" the "N" is subscript.

Km. K is italicized, it is a thermodynamic constant, not a temperature unit. m should be subscript.

Latin is italicized, be it for species name (*E. coli* – also it is not *E. Coli*, tobacco etch virus), in vitro/vivo, e.g./i.e/etc.

Ranges and minus are en-dash, not hyphen.

L320 β -D-thiogalactoside "D" is small caps

Minor, along the text.

L16 NAc, not nac.

L22, 202, 270 Sialyate sialylate

L26-27 "This systems modelling unpacks the molecular details of enzyme localization and substrate glycan occupancy that is fundamental to the regulatory mechanisms that gives rise to tumor-associated MUC1 antigens."
Reformulate.

Figure 1 Move the SNGF labels to separate better A and B, indicate more clearly the differences between healthy and tumor in B. Ideally also show the glycan on the peptide along the arrow, not just unglycosylated and endpoint. The authors have space on the width as they went double column.

Figure 1 caption, "between normal epithelial tumor epithelial cells attributed to COP-1"

At least a word is missing.

L59-60: "such as the relocation of GALNTs"

Could the authors expand here? It is a major point behind the rationale of their modelling. Are all GALNTs moving equally?

Are all of the other GTs at the same place? In Fig 1B there appear to be more sialyltransferases on tumor than healthy, is it by design?

L65 "parameters governing glycosyltransferase (GT) activities and substrate specificities using the UGC assay¹⁴"

Ref 14 is a 2024 original paper by the authors, maybe they could expand here and explain what is the UGC assay, as the assay was proposed less than a year ago it is not necessarily common knowledge. Just saying that it detects nucleoside mono- and di- phosphate formation through nucleotide kinases and NADH consumption, etc.

Fig 2C is not clear enough as it describes the process for the article. I imagine the GTs are not his tagged, the TEV is, and cleavage happens on bead? Conversely, the authors can forego the part when they represent heterologous protein expression (top of Fig 2C), readers could be expected to understand that.

L80" sfGFP was selected for its folding efficiency, minimized dimerization, and enhanced solubility."

Precise that GFP is not used not for its fluorescence properties. When presented with an essay using GFP, readers may assume that fluorescence is a thing - and it is actually for the UGC assay.

L81-82 "with the MUC1 peptide, sfGFP was separated from the peptide using a linker (Linker 2) comprising of three rigid and three flexible units (from the N to C direction). Additionally, a rigid linker (Linker 1) was incorporated to improve steric presentation to the N-terminus His"

Linker 1 and 2 are not shown on Fig 1A, and I do not see three rigid and three flexible units as linker 2 in Fig 1A.

L129 "LC-MS results confirmed the structure and purity of the peptide (i)."

A reference to the appropriate supplementary figure – which should be indexed – has to be included here.

L145, "over expression", in one word.

Fig 3 overall separate the panels better., eg. The representation of the enzymes and their domains appear to be in C and not in B, etc..

Fig 3A, no need to write "no reaction", that is visible. Do not represent the reaction with the 23 mer and GALNT2+GALNT4

differently if that is not commented on.

Fig 3B is too cramped, it makes it hard to read.

Fig 3C I The “tag removal” to replace TEV+IMAC does not help, just make the figure slightly confusing. What would be indicating whether the product is after elution or the flow through. I assume the first ones with black arrows are always after elution, while the last ones in light brown the flow through is collect. Whether I am right or wrong in assuming so, it should be specified in the caption. Or just do like in Fig 4, more consistent and clearer.

Fig 3C III Remove the arrows and align the glycosylated peptide products. It would be consistent with Fig 4, which is much clearer in this regard despite dealing with more complicated structures.

Fig 3C If it has been done (I assume so), it lacks a line with GALNT4 alone yielding non-glycosylated peptide i)

L161 “Within this context the core 3 is synthesized by B3GNT6 catalyses of the GlcNAc covalent β 1-3 bond of to the Tn antigen.”

Reformulate

Fig 4A, Vmax as a value, kcat could be included, and kcat/Km.

Fig 4BI, first chromatogram shows at least two peaks for peptide labelled (v), may the authors comment on that?

L 171 “indicates that the enzyme has a lower affinity (Km value of 0.114 mM for GalNAc5-23mer vs 0.062 mM for GalNAc3-23mer) but higher turnover (Vmax for GalNAc5-23mer is 180 % of the Vmax of GalNAc3-23mer) of the fully glycosylated MUC1 compared with the semi glycosylated MUC1 (table in Figure 4 A). On the other hand, no significant difference in the Km values of C1GALT1 were observed between the semi-saturated and completely saturated MUC1. This indicates that C1GALT1 prefers glycosylation of the GALNT_D-sites over the GALNT_L-sites at any concentration of MUC1 and ST6GALNAC1 prefers glycosylation of the GALNT_L-sites over the GALNT_D-sites at high concentrations of MUC1.” Actually in both systems in Table 4A it shows that both peptides show the same Vmax over Km. As I assume natural concentrations to be far below Km, it means no difference in activity? Or at least no site preference for both enzymes in dilute conditions. Anyway it should be commented.

L177 The authors may add a reference for the use of asialofetuin as a model protein for 2,6 sialylation, eg.

<https://academic.oup.com/glycob/article/29/10/735/5528790>

L200 overcomes a second transition state 2 (TS2) barrier of 24.82 kcal/mol.

L206 a transition state 2 (TS2) energy barrier of 17.46 kcal/mol.

That is a difference of 7.36 kcal/mol, ie. c. factor 100 000 in rates. According to table if figure 4A, the difference is not even of a factor 2, nowhere close to 100 0000. The computations are interesting, but maybe the authors should modulate their interpretation on the relation between experimental rates and computed energy barriers. Their computational method is not accurate enough. On the other hand, 17.46 kcal/mol seems perfectly on line with what is expected for a GT reaction.

Fig 5B alone is a bit poor to see the interaction. I would like to see zoom-in of all the states schematized in Fig 5D for each system in a (or several) supplementary figure. Would also help to follow the discussion on molecular rationale.

Fig 5C, the x-axis is unclear. It reads a bit like the distance of the bond to be formed (O6-C1) minus the distance of the bond to be cleave (UDP-sugar). In that cases the signs seem wrong? Anyway has to be made clearer.

Fig 5C, the first two energy profiles would be indicative of a reaction not occurring (activation energy above 25 kcal/mol). I am also mildly surprised to see such a large difference in the energies of the products (I and II vs III). Could the authors comment on it?

Fig 5 period is missing at the end of the caption. Actually, it looks like more than that s missing, with the rest of the color coding?

L196 The sialylation of the T20 residue on partially glycosylated peptides (I) exhibited an energetic profile that is distinct from its fully glycosylated (II) counterpart (Figure 5 C).

I do not really see a difference, could the authors explain what they mean by “distinct”? Also given the errors, I am not sure this distinction would be meaningful. I would replace “distinct” by “nearly identical”.

L208 “The molecular reasons for this are that the sugar conformation necessary for the nucleophilic attack is critical to the sialylation reaction. If the sugar conformation is not within the near attack conformation cone angle, the histidine may covalently bind the anomeric carbon after the disassociation of the phosphate.”

On one hand, as enzyme and donor are identical for all three reactions it is expected that the nucleophile should be the difference. Yet it appears the biggest difference between the profiles I / II and III in Fig 5C to be the formation of the TS1 and the OC intermediate. At this stage, the nucleophilic sugar should not have such a drastic importance as they claim it to be SN1 like? It may be commented a bit more here?

L 210 “This is the scenario in the T20 I and II T20 case where a side product forms when the catalytic histidine (HIS657) is covalently bonded to the anomeric carbon of the sialic acid lead.”

Is that something the authors saw experimentally? It would be surprising, the donor substrate acting as suicide inhibitor of

the enzyme. Did the authors observe a loss of activity upon incubation with the donor, or a mass increase? If it is only observed during simulations and no experimental clue, the authors may suspect it to be a computational artefact, or justify why they imagine it is not so. Again here, having some snapshots of the modelled reactions would be useful.

L 226 C1GALT1 competes with these enzymes' lectin domains for the modified sites. This competition results in the synthesis of Core 1 structures and inhibits subsequent GalNAc glycosylation at other sites via the lectin-dependent mechanism of GALNTs.

That is pretty interesting, do the authors have some data (possibly from literature) on the affinity of the lectins of GALNT1 and GALNT2 for the Tn and just GalNAc?

L287 "Figure 1C"

There is no Fig 1C. Fig 2C?

L289 "AlphaFold structure prediction tools"

Version of AlphaFold and corresponding reference should be added.

L294, "(i) glycosylation of the naked peptide using only the catalytic domain, (ii) glycosylation of pre-glycosylated peptides through the lectin domain, and (ii) a combination of the two mechanisms occurring sequentially."

The second "ii" should be "iii".

L304 The optimal sequence must be inclusive of all the variables that determine GALNT specificity, and each variable must only be represented once in the peptide sequence.

Why only once? What if a glycosylation at S19 influence the glycosylation of S19 on the next repeat? I understand that this made modelling easier, but.

L358 "concertation".

"concentration"?

L410 The QM region comprised HIS567 (catalytic base) sidechain atoms, HIS552 (stabilizing the CMP phosphate group), the Neu5Ac moiety, and the 12 CMP phosphate group. The nonpolar C-C bonds were treated with hydrogen link atoms, and the polar C-O bonds were treated with the Simple Link Atom Saccharide Hybrid (SLASH) 41 method.

Aren't the authors missing the nucleophile? In Fig 5D the GalNAc was included in the QM region including the glycosidic O1 and is actually the only bond treated with SLASH (which seems sound).

Ref 12 format to be modified.

After all these sometimes-critical comments, I want to reemphasize that I enjoyed reading this manuscript, which I found informative and to constitute a sizeable addition to the field. I think the overall assay to be cleverly design, and that important conclusions are drawn from the results. I want to congratulate all the authors for this very good piece of work.

Reviewer #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3

(Remarks to the Author)

REVIEWER COMMENTS

This manuscript presents an innovative one-pot synthetic biology strategy to reconstitute and dissect the complex glycosylation pathways of MUC1, with a particular focus on enzyme specificity and kinetics involved in the generation of tumor-associated glycoforms such as Tn, T, and their sialylated derivatives. The study provides valuable insights into how enzyme localization (ER vs. Golgi) and competitive activities among GALNT, C1GALT1, and ST6GALNAC1 govern MUC1

glycoform heterogeneity. Furthermore, the integration of QM/MM simulations with experimental kinetic data reveals site-selective sialylation of T13 and T20 residues by ST6GALNAC1 under high MUC1 concentrations. This offers a novel mechanistic perspective on the glycosylation heterogeneity of MUC1 in cancer, with potential implications for designing MUC1-targeted cancer immunotherapies.

The manuscript meets the high standards of novelty, significance, and scientific rigor expected by Nature Communications. However, prior to publication, the authors should consider the following points to further strengthen their work:

Major Comments:

1. Energetic Profiles and Reaction Feasibility:

The computed product states (P) for reactions I and II are significantly higher in energy than the corresponding reactant complexes (RC), by approximately 20 kcal/mol. Similarly, in reaction III, the product remains more than 10 kcal/mol above the RC (Figure 5C), suggesting these transformations are thermodynamically unfavorable (i.e., endothermic). Moreover, the activation barriers for TS2 in reactions I and II (24.82–25.21 kcal/mol) exceed typical enzymatic reaction barriers (<21 kcal/mol). While the authors mention "relatively slow reactivity" in the experimental data, a quantitative comparison between the computed activation energies and the experimental rate constants is needed to confirm whether the predicted energetics are consistent with the observed kinetics.

These findings raise concerns about the reliability of the computational model. The authors are encouraged to address the following:

Model Quality: While ST6GALNAC1 was modeled using SWISS-MODEL, the authors should provide quality metrics (e.g., GMQE, QMEAN scores, Ramachandran plot) and specify the sequence identity with the template ST6GALNAC2 (PDB ID: 6APL). If the identity is below 30%, the model's reliability may be limited. A multiple sequence alignment could be added to support this.

Comparison with AlphaFold3: Have the authors compared the SWISS-MODEL structure with a model predicted by AlphaFold3? If so, please report RMSD values and discuss any structural differences, particularly within the active site.

2. Docking Procedure and Pose Validation:

While selecting the lowest-energy structure from the top HADDOCK clusters is a valid approach, further validation is needed for catalytically relevant poses. Do the selected conformations meet the geometric criteria for nucleophilic attack (e.g., distance and angle)? The authors should explicitly state whether the poses are catalytically competent.

3. QM/MM Methodology:

The use of DFTB3 as a semi-empirical QM method raises concerns about its ability to accurately capture transition state energetics for GTs reactions. Please cite relevant literature supporting the applicability of DFTB3 in modeling this class of enzymatic mechanisms.

Given the relatively high energies of the product states, the authors should discuss whether any potentially stabilizing interactions (e.g., hydrogen bonding, π -stacking, ionic interactions) were omitted or not captured in the model. Additionally:

Clarify the selection criteria for representative MD snapshots used in QM/MM calculations.

Confirm whether interactions with key catalytic residues are preserved in these snapshots.

Address the influence of sugar ring conformations on reactivity. Have conformational analyses been performed to ensure ring stability?

4. Catalytic Mechanism Interpretation :

The manuscript proposes a covalent adduct formation between the catalytic histidine and the anomeric carbon during side product formation, diverging from the widely accepted role of histidine as a general base in GTs. Without direct experimental evidence (e.g., mass spectrometry), this interpretation remains speculative. The authors should either provide supporting evidence or clearly state that this mechanistic proposal is hypothetical and subject to further investigation or simply not discuss this.

5. Figure Organization and Molecular Detail:

Figure 5 could benefit from improved organization:

Consider placing the general catalytic mechanism (currently panel D) as panel A, followed by the free energy profiles (panel C).

Previous panels A and B appear redundant; one could be moved to the Supplementary Information.

Additional molecular graphics showing close-up views of the active site would enhance mechanistic clarity. These figures should include annotated catalytic residues and key distances in the RC and TS structures.

6. T13 vs. T20 Sialylation Specificity:

Can the authors provide more mechanistic detail on the observed specificity?

At the QM/MM level, do differences in anomeric carbon charge or sugar ring conformations between T13 and T20 explain the preferential reactivity?

At the MD level, the discussion regarding the proportion of the near attack conformation (NAC) is particularly insightful. A statistical analysis comparing nucleophile–anomeric carbon distances and angles at T13 and T20 would provide quantitative support for this observation.

Minor Comments:

1. Line 15 – Abbreviation Definition:

On first use, spell out all abbreviations. For example, “GALNT” should be introduced as “N-acetyl-galactosaminyltransferase (GALNT).”

2. Line 16 – Consistency in Terminology:

Ensure correct and consistent capitalization of scientific terms such as “GalNAc.” Please correct all instances of “Galnac” to “GalNAc.”

Reviewer #4

(Remarks to the Author)

Key Results

The authors developed an in vitro synthetic biology approach to investigate how the presence and localization of glycosyltransferase (GT) enzymes across the endoplasmic reticulum (ER)–Golgi system influence the glycosylation patterns of proteins. MUC1 was used as a model protein; however, the approach can be readily adapted to other O-glycoproteins. Given that GT relocation during malignant processes is well documented in the scientific literature, this method also enables the elucidation of mechanisms underlying the formation of cancer-associated glycan structures.

Validity

The fundamental concept of the study is questionable, as the authors claim to have developed a one-pot method, which implies that the entire reaction takes place in a single vessel and that there is no need to isolate or purify the products between steps. The presented workflow does not meet these criteria, especially when the GTs are applied sequentially. The structural interpretation of the synthetic products (glycoforms) relies heavily on the results of the HPLC-MS measurements. However, the data provided in the supplementary information do not sufficiently support the structures presented in the manuscript, rendering the findings speculative and casting serious doubt on the conclusions. Below, we outline the main shortcomings in bullet points.

- The mass accuracy of the peptides and glycopeptides is not provided, and in most cases, it cannot be calculated based on the data given.
- Figures 3 and 4 depict a total of eight structures (i–viii); however, tandem mass spectra supporting the proposed structures are provided only for the naked peptide (i) and a single glycopeptide (ii).
- In some cases, peak shouldering can be observed in the chromatograms (see line 69, peak 1 in the supplementary information). This may indicate the separation of isomers, suggesting that more products were formed than reported. In other instances (see line 71, peak 1 in the supplementary information), several peaks with higher intensity than the identified component are present, for which no explanation or characterization is provided.

Significance

Considering that carbohydrates remain among the least understood contributors to cancer, the conclusions of this work could significantly advance our understanding of the O-glycosylation process and the formation of cancer-associated glycan structures. Moreover, the approach presented here is thoughtfully designed and could be adopted by many research groups for the in vitro simulation of the glycosylation machinery in the ER–Golgi system.

Data and methodology

Aside from the execution of the HPLC-MS measurements, the approach appears valid, and the data quality is good. The presentation is clear and well-organized. Please note that the computational modeling is outside the reviewers' expertise.

Analytical approach

Apart from the HPLC-MS measurements, from which the authors drew overly far-reaching conclusions, the analytical approach is generally sound, and the statistical tests are appropriate. Please note that the computational modeling is outside the reviewers' expertise.

Suggested improvements

The reviewers suggest several substantial improvements before publication, mainly based on the comments of the Validity

section of the review report.

1. The mass accuracy of all proposed peptides/glycopeptides must be reported (in ppm), and the data should allow readers to verify this by providing both the theoretical exact mass and the observed mass to at least four decimal places. Raw data files should be uploaded to a public repository (e.g. PRIDE, Massive).
2. The structures of all reported glycopeptides, including sialylated glycoforms, must be validated using tandem MS analyses.
3. Peak shouldering should be investigated, and the major chromatographic peaks must be accounted for in the base peak chromatogram to ensure that the results adequately support the conclusions. For these first three points, the reviewers recommend a workflow that includes Electron Transfer Dissociation (ETD) for O-glycan site localization, Higher-energy C-trap Dissociation (HCD) for glycan structure elucidation, and a database search (e.g., Byonic, MSFragger, or pGlyco) to efficiently identify both the main components and potential side products.
4. The Roepstorff–Fohlman nomenclature should be used for naming fragment ions, as it is the most widely accepted system in peptide mass spectrometry.
5. Some details of the HPLC-MS measurements necessary for reproducing the experiments are missing, including injection volume, trapping method parameters, and the column thermostat temperature.
6. In lines 36–38 of the Supporting Information, the authors state that measurements were performed using either a Q Exactive quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) or a SYNAPT G2 mass spectrometer (Waters, USA); however, no experimental details are provided for the latter instrument.
7. The SDS-PAGE workflow should be presented in the Methods section or in the Supporting Information.
8. The expression “one-pot” should be removed from the title and the text.

Clarity and context

Overall, the reviewers found the text clear and accessible, and the figures help to clarify the approach and workflow. However, only few details are provided on how the results align with or contradict current knowledge of the O-glycosylation machinery.

References

The manuscript references the relevant literature appropriately. Key studies are cited to provide context for the current research, and the references support the rationale and methodology presented.

Your expertise

The reviewers state that the computational modeling presented in the manuscript falls outside their area of expertise.

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Note: as colour markings are not visible in the provided text box, please refer to the enclosed word document “Review2.0.docx”. Beside this note and colour markings, it contains the exact same information.

“An in vitro one-pot synthetic biology approach to simulating diverging Golgi O-glycosylation of tumor-associated MUC1 from normal tissue MUC1” by Abdullateef Nashed and coauthors is a very interesting paper in the field of protein O-glycosylation. The authors addressed well the various reviewer’s requests. Below are a few comments in green (black for previous review, blue and red for authors responses), mostly on the representation of the QM/MM data.

[...]

QM/MM simulations. The first two glycosylation reactions show activation energy of 25 kcal/mol, which means that there is basically no reaction. The authors justify an observed difference of less than 2-fold in rate by modelling showing a 7.5 kcal/mol difference, which amount for a 100 000 fold difference in rate. They justify the slow rate by some far-fetched mechanism in which the catalytic base would actually be covalently bond to the sugar donor, making it a suicide inhibitor. Worse, they actually do not show anything of their modelling. There is not a single figure in the SI, and in the main text only one figure that does not show any conformation, distance, etc. The authors should at least provide some of those. As well, they should tune down some interpretation on their simulations given the results they obtain. They could also remove the two simulations on T20 glycosylation and only keep the one on T23, which provides a reasonable energy barrier, and expand its description.

AUTHOR RESPONSE:

We agree with reviewer's comments regarding the unrealistic high activation energies. We therefore repeated the QM/MM study focusing only on the cancer-associated fully glycosylated MUC1. Reaction dynamics calculations were performed with increased sampling and corrected starting structure to obtain reasonable barrier heights for the glycosylation reactions. After these improvements have been made, we report activation energies of 17.11 kcal/mol for the T20 site, 23.87 kcal/mol for the T8 site, and 16.73 kcal/mol for the T13 site sialylation reactions, as seen in Figure 5C.

Necessary figures have been included in Figure 5 and SI. There is no need for chemical detail as this is not a manuscript that describes chemical mechanistic differences between sites. The function the simulations was to compare and determine the viability of ST6GALNAC1 catalysis at each of the 5 sites. The comparison is made through simple docking experiments, which are limited on their own, as well as sophisticated computational modelling to determine barrier heights. While we have in previous works (e.g., <https://pubs.acs.org/doi/abs/10.1021/acs.accounts.1c00477>) performed in depth electronic and atomic level analysis in pursuit of chemical mechanistic analysis this would be a distraction from the objective to build and demonstrate in vitro molecular models of normal vs. tumour golgi construction.

In line with the reviewer's concern, we have provided additional figures in the Supporting Information Figure S21, S22, and S23. Close up images of important residues involved in the reaction and relevant distances and angles across the QM/MM simulation are provided in Figure 5B Additional. The reported distances and angles are discussed in Section 2.3.1 to clearly illustrate the mechanistic features observed in our simulations.

It is a fantastic result to have managed to refine the energy barriers to feasible ones, and we would like to congratulate the authors for doing so. However, "no chemical detail" is definitely not satisfactory. Please add some clear and readable snapshot along the QM/MM simulation, at least for each trajectory with a reported barrier of initial state, transition state, and product.

I understand that this is not the focus of the authors, but they are currently reporting trajectories and energy barriers that we cannot evaluate. Those can easily be provided in 1-3 additional SI figures.

The authors added some information and lengthy discussion on their very low-level calculations (protein model generation and docking), but absolutely nothing on the high-end computations they did. I would also propose that the authors upload their full trajectories on a public repository, e.g. Zenodo, with a doi provided in the "data availability section".

Table S2 is also not satisfactory: "Table S2. Distances between centers of masses of the possible glycosylation sites and key residues involved in the reaction 252 i.e. the CMP-Neu5Ac donor and catalytic base HIS567."

"Centers of mass" is not an appropriate measure point. You would get a lower distance if the donor was in a totally not reactive position, but with the phosphate close to the base. Also, present in the caption that those measures are for the TS*. If a distance should be measured between the catalytic base and the Neu5Ac donor, I suggest between the reactive nitrogen of His567 and the C1 of the donor. Ideally, I would also like to see the distance between the reactive nitrogen of His567 and the proton to be abstracted on the Ser/Thr.

That being said, we would still like to stress that it is only a minor disagreement on how to present the data, and we would like our main message as an answer to the QM/MM part to be commending the authors for a great revision and refinement of their work.

[...]

Fig 4A, Vmax as a value, kcat could be included, and kcat/Km.

AUTHOR RESPONSE: We thank the reviewer for pointing this out. In this case, the substrate peptides contain multiple reactive sites, making it impossible to attribute the kinetic parameters to a single reaction per molecule. Therefore, kcat is not meaningful in this context, as the turnover number cannot be linked to a defined thermodynamic step. Instead, we report apparent Km and Vmax values, which serve as averaged proxies for substrate affinity and maximal reaction rate, respectively. This provides a more appropriate comparative measure across enzymes and substrates. We have clarified this in the caption of Fig. 4A

There is no logic in making this argument for kcat and not for Km or Vmax. Some argue that any kcat is an averaged result of all potential substrates. Substrate heterogeneity by itself does influence the kinetic profile but does not precludes to obtain an analytical solution to the Michaelis Menten equation, see:

<https://www.biorxiv.org/content/biorxiv/early/2025/08/24/2025.08.21.671619.full.pdf>

This is a very minor point though.

Additional points

The mass spectrometry nomenclature is quite confusing, as the authors move between "MS² analysis" (SI 114 and 116), MS2 (main text 1508), "LC-MS" or "LC-MS" (please chose one, better "LC-MS", e.g. Figs S4, S5 and S11, S13, S14, S15, S16, S17); "MS1" (Figure S6), also MS1 (Figure 6A, 7A, line 508), MS 1, M1, etc. Please homogenise and check all instances thoroughly.

L162: "However, the observed glycosylation of the remaining two sites cannot be attributed", the discussion would be easier to follow if the sites were immediately specified, e.g. "However, the observed glycosylation of the remaining two sites (S9

and T13) cannot be attributed”.

The following request has been ignored by the authors, or did we miss the response?

“Lectin dependent mechanism. To further confirm which glycosylation are lectin-dependent or not, why not express only the catalytic domain? Or at least a single point mutant of the lectin binding site? If it is known it cannot be done, the authors should explain why.”

We would also like to suggest that the authors group figure 6A with 7A in figure 6, and to group figure 6B and 7B in figure 7. Having the site-to-site comparison of the localization of GALANT's stresses the point and makes the message clearer for the reader.

Please make sure, that the figures are exported in high quality, as the details in some of the figures are either not visible or difficult to see. This issue concerns the text in figure 3B, the charges in figure 5A, and S' and T's in figure 7B.

Reviewer #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3

(Remarks to the Author)

The authors have addressed most of the issues raised in my previous review, and the revised manuscript is significantly improved. However, several aspects related to the catalytic mechanism, docking analysis, and data presentation still require clarification before the manuscript can be accepted. My comments are summarized below.

1. Catalytic mechanism and QM/MM presentation

1.1 Nomenclature and atom assignment. Histidine should be abbreviated as His (not HIS) in the main text and figures. The mechanistic description must also confirm whether the reactive atom of the GalNAc acceptor is O6 rather than C6. In ST6GALNAC1-catalyzed sialylation, the nucleophile is generally the O6 atom attacking the donor anomeric carbon (C2). Distances involving C6 are chemically inconsistent and suggest an incorrect reaction coordinate.

1.2 Missing catalytic atom labels in Figure 5B. Key atoms such as the donor C2, the acceptor O6, and the NE2 atom of His567 are not labeled. Proper labeling is necessary for readers to follow the bond-breaking and bond-forming steps.

1.3 Incomplete reaction pathway in Figure 5B. Only reactant snapshots are shown for the T20, T8, and T13 systems. Representative structures of the TS and Product are needed and should be provided either in the figure or Supplementary Information.

1.4 Figure clarity and visual quality. In Figure 5B, the distance indicators shown as dashed lines are too thin to be legible and should be drawn thicker. All relevant interatomic distances should be clearly marked. In Figure 5A, certain structural lines, particularly the leftmost GalNAc 4-OH group, are too faint and should be redrawn to improve visibility.

1.5 Data availability and reproducibility. To allow structural validation, the Cartesian coordinates of the QM region for the Reactant, TS, and Product structures should be provided in the Supplementary Information.

2. Molecular Docking Analysis

2.1 Unrealistic distances and catalytic relevance. COM distances in Table S2 (13.66 Å T13 to donor; 9.72 Å T13 to His567) indicate significant separation, whereas reactive atoms should be within ~3–4 Å. These data do not support the statement that T13 is “orientated favourably to react.” To support this claim, the authors should either provide close-up views of the relevant docking poses showing Thr side chain orientation, the catalytic pocket, and atom-to-atom distances, or clearly qualify the text to indicate that the docking only suggests a potential binding region, not a catalytic mechanism.

2.2 Figure S23 caption and panel discrepancies. The caption for Panel III describes T20, but the image corresponds to T13. The main text also references Figure S23 IV, which does not exist. The caption should be corrected, and the reference to the nonexistent panel removed.

Reviewer #4

(Remarks to the Author)

The reviewers suggest improvements before publication, mainly related to the LC-MS data confirming the synthetic results. This is of critical importance considering that LC-MS was the only analytical tool to validate their computational and synthetic results.

1. In comment #2 of our review, tandem MS experiments were requested to confirm the structures of the proposed glycopeptides. In the revised version of the supporting materials, the authors state: "The identity of the target peptide peaks was confirmed through MS² analysis (not shown), in which the corresponding precursor ions exhibited characteristic fragment ions, including b- and y-ion series and diagnostic glycan oxonium ions." However, in their reply, the authors say: "MS² spectra were not included, and only primary ions were used in the analysis. However, the site-specific digestion with IMPa used with the synthesis carried out on the double-repeat MUC1 (added to this revised version) improved the site occupancy analysis to a level sufficient to draw the conclusions made in the manuscript." While the use of IMPa digestion and subsequent LC-MS analysis indeed improves the reliability of the results, some of the conclusions regarding the glycopeptide structures remain overstated. For example, IMPa digestion apparently generated species with identical sequences (e.g., see the M1 sequences in Figures S15–20), which raises the question: how can the authors be certain that both of potential glycosylation sites are occupied?

2. Although it has been advised to use the Roepstorff–Fohlman nomenclature for naming fragment ions (b- and y-ions) to improve comprehensibility, it has not been performed on Figure S11.

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This is now the third time this manuscript has been presented to review. The authors seem to largely ignore comments that have been written twice to them. Sadly enough, this review round consists of just repeating previously said comments in an unnecessary harsher tone.

Overall, the QM/MM reporting is as unsatisfactory as previously.

It is puzzling why the authors refuse to report the data they already acquired, at least in SI. As is, it is deeply unsatisfactory. It is not informative enough and can hardly be evaluated. There is only Fig 5, in which the caption does not fully inform on what is displayed, and S24.

As a bare minimum, the authors should report on all the key distances and angles, for each trajectory, at each of the key points (reactants, transition states, intermediates, products). They should also display readable snapshot figures of each of said key points.

Below are in green previous comments, and in blue the authors answer to those. Black is for the new answer (colours are visible only on the enclosed word document).

It is a fantastic result to have managed to refine the energy barriers to feasible ones, and we would like to congratulate the authors for doing so. However, "no chemical detail" is definitely not satisfactory. Please add some clear and readable snapshot along the QM/MM simulation, at least for each trajectory with a reported barrier of initial state, transition state, and product.

I understand that this is not the focus of the authors, but they are currently reporting trajectories and energy barriers that we cannot evaluate. Those can easily be provided in 1-3 additional SI figures.

The authors added some information and lengthy discussion on their very low-level calculations (protein model generation and docking), but absolutely nothing on the high-end computations they did. I would also propose that the authors upload their full trajectories on a public repository, e.g. Zenodo, with a doi provided in the "data availability section".

AUTHOR RESPONSE:

Initial reactant coordinates and an exported trajectory video of the product formation are available on Github (<https://github.com/scientificcomputing/NCOMMS-25-06286-T-MUC1-in-vitro-Synthesis-tree/main>).

In the repository, there are a couple of movies I cannot read, and a few files without any guidance / information. I did not find any actual coordinates.

Table S2 is also not satisfactory: "Table S2. Distances between centers of masses of the possible glycosylation sites and key residues involved in the reaction 252 i.e. the CMP-Neu5Ac donor and catalytic base HIS567." "Centers of mass" is not an appropriate measure point. You would get a lower distance if the donor was in a totally not reactive position, but with the phosphate close to the base. Also, present in the caption that those measures are for the TS*. If a distance should be measured between the catalytic base and the Neu5Ac donor, I suggest between the reactive nitrogen of His567 and the C1 of the donor. Ideally, I would also like to see the distance between the reactive nitrogen of His567 and the proton to be

abstracted on the Ser/Thr.

That being said, we would still like to stress that it is only a minor disagreement on how to present the data, and we would like our main message as an answer to the QM/MM part to be commending the authors for a great revision and refinement of their work.

AUTHOR RESPONSE:

Table S2 does not discuss any distances related to the TS as the GalNAc residues are not considered. The protein-protein docking performed is meant to determine the initial peptide binding poses and potential binding region for the MUC1 peptide, not the molecular catalytic mechanism.

Distances involving catalytically relevant atoms are reported in Figure 5. The reference to Figure S2 was removed.

Again, the authors should spend more of the reader's time on QM/MM results, and less on docking ones.

There is no logic in making this argument for k_{cat} and not for K_m or V_{max} . Some argue that any k_{cat} is an averaged result of all potential substrates. Substrate heterogeneity by itself does influence the kinetic profile but does not precludes to obtain an analytical solution to the Michaelis Menten equation, see:

<https://www.biorxiv.org/content/biorxiv/early/2025/08/24/2025.08.21.671619.full.pdf>

This is a very minor point though.

AUTHOR RESPONSE: We thank the reviewer for this comment and the reference provided. While k_{cat} and k_{cat}/K_m could, in principle, be calculated as averages across multiple reactive sites, we deliberately chose not to report these values for our multi-site substrates. Calculating k_{cat} for such peptides could be misleading because not all sites are equally active—some are kinetically inactive or have different turnover rates, as later revealed by our structural and computational analyses. Reporting a single turnover number or catalytic efficiency could be misinterpreted as reflecting site-specific reaction energy barriers or enzyme preference, which we explicitly address through detailed LC-MS and computational studies. Instead, we report apparent K_m and V_{max} values as empirical measures of overall reaction behavior. These initial parameters served the purpose of guiding our subsequent site-specific analysis, where K_m was recalculated based on the number of reactive sites, which eventually led to providing accurate mechanistic insights without introducing potentially misleading interpretations.

I feel that the authors should read the comments made to them before answering them, and read the provided references rather than just thanking the referee for having provided them.

We would also like to suggest that the authors group figure 6A with 7A in figure 6, and to group figure 6B and 7B in figure 7. Having the site-to-site comparison of the localization of GALANT's stresses the point and makes the message clearer for the reader.

AUTHOR RESPONSE:

We appreciate the reviewer's thoughtful suggestion to group Figures 6A with 7A and 6B with 7B, to present the experimental panels side by side and the corresponding schematic conclusions side by side. We fully understand how such an arrangement could emphasize the site-to-site comparison of the products associated with the two modes of GALNT localization in normal versus cancer-associated models and potentially enhance clarity.

However, after evaluating this option, we found that combining these panels would result in an excessively large and dense figure, particularly for Figure 6, which contains extensive experimental data from both models. The amount of detail—both in volume and visual complexity—would overwhelm the reader and compromise the readability and visual clarity of the experimental results.

For these reasons, we opted to maintain the separation between Figures 6 and 7, where Figure 6 presents both the experimental results and the derived schematic model for the normal system, and Figure 7 presents the corresponding experimental results and schematic model for the cancer-associated system. This layout allows a clear distinction between the normal and cancer-associated systems, preserves the intended narrative flow, and improves overall readability by presenting each system in a self-contained, coherent figure.

We hope the reviewer agrees that this organization best supports clarity and interpretability of the presented data.

Maybe the authors could show the reviewers these excessively dense figures. Given the level of efforts the authors.

Reviewer #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3

(Remarks to the Author)

The authors have satisfactorily addressed all comments from the previous review. The manuscript has been significantly improved, and I recommend acceptance.

Reviewer #4

(Remarks to the Author)

The Authors have addressed all the concerns raised. I now recommend the manuscript for publication.

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 3:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I find it disappointing that the authors would not share more on their QM/MM results, but I do understand the arguments they provide. As it stems from a conflict of interests between providing their readers with full information and protecting their intellectual property, this should be stated somewhere.

Otherwise, I do find the figures 6 and 7 they provide in the response to reviewers definitely clearer than the ones in the current version of the manuscript, particularly the new Fig. 7.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain

permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer: 1

“An in vitro one-pot synthetic biology approach to simulating diverging Golgi O-glycosylation of tumor-associated MUC1 from normal tissue MUC1” by Abdullateef Nashed and coauthors is a very interesting paper in the field of protein O-glycosylation. They develop an assay to study the effect of the order in which glycosyltransferases act, mimicking spatial organization of the Golgi. As GALNTs are known to migrate from the cis Golgi to the ER in tumors, they apply GALNTs before other GTs leads to higher site occupancy and shortened glycans, similarly to what is seen in cancer tissues. Their assay also allows for investigation of lectin dependent versus lectin independent mechanisms, exact substrate specificity of GTs. Finally, they model the sialylation reaction by ST6GALNAC1 to rationalize site specificity. The paper is well-written and easy to follow, altogether an informative and enjoyable read that definitely add to the understanding of the protein O-glycosylation field, and provide an assay that seems quite straightforward and practical to use.

However, there are some issues relative to the peptide design, the study of the lectin-dependent mechanism, and the QM/MM modelling and the ensuing rationale.

Peptide design. To me the biggest limitation is to have limited the analysis to a single repeat unit of MUC-1. While the design of the 23 and 27-mer can a little bit differentiate between the effects of intra and inter repeats, it is limited. The authors just state that the optimal sequence should only have each glycosylation site once, without justifying it. They conclude by stating that the commonly used TAP24MUC1 is unsuitable, using the principle that they propose without justifying them. They also only have this discussion in the methods section. Is it impossible that the glycosylation on one site affects the same site on unit $n+1$ or $n-1$? The kinetics with their UGS assay should be all doable with several repeats without having to adjust it, and it should be doable to express their constructs with several repeats to see if there are any differences. Did the authors attempted that, or why did they not do it? They should at least provide some rationale for why it is fine to study single repeat units rather than the full repeat, and provide it early in the manuscript, when they explain the design of their assay.

Lectin dependent mechanism. To further confirm which glycosylation are lectin-dependent or not, why not express only the catalytic domain? Or at least a single point mutant of the lectin binding site? If it is known it cannot be done, the authors should explain why.

QM/MM simulations. The first two glycosylation reactions show activation energy of 25 kcal/mol, which means that there is basically no reaction. The authors justify an observed difference of less than 2-fold in rate by modelling showing a 7.5 kcal/mol difference, which amount for a 100 000 fold difference in rate. They justify the slow rate by some far-fetched mechanism in which the catalytic base would actually be covalently bond to the sugar donor, making it a suicide inhibitor. Worse, they actually do not show anything of their modelling. There is not a single figure in the SI, and in the main text

only one figure that does not show any conformation, distance, etc. The authors should at least provide some of those. As well, they should tune down some interpretation on their simulations given the results they obtain. They could also remove the two simulations on T20 glycosylation and only keep the one on T23, which provides a reasonable energy barrier, and expand its description.

AUTHOR RESPONSE:

We agree with reviewer's comments regarding the unrealistic high activation energies. We therefore repeated the QM/MM study focusing only on the cancer-associated fully glycosylated MUC1. Reaction dynamics calculations were performed with increased sampling and corrected starting structure to obtain reasonable barrier heights for the glycosylation reactions. After these improvements have been made, we report activation energies of 17.11 kcal/mol for the T20 site, 23.87 kcal/mol for the T8 site, and 16.73 kcal/mol for the T13 site sialylation reactions, as seen in Figure 5C.

Necessary figures have been included in Figure 5 and SI. There is no need for chemical detail as this is not a manuscript that describes chemical mechanistic differences between sites. The function the simulations was to compare and determine the viability of ST6GALNAC1 catalysis at each of the 5 sites. The comparison is made through simple docking experiments, which are limited on their own, as well as sophisticated computational modelling to determine barrier heights. While we have in previous works (e.g., <https://pubs.acs.org/doi/abs/10.1021/acs.accounts.1c00477>) performed in depth electronic and atomic level analysis in pursuit of chemical mechanistic analysis this would be a distraction from the objective to build and demonstrate in vitro molecular models of normal vs. tumour golgi construction.

In line with the reviewer's concern, we have provided additional figures in the Supporting Information Figure S21, S22, and S23. Close up images of important residues involved in the reaction and relevant distances and angles across the QM/MM simulation are provided in Figure 5B Additional. The reported distances and angles are discussed in Section 2.3.1 to clearly illustrate the mechanistic features observed in our simulations.

Also, the SI needs to be formatted. There are not even supplementary figure numbers.

AUTHOR RESPONSE: SI has been updated and formatted

Minor, throughout the text.

There are also formatting issues:

-L13 O-GalNAc, Galnac "O" is positional and should be italicized, L47 N-acetyl "N" is positional and should be italicized. L94 N-acetylglucosaminylation, L103 O-GalNAc,

Other occurrences.

AUTHOR RESPONSE: All corrected

-Numbers for stoichiometry are subscript, e.g. MnCl_2 / MgCl_2 L371.

AUTHOR RESPONSE: All corrected

“Core” is written either as “core” or Core”, chose one and be consistent. Same thing for “one-pot” and “One-pot”, Serine/serine, Threonine/threonine, etc.

“SN1” the “N” is subscript.

Km. K is italicized, it is a thermodynamic constant, not a temperature unit. m should be subscript.

Latin is italicized, be it for species name (*E. coli* – also it is not *E. Coli*, tobacco etch virus), in vitro/vivo, e.g./i.e/etc.

AUTHOR RESPONSE: All corrected. However, Nature portfolio style guide mandates that these common Latin terms should be written in roman font, not italics

Ranges and minus are en-dash, not hyphen.

L320 β -D-thiogalactoside “D” is small caps

AUTHOR RESPONSE: All corrected

Minor, along the text.

L16 NAc, not nac.

AUTHOR RESPONSE: All corrected

L22, 202, 270 Sialyate $\diamond$ sialylate

AUTHOR RESPONSE: All corrected

L26-27 “This systems modelling unpacks the molecular details of enzyme localization and substrate glycan occupancy that is fundamental to the regulatory mechanisms that gives rise to tumor-associated MUC1 antigens.”

Reformulate.

AUTHOR RESPONSE: The sentence was revised

Figure 1 Move the SNGF labels to separate better A and B, indicate more clearly the differences between healthy and tumor in B. Ideally also show the glycan on the peptide along the arrow, not just unglycosylated and endpoint. The authors have space on the width as they went double column.

AUTHOR RESPONSE: SNGF labels were moved to the side. As for the healthy vs tumor model, the figure is a schematic demonstration of the differential localization of GALNTs and the glycan structures associated with it. The mechanistic details of the step-wise biosynthesis are investigated in the study and revealed later in the manuscript.

Figure 1 caption, “between normal epithelial tumor epithelial cells attributed to COP-1”
At least a word is missing.

AUTHOR RESPONSE: revised

L59-60: “such as the relocation of GALNTs”

Could the authors expand here? It is a major point behind the rationale of their modelling. Are all GALNTs moving equally? Are all of the other GTs at the same place? In Fig 1B there appear to be more sialyltransferases on tumor than healthy, is it by design?

AUTHOR RESPONSE: The main objective in this work is to study the impact of GALNT relocation independently of changes in GT expression levels.

As shown by Gill et al. (2010), several members of the GALNT family—including GALNT1, GALNT2, GALNT3, GALNT4, and GALNT6—were demonstrated to undergo relocation from the cis-Golgi to the ER, which supports our use of the collective term “GALNTs” in the manuscript.

With regard to the reviewer’s observation on Figure 1B, the intention was indeed to depict both aspects of deregulation: GALNT relocation and sialyltransferase upregulation. To avoid confusion, we have modified the figure to make the upregulation of sialyltransferases clearer and revised the caption accordingly.

L65 “parameters governing glycosyltransferase (GT) activities and substrate specificities using the UGC assay¹⁴”

Ref 14 is a 2024 original paper by the authors, maybe they could expand here and explain what is the UGC assay, as the assay was proposed less than a year ago it is not necessarily common knowledge. Just saying that it detects nucleoside mono- and di- phosphate formation through nucleotide kinases and NADH consumption, etc.

AUTHOR RESPONSE: The details of the UGC assay were added to the relevant section

Fig 2C is not clear enough as it describes the process for the article. I imagine the GTs are not his tagged, the TEV is, and cleavage happens on bead? Conversely, the authors can forego the part when they represent heterologous protein expression (top of Fig 2C), readers could be expected to understand that.

AUTHOR RESPONSE: We have revised Figure 2C to improve clarity. His-tags are now indicated on sfGFP and TEV.P, and the flexible units on Linker 2 are also marked. The “No reactions” note in panel B has been removed. Regarding cleavage, we clarify that TEV protease acts on the purified fusion protein in solution, not on bead, as indicated by the reaction tube. Following cleavage, the His-tagged components (TEV.P, sfGFP, and the linker) are removed from the mixture. We have also expanded the figure caption to provide additional detail on the pipeline. Finally, we retained the cloning and transformation steps to highlight the simplicity of recombinant peptide expression in comparison to chemical synthesis.

L80” sfGFP was selected for its folding efficiency, minimized dimerization, and enhanced solubility.”

Precise that GFP is not used not for its fluorescence properties. When presented with an assay using GFP, readers may assume that fluorescence is a thing - and it is actually for the UGC assay.

AUTHOR RESPONSE: We thank the reviewer for this observation and agree that sfGFP is not used for its fluorescence properties in the UGCA. Instead, sfGFP was included in the construct for the reasons mentioned in the manuscript. However, the fluorescence facilitates visualization of the fusion protein during purification and enables accurate quantification. We have revised the text accordingly to avoid any misunderstanding.

L81-82 “with the MUC1 peptide, sfGFP was separated from the peptide using a linker (Linker 2) comprising of three rigid and three flexible units (from the N to C direction). Additionally, a rigid linker (Linker 1) was incorporated to improve steric presentation to the N-terminus His”

Linker 1 and 2 are not shown on Fig 1A, and I do not see three rigid and three flexible units as linker 2 in Fig 1A.

AUTHOR RESPONSE: Figure 1A was modified to illustrate the flexible and rigid units. The text was edited to refer to the detailed structure and sequence in the supportive information (Figure S1)

L129 “LC-MS results confirmed the structure and purity of the peptide (i).”

A reference to the appropriate supplementary figure – which should be indexed – has to be included here.

AUTHOR RESPONSE: All supplementary figures were indexed and referred to when relevant

L145, “over expression”, in one word.

AUTHOR RESPONSE: Corrected

Fig 3 overall separate the panels better., eg. The representation of the enzymes and their domains appear to be in C and not in B, etc..

AUTHOR RESPONSE: Panels in Figure 3 were separated as suggested by the reviewer.

Fig 3A, no need to write “no reaction”, that is visible. Do not represent the reaction with the 23 mer and GALNT2+GALNT4 differently if that is not commented on.

AUTHOR RESPONSE: “no reaction” note was removed. The activity of GALNT4 in combination with GALNT3 with the 23-mer peptide was commented on in the figure caption.

Fig 3B is too cramped, it makes it hard to read.

AUTHOR RESPONSE: Panel B was spread to improve its readability.

Fig 3C I The “tag removal” to replace TEV+IMAC does not help, just make the figure slightly confusing. What would be an indicator whether the product is after elution or the flow through. I assume the first ones with black arrows are always after elution, while the last ones in light brown the flow through is collected. Whether I am right or wrong in assuming so, it should be specified in the caption. Or just do like in Fig 4, more consistent and clearer.

AUTHOR RESPONSE: To improve clarity, we have revised Fig. 3C to follow the same style as Fig. 4, as recommended. This ensures consistency across figures and avoids potential confusion

Fig 3C III Remove the arrows and align the glycosylated peptide products. It would be consistent with Fig 4, which is much clearer in this regard despite dealing with more complicated structures.

AUTHOR RESPONSE: The arrows were removed and the structures aligned.

Fig 3C If it has been done (I assume so), it lacks a line with GALNT4 alone yielding non-glycosylated peptide i)

AUTHOR RESPONSE: GALNT4 alone (as well as GALNT7) did not show activity in the UGCA and was therefore excluded from the in vitro synthesis experiments. This has been clarified in Section 2.2 (“In vitro synthesis of the Tn antigen”).

L161 “Within this context the core 3 is synthesized by B3GNT6 catalyses of the GlcNAc covalent β 1-3 bond of to the Tn antigen.”
Reformulate

AUTHOR RESPONSE: revised

Fig 4A, Vmax as a value, kcat could be included, and kcat/Km.

AUTHOR RESPONSE: We thank the reviewer for pointing this out. In this case, the substrate peptides contain multiple reactive sites, making it impossible to attribute the kinetic parameters to a single reaction per molecule. Therefore, kcat is not meaningful in this context, as the turnover number cannot be linked to a defined thermodynamic step. Instead, we report apparent Km and Vmax values, which serve as averaged proxies for substrate affinity and maximal reaction rate, respectively. This provides a more appropriate comparative measure across enzymes and substrates. We have clarified this in the caption of Fig. 4A

Fig 4BI, first chromatogram shows at least two peaks for peptide labelled (v), may the authors comment on that?

AUTHOR RESPONSE: Closer inspection of the LC–MS data confirmed that the two partially overlapping peaks correspond to glycopeptide isomers, arising from alternative sites of glycosylation (two GalNAc residues). The integrated MS spectra of both peaks show the same m/z, consistent with this interpretation. While we did not perform detailed site-specific structural assignment in this study, our interpretation is supported by literature precedence, as outlined in the text.

L 171 “indicates that the enzyme has a lower affinity (Km value of 0.114 mM for GalNAc5-23mer vs 0.062 mM for GalNAc3-23mer) but higher turnover (Vmax for GalNAc5-23mer is 180 % of the Vmax of GalNAc3-23mer) of the fully glycosylated MUC1 compared with the semi glycosylated MUC1 (table in Figure 4 A). On the other hand, no significant difference in the Km values of C1GALT1 were observed between the semi-saturated and completely saturated MUC1. This indicates that C1GALT1 prefers glycosylation of the GALNT_D-sites over the GALNT_L-sites at any concentration of MUC1 and ST6GALNAC1 prefers glycosylation of the GALNT_L-sites over the GALNT_D-sites at high concentrations of MUC1.”
Actually in both systems in Table 4A it shows that both peptides show the same Vmax over Km. As I assume natural concentrations to be far below Km, it means no difference in activity? Or at least no site preference for both enzymes in dilute conditions. Anyway it should be commented.

AUTHOR RESPONSE: We agree with the reviewer that in vitro kinetic parameters cannot be directly extrapolated to in vivo Golgi conditions, where enzyme membrane association, complete substrate structure and complexity, compartmentalization, and regulatory factors strongly influence glycosylation. Our data therefore provide apparent

kinetic values that reflect relative enzyme behaviors on defined substrates, rather than absolute measures of site preference under physiological conditions. We have clarified this limitation in the revised text. In addition, the synthesis performed on the double-repeat MUC1 construct in this revised version revealed that not all sites within the GALNT_D and GALNT_L regions participate in the sialylation reaction catalyzed by ST6GALNAC1. Taking this into account, the K_m values presented in the table were recalculated, and the revised values showed no significant differences between the two substrates. The text was edited to clarify this.

sialylation, eg. [HYPERLINK](#)

"<https://url.za.m.mimecastprotect.com/s/JGXaCj2gl8ijlYnx1I5iwfmVXvO>"<https://academic.oup.com/glycobiology/article/29/10/735/5528790>

AUTHOR RESPONSE: Reference was added.

L200 overcomes a second transition state 2 (TS2) barrier of 24.82 kcal/mol.

L206 a transition state 2 (TS2) energy barrier of 17.46 kcal/mol.

That is a difference of 7.36 kcal/mol, ie. c. factor 100 000 in rates. According to table if figure 4A, the difference is not even of a factor 2, nowhere close to 100 0000. The computations are interesting, but maybe the authors should modulate their interpretation on the relation between experimental rates and computed energy barriers. Their computational method is not accurate enough. On the other hand, 17.46 kcal/mol seems perfectly on line with what is expected for a GT reaction.

AUTHOR RESPONSE:

QM/MM simulations and values have been revised and addressed in accordance with previous review comments.

Fig 5B alone is a bit poor to see the interaction. I would like to see zoom-in of all the states schematized in Fig 5D for each system in a (or several) supplementary figure. Would also help to follow the discussion on molecular rationale.

AUTHOR RESPONSE:

Figure 5 has been updated to reflect these, see panel B

Fig 5C, the x-axis is unclear. It reads a bit like the distance of the bond to be formed (O6-C1) minus the distance of the bond to be cleave (UDP-sugar). In that cases the signs seem wrong? Anyway has to be made clearer.

AUTHOR RESPONSE:

Figure 5C x-axis label updated to correctly reflect the correct sign of values on the x-axis. An explanation of the reaction coordinates is detailed in the computational methods Section 4.9.

Fig 5C, the first two energy profiles would be indicative of a reaction not occurring (activation energy above 25 kcal/mol). I am also mildly surprised to see such a large difference in the energies of the products (I and II vs III). Could the authors comment on it?

AUTHOR RESPONSE:

QM/MM simulations and values have been revised and addressed in accordance with previous review comments. In revising the reaction coordinates, the difference between product and reactants has largely been reduced.

Fig 5 period is missing at the end of the caption. Actually, it looks like more than that s missing, with the rest of the color coding?

AUTHOR RESPONSE:

Missing text in Figure 5 captions added.

L196 The sialylation of the T20 residue on partially glycosylated peptides (I) exhibited an energetic profile that is distinct from its fully glycosylated (II) counterpart (Figure 5 C). I do not really see a difference, could the authors explain what they mean by “distinct”? Also given the errors, I am not sure this distinction would be meaningful. I would replace “distinct” by “nearly identical”.

AUTHOR RESPONSE:

Discussion has been edited to now discuss the nearly identical reaction profiles of T13 and T20 in Section 2.3.1.

L208 “The molecular reasons for this are that the sugar conformation necessary for the nucleophilic attack is critical to the sialylation reaction. If the sugar conformation is not within the near attack conformation cone angle, the histidine may covalently bind the anomeric carbon after the disassociation of the phosphate.”

On one hand, as enzyme and donor are identical for all three reactions it is expected that the nucleophile should be the difference. Yet it appears the biggest difference between the profiles I / II and III in Fig 5C to be the formation of the TS1 and the OC intermediate. At this stage, the nucleophilic sugar should not have such a drastic

importance as they claim it to be SN1 like? It may be commented a bit more here?

AUTHOR RESPONSE:

As discussed in Section 2.3.1, the formation of the OC intermediate (and thereby TS1) is influenced by the hydrogen bond formation between the CMP moiety and HIS552. The T8 system demonstrates a short-lived hydrogen bond across the simulation accounting for the high TS1 value reported.

L 210 “This is the scenario in the T20 I and II T20 case where a side product forms when the catalytic histidine (HIS657) is covalently bonded to the anomeric carbon of the sialic acid lead.”

Is that something the authors saw experimentally? It would be surprising, the donor substrate acting as suicide inhibitor of the enzyme. Did the authors observe a loss of activity upon incubation with the donor, or a mass increase? If it is only observed during simulations and no experimental clue, the authors may suspect it to be a computational artefact, or justify why they imagine it is not so. Again here, having some snapshots of the modelled reactions would be useful.

AUTHOR RESPONSE: We have run entirely new experiments simulations on all five sites and this point and this is now moot.

L 226 C1GALT1 competes with these enzymes' lectin domains for the modified sites. This competition results in the synthesis of Core 1 structures and inhibits subsequent GalNAc glycosylation at other sites via the lectin-dependent mechanism of GALNTs. That is pretty interesting, do the authors have some data (possibly from literature) on the affinity of the lectins of GALNT1 and GALNT2 for the Tn and just GalNAc?

AUTHOR RESPONSE: The study by Pederson et al. 2011 investigated this specificity in detail. The study was also cited in this revised version of the manuscript. Additionally, the synthesis carried out on the double repeat MUC1 in the revised version when ST6GALNAC or C1GALT1 were combined with GALNTs produced the same sites occupancy. This indicates that the lectin-dependent mechanism of GALNTs is impacted similarly by both T and sTn structures.

L287 “Figure 1C”

There is no Fig 1C. Fig 2C?

AUTHOR RESPONSE: Corrected

L289 “AlphaFold structure prediction tools”

Version of AlphaFold and corresponding reference should be added.

AUTHOR RESPONSE: reference added

L294, “(i) glycosylation of the naked peptide using only the catalytic domain, (ii) glycosylation of pre-glycosylated peptides through the lectin domain, and (ii) a combination of the two mechanisms occurring sequentially.”

The second “ii” should be “iii”.

AUTHOR RESPONSE: corrected

L304 The optimal sequence must be inclusive of all the variables that determine GALNT specificity, and each variable must only be represented once in the peptide sequence. Why only once? What if a glycosylation at S19 influence the glycosylation of S19 on the next repeat? I understand that this made modelling easier, but.

AUTHOR RESPONSE: We thank the reviewer for this valuable comment and for highlighting the importance of potential inter-repeat effects. In this revised version, the peptide design rationale has been expanded and moved from the Methods to a new Results subsection, “2.1.1 Peptide design and fusion protein expression.” The criterion of representing each potential glycosylation site only once was adopted to isolate the intrinsic site-specific preferences of each GALNT isoform and to derive apparent kinetic parameters that reflect the cumulative activity across all unique sites without structural redundancy (explained in more details in the section). Including multiple tandem repeats introduces additional variables such as inter-repeat interactions, altered spatial orientation, and lectin-domain directionality, which could confound kinetic interpretation. To partially capture these structural variables, two sequence frames (27-mer and 23-mer) were designed and compared, with the 23-mer revealing the distinct activity of GALNT4. Furthermore, in response to the reviewer’s suggestion, we have now extended the study to include a double-repeat MUC1 construct to assess inter-repeat cooperativity after establishing the baseline kinetic profile of individual sites.

L358 “concertation”.

AUTHOR RESPONSE: corrected

L410 The QM region comprised HIS567 (catalytic base) sidechain atoms, HIS552 (stabilizing the CMP phosphate group), the Neu5Ac moiety, and the 12 CMP phosphate group. The nonpolar C-C bonds were treated with hydrogen link atoms, and the polar C-O bonds were treated with the Simple Link Atom Saccharide Hybrid (SLASH) 41 method. Aren’t the authors missing the nucleophile? In Fig 5D the GalNAc was included in the QM region including the glycosidic O1 and is actually the only bond treated with SLASH (which seems sound).

AUTHOR RESPONSE:

GalNAc was included in the QM region. This was updated in text Section 4.9

Ref 12 format to be modified.

AUTHOR RESPONSE: All references were revised and formatted according to the journal guideline

After all these sometimes-critical comments, I want to reemphasize that I enjoyed reading this manuscript, which I found informative and to constitute a sizeable addition to the field. I think the overall assay to be cleverly design, and that important conclusions are drawn from the results. I want to congratulate all the authors for this very good piece of work.

Reviewer: 2

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer: 3

This manuscript presents an innovative one-pot synthetic biology strategy to reconstitute and dissect the complex glycosylation pathways of MUC1, with a particular focus on enzyme specificity and kinetics involved in the generation of tumor-associated glycoforms such as Tn, T, and their sialylated derivatives. The study provides valuable insights into how enzyme localization (ER vs. Golgi) and competitive activities among GALNT, C1GALT1, and ST6GALNAC1 govern MUC1 glycoform heterogeneity. Furthermore, the integration of QM/MM simulations with experimental kinetic data reveals site-selective sialylation of T13 and T20 residues by ST6GALNAC1 under high MUC1 concentrations. This offers a novel mechanistic perspective on the glycosylation heterogeneity of MUC1 in cancer, with potential implications for designing MUC1-targeted cancer immunotherapies.

The manuscript meets the high standards of novelty, significance, and scientific rigor expected by Nature Communications.

AUTHOR RESPONSE: Thank you

However, prior to publication, the authors should consider the following points to further strengthen their work:

Major Comments:

1. Energetic Profiles and Reaction Feasibility:

The computed product states (P) for reactions I and II are significantly higher in energy than the corresponding reactant complexes (RC), by approximately 20 kcal/mol. Similarly, in reaction III, the product remains more than 10 kcal/mol above the RC (Figure 5C), suggesting these transformations are thermodynamically unfavorable (i.e., endothermic). Moreover, the activation barriers for TS2 in reactions I and II (24.82–25.21 kcal/mol) exceed typical enzymatic reaction barriers (<21 kcal/mol). While the authors mention "relatively slow reactivity" in the experimental data, a quantitative comparison between the computed activation energies and the experimental rate constants is needed to confirm whether the predicted energetics are consistent with the observed kinetics.

AUTHOR RESPONSE

We acknowledge the reviewer's concern regarding the high relative energies of the product states and activation barriers in our initial calculations. To address this, we repeated the QM/MM and reaction dynamics simulations with corrected settings. The revised results now yield barrier heights within the expected enzymatic range. We have updated Figure 5 and the Supporting Information accordingly, and we now include a quantitative comparison of the computed barriers.

These findings raise concerns about the reliability of the computational model. The authors are encouraged to address the following:

Model Quality: While ST6GALNAC1 was modeled using SWISS-MODEL, the authors should provide quality metrics (e.g., GMQE, QMEAN scores, Ramachandran plot) and specify the sequence identity with the template ST6GALNAC2 (PDB ID: 6APL). If the identity is below 30%, the model's reliability may be limited. A multiple sequence alignment could be added to support this.

AUTHOR RESPONSE:

We now provide detailed quality metrics for the ST6GALNAC1 model generated by SWISS-MODEL, including QMEAN scores, and a Ramachandran plot (Figure S21B). The sequence identity between ST6GALNAC1 and the ST6GALNAC2 template (PDB ID: 6APL) is 41.6%, above the 30% reliability threshold. Additionally, we include a multiple sequence alignment (Figure S21A) to highlight conserved residues.

Comparison with AlphaFold3: Have the authors compared the SWISS-MODEL structure with a model predicted by AlphaFold3? If so, please report RMSD values and discuss any structural differences, particularly within the active site.

AUTHOR RESPONSE:

Both AlphaFold3 and AlphaFold2 models of ST6GALNAC1 were compared to the SWISSMODEL structure used via protein structural alignment. The SWISS-MODEL

structure reported an RMSD of 0.53 Å when compared to the ST6GALNAC2 template, substantially lower than the AlphaFold2 (1.74 Å) and AlphaFold3 (1.83 Å) models. Due to this high structural fidelity, the SWISS-Model structure of ST6GALNAC1 was used for molecular dynamics simulations and downstream reaction dynamics. This information is provided in the Supplementary Information, see Figure S22.

2. Docking Procedure and Pose Validation:

While selecting the lowest-energy structure from the top HADDOCK clusters is a valid approach, further validation is needed for catalytically relevant poses. Do the selected conformations meet the geometric criteria for nucleophilic attack (e.g., distance and angle)? The authors should explicitly state whether the poses are catalytically competent.

AUTHOR RESPONSE:

HADDOCK proved unable to dock the large glycopeptide and was only used to determine the initial MUC1 peptide binding pose. The acceptor GalNAc was modeled in on the reported peptide pose form HADDOCK into a catalytically relevant pose (the nucleophile–anomeric carbon distance is within 3.0–3.3 Å and the approach angle lies within 155–170°, values consistent with established criteria for productive glycosyl transfer reactions). Based on the reviewer's comment, additional HADDOCK clusters were analysed and used for simulations as reported in the supplementary information section on Molecular Docking, see Figure S23.

3. QM/MM Methodology:

The use of DFTB3 as a semi-empirical QM method raises concerns about its ability to accurately capture transition state energetics for GTs reactions. Please cite relevant literature supporting the applicability of DFTB3 in modeling this class of enzymatic mechanisms.

Given the relatively high energies of the product states, the authors should discuss whether any potentially stabilizing interactions (e.g., hydrogen bonding, π -stacking, ionic interactions) were omitted or not captured in the model. Additionally:

Clarify the selection criteria for representative MD snapshots used in QM/MM calculations.

Confirm whether interactions with key catalytic residues are preserved in these snapshots.

Address the influence of sugar ring conformations on reactivity. Have conformational analyses been performed to ensure ring stability?

AUTHOR RESPONSE:

We selected DFTB3 for the QM region because it provides an excellent cost/accuracy trade-off for QM/MM sampling of enzymatic reaction pathways where extensive conformational sampling is required. DFTB3 and its parameter sets have been widely used and benchmarked for biochemical systems and carbohydrates, and parameter extensions exist for phosphate and sulfur species relevant to nucleotide-sugar donors. We now cite the DFTB3 parametrization and methodological literature and discuss known limitations of semi-empirical approaches (uncertainties of a few kcal·mol⁻¹ in absolute barrier heights). Key references added to the Methods are: Gaus et al. (DFTB3 parametrization for S/P), performance comparisons of DFTB versus DFT in biochemical systems, and reviews on QM/MM studies of glycosyltransferases and carbohydrate-active enzymes

<https://pubs.acs.org/doi/10.1021/ct401002w>

<https://pubs.acs.org/doi/10.1021/acs.jctc.6b00531>

<https://pubs.acs.org/doi/10.1021/jacs.5b01156>

<https://pubs.acs.org/doi/abs/10.1021/jp107620h>

We verified that all local enzyme-substrate interactions likely to stabilize reactive states were included in the model and sampled: the full nucleotide-sugar donor (CMP-Neu5Ac / UDP-Gal variants) and its phosphate interactions with nearby basic residues were included in the QM region or treated explicitly in the MM region with explicit counterions as appropriate.

4. Catalytic Mechanism Interpretation :

The manuscript proposes a covalent adduct formation between the catalytic histidine and the anomeric carbon during side product formation, diverging from the widely accepted role of histidine as a general base in GTs. Without direct experimental evidence (e.g., mass spectrometry), this interpretation remains speculative. The authors should either provide supporting evidence or clearly state that this mechanistic proposal is hypothetical and subject to further investigation or simply not discuss this.

AUTHOR RESPONSE: A new set of simulations were run only the T sites showed catalytic reactivity. The S sites were excluded. Analysis or comments re: covalent binding to the enzyme were not necessary.

5. Figure Organization and Molecular Detail:

Figure 5 could benefit from improved organization: Consider placing the general catalytic mechanism (currently panel D) as panel A, followed by the free energy profiles (panel C).

Previous panels A and B appear redundant; one could be moved to the Supplementary Information.

Additional molecular graphics showing close-up views of the active site would enhance mechanistic clarity. These figures should include annotated catalytic residues and key distances in the RC and TS structures.

AUTHOR RESPONSE:

Figure 5 updated as suggested.

6. T13 vs. T20 Sialylation Specificity:

Can the authors provide more mechanistic detail on the observed specificity?

At the QM/MM level, do differences in anomeric carbon charge or sugar ring conformations between T13 and T20 explain the preferential reactivity?

At the MD level, the discussion regarding the proportion of the near attack conformation (NAC) is particularly insightful. A statistical analysis comparing nucleophile–anomeric carbon distances and angles at T13 and T20 would provide quantitative support for this observation.

AUTHOR RESPONSE:

The discussion has been revised based on the updated values and Section 2.3.1 provides details on the mechanistic details explaining the difference in reactivity across the T20, T8 and T13 sites.

Minor Comments:

1. Line 15 – Abbreviation Definition:

On first use, spell out all abbreviations. For example, “GALNT” should be introduced as “N-acetyl-galactosaminyltransferase (GALNT).”

AUTHOR RESPONSE: corrected

2. Line 16 – Consistency in Terminology:

Ensure correct and consistent capitalization of scientific terms such as “GalNAc.” Please correct all instances of “Galnac” to “GalNAc.”

AUTHOR RESPONSE: corrected

Reviewer: 4

Key Results

The authors developed an in vitro synthetic biology approach to investigate how the presence and localization of glycosyltransferase (GT) enzymes across the endoplasmic reticulum (ER)–Golgi system influence the glycosylation patterns of proteins. MUC1 was used as a model protein; however, the approach can be readily adapted to other O-glycoproteins. Given that GT relocation during malignant processes is well documented in the scientific literature, this method also enables the elucidation of mechanisms underlying the formation of cancer-associated glycan structures.

Validity

The fundamental concept of the study is questionable, as the authors claim to have developed a one-pot method, which implies that the entire reaction takes place in a single vessel and that there is no need to isolate or purify the products between steps. The presented workflow does not meet these criteria, especially when the GTs are applied sequentially.

The structural interpretation of the synthetic products (glycoforms) relies heavily on the results of the HPLC-MS measurements. However, the data provided in the supplementary information do not sufficiently support the structures presented in the manuscript, rendering the findings speculative and casting serious doubt on the conclusions. Below, we outline the main shortcomings in bullet points.

- The mass accuracy of the peptides and glycopeptides is not provided, and in most cases, it cannot be calculated based on the data given.
- Figures 3 and 4 depict a total of eight structures (i–viii); however, tandem mass spectra supporting the proposed structures are provided only for the naked peptide (i) and a single glycopeptide (ii).
- In some cases, peak shouldering can be observed in the chromatograms (see line 69, peak 1 in the supplementary information). This may indicate the separation of isomers, suggesting that more products were formed than reported. In other instances (see line 71, peak 1 in the supplementary information), several peaks with higher intensity than the identified component are present, for which no explanation or characterization is provided.

Significance

Considering that carbohydrates remain among the least understood contributors to cancer, the conclusions of this work could significantly advance our understanding of the O-glycosylation process and the formation of cancer-associated glycan structures. Moreover, the approach presented here is thoughtfully designed and could be adopted by many research groups for the in vitro simulation of the glycosylation machinery in the ER–Golgi system.

Data and methodology

Aside from the execution of the HPLC-MS measurements, the approach appears valid, and the data quality is good. The presentation is clear and well-organized. Please note that the computational modeling is outside the reviewers' expertise.

Analytical approach

Apart from the HPLC-MS measurements, from which the authors drew overly far-reaching conclusions, the analytical approach is generally sound, and the statistical tests are appropriate. Please note that the computational modeling is outside the reviewers' expertise.

Suggested improvements

The reviewers suggest several substantial improvements before publication, mainly based on the comments of the Validity section of the review report.

1. The mass accuracy of all proposed peptides/glycopeptides must be reported (in ppm), and the data should allow readers to verify this by providing both the theoretical exact mass and the observed mass to at least four decimal places. Raw data files should be uploaded to a public repository (e.g. PRIDE, Massive).

AUTHOR RESPONSE: All theoretical and experimental m/z values for the proposed peptides and glycopeptides are reported to at least four decimal places, and the corresponding mass accuracy (in ppm) has been calculated for each peak. Theoretical m/z values and fragment ions were predicted using the MS-Product tool (version 6.6.6) from the Protein Prospector suite (Baker and Clauser, <http://prospector.ucsf.edu>), and theoretical isotope distributions were generated using the MS-Isotope tool from the same suite. The raw data files have been deposited in the MassIVE repository.

2. The structures of all reported glycopeptides, including sialylated glycoforms, must be validated using tandem MS analyses.

AUTHOR RESPONSE: MS analyses were performed using two instruments: a Q-Exactive quadrupole–Orbitrap and a SYNAPT G2 system employing HCD fragmentation. While these methods provide high mass accuracy, they are not always ideal for precise glycosylation site localization, particularly for sialylated glycopeptides. The inherent lability of sialic acids often results in their loss prior to peptide backbone fragmentation, and in cases involving long peptide sequences with multiple potential glycosylation sites, site assignment becomes inconclusive. An example of glycopeptides with uncertain site localization is presented in Figures S11. Therefore, MS² spectra were not included, and only primary ions were used in the analysis. However, the site-specific digestion with IMPa used with the synthesis carried out on the double-repeat MUC1 (added to this revised version) improved the site occupancy analysis to a level sufficient to draw the conclusions made in the manuscript.

3. Peak shouldering should be investigated, and the major chromatographic peaks must be accounted for in the base peak chromatogram to ensure that the results adequately support the conclusions. For these first three points, the reviewers recommend a workflow that includes Electron Transfer Dissociation (ETD) for O-glycan site localization, Higher-energy C-trap Dissociation (HCD) for glycan structure elucidation, and a database search (e.g., Byonic, MSFragger, or pGlyco) to efficiently identify both the main components and potential side products.

Author Response: The issue of peak shouldering and overlapping chromatographic features has been addressed. Closer inspection of the LC–MS data confirmed that these partially overlapping peaks correspond to glycopeptide isomers arising from alternative glycosylation sites (two GalNAc residues), consistent with their identical m/z values and retention behavior.

We appreciate the reviewer's valuable suggestions regarding the use of ETD/HCD workflows and database-assisted analyses (e.g., Byonic, MSFragger, or pGlyco) for detailed O-glycan site localization and glycopeptide identification. These methods indeed represent powerful tools for in-depth structural characterization and will be highly relevant for future studies aimed at comprehensive site mapping.

In the current work, however, our focus was on defining the enzymatic specificity and divergence of glycosylation patterns between tumor-associated and normal GALNT distributions, rather than exhaustive structural elucidation. The LC–MS analyses—supported by site-specific digestion using IMPa and validated by kinetic and computational modeling—were sufficient to establish the observed differences in glycosylation outcomes. Moreover, the kinetic and computational data provided deeper insight into site preference and enzyme behavior that could not be fully resolved through MS-based analysis alone, particularly for extended glycosylation reactions where multiple positional isomers are expected.

Therefore, while we acknowledge the value of the suggested analytical extensions, the current level of MS analysis adequately supports the study's conclusions and aligns with the defined scope of the manuscript.

4. The Roepstorff–Fohlman nomenclature should be used for naming fragment ions, as it is the most widely accepted system in peptide mass spectrometry.
5. Some details of the HPLC-MS measurements necessary for reproducing the experiments are missing, including injection volume, trapping method parameters, and the column thermostat temperature.

AUTHOR RESPONSE: More details were added to the corresponding section in the Supporting Information, including injection volumes (2 μL for the Q-Exactive and 7.5 μL for the SYNAPT G2), trapping parameters (trap column specifications, 300 nL/min flow rate, 3 min trapping duration), and column temperature (ambient, $\sim 25^\circ\text{C}$). These additions ensure full reproducibility of the LC–MS measurements.

6. In lines 36–38 of the Supporting Information, the authors state that measurements were performed using either a Q Exactive quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) or a SYNAPT G2 mass spectrometer (Waters, USA); however, no experimental details are provided for the latter instrument.

AUTHOR RESPONSE: Details were added to the corresponding section in the Supporting information

7. The SDS-PAGE workflow should be presented in the Methods section or in the Supporting Information.

AUTHOR RESPONSE: A detailed description of the SDS-PAGE workflow has now been added to the Supporting Information

8. The expression “one-pot” should be removed from the title and the text.

AUTHOR RESPONSE: We appreciate the reviewer’s comment and understand the concern regarding the term “one-pot.” In this work, the expression “one-pot glycosylation” refers specifically to reactions where multiple glycosyltransferases (GTs) were co-incubated in a single reaction mixture without intermediate purification steps. These reactions were designed to simulate enzyme co-localization scenarios inferred from our GT localization model, such as GALNTs acting in combination with C1GALT1 and ST6GALNAC1. In contrast, sequential reactions were performed to represent cases where these enzymes act in isolation and for a stepwise mechanistic study. When investigating the the intermediate structures using the 23-mer peptide, the stepwise synthesis was had to be validated via performing intermediate purification and characterization. However, when the double repeat MUC1 peptide was used as a substrate, the sequential (stepwise) synthesis modeling the tumor mode (figure 7) was performed without intermediate purification steps (when ST6GAL1 and C1GALT1 added after GALNTs). Thus, the use of the term “one-pot” accurately reflects a subset of the experimental workflow in which enzymatic steps were combined in a single reaction vessel.

Clarity and context

Overall, the reviewers found the text clear and accessible, and the figures help to clarify the approach and workflow. However, only few details are provided on how the results align with or contradict current knowledge of the O-glycosylation machinery.

References

The manuscript references the relevant literature appropriately. Key studies are cited to provide context for the current research, and the references support the rationale and methodology presented.

Your expertise

The reviewers state that the computational modeling presented in the manuscript falls outside their area of expertise.

Reviewer: 5

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Note: as colour markings are not visible in the provided text box, please refer to the enclosed word document "Review2.0.docx". Beside this note and colour markings, it contains the exact same information.

"An in vitro one-pot synthetic biology approach to simulating diverging Golgi O-glycosylation of tumor-associated MUC1 from normal tissue MUC1" by Abdullateef Nashed and coauthors is a very interesting paper in the field of protein O-glycosylation. The authors addressed well the various reviewer's requests. Below are a few comments in green (black for previous review, blue and red for authors responses), mostly on the representation of the QM/MM data.

[...]

QM/MM simulations. The first two glycosylation reactions show activation energy of 25 kcal/mol, which means that there is basically no reaction. The authors justify an observed difference of less than 2-fold in rate by modelling showing a 7.5 kcal/mol difference, which amount for a 100 000 fold difference in rate. They justify the slow rate by some far-fetched mechanism in which the catalytic base would actually be covalently bond to the sugar donor, making it a suicide inhibitor. Worse, they actually do not show anything of their modelling. There is not a single figure in the SI, and in the main text only one figure that does not show any conformation, distance, etc. The authors should at least provide some of those. As well, they should tune down some interpretation on their simulations given the results they obtain. They could also remove the two simulations on T20 glycosylation and only keep the one on T23, which provides a reasonable energy barrier, and expand its description.

AUTHOR RESPONSE:

We agree with reviewer's comments regarding the unrealistic high activation energies. We therefore repeated the QM/MM study focusing only on the cancer-associated fully glycosylated MUC1. Reaction dynamics calculations were performed with increased sampling and corrected starting structure to obtain reasonable barrier heights for the glycosylation reactions. After these improvements have been made, we report activation energies of 17.11 kcal/mol for the T20 site, 23.87 kcal/mol for the T8 site, and 16.73 kcal/mol for the T13 site sialylation reactions, as seen in Figure 5C.

Necessary figures have been included in Figure 5 and SI. There is no need for chemical detail as this is not a manuscript that describes chemical mechanistic differences between sites. The function the simulations was to compare and determine the viability of ST6GALNAC1 catalysis at each of the 5 sites. The comparison is made through simple docking experiments, which are limited on their own, as well as sophisticated computational modelling to determine barrier heights. While we have in previous works (e.g.,

<https://pubs.acs.org/doi/abs/10.1021/acs.accounts.1c00477>) performed in depth electronic and atomic level analysis in pursuit of chemical mechanistic analysis this would be a distraction from the objective to build and demonstrate in vitro molecular models of normal vs. tumour golgi construction.

In line with the reviewer's concern, we have provided additional figures in the Supporting Information Figure S21, S22, and S23. Close up images of important residues involved in the reaction and relevant distances and angles across the QM/MM simulation are provided in Figure 5B Additional. The reported distances and angles are discussed in Section 2.3.1 to clearly illustrate the mechanistic features observed in our simulations.

It is a fantastic result to have managed to refine the energy barriers to feasible ones, and we would like to congratulate the authors for doing so. However, "no chemical detail" is definitely not satisfactory. Please add some clear and readable snapshot along the QM/MM simulation, at least for each trajectory with a reported barrier of initial state, transition state, and product. I understand that this is not the focus of the authors, but they are currently reporting trajectories and energy barriers that we cannot evaluate. Those can easily be provided in 1-3 additional SI figures.

The authors added some information and lengthy discussion on their very low-level calculations (protein model generation and docking), but absolutely nothing on the high-end computations they did. I would also propose that the authors upload their full trajectories on a public repository, e.g. Zenodo, with a doi provided in the "data availability section".

AUTHOR RESPONSE:

Initial reactant coordinates and an exported trajectory video of the product formation are available on Github (<https://github.com/scientificcomputing/NCOMMS-25-06286-T-MUC1-in-vitro-Synthesis-/tree/main>).

Table S2 is also not satisfactory: "Table S2. Distances between centers of masses of the possible glycosylation sites and key residues involved in the reaction 252 i.e. the CMP-Neu5Ac donor and catalytic base HIS567." "Centers of mass" is not an appropriate measure point. You would get a lower distance if the donor was in a totally not reactive position, but with the phosphate close to the base. Also, present in the caption that those measures are for the TS*. If a distance should be measured between the catalytic base and the Neu5Ac donor, I suggest between the reactive nitrogen of His567 and the C1 of the donor. Ideally, I would also like to see the distance between the reactive nitrogen of His567 and the proton to be abstracted on the Ser/Thr. That being said, we would still like to stress that it is only a minor disagreement on how to present the data, and we would like our main message as an answer to the QM/MM part to be commending the authors for a great revision and refinement of their work.

AUTHOR RESPONSE:

Table S2 does not discuss any distances related to the TS as the GalNAc residues are not considered. The protein-protein docking performed is meant to determine the initial peptide

binding poses and potential binding region for the MUC1 peptide, not the molecular catalytic mechanism.

Distances involving catalytically relevant atoms are reported in Figure 5. The reference to Figure S2 was removed.

[...]

Fig 4A, V_{max} as a value, k_{cat} could be included, and k_{cat}/K_m .

AUTHOR RESPONSE: We thank the reviewer for pointing this out. In this case, the substrate peptides contain multiple reactive sites, making it impossible to attribute the kinetic parameters to a single reaction per molecule. Therefore, k_{cat} is not meaningful in this context, as the turnover number cannot be linked to a defined thermodynamic step. Instead, we report apparent K_m and V_{max} values, which serve as averaged proxies for substrate affinity and maximal reaction rate, respectively. This provides a more appropriate comparative measure across enzymes and substrates. We have clarified this in the caption of Fig. 4A

There is no logic in making this argument for k_{cat} and not for K_m or V_{max} . Some argue that any k_{cat} is an averaged result of all potential substrates. Substrate heterogeneity by itself does influence the kinetic profile but does not precludes to obtain an analytical solution to the Michaelis Menten equation, see:

<https://www.biorxiv.org/content/biorxiv/early/2025/08/24/2025.08.21.671619.full.pdf>

This is a very minor point though.

AUTHOR RESPONSE: We thank the reviewer for this comment and the reference provided. While k_{cat} and k_{cat}/K_m could, in principle, be calculated as averages across multiple reactive sites, we deliberately chose not to report these values for our multi-site substrates. Calculating k_{cat} for such peptides could be misleading because not all sites are equally active—some are kinetically inactive or have different turnover rates, as later revealed by our structural and computational analyses. Reporting a single turnover number or catalytic efficiency could be misinterpreted as reflecting site-specific reaction energy barriers or enzyme preference, which we explicitly address through detailed LC-MS and computational studies. Instead, we report apparent K_m and V_{max} values as empirical measures of overall reaction behavior. These initial parameters served the purpose of guiding our subsequent site-specific analysis, where K_m was recalculated based on the number of reactive sites, which eventually led to providing accurate mechanistic insights without introducing potentially misleading interpretations.

Additional points

The mass spectrometry nomenclature is quite confusing, as the authors move between "MS² analysis" (SI I114 and I116), MS2 (main text I508), "Lc-MS" or "LC-MS" (please chose one, better "LC-MS", e.g. Figs S4, S5 and S11, S13, S14, S15, S16, S17); " MS¹" (Figure S6), also MS1 (Figure 6A, 7A, line 508), MS 1 , M1, etc. Please homogenise and check all instances thoroughly.

AUTHOR RESPONSE: We have carefully reviewed all mass spectrometry nomenclature throughout the main text and Supporting Information. All instances have now been standardized to MS1 and MS2, and "LC-MS" is used consistently across the manuscript and figures

L162: "However, the observed glycosylation of the remaining two sites cannot be attributed", the discussion would be easier to follow if the sites were immediately specified, e.g. "However, the observed glycosylation of the remaining two sites (S9 and T13) cannot be attributed".

AUTHOR RESPONSE:

We thank the reviewer for the suggestion. We have revised the text to clearly distinguish the primary glycosylation sites for GALNT1 (T8) and GALNT2 (T20) from the additional secondary sites. The updated wording clarifies that these secondary glycosylation sites do not correspond to the previously reported lectin-dependent mechanisms and instead likely reflect a less specific mode of glycosylation.

The following request has been ignored by the authors, or did we miss the response?

"Lectin dependent mechanism. To further confirm which glycosylation are lectin-dependent or not, why not express only the catalytic domain? Or at least a single point mutant of the lectin binding site? If it is known it cannot be done, the authors should explain why."

AUTHOR RESPONSE:

We apologize for not explicitly addressing the reviewer's comment in our previous response, although the requested clarifications were incorporated into the revised manuscript during the first-round revision.

Specifically, we added two passages noting that determining whether the additional glycosylation sites arise from lectin-dependent versus direct mechanisms would require lectin-domain mutagenesis, which lies outside the focus of our biosynthesis-oriented one-pot model. These in-text additions state:

"Whether the additional two sites represent less-preferable direct glycosylation or lectin-dependent events remains to be proven. While this can potentially be resolved by mutating the lectin domains, such an investigation is outside the focus of the biosynthesis one-pot model being developed here."

and

"revealing and quantifying the exact contributions of lectin-domain inhibition versus the proposed steric hindrance would require a more rigorous investigation, potentially involving mutagenesis of the GALNT lectin domains and site-specific LC-MS/MS analysis, such as Electron Transfer Dissociation (ETD) for precise O-glycan site localization. Such analyses is out the scope as it distracts from the focus the present work to model the system of glycosylation biosynthesis within normal or diseased state golgi."

Furthermore, we note that dissecting the precise mechanistic contributions of the lectin domains does not affect the conclusions of this study, which focuses on modelling the overall glycosylation output rather than specifying individual GALNT catalytic modes. Detailed investigations of lectin-dependent mechanisms have been reported by several groups and were cited in our manuscript where relevant to the argument.

In summary the reviewer's suggestion was implemented in the manuscript text in the earlier revision, even though it was inadvertently omitted from our prior point-by-point response.

We would also like to suggest that the authors group figure 6A with 7A in figure 6, and to group figure 6B and 7B in figure 7. Having the site-to-site comparison of the localization of GALANT's stresses the point and makes the message clearer for the reader.

AUTHOR RESPONSE:

We appreciate the reviewer's thoughtful suggestion to group Figures 6A with 7A and 6B with 7B, to present the experimental panels side by side and the corresponding schematic conclusions side by side. We fully understand how such an arrangement could emphasize the site-to-site comparison of the products associated with the two modes of GALNT localization in normal versus cancer-associated models and potentially enhance clarity.

However, after evaluating this option, we found that combining these panels would result in an excessively large and dense figure, particularly for Figure 6, which contains extensive experimental data from both models. The amount of detail—both in volume and visual complexity—would overwhelm the reader and compromise the readability and visual clarity of the experimental results.

For these reasons, we opted to maintain the separation between Figures 6 and 7, where Figure 6 presents both the experimental results and the derived schematic model for the normal system, and Figure 7 presents the corresponding experimental results and schematic model for the cancer-associated system. This layout allows a clear distinction between the normal and cancer-associated systems, preserves the intended narrative flow, and improves overall readability by presenting each system in a self-contained, coherent figure.

We hope the reviewer agrees that this organization best supports clarity and interpretability of the presented data.

Please make sure, that the figures are exported in high quality, as the details in some of the figures are either not visible or difficult to see. This issue concerns the text in figure 3B, the charges in figure 5A, and S' and T's in figure 7B.

AUTHOR RESPONSE:

We thank the reviewer for noting this. The low resolution is due to the the PDF conversion; the original figures are high resolution in the word submission. Further, the referred to panels were edited to enhance visibility. All figures will be uploaded at full resolution during the final submission.

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3 (Remarks to the Author):

The authors have addressed most of the issues raised in my previous review, and the revised manuscript is significantly improved. However, several aspects related to the catalytic mechanism, docking analysis, and data presentation still require clarification before the manuscript can be accepted. My comments are summarized below.

1. Catalytic mechanism and QM/MM presentation

1.1 Nomenclature and atom assignment. Histidine should be abbreviated as His (not HIS) in the main text and figures. The mechanistic description must also confirm whether the reactive atom of the GalNAc acceptor is O6 rather than C6. In ST6GALNAC1-catalyzed sialylation, the nucleophile is generally the O6 atom attacking the donor anomeric carbon (C2). Distances involving C6 are chemically inconsistent and suggest an incorrect reaction coordinate.

AUTHOR RESPONSE:

Histidine nomenclature was corrected. The figure 5 caption was updated to clarify that the reactive atom of the GalNAc acceptor is O6 rather than C6. Atom-based distances were measured using the GalNAc C6 atom rather than O6, because the hydroxyl group exhibits substantial rotational mobility and therefore does not provide a stable reference point. The C6 carbon, that is covalently bonded to the rotatable (GT-GG-TG conformers) reactive O6 hydroxyl, provides a more reliable donor -acceptor distance measure point for comparing the glycosylation sites reaction readiness between T20, T8 and T13 (clarified in figure caption as well). The reviewer may have missed the mentioned in Section 4.7 paragraph 4, that the reaction coordinates used for QM/MM simulations involved bond forming (C2-O6) and bond breaking (C2-O2) and not C6.

1.2 Missing catalytic atom labels in Figure 5B. Key atoms such as the donor C2, the acceptor O6, and the NE2 atom of His567 are not labeled. Proper labeling is necessary for readers to follow the bond-breaking and bond-forming steps.

AUTHOR RESPONSE:

Figure 5B had been replaced with a 2D schematic representation of the reaction configuration that where atoms and distances are clearly labelled to make the table and discussion easier for the reader to follow. Bond breaking and bond forming steps are indicated in the mechanism presented in Figure 5A.

1.3 Incomplete reaction pathway in Figure 5B. Only reactant snapshots are shown for the T20, T8, and T13 systems. Representative structures of the TS and Product are needed and should be provided either in the figure or Supplementary Information.

AUTHOR RESPONSE:

Figure 5B was updated to a generalized 2D schematic diagram. Snapshots of the TS and products as well as the oxocarbenium intermediate are provided in the Supplementary Information Figure S24

1.4 Figure clarity and visual quality. In Figure 5B, the distance indicators shown as dashed lines are too thin to be legible and should be drawn thicker. All relevant interatomic distances should be clearly marked. In Figure 5A, certain structural lines, particularly the leftmost GalNAc 4-OH group, are too faint and should be redrawn to improve visibility.

AUTHOR RESPONSE:

Figure 5B was changed to an explicit 2D representation. Figure 5A was redrawn and exported at a higher resolution to improve visibility.

1.5 Data availability and reproducibility. To allow structural validation, the Cartesian coordinates of the QM region for the Reactant, TS, and Product structures should be provided in the Supplementary Information.

AUTHOR RESPONSE:

We have not provided Cartesian coordinates as they are not requisite for the interpretation of these results. The plots in Figure 5C display minimum energy pathways computed from an ensemble of trajectories (detailed in Section 4.7, paragraph 4). These plots serve to compare reaction barrier heights across the three T glycosylation sites, utilizing the well-documented FEARCF method to compute free energy curves.

Chemically detailed structures are unnecessary for a high-level average thermodynamic and kinetic study of this cellular model. The computational models support the LC-MS analysis regarding T site sialylation but are not essential for validating the Golgi experimental model. Finally, the raw trajectory data cannot be disclosed at this time as the work is currently undergoing an IP protection evaluation process.

2. Molecular Docking Analysis

2.1 Unrealistic distances and catalytic relevance. COM distances in Table S2 (13.66 Å T13 to donor; 9.72 Å T13 to His567) indicate significant separation, whereas reactive atoms should be within ~3–4 Å. These data do not support the statement that T13 is “orientated favourably to react.” To support this claim, the authors should either provide close-up views of the relevant docking poses showing Thr side chain orientation, the catalytic pocket, and atom-to-atom distances, or clearly qualify the text to indicate that the docking only suggests a potential binding region, not a catalytic mechanism.

AUTHOR RESPONSE:

Text in the supplementary information was updated to clarify that the COM distance was used to draw attention to potential substrate binding region and not catalytic mechanism.

2.2 Figure S23 caption and panel discrepancies. The caption for Panel III describes T20, but the image corresponds to T13. The main text also references Figure S23 IV, which does not exist. The caption should be corrected, and the reference to the nonexistent panel removed.

AUTHOR RESPONSE:

Discrepancy in caption corrected. Reference to nonexistent panel replaced with correct reference in

main text.

Reviewer #4 (Remarks to the Author):

The reviewers suggest improvements before publication, mainly related to the LC-MS data confirming the synthetic results. This is of critical importance considering that LC-MS was the only analytical tool to validate their computational and synthetic results.

1. In comment #2 of our review, tandem MS experiments were requested to confirm the structures of the proposed glycopeptides. In the revised version of the supporting materials, the authors state: "The identity of the target peptide peaks was confirmed through MS² analysis (not shown), in which the corresponding precursor ions exhibited characteristic fragment ions, including b- and y-ion series and diagnostic glycan oxonium ions." However, in their reply, the authors say: "MS² spectra were not included, and only primary ions were used in the analysis. However, the site-specific digestion with IMPa used with the synthesis carried out on the double-repeat MUC1 (added to this revised version) improved the site occupancy analysis to a level sufficient to draw the conclusions made in the manuscript." While the use of IMPa digestion and subsequent LC-MS analysis indeed improves the reliability of the results, some of the conclusions regarding the glycopeptide structures remain overstated. For example, IMPa digestion apparently generated species with identical sequences (e.g., see the M1 sequences in Figures S15–20), which raises the question: how can the authors be certain that both of potential glycosylation sites are occupied?

Author response :

We thank the reviewer for this important point and fully agree with the concern. The limitation arises inherently from the use of an identical double-repeat MUC1 sequence, where IMPa digestion followed by LC-MS generates some fragments that cannot be uniquely assigned to either the N- or C-terminal repeat. For this reason, whenever ambiguity identified, all corresponding potential glycosylation sites were annotated for both repeats (this statement was added to the captions in Figure 6 and Figure 7 and as a statement in the LC-MS analysis section in SI file).

Importantly, the observed glycopeptide compositions and site-occupancy patterns—though not always assignable to a specific repeat—matched in full the catalytic preferences inferred from the single-repeat 23-mer data and previously established GALNT mechanisms. And most importantly, solidified the conclusion of the structural deviation influenced by the tumor vs normal distribution pattern of GALNTs. Thus, while the 23-mer enabled clear mechanistic interpretation, the double repeat reproduced the same expected structural pattern—even though repeat preference could not be unambiguously resolved in some cases. Taken together, these concordant observations support the structural interpretations and mechanistic conclusions presented in the manuscript.

2. Although it has been advised to use the Roepstorff–Fohlman nomenclature for naming fragment ions (b- and y-ions) to improve comprehensibility, it has not been performed on Figure S11.

Author response : Roepstorff–Fohlman nomenclature for naming fragment ions (b- and y-ions) was performed on the MS2 spectrum of figure S11

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer's response

This is now the third time this manuscript has been presented to review. The authors seem to largely ignore comments that have been written twice to them. Sadly enough, this review round consists of just repeating previously said comments in an unnecessary harsher tone.

Overall, the QM/MM reporting is as unsatisfactory as previously.

It is puzzling why the authors refuse to report the data they already acquired, at least in SI. As is, it is deeply unsatisfactory. It is not informative enough and can hardly be evaluated. There is only Fig 5, in which the caption does not fully inform on what is displayed, and S24.

As a bare minimum, the authors should report on all the key distances and angles, for each trajectory, at each of the key points (reactants, transition states, intermediates, products). They should also display readable snapshot figures of each of said key points.

Author response:

Our latest response is in red, and changes in the main text or SI are highlighted in yellow

We would like to provide additional context regarding the scope of mechanistic detail presented in this manuscript. In earlier work on OGT, a cancer target, we fully described and presented details of the enzymatic reaction mechanism and transition state for that glycosyltransferase system using QM/MM analysis (see Naidoo, et al., "Multidimensional Free Energy and Accelerated Quantum Library Methods Provide a Gateway to Glycoenzyme Conformational, Electronic, and Reaction Mechanisms," *Accounts of Chemical Research* 54, 2021, DOI: 10.1021/acs.accounts.1c00477). This detailed mechanistic framework has been cited and used in subsequent efforts by others to develop small-molecule inhibitors (e.g., Wang, et al., "Insights into the missing apiosylation step in flavonoid apiosides biosynthesis of Leguminosae plants," *Nature Communications* 14, 2023, DOI: 10.1038/s41467-023-42393-1). We have been advising that the commercialization of uniquely developed OGT products could not be fully protected by patents as we had disclosed key elements of our design process. Examples such as this demonstrate our commitment to mechanistic transparency while also illustrating the challenges of balancing detailed disclosure of structures and TS data that form part of our in-house downstream medicinal chemistry objectives. In the present study, we use the relative reaction rates derived from ensembled averages across a conflated reaction coordinate that is relevant to the experimental kinetics. Here we do not want to distract from the thermodynamic and kinetic observations with a detour into detailed chemical QM/MM analysis that provides mechanistic insight supported by detailed chemical structural analysis. We have clarified this scope and its limitations in the revised manuscript.

Our points are two fold 1) To provide the reader with a full disclosure of all QM/MM trajectories and transition-state geometries would go beyond what is necessary to support the central conclusions of this work and is not essential for assessing the validity of the mechanistic arguments we advance. 2) An unnecessary inclusion of chemical details such as labelled snapshots, TS distance and angles descends into chemical details that are not necessary and moreover places important laboratory projects at risk.

That said, we are committed to transparency and reproducibility within these constraints. To this end, we provide exported MP4 visualizations of representative product-forming trajectories (available on GitHub), which provides visual evidence for the key chemical events and mechanistic features discussed in the manuscript. These materials provide direct evidence of the simulations performed while avoiding unnecessary disclosure of sensitive quantitative details. Furthermore, the request for full trajectory data is no different from asking for angles and distances as the latter can be computed/derived from the former.

We have with this revision, and the previous three submissions resolved the technical concerns related to the in vitro Golgi models we present. We believe we have now provided the maximum level of transparency possible with a focus on relevant evidence while avoiding the over sharing of sensitive data. However, if these revisions are still considered insufficient because of the suggestion that there be an inclusion of structural chemical reaction atomistic detail, we will remove and make no mention of the QM/MM work. In that scenario, we would remove the reaction dynamics content and retain only the minimal docking/binding data that supports the preference for the PDTR region. This would focus the report on the in vitro Golgi models, the primary claim of this study, with any hints of atomistic level chemistry, that may distract readers, removed.

Author response with regard to Figure 5 caption:

The figure 5 caption has been revised to provide improve clarity page 9-10.

Reviewer 1: Below are in green previous comments, and in blue the authors answer to those. Black is for the new answer.

It is a fantastic result to have managed to refine the energy barriers to feasible ones, and we would like to congratulate the authors for doing so. However, “no chemical detail” is definitely not satisfactory. Please add some clear and readable snapshot along the QM/MM simulation, at least for each trajectory with a reported barrier of initial state, transition state, and product. I understand that this is not the focus of the authors, but they are currently reporting trajectories and energy barriers that we cannot evaluate. Those can easily be provided in 1-3 additional SI figures. The authors added some information and lengthy discussion on their very low-level calculations (protein model generation and docking), but absolutely nothing on the high-end computations they did. I would also propose that the authors upload their full trajectories on a public repository, e.g. Zenodo, with a doi provided in the “data availability section”.

Author response:

Initial reactant coordinates and an exported trajectory video of the product formation are available on Github (<https://github.com/scientificcomputing/NCOMMS-25-06286-T-MUC1-in-vitro-Synthesis-/tree/main>).

Reviewer 1: In the repository, there are a couple of movies I cannot read, and a few files without any guidance / information. I did not find any actual coordinates.

Latest author response:

The GitHub repository was updated. PDB files of the initial coordinates from which QM/MM reaction dynamic simulations were run are shared for each system as well as an mp4 movie of a product forming trajectory. This description was also added to a README file in the GitHub repository.

-Data availability statement was updated in the main text and reporting summary

Reviewer 1: Table S2 is also not satisfactory: “Table S2. Distances between centers of masses of the possible glycosylation sites and key residues involved in the reaction 252 i.e. the CMP-Neu5Ac donor and catalytic base HIS567.” “Centers of mass” is not an appropriate measure point. You would get a lower distance if the donor was in a totally not reactive position, but with the phosphate close to the base. Also, present in the caption that those measures are for the TS*. If a distance should be measured between the catalytic base and the Neu5Ac donor, I suggest between the reactive nitrogen of His567 and the C1 of the donor. Ideally, I would also like to see the distance between the reactive nitrogen of His567 and the proton to be abstracted on the Ser/Thr. That being said, we would still like to stress that it is only a minor disagreement on how to present the data, and we would like our main message as an answer to the QM/MM part to be commending the authors for a great revision and refinement of their work.

Author response :

Table S2 does not discuss any distances related to the TS as the GalNAc residues are not considered. The protein-protein docking performed is used to determine the initial peptide binding poses and relative preferences of binding region for the MUC1 peptide. This is not intended to be part of a molecular catalytic mechanism study. The docking study undertaken in following suggestions by previous reviews' comments.

Distances involving catalytically relevant atoms are reported in Figure 5. The reference to Figure S2 was removed.

Reviewer 1: Again, the authors should spend more of the reader's time on QM/MM results, and less on docking ones.

Latest Author response to previous comments regarding COM:

We agree that a center-of-mass (COM) distance is not, by itself, a chemically rigorous descriptor of catalytic competence. COM distance was therefore not intended as a proxy for reactivity or transition-state proximity. Rather, it was used as a average area of the peptide to identify poses generated by protein-protein docking that place the donor sugar in the vicinity of the acceptor site. This is now clarified in the Molecular docking section of the Supplementary Information **Page S25**. All conclusions regarding catalytic relevance and reaction geometry were based on subsequent QM/MM simulations with atomistic distances and angles calculated and reported in Figure 5B.

Author response to spending more of the readers time on QM/MM results:

We have expanded the QM/MM details to improve the reproducibility and to clarify the scope and purpose of the QM/MM simulations. The main purpose of the QMMM and reaction dynamics simulations in this work is to provide qualitative and comparative mechanistic insight into the site preference rather than absolute and in-depth calculations.

To improve transparency and reproducibility, we have added detailed descriptions of the QM/MM partitioning scheme, force fields, QM method, etc. (**Main text section 2.3.1 page 8**). These additions ensure that the simulations can be independently evaluated and reproduced.

Reviewer 1: There is no logic in making this argument for k_{cat} and not for K_m or V_{max} . Some argue that any k_{cat} is an averaged result of all potential substrates. Substrate heterogeneity by itself does influence the kinetic profile but does not precludes to obtain an analytical solution to the Michaelis Menten equation, see:

<https://www.biorxiv.org/content/biorxiv/early/2025/08/24/2025.08.21.671619.full.pdf> This is a very minor point though.

Author response: We thank the reviewer for this comment and the reference provided. While k_{cat} and k_{cat}/K_m could, in principle, be calculated as averages across multiple reactive sites, we deliberately chose not to report these values for our multi-site substrates. Calculating k_{cat} for such peptides could be misleading because not all sites are equally active—some are kinetically inactive or have different turnover rates, as later revealed by our structural and computational analyses. Reporting a single turnover number or catalytic efficiency could be misinterpreted as reflecting site-specific reaction energy barriers or enzyme preference, which we explicitly address through detailed LC-MS and computational studies. Instead, we report apparent K_m and V_{max} values as empirical measures of overall reaction behavior. These initial parameters served the purpose of guiding our subsequent site-specific analysis, where K_m was recalculated based on the number of reactive sites, which eventually led to providing accurate mechanistic insights without introducing potentially misleading interpretations.

Reviewer 1: I feel that the authors should read the comments made to them before answering them, and read the provided references rather than just thanking the referee for having provided them.

Latest author response:

The kinetics table in Figure 4 has been updated in this revision to include k_{cat} and catalytic efficiency (k_{cat}/K_m) values, which were newly calculated and added. In line with the reviewer's suggestion, we now explicitly state in the figure caption that all kinetic parameters represent apparent values, reflecting the combined contributions of multiple glycosylation sites. This interpretation is supported by, and now explicitly cites, the reference recommended by the reviewer.

These changes were implemented in agreement with the reviewer's conceptual point raised in the previous round of comments. Accordingly, the discussion in Sections 3.2 and 2.5.1 has been revised to interpret the kinetic behavior primarily using k_{cat} and k_{cat}/K_m , rather than relying on V_{max} and K_m alone.

We believe that this revision provides a more standardized and intrinsic characterization of enzyme performance, while remaining transparent about the limitations imposed by multi-site substrate heterogeneity.

Changes in main text:

-section 2.3 page 6-7

- Figure 4: Kinetics table and figure captions

-Section 2.5.2 last paragraph page 12-13

Reviewer 1: We would also like to suggest that the authors group figure 6A with 7A in figure 6, and to group figure 6B and 7B in figure 7. Having the site-to-site comparison of the localization of GALNT's stresses the point and makes the message clearer for the reader.

Author response: We appreciate the reviewer's thoughtful suggestion to group Figures 6A with 7A and 6B with 7B, to present the experimental panels side by side and the corresponding schematic conclusions side by side. We fully understand how such an arrangement could emphasize the site-to-site comparison of the products associated with the two modes of GALNT localization in normal versus cancer-associated models and potentially enhance clarity.

However, after evaluating this option, we found that combining these panels would result in an excessively large and dense figure, particularly for Figure 6, which contains extensive experimental data from both models. The amount of detail—both in volume and visual complexity—would overwhelm the reader and compromise the readability and visual clarity of the experimental results.

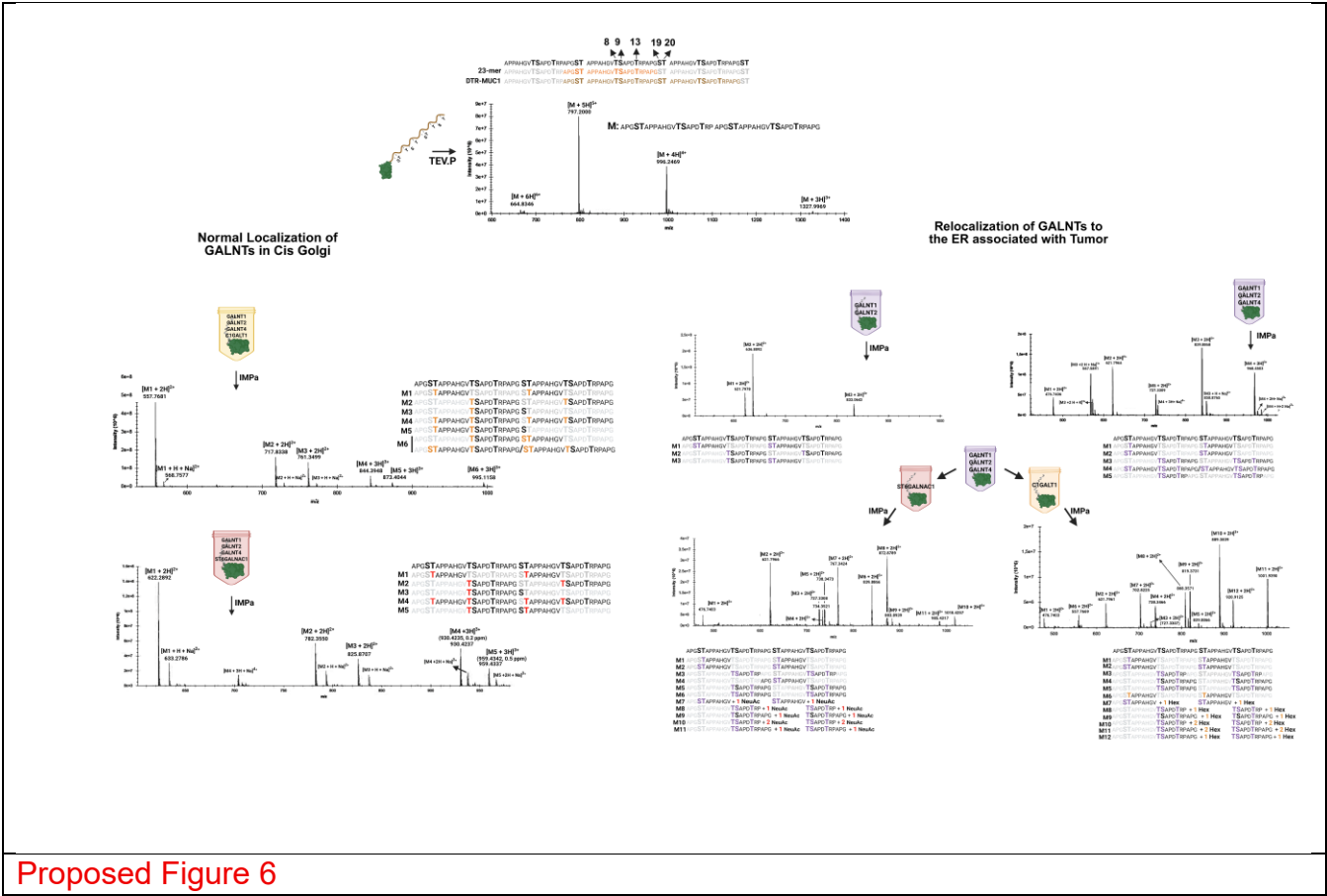
For these reasons, we opted to maintain the separation between Figures 6 and 7, where Figure 6 presents both the experimental results and the derived schematic model for the normal system, and Figure 7 presents the corresponding experimental results and schematic model for the cancer-associated system. This layout allows a clear distinction between the normal and cancer-associated systems, preserves the intended narrative flow, and improves overall readability by presenting each system in a self-contained, coherent figure.

We hope the reviewer agrees that this organization best supports clarity and interpretability of the presented data.

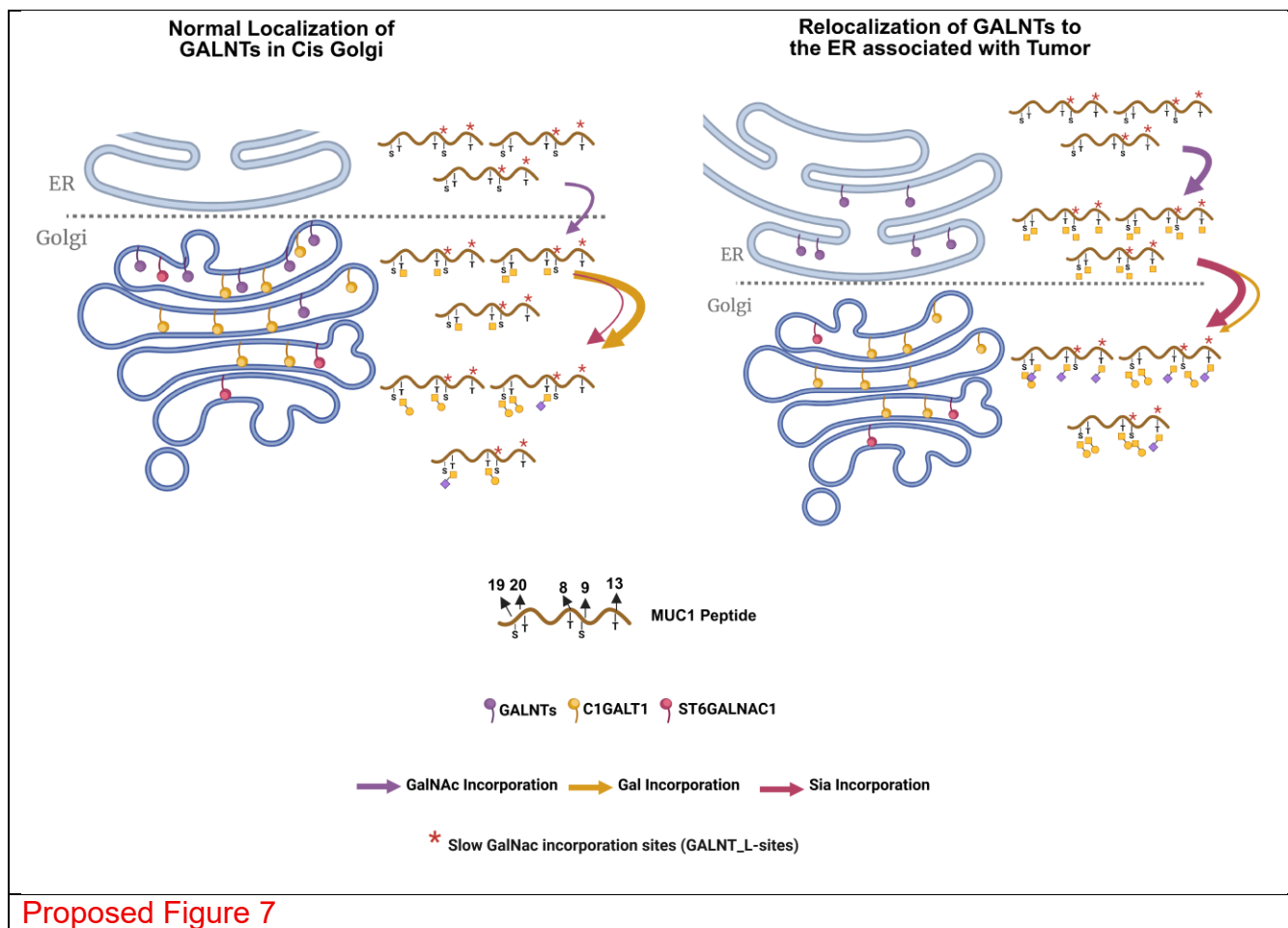
Reviewer 1: Maybe the authors could show the reviewers these excessively dense figures. Given the level of efforts the authors.

Latest Author response:

We provided the regrouped Figures as requested (see below). However, Figures 6 and 7 are each designed as self-contained representations of two distinct biological models (normal vs cancer-associated). This pairing is essential to preserve the logical flow from data to interpretation within each biological context. Additionally, combining Panels 6A and 7A would require assembling two highly information-dense LC–MS datasets into a single, wide and cumbersome figure, which would substantially reduce readability and visually separate experimental data from its corresponding interpretation. Given that Sections 2.5.1 and 2.5.2 are structured to treat the two systems independently, we believe the figure organization in its normal (figure 6) vs cancer (Figure 7) format best supports clarity, narrative flow, and data interpretability.



Created in BioRender. Naidoo, K. (2026) <https://BioRender.com/qban85z>
Created in BioRender. Naidoo, K. (2026) <https://BioRender.com/d7iikxz>



Proposed Figure 7

Created in BioRender. Naidoo, K. (2026) <https://BioRender.com/qban85z>

Created in BioRender. Naidoo, K. (2026) <https://BioRender.com/d7iikxz>

“An in vitro one-pot synthetic biology approach to simulating diverging Golgi O-glycosylation of tumor-associated MUC1 from normal tissue MUC1” by Abdullateef Nashed and coauthors is a very interesting paper in the field of protein O-glycosylation. The authors addressed well the various reviewer’s requests. Below are a few comments in green (black for previous review, blue and red for authors responses), mostly on the representation of the QM/MM data.

[...]

QM/MM simulations. The first two glycosylation reactions show activation energy of 25 kcal/mol, which means that there is basically no reaction. The authors justify an observed difference of less than 2-fold in rate by modelling showing a 7.5 kcal/mol difference, which amount for a 100 000 fold difference in rate. They justify the slow rate by some far-fetched mechanism in which the catalytic base would actually be covalently bond to the sugar donor, making it a suicide inhibitor. Worse, they actually do not show anything of their modelling. There is not a single figure in the SI, and in the main text only one figure that does not show any conformation, distance, etc. The authors should at least provide some of those. As well, they should tune down some interpretation on their simulations given the results they obtain. They could also remove the two simulations on T20 glycosylation and only keep the one on T23, which provides a reasonable energy barrier, and expand its description.

AUTHOR RESPONSE:

We agree with reviewer's comments regarding the unrealistic high activation energies. We therefore repeated the QM/MM study focusing only on the cancer-associated fully glycosylated MUC1. Reaction dynamics calculations were performed with increased sampling and corrected starting structure to obtain reasonable barrier heights for the glycosylation reactions. After these improvements have been made, we report activation energies of 17.11 kcal/mol for the T20 site, 23.87 kcal/mol for the T8 site, and 16.73 kcal/mol for the T13 site sialylation reactions, as seen in Figure 5C.

Necessary figures have been included in Figure 5 and SI. There is no need for chemical detail as this is not a manuscript that describes chemical mechanistic differences between sites. The function the simulations was to compare and determine the viability of ST6GALNAC1 catalysis at each of the 5 sites. The comparison is made through simple docking experiments, which are limited on their own, as well as sophisticated computational modelling to determine barrier heights. While we have in previous works (e.g., <https://pubs.acs.org/doi/abs/10.1021/acs.accounts.1c00477>) performed in depth electronic and atomic level analysis in pursuit of chemical mechanistic analysis this would be a distraction from the objective to build and demonstrate in vitro molecular models of normal vs. tumour golgi construction.

In line with the reviewer's concern, we have provided additional figures in the Supporting Information Figure S21, S22, and S23. Close up images of important residues involved in the reaction and relevant distances and angles across the QM/MM simulation are provided in Figure 5B Additional. The reported distances and angles are discussed in Section 2.3.1 to clearly illustrate the mechanistic features observed in our simulations.

It is a fantastic result to have managed to refine the energy barriers to feasible ones, and we would like to congratulate the authors for doing so. However, "no chemical detail" is definitely not satisfactory. Please add some clear and readable snapshot along the QM/MM simulation, at least for each trajectory with a reported barrier of initial state, transition state, and product.

I understand that this is not the focus of the authors, but they are currently reporting trajectories and energy barriers that we cannot evaluate. Those can easily be provided in 1-3 additional SI figures.

The authors added some information and lengthy discussion on their very low-level calculations (protein model generation and docking), but absolutely nothing on the high-end computations they did. I would also propose that the authors upload their full trajectories on a public repository, e.g. Zenodo, with a doi provided in the "data availability section".

Table S2 is also not satisfactory: "Table S2. Distances between centers of masses of the possible glycosylation sites and key residues involved in the reaction 252 i.e. the CMP-Neu5Ac donor and catalytic base HIS567."

"Centers of mass" is not an appropriate measure point. You would get a lower distance if the donor was in a totally not reactive position, but with the phosphate close to the base. Also, present in the caption that those measures are for the TS*. If a distance should be measured between the catalytic base and the Neu5Ac donor, I suggest between the reactive nitrogen of His567 and the C1 of the donor. Ideally, I would also like to see the distance between the reactive nitrogen of His567 and the proton to be abstracted on the Ser/Thr.

That being said, we would still like to stress that it is only a minor disagreement on how to present the data, and we would like our main message as an answer to the QM/MM part to be commending the authors for a great revision and refinement of their work.

[...]

Fig 4A, Vmax as a value, kcat could be included, and kcat/Km.

AUTHOR RESPONSE: We thank the reviewer for pointing this out. In this case, the substrate peptides contain multiple reactive sites, making it impossible to attribute the kinetic parameters to a single reaction per molecule. Therefore, kcat is not meaningful

in this context, as the turnover number cannot be linked to a defined thermodynamic step. Instead, we report apparent K_m and V_{max} values, which serve as averaged proxies for substrate affinity and maximal reaction rate, respectively. This provides a more appropriate comparative measure across enzymes and substrates. We have clarified this in the caption of Fig. 4A

There is no logic in making this argument for k_{cat} and not for K_m or V_{max} . Some argue that any k_{cat} is an averaged result of all potential substrates. Substrate heterogeneity by itself does influence the kinetic profile but does not precludes to obtain an analytical solution to the Michaelis Menten equation, see:

<https://www.biorxiv.org/content/biorxiv/early/2025/08/24/2025.08.21.671619.full.pdf>

This is a very minor point though.

Additional points

The mass spectrometry nomenclature is quite confusing, as the authors move between "MS² analysis" (SI I114 and I116), MS2 (main text I508), "Lc-MS" or "LC-MS" (please chose one, better "LC-MS", e.g. Figs S4, S5 and S11, S13, S14, S15, S16, S17); "MS¹" (Figure S6), also MS1 (Figure 6A, 7A, line 508), MS 1 , M1, etc. Please homogenise and check all instances thoroughly.

L162: "However, the observed glycosylation of the remaining two sites cannot be attributed", the discussion would be easier to follow if the sites were immediately specified, e.g. "However, the observed glycosylation of the remaining two sites (S9 and T13) cannot be attributed".

The following request has been ignored by the authors, or did we miss the response?

"Lectin dependent mechanism. To further confirm which glycosylation are lectin-dependent or not, why not express only the catalytic domain? Or at least a single point mutant of the lectin binding site? If it is known it cannot be done, the authors should explain why."

We would also like to suggest that the authors group figure 6A with 7A in figure 6, and to group figure 6B and 7B in figure 7. Having the site-to-site comparison of the localization of GALANT's stresses the point and makes the message clearer for the reader.

Please make sure, that the figures are exported in high quality, as the details in some of the figures are either not visible or difficult to see. This issue concerns the text in figure 3B, the charges in figure 5A, and S' and T's in figure 7B.

This is now the third time this manuscript has been presented to review. The authors seem to largely ignore comments that have been written twice to them. Sadly enough, this review round consists of just repeating previously said comments in an unnecessary harsher tone.

Overall, the QM/MM reporting is as unsatisfactory as previously.

It is puzzling why the authors refuse to report the data they already acquired, at least in SI. As is, it is deeply unsatisfactory. It is not informative enough and can hardly be evaluated. There is only Fig 5, in which the caption does not fully inform on what is displayed, and S24.

As a bare minimum, the authors should report on all the key distances and angles, for each trajectory, at each of the key points (reactants, transition states, intermediates, products). They should also display readable snapshot figures of each of said key points.

Below are in green previous comments, and in blue the authors answer to those. Black is for the new answer.

It is a fantastic result to have managed to refine the energy barriers to feasible ones, and we would like to congratulate the authors for doing so. However, "no chemical detail" is definitely not satisfactory. Please add some clear and readable snapshot along the QM/MM simulation, at least for each trajectory with a reported barrier of initial state, transition state, and product.

I understand that this is not the focus of the authors, but they are currently reporting trajectories and energy barriers that we cannot evaluate. Those can easily be provided in 1-3 additional SI figures.

The authors added some information and lengthy discussion on their very low-level calculations (protein model generation and docking), but absolutely nothing on the high-end computations they did. I would also propose that the authors upload their full trajectories on a public repository, e.g. Zenodo, with a doi provided in the "data availability section".

AUTHOR RESPONSE:

Initial reactant coordinates and an exported trajectory video of the product formation are available on Github (<https://github.com/scientificcomputing/NCOMMS-25-06286-T-MUC1-in-vitro-Synthesis-/tree/main>).

In the repository, there are a couple of movies I cannot read, and a few files without any guidance / information. I did not find any actual coordinates.

Table S2 is also not satisfactory: "Table S2. Distances between centers of masses of the possible glycosylation sites and key residues involved in the reaction 252 i.e. the CMP-Neu5Ac donor and catalytic base HIS567." "Centers of mass" is not an appropriate measure point. You would get a lower distance if the donor was in a totally not reactive position, but with the phosphate close to the base. Also, present in the caption that those measures are for the TS*. If a distance should be measured between the catalytic base and the Neu5Ac donor, I suggest between the reactive

nitrogen of His567 and the C1 of the donor. Ideally, I would also like to see the distance between the reactive nitrogen of His567 and the proton to be abstracted on the Ser/Thr. That being said, we would still like to stress that it is only a minor disagreement on how to present the data, and we would like our main message as an answer to the QM/MM part to be commending the authors for a great revision and refinement of their work.

AUTHOR RESPONSE:

Table S2 does not discuss any distances related to the TS as the GalNAc residues are not considered. The protein-protein docking performed is meant to determine the initial peptide binding poses and potential binding region for the MUC1 peptide, not the molecular catalytic mechanism.

Distances involving catalytically relevant atoms are reported in Figure 5. The reference to Figure S2 was removed.

Again, the authors should spend more of the reader's time on QM/MM results, and less on docking ones.

There is no logic in making this argument for k_{cat} and not for K_m or V_{max} . Some argue that any k_{cat} is an averaged result of all potential substrates. Substrate heterogeneity by itself does influence the kinetic profile but does not precludes to obtain an analytical solution to the Michaelis Menten equation, see:

<https://www.biorxiv.org/content/biorxiv/early/2025/08/24/2025.08.21.671619.full.pdf>

This is a very minor point though.

AUTHOR RESPONSE: We thank the reviewer for this comment and the reference provided. While k_{cat} and k_{cat}/K_m could, in principle, be calculated as averages across multiple reactive sites, we deliberately chose not to report these values for our multi-site substrates. Calculating k_{cat} for such peptides could be misleading because not all sites are equally active—some are kinetically inactive or have different turnover rates, as later revealed by our structural and computational analyses. Reporting a single turnover number or catalytic efficiency could be misinterpreted as reflecting site-specific reaction energy barriers or enzyme preference, which we explicitly address through detailed LC-MS and computational studies. Instead, we report apparent K_m and V_{max} values as empirical measures of overall reaction behavior. These initial parameters served the purpose of guiding our subsequent site-specific analysis, where K_m was recalculated based on the number of reactive sites, which eventually led to providing accurate mechanistic insights without introducing potentially misleading interpretations.

I feel that the authors should read the comments made to them before answering them, and read the provided references rather than just thanking the referee for having provided them.

We would also like to suggest that the authors group figure 6A with 7A in figure 6, and to group figure 6B and 7B in figure 7. Having the site-to-site comparison of the localization of GALNT's stresses the point and makes the message clearer for the reader.

AUTHOR RESPONSE:

We appreciate the reviewer's thoughtful suggestion to group Figures 6A with 7A and 6B with 7B, to present the experimental panels side by side and the corresponding schematic conclusions side by side. We fully understand how such an arrangement could emphasize the site-to-site comparison of the products associated with the two modes of GALNT localization in normal versus cancer-associated models and potentially enhance clarity.

However, after evaluating this option, we found that combining these panels would result in an excessively large and dense figure, particularly for Figure 6, which contains extensive experimental data from both models. The amount of detail—both in volume and visual complexity—would overwhelm the reader and compromise the readability and visual clarity of the experimental results.

For these reasons, we opted to maintain the separation between Figures 6 and 7, where Figure 6 presents both the experimental results and the derived schematic model for the normal system, and Figure 7 presents the corresponding experimental results and schematic model for the cancer-associated system. This layout allows a clear distinction between the normal and cancer-associated systems, preserves the intended narrative flow, and improves overall readability by presenting each system in a self-contained, coherent figure.

We hope the reviewer agrees that this organization best supports clarity and interpretability of the presented data.

Maybe the authors could show the reviewers these excessively dense figures. Given the level of efforts the authors.