Article:
Bilyard, Matthew K, Bailey, Henry J, Raich, Lluís et al. (8 more authors) (2018) Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. Nature. 235–240. ISSN: 0028-0836

https://doi.org/10.1038/s41586-018-0644-7

Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis

Matthew K. Bilyard[1], Henry Bailey[2], Lluís Raich,[3] Maria Gafitescu[1]†, Takuya Machida[1]†, Javier Iglésias-Fernández,[3,5] Seung Seo Lee[1,6], Christopher D. Spicer[1], Carme Rovira[3,4], Wyatt. W. Yue[2*] and Benjamin G. Davis[1*]

[1]Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford, OX1 3TA, UK

[2]Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7DQ, UK

[3] Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) and Institut de Química Teòrica i Computacional (IQTC), Universitat de Barcelona, Barcelona, Spain

[4] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, 08020 Barcelona, Spain.

[5] Current Address: Institut de Química Computacional i Catalisi and Departament de Química, Universitat de Girona, Girona, Spain

[6] School of Chemistry, University of Southampton, Southampton, SO17 1BJ

† These authors contributed equally.

*e-mail: wyatt.yue@sgc.ox.ac.uk, ben.davis@chem.ox.ac.uk

Dedication: This paper is dedicated to the memory of Prof. Peter Reilly (Iowa State University), a friend, mentor and colleague who is still very much missed.

Biosynthesis of glycogen, the essential glucose (and hence energy) storage molecule in humans, animals and fungi,[1] is initiated by glycosyltransferase enzyme glycogenin (GYG) (ED Figure 1). Deficiencies in glycogen formation cause neurodegenerative and metabolic disease (ED Figure 1b).[2-4] Mouse knockout[5] and inherited human mutations[6] of GYG impair glycogen synthesis. GYG acts as a 'seed core' for the formation of the glycogen particle by catalyzing its own stepwise auto-glucosylation to form a covalently-bound gluco-oligosaccharide chain at initiation site Tyr195. To date, an inability to access homogeneous glycoforms of this protein, which unusually acts as both catalyst and substrate, has precluded precise mechanistic studies. Here we show that, unprecedented, direct access to different, homogeneously glucosylated states of GYG can be accomplished through a palladium-mediated enzyme activation 'shunt' process using on-protein C–C bond-formation. Careful mimicry of GYG intermediates recapitulates catalytic activity at distinct stages, which in turn allows discovery of tri-phasic kinetics and substrate plasticity in GYG's use of sugar substrates. This reveals a tolerant but 'proof-read' mechanism that underlies the precision of this vital metabolic process. This demonstration of direct, chemically-controlled access to intermediate states of active enzymes suggests that such ligation-dependent activation could be a powerful tool in the study of mechanism.

The initial anchor point for the dendron-like structures that make up glycogen is the Tyr195 residue of GYG (GYG1 numbering); glycogenesis is therefore a striking example of α-linked protein autoglucosylation.[7] Prior studies have suggested GYG to be a dimeric,[8] Mn^{2+}-dependent enzyme belonging to the GT-8 family of retaining glycosyltransferases.[9,10] GYG is – by virtue of its self-modifying nature – non-identical for each glucosylation step; i.e. GYG, unlike nearly all biosynthetic enzymes, is strictly *not* a catalyst since it is itself changed at each step. This leads potentially to altered activity for each intermediate state and presumably to eventual inactivity once 'buried' in glycan. This opens up the unusual possibility of distinct sub-phases and mechanisms occurring at different oligosaccharide chain lengths; crystal structures suggest possible intra-monomeric and inter-monomeric glucosylation modes within the GYG protein dimer.[10] Whilst bespoke biosynthetically-deficient expression host strains can generate a glycan-free, starting form of GYG,[11] this allows access to only one catalyst state (**Supplementary Text**). As a result, any possible '(sub)phases' subsequent to this starting state may be obscured if they follow faster kinetics. A lack of access to homogeneous GYG catalyst states therefore restricts our current understanding.

We reasoned that chemical construction of pure GYG in different glucosylation states might allow a strategy for direct, guided ('shunted') activation (and hence interrogation) of chosen intermediate states (**Figure 1** and **ED Figures 1c,2a**). The unusual hybrid nature of these catalyst states – part-catalyst-part-substrate – suggested a convergent (tag-and-modify[12]) construction

process in which the desired (glycosyl acceptor) glycan moiety would be covalently attached in one-step to key catalytic site 195 (**ED Figure 1c**). We have previously demonstrated that Pd(0)-mediated C–C-bond forming ligation is feasible and benign in certain biological contexts.[13-17] Pd-mediated approaches in biology have since been elegantly exploited by various groups.[18-20] However, GYG is a testing target biomolecule on which to apply this method. Not only is site 195 in the heart of the active site, but GYG is also metal-dependent, raising the possibility of inhibitory 'poisoning' cross-competition[21,22] by Pd at the metal co-factor site.

A suitable precursor GYG1 bearing a reactive 'tag' for Pd(0)-mediated C–C bond formation was generated via site-specific unnatural amino acid incorporation,[14,23,24] giving a variant in which the *para*-hydroxy group of the natural, wild-type (wt) tyrosine residue at site 195 was exchanged for an iodide atom (OH→I, GYG-Tyr195→GYG-*p*IPhe195, **Figure 1**). Characterization confirmed no deleterious effects on overall enzyme structure. The structure of GYG-Y195*p*IPhe, determined in both *apo* (2.2 Å) and Mn^{2+}+UDP bound (2.4 Å) states (**Supplementary Table S1**), revealed highly superimposable dimers to those in GYG-wt[10] (**ED Figure 2c**). In the ligand-bound state, the *p*IPhe195 group from one monomer is clearly visible (**ED Figure 2d**, inset), located within a partially unwound helix that adopts a catalytically poised position equidistant to either active site of the dimer (**ED Figure 2d**, red). Asymmetry at the dimer interface, consistent with previous unglucosylated GYG-wt structures[10], suggested likely conformational flexibility needed as GYG transitions from unconjugated to differently glucosylated forms.

Studies on wild-type GYG (GYG-wt / GYG-Tyr195) revealed concentration-dependence of Pd inhibition and hence determination of essentially benign Pd concentrations that would successfully allow preservation of enzymatic activity (**ED Figure 3** and **Supplementary Note**); other cross-coupling components had minimal effect. These conditions allowed successful Pd-mediated $C(sp^2)$–$C(sp^2)$ ligation of GYG-*p*IPhe195 to a variety of designed, systematically-altered 'substrate templates' (**Figure 1** and **ED Figure 2a, 4**); all bore nucleophilic, hydroxyl groups as possible reaction sites for auto-glucosylation (readily prepared as their corresponding $C(sp^2)$ boronic acid derivatives **1**, see **Supplementary Methods**). Small amounts of side-products were also identified (**ED Figure 5**): for example, unreacted GYG-*p*IPhe195 or species attributable to dehalogenation[17] using LC-MS analysis and negative control studies (**Supplementary Methods** and **Supplementary Text**). Despite successful Pd-mediated ligation, 'simple' glycan-mimic templates (**ED Figure 4**) provided ineffective mimicry: irrespective of systematically varied nature (orientation, length or pK_A), none led to activation of autoglucosylation. Activation of GYG requires more than just an available hydroxyl nucleophile positioned in the active site. However, for more complex substrate templates displaying glucosyl moieties inside GYG not only did protein LC-MS analysis reveal successful C–C ligation but also

concomitant activation and clear auto-glucosylation activity in the resulting 'shunted' intermediate product GYG-**Glc** as well as the more advanced, chain-extended shunted states GYG-**Glc-Glc** and GYG-$\mathbf{Glc_6}$ (**Figure 2** and **Supplementary Information**). Side-products from cross-coupling (**ED Figure 5**) were inactive to autoglucosylation and thus did not interfere in the assay.

This chemically-generated access to shunted functionally-active intermediate states of GYG along the glycogen biosynthetic pathway allowed us to uniquely probe and compare activity using LC-MS-monitoring of the sugars attached over time (**Figure 2c,d**, **ED Figure 6,7** and **Supplementary Methods**). Immediately contrasting behaviours from different states were observed. For more extended GYG-**Glc**-**Glc**, two distinct glucosylation phases were apparent: rapid glucosylation from 2 until ~4-5 **Glc** total, then significantly slower catalysis thereafter (**Figure 2c,d**). Indeed, the initial step (GYG-**Glc**-**Glc**→GYG-**Glc**-**Glc**-**Glc**) was extremely rapid; on-protein kinetic analyses conducted in replicate (see **Supplementary Methods**) revealed that ~90% of starting GYG-**Glc**-**Glc** was consumed within 20 seconds. In striking contrast, GYG-**Glc** exhibited a more gradual decline in glucosylation rate with increasing oligosaccharide length (**Figure 2c,d**), consistent with a significantly slower initiation sub-phase for GYG-**Glc** (GYG-**Glc**→GYG-**Glc**-**Glc**) that thus obscures the rapid phase immediately following (**ED Figure 6d**). Taken together, these data suggested a triphasic mechanism, in which a rapid intermediate phase is flanked by significantly slower initiation (<2 glucoses) and elongation (> 4/5 glucoses) phases (**ED Figure 6d**). Notably, only through the direct 'shunt' formation of intermediates (GYG-**Glc**, GYG-**Glc**-**Glc** *etc*) achieved through Pd-mediated ligation, was unobscured analysis of each sub-phase made possible (**ED Figure 7**). Clear visualisation of this kinetic profile was a consequence of our ability to both circumvent initial slow Tyr195 glucosylation and also probe discrete glucosylation states immediately after this. The presence of distinct (sub)phases is consistent with the proposed existence of different glucosylation mechanisms for GYG[10,25,26].

Use of 'shunted' intermediates GYG-**Glc**, GYG-**Glc**-**Glc** and GYG-$\mathbf{Glc_6}$ allowed the determination of initial rates that gave apparent rate constants for each associated phase of k_{app} = 0.016, 0.126 and 0.003 s^{-1}, respectively (**ED Figure 7b**). These were also compared directly with kinetics determined from analysis of wild-type GYG in unglucosylated form (GYG-wt-Glc0, **ED Figure 7**). As expected, the inability to access intermediate states for GYG-wt failed to reveal the distinct phases shown by our chemically 'shunted' system. Nonetheless, global values for turnover proved consistent; we now show that one consequence of the triphasic regime is an accumulation of glucosylation at the end of the fast phase 2 mechanism regime (lengths 5-6 Glc) going into the slower phase 3. Taken together, this confirmed quantitative mimicry at similar activity levels and highlighted the need for the chemical 'shunted' approach in revealing detailed mechanism.

Quantum mechanics/molecular mechanics (QM/MM) metadynamics[27,28] simulations (see **Supplementary Methods**) allowed further insight through detailed reconstruction of the free-energy surface of reaction as a function of a few selected degrees of freedom (collective variables, CVs, **Supplementary Methods**). Michaelis complexes equivalent to GYG-**Glc**-**Glc**→GYG-**Glc**-**Glc**-**Glc** (both in wt, GYG-wt-Glc3→GYG-wt-Glc4, and shunted, GYG-**Glc**-**Glc**→GYG-**Glc**-**Glc**-**Glc,** form) were reconstructed from the structures determined here and of those in complex with UDP-Glc and cellotetraose.[10] Both wild-type and shunted forms gave similar results (**ED Figure 8**), consistent with kinetic parameters. The free energy surface revealed a short-lived intermediate (**ED Figure 9**) along the minimum free energy pathway indicative of a front-face, 'S_Ni-like' reaction mechanism (see **Supplementary Video**).[29,30] Notably, the free energy barrier ~10 kcal/mol was very low compared with typical values obtained previously for similar 'S_Ni-like' glucosyl transfer reactions (~20 kcal/mol[30]). Thus, together our kinetic and QM/MM experimental data reveal unprecedentedly fast glycosyl-transfer for the second sub-phase of glycogen formation. The Michaelis complex (R' in **ED Figure 9**) exhibits a near-perfect approach between the O4'-H acceptor bond and the C1–O_P donor bond to assist the departure of UDP. The resulting very short C1···O4' and H···O_P distances (3.3 and 2.0 Å, respectively, *cf* 3.2 and 2.5 Å in prior, representative systems[30]) for formed bonds provide excellent stabilization of charge developed at the phosphate, together with proper orientation for forthcoming front-face nucleophilic attack of O4' onto C1 of Glc. The acceptor O–H in GYG thus creates a direct hydrogen bond H···O_P, unlike prior systems, resulting in a more stretched sugar-phosphate bond (C1–O_P) in GYG (1.58 Å *cf* 1.51 Å[30]) with a much lower associated bond-energy (~10 kcal/mol *cf* ~18 kcal/mol).

To probe the selectivities of this multiphasic GYG mechanism, we next investigated the potential of GYG to use non-glucose sugar substrates.[31] The potential for GYG to use non-glucose *acceptor* sugar moieties has not been examined due to the inability, until now, to directly access requisite intermediate enzyme states and to insert into those states non-glucose sugars. GYG-**Glc** and GYG-**Glc**-**Glc** generated by Pd-mediated ligation were capable of utilising the non-glucose *donor* sugar UDP-Galactose with kinetic profiles essentially qualitatively similar to analogous auto-glucosylation reactions (**Figure 3a**) thereby forming GYG-**Glc**-(**Gal**)$_n$ and GYG-**Glc**-**Glc**-(**Gal**)$_n$. Notably, however, the third kinetic (sub)phase observed for auto-glucosylation was curtailed for auto-*galacto*sylation (**ED Figure 6**). Shunted access to GYG bearing common non-glucose but naturally-occurring mammalian monosaccharides D-galactose (GYG-**Gal**) and D-mannose (GYG-**Man**) (**Figure 3b** and **Supplementary Information**, using boronic acid reagents **1-Gal**, **1-Man**) revealed, remarkably, that both were capable of auto-glucosylation to form both GYG-**Gal**-(**Glc**)$_n$ etc and GYG-**Man**-(**Glc**)$_n$ etc (**Figure 3c**). Kinetic analyses of this non-glucose *acceptor* activity of GYG revealed

glucosylation rates for GYG-**Gal** and GYG-**Man** that are initially lower as a consequence of a significantly slower initiation step / (sub)phase. In contrast to their plasticity towards glucosylation, the non-glucose enzyme states GYG-**Gal** and GYG-**Man** did not catalyse auto-galactosylation to any significant extent (**Supplementary Table 12**). Molecular dynamics (MD) simulations (**ED Figure 9**) suggested that, strikingly, the altered configurations of non-glucose sugars, e.g. **Gal** in GYG-**Glc**-**Gal** or UDP-**Gal**, necessitated slight reorientations but could be accommodated without significantly altering the interactions at the active site with key hydroxyl-binding residues. The result is that the distance of the putative nucleophile (OH-4) from the electrophilic anomeric carbon (in UDP-**Gal** or UDP-**Glc**, respectively) is not greatly perturbed (O···C1 change < 0.5Å) and O–C bond formation can thus, unusually for GTs, evolve essentially 'normally' despite such changes, reflecting this experimentally observed plasticity.

Taken together, distinct mechanistic phases of GYG (**Figure 4**) are therefore defined not only by differential rates but also different donor/acceptor tolerance. Whilst the second, rapid phase (2-4/5 sugars) readily tolerates Gal-to-Gal transfer throughout (species with up to 5 sugars are quickly generated from GYG-**Glc**-**Glc**), the first and third phases show similarities in being linked by not only their slower glucosylation rates but also their apparent lower tolerance of non-glucose in both acceptor and donor at the same time. A plastic and rapid second phase is thus seemingly preceded by a slower step that can nonetheless be primed with unnatural sugars – immediately surprising given the presumed specific role of glycogen as a glucose-storage polymer – and is followed by a slower and much more selective third phase. Together these three phases appear to allow 'priming' with non-glucose sugars in the first phase (e.g. Gal, Man) followed by more rapid and more plastic 'extension' in the second phase (with either UDP-Glc or UDP-Gal) before a third 'refining' phase that ensures use of only glucose in the more extended portions of the inner core of glycogen.

From data gathered here and before,[10] we speculate that these phases may reflect, in part, transitions between intra-monomeric and inter-monomeric modes of glucosylation within the active GYG protein dimer. From our structure of GYG-*p*IPhe195 the anchor point for the oligosaccharide chain of glycogen is essentially equidistant from the two active sites in GYG dimer. Molecular dynamics (MD) simulations (**ED Figure 8**) with GYG bearing Glc-oligomer chains of different lengths (GYG-$\mathbf{Glc_n}$, n = 0-5) and conformations (intra- / inter-monomeric) suggest that the first glucosylation steps (n = 0, 1) are preferentially inter-monomeric. A 'blocking loop' inbetween the acceptor arm and the active site of the same subunit hampers intra-monomeric conformations. In contrast, sugar chains of subsequent steps (n = 2, 3) circumvent the blocking loop, allowing intramonomeric conformations. A key positioning residue GYG-Asp125 binds the nucleophilic acceptor Glc terminus allowing equilibration into a productive Michaelis complex and guides the OH-4 to the donor site

from the alpha face of UDP-Glc, ready for the front-face attack (optimal for intermediate levels of GYG glucosylation). Key to this process is a striking flexibility of GYG-Tyr195, which steadily recoils step-by-step by the distance of one sugar ring to accommodate acceptor **Glc**$_n$ chains of increasing lengths (**Figure 4**).

Together these data suggest a first inter-monomer phase where the nascent oligosaccharide chain is of insufficient length to easily provide the right orientation to be processed by the active site but can eventually equilibrate ('hooked' into place by Asp125) to a productive Michaelis complex due to flexibility of Tyr195. In the second phase, sufficient flexibility of the oligosaccharide chain allows correct orientation and a rapid intra-monomer extension, yet with low selectivity. Finally, in the third 'refining' phase, extension of the nascent oligosaccharide chain past the active site of its own protein monomer requires extension by the active site of another monomer in a much more closely linked dimer requiring careful alignment of donor substrate (UDP-Glc only) recruitment with binding of the extending chain. Eventually, this chain too processes past the point of the second active site and GYG's activity ceases at a longer chain length of >12 Glc units. This presents a Glc-terminated core-glycan particle ready for elaboration by Glycogen Synthase (GYS) and Glycogen Branching Enzyme (GBE), respectively (**ED Figure 1**).[1,32]

The plasticity of GYG raises the question of whether non-glucose sugars can ever be incorporated into mature glycogen particles. Whilst natural incorporation of mannose from its most abundant nucleotide GDP-mannose is not feasible owing to the known specificity of GYG for pyrimidine nucleotide sugar donors,[32] UDP-Gal is readily available *in vivo*. In this light, the limited final kinetic phase for auto-galactosylation is consistent with a 'refining' mechanism that prevents mis-formed glycogen particles due to, for example, poly-Gal incorporation (**Figure 4**). At the same time, GYG's ability to utilise UDP-Gal in earlier phases may facilitate early glycogenesis during times in which UDP-Glc supplies are scarce. Fascinatingly, this suggests that the core of glycogen can carry priming glycans that may be non-glucose in nature. Our work here also highlights that whilst non-glucose sugars might serve this role, other simpler hydroxyl-only templates fail. This, in turn, suggests that this core region does not serve a role as an energy storage polymer (since it would release incorrect sugars for metabolism) but instead acts to anchor glycogen to the glycogenin core protein. Together, these three phases – prime-extend-refine – therefore appear to represent a mechanistic solution to the exquisite evolutionary balance between the difficult-to-achieve need to anchor glucose-polymer to a protein with the need to ensure precise glucose-only particle formation at its outer regions.

The chemical ligation approach used here has shown that, whilst natural C–O Tyr195-to-glucose linkages cannot be accessed via any current chemical modification approach (**ED Figure 10**), Pd-mediated formation of an irreversible C–C bond can yield sufficiently similar motifs to allow functional mimicry of GYG in glycogenesis. They reveal that GYG's catalytic activity does indeed vary through these intermediate states and highlight how this 'self-modulation' seems to be exploited by nature in three phases with different function. We anticipate that this methodology may ultimately be expanded to now access a wider range of precise glycogen structures, enabling study of other glucosylation and associated processing steps that will shine further light on the significant and ever-expanding number of glycogen-associated diseases[1,2,4,33].

More broadly, the demonstration of successful mimicry that we have achieved here by using Chemistry to covalently and directly 'bolt in' a key residue alteration to create an intermediate catalytic state highlights that new protein chemistries are becoming precise and subtle enough to allow precise (e.g. 'shunt') mechanistic experiments that would be difficult through classical biochemical means. Although strategies for chemical rescue of enzymes via unmasking of caged natural residues have been elegantly explored,[19,34,35] to our knowledge these mark rare application of Pd-mediated C–C-bond-forming ligation as a mode of chemical enzyme activation. It suggests that such ligation-dependent activation (here using catalytic metal Pd(0) as a 'switch') could be a powerful tool not only in the study of mechanism but even potentially in the future 'rescue' of deficient enzymes.

Figure Legends

Figure 1 | Palladium-mediated C(sp^2)–C(sp^2) ligation as a strategy for mechanistic investigation of Glycogenin. Amber codon suppression enables "OH→I" replacement of native Tyr195 acceptor of GYG-wt with an unnatural L-*p*-iodophenylalanine residue. This GYG-*p*IPhe195 enzyme, which lacks a native glycosyl acceptor and thus cannot undergo glucosylation, represents a suitable substrate for Suzuki-Miyaura cross-coupling to a range of boronic acids sugar mimic templates, to generate potentially active enzyme species that mimic defined GYG glycoforms. In this way, inactive GYG-*p*IPhe195 might be activated through C–C bond forming ligation allowing pre-determined, 'shunted' access to intermediate catalyst states of GYG. See **ED Figures 2,4** for templates.

Figure 2 | Generation of homogeneously glucosylated, catalytically-active GYG glycoforms and kinetic studies of GYG-Glc and GYG-GlcGlc. (**a**) Pd-mediated C–C bond forming ligation of glucose-derived boronic acid **1-Glc** to GYG-Y195*p*IPhe generates in good yield the homogeneous glycoform GYG-**Glc,** which exhibits catalytic activity, as shown by LC-MS analysis. Similar results were obtained for at least 4 independent repeats. In all cases, non-glucosylated side-products present show no activity in the assay. (**b**) Cross-coupling to **1-GlcGlc** instead enables direct, shunted access to a further catalytic intermediate of GYG-**Glc**, GYG-**GlcGlc**, which also proved catalytically active. Similar results were obtained for at least 5 independent repeats. In all cases, non-glucosylated side-products present show no activity in the assay. (**c,d**) Kinetic profiles of overall glucosylation (**c**) and initial glucosylation step as monitored through consumption of starting enzyme (**d** and inset) for GYG-**Glc** and GYG-**GlcGlc**. Glucosylation levels and abundance of starting enzyme were determined through LC-MS analysis (see also **ED Figure 6** and **Supplementary Methods**). Whilst GYG-**GlcGlc** exhibits a marked "fast → slow" biphasic profile, these same phases, whilst necessarily present for GYG-**Glc**, are not visible, being instead obscured by a slower initiation step. For both (c) and (d), data points represent mean averages of n independent replicate kinetic runs; n = 4 (GYG-Glc) and n = 5 (GYG-GlcGlc). Error bars are ± s.d.

Figure 3 | Donor and acceptor plasticity of GYG. (**a**) GYG-**Glc** and GYG-**GlcGlc** are capable of utilising the unnatural donor UDP-Galactose. Kinetic profiles for overall galactosylation (left) and rate of initial galactosylation step (right) are illustrated. Data points represent mean averages of n = 3 independent replicate kinetic runs for both GYG-Glc and GYG-GlcGlc galactosylation. Error bars are ± s.d. (**b**) Generation of non-natural GYG glycosyl acceptors GYG-**Gal** and GYG-**Man**. Similar results were obtained for at least 3 independent repeats. (**c**) Autoglucosylation activity of GYG-**Man** and GYG-**Gal**, compared to GYG-**Glc**. Kinetic profiles analogous to those in (a), overall glucosylation (left) and rate of initial glucosylation step (right) are shown. Data points represent mean averages of n independent replicate kinetic runs; n = 4 (GYG-Glc) and n = 3 (GYG-Gal, GYG-Man). Error bars are ± s.d. In all cases, non-glycosylated side-products present show no activity in the assay.

Figure 4 | Natural and unnatural pathways of GYG catalysis further delineate triphasic mechanism and reveal possible proofreading step. (**a**) Motion of Tyr195 to accommodate acceptor substrates of various lengths and conformations (intra- and inter-monomeric. Results obtained from MD

simulations for each Michaelis complex. Acceptor sugar units have been omitted for clarity. The orange loop corresponds to the acceptor arm of the same subunit of the displayed active site (i.e. intra), whereas the white loop is the acceptor arm of the opposite subunit (inter). The tyrosine residue represented as transparent indicate an unstable conformation due to steric hindrance with the 'blocking loop' coloured in blue. Notice that the tyrosine residue recoils one position for each sugar that is attached to it. Hydrogen atoms and acceptor glucose units have been omitted for clarity. (**b**) Comparison of the natural autoglucosylation pathway [and unnatural autogalactosylation pathway for various GYG substrates] reveals that, whilst the slower 1st and 3rd phases (which we speculate are inter-monomer) display limited Gal-Gal transfer, this reaction readily proceeds throughout the fast 2nd phase (which we speculate is intra-monomer). The consequent absence of a 3rd phase for autogalactosylation may function as a 'refining' step, preventing incorporation of poly-Gal oligosaccharides into glycogen and thus preventing accumulation of misformed, potentially toxic, glycogen particles.

Acknowledgements

This work was supported by grants from the EPSRC (DTA to CDS, MKB), MINECO (CTQ2017-85496-P to CR), AGAUR (2017SGR-1189 to CR), EU Horizon 2020 programme, Marie Skłodowska-Curie (67507) and the Royal Society (Wolfson Research Merit Award to BGD). We thank Gustaf Hemberg for useful discussions, Prof Schultz for provision of initial pEVOL-pIPhe plasmid and BSC-CNS for computer resources and technical support at *MareNostrum* (RES-QCM-2016-3-0017).

Author Contributions

MKB, SSL, CDS WWY, BGD designed the project. MKB, TM, MG carried out chemical synthesis, protein modification reactions, enzymatic assays and associated analysis. HB, SSL, CDS, MKB, TM, MG carried out protein expression. HB performed protein expression optimisation and crystallography experiments. LR, JI-F, CR performed computational experiments. MKB, CR, WWY, BGD wrote the manuscript. All authors read and commented on the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Data Availability

Crystallographic data have been deposited and made available under PDB accession codes: 6EQJ (apo) and 6EQL (with ligands). All raw MS data supporting Figures are given in the SI and/or is available on request. All primary numerical data for graphical plots in Figures is available as spreadsheets. Key data deposited in open access depository ORA-data.

References

1 Roach, P. J., Depaoli-Roach, A. A., Hurley, T. D. & Tagliabracci, V. S. Glycogen and its metabolism: some new developments and old themes. *Biochemical J.* **441**, 763-787, doi:10.1042/BJ20111416 (2012).

2 Adeva-Andany, M. M., González-Lucán, M., Donapetry-García, C., Fernández-Fernández, C. & Ameneiros-Rodríguez, E. Glycogen metabolism in humans. *BBA Clinical* **5**, 85-100, doi:10.1016/j.bbacli.2016.02.001 (2016).

3 Roach, P. J. Are there errors in glycogen biosynthesis and is laforin a repair enzyme? *FEBS Lett.* **585**, 3216-3218, doi:10.1016/j.febslet.2011.09.009 (2011).

4 Zois, C. E., Favaro, E. & Harris, A. L. Glycogen metabolism in cancer. *Biochemical Pharmacology* **92**, 3-11, doi:10.1016/j.bcp.2014.09.001 (2014).

5 Testoni, G. *et al.* Lack of Glycogenin Causes Glycogen Accumulation and Muscle Function Impairment. *Cell Metab* **26**, 256-266.e254, doi:10.1016/j.cmet.2017.06.008 (2017).

6 Moslemi, A.-R. *et al.* Glycogenin-1 Deficiency and Inactivated Priming of Glycogen Synthesis. *New England Journal of Medicine* **362**, 1203-1210, doi:10.1056/NEJMoa0900661 (2010).

7 Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. Tyrosine-194 of glycogenin undergoes autocatalytic glucosylation but is not essential for catalytic function and activity. *FEBS Lett.* **342**, 38-42 (1994).

8 Hurley, T. D., Stout, S., Miner, E., Zhou, J. & Roach, P. J. Requirements for catalysis in mammalian glycogenin. *J. Biol. Chem.* **280**, 23892-23899, doi:10.1074/jbc.M502344200 (2005).

9 Gibbons, B. J., Roach, P. J. & Hurley, T. D. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* **319**, 463-477, doi:10.1016/S0022-2836(02)00305-4 (2002).

10 Chaikuad, A. *et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proc. Natl. Acad. Sci. USA* **108**, 21028-21033, doi:10.1073/pnas.1113921108 (2011).

11 Hurley, T. D., Walls, C., Bennett, J. R., Roach, P. J. & Wang, M. Direct detection of glycogenin reaction products during glycoden initiation. *Biochem Biophys Res Commun.* **348**, 374-378 (2006).

12 Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A "Tag-and-Modify" Approach to Site-Selective Protein Modification. *Accounts of Chemical Research* **44**, 730-741, doi:10.1021/ar200056q (2011).

13 Chalker, J. M., Wood, C. S. C. & Davis, B. G. A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *J. Am. Chem. Soc.* **131**, 16346-16347, doi:10.1021/ja907150m (2009).

14 Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun.* **47**, 1698-1700, doi:10.1039/c0cc04970k (2011).

15 Spicer, C. D., Triemer, T. & Davis, B. G. Palladium-mediated cell-surface labeling. *J. Am. Chem. Soc,* **134**, 800-803, doi:10.1021/ja209352s (2012).

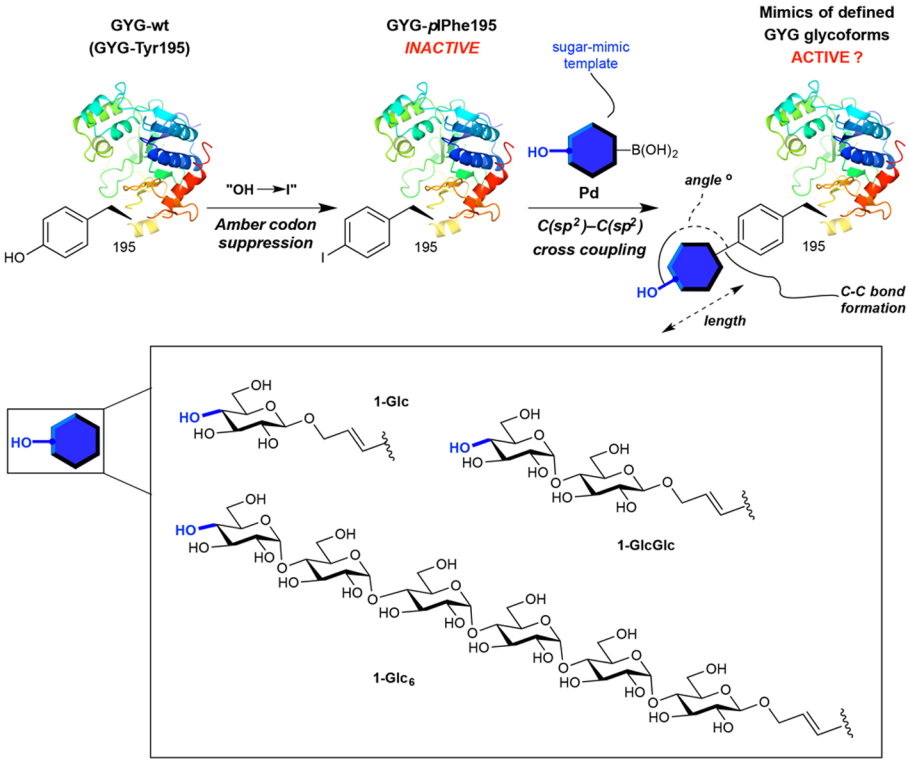
16 Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki-Miyaura cross-coupling. *Chem. Commun.* **49**, 2747-2749, doi:10.1039/c3cc38824g (2013).

17 Dumas, A. *et al.* Self-liganded Suzuki-Miyaura coupling for site-selective protein PEGylation. *Angew. Chem. Intl Ed. Engl.* **52**, 3916-3921, doi:10.1002/anie.201208626 (2013).

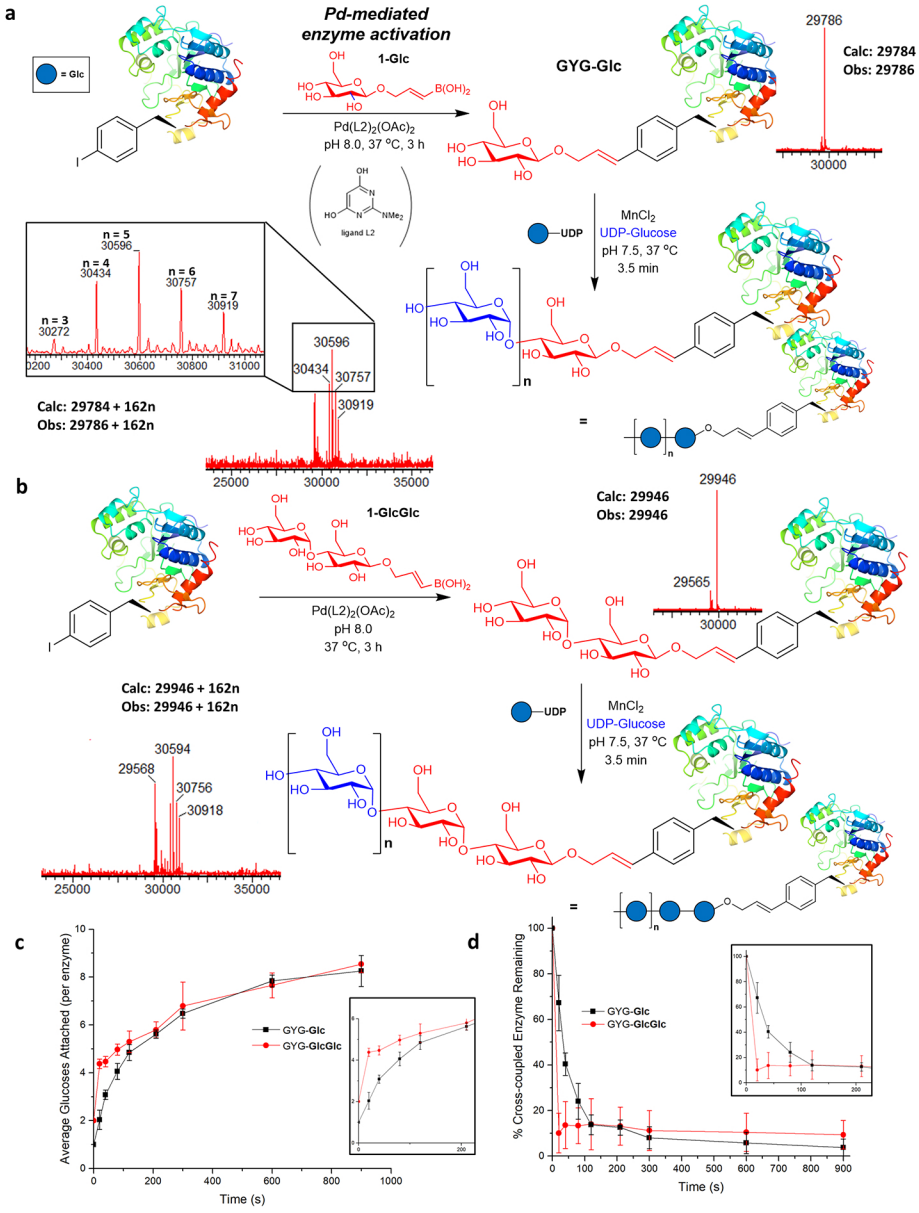
18 Li, J. & Chen, P. R. Moving Pd-Mediated Protein Cross Coupling to Living Systems. *ChemBioChem* **13**, 1728-1731, doi:10.1002/cbic.201200353 (2012).

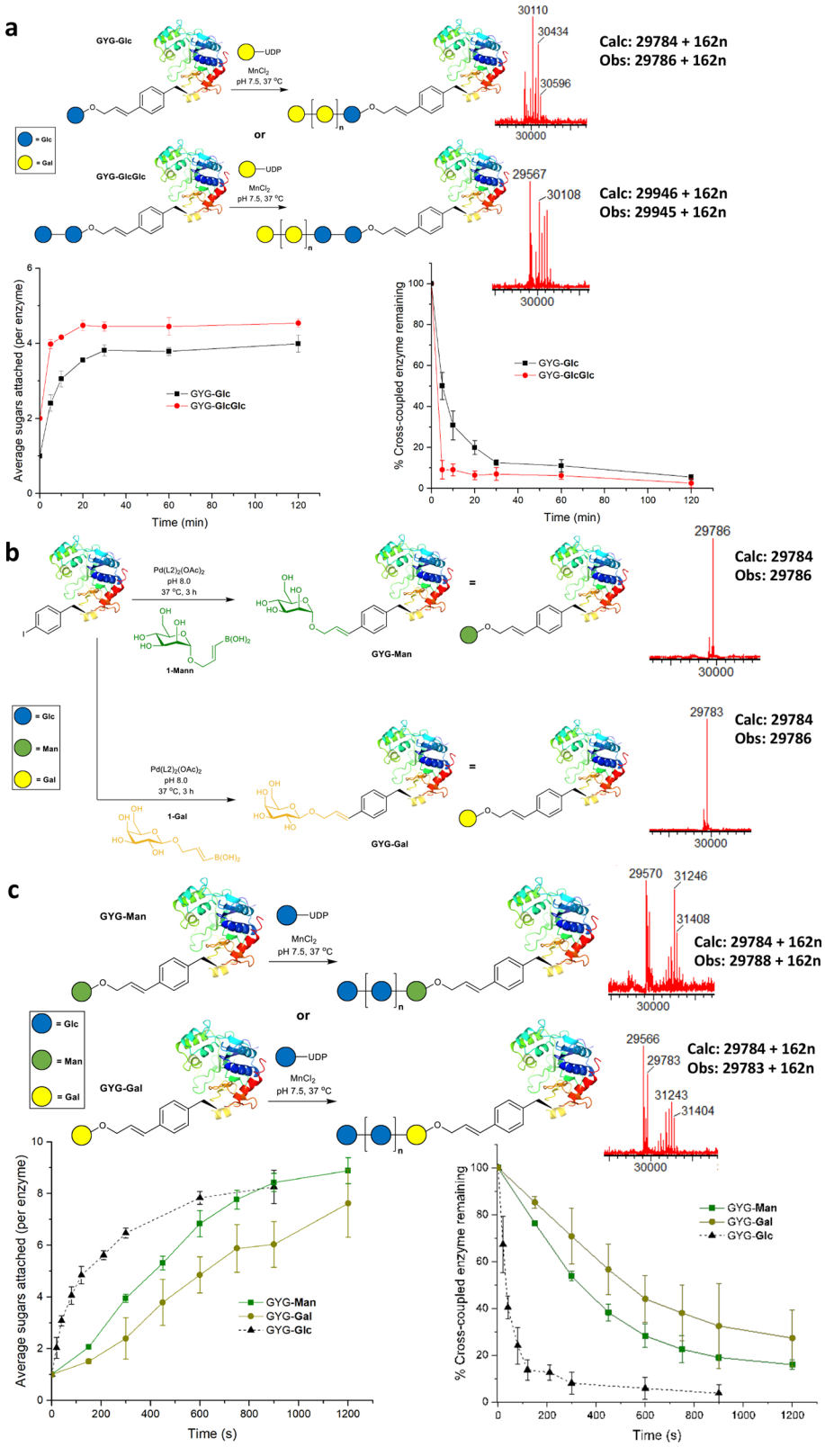
19 Yang, M., Li, J. & Chen, P. R. Transition metal-mediated bioorthogonal protein chemistry in living cells. *Chemical Society Reviews* **43**, 6511-6511, doi:10.1039/C4CS00117F (2014).

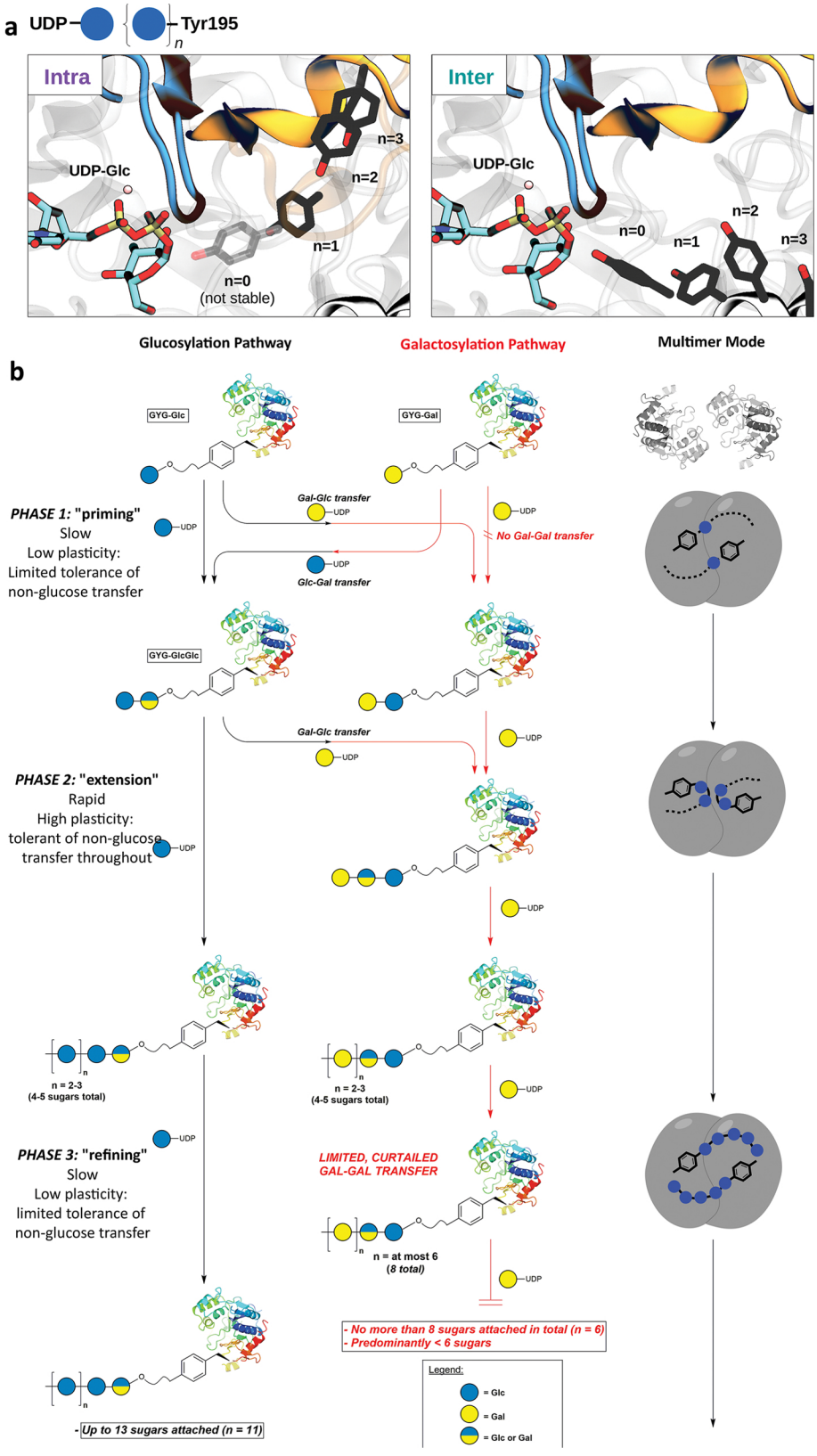
20 Muhammad, J., Kumar, M. S. & Ashraf, B. Palladium in the Chemical Synthesis and Modification of Proteins. *Angewandte Chemie International Edition* **56**, 10644-10655, doi:doi:10.1002/anie.201702370 (2017).
21 Boeggeman, E. & Qasba, P. K. Studies on the metal binding sites in the catalytic domain of β1,4-galactosyltransferase. *Glycobiology* **12**, 395-407, doi:10.1093/glycob/cwf045 (2002).
22 Nielsen, M. M. *et al.* Substrate and Metal Ion Promiscuity in Mannosylglycerate Synthase. *Journal of Biological Chemistry* **286**, 15155-15164, doi:10.1074/jbc.M110.199844 (2011).
23 Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.* **395**, 361-374, doi:10.1016/j.jmb.2009.10.030 (2010).
24 Davis, L. & Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Molecular Cell Biology* **13**, 168-182, doi:10.1038/nrm3286 (2012).
25 Issoglio, F. M., Carrizo, M. E., Romero, J. M. & Curtino, J. a. Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J. Biol. Chem.* **287**, 1955-1961, doi:10.1074/jbc.M111.287813 (2012).
26 Bazán, S., Issoglio, F. M., Carrizo, M. E. & Curtino, J. a. The intramolecular autoglucosylation of monomeric glycogenin. *Biochem. Biophys. Res. Commun.* **371**, 328-332, doi:10.1016/j.bbrc.2008.04.076 (2008).
27 Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **99**, 12562-12566, doi:10.1073/pnas.202427399 (2002).
28 Ardèvol, A. & Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* **137**, 7528-7547, doi:10.1021/jacs.5b01156 (2015).
29 Lee, S. S. *et al.* Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. *Nat Chem Biol* **7**, 631-638, doi:10.1038/nchembio.628 (2011).
30 Ardèvol, A. & Rovira, C. The molecular mechanism of enzymatic glycosyl transfer with retention of configuration: evidence for a short-lived oxocarbenium-like species. *Angew. Chem. Int. Ed. Engl.* **50**, 10897-10901, doi:10.1002/anie.201104623 (2011).
31 Carrizo, M. E., Miozzo, M. C., Goldraij, A. & Curtino, J. A. Purification of rabbit skeletal muscle proteoglycogen: Studies on the glucosyltransferase activity of polysaccharide-free and -bound glycogenin. *Glycobiology* **7**, 571-578, doi:10.1093/glycob/7.4.571 (1997).
32 Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. A new look at the biogenesis of glycogen. *FASEB J.* **9**, 1126-1137 (1995).
33 Ashcroft, F. M., Rohm, M., Clark, A. & Brereton, M. F. Is Type 2 Diabetes a Glycogen Storage Disease of Pancreatic β Cells? *Cell Metabolism* **26**, 17-23, doi:10.1016/j.cmet.2017.05.014 (2017).
34 Lee, H.-M., Larson, D. R. & Lawrence, D. S. Illuminating the Chemistry of Life: Design, Synthesis, and Applications of "Caged" and Related Photoresponsive Compounds. *ACS Chemical Biology* **4**, 409-427, doi:10.1021/cb900036s (2009).
35 Li, J. *et al.* Palladium-triggered deprotection chemistry for protein activation in living cells. *Nat. Chem.* **6**, 352-361, doi:10.1038/nchem.1887 (2014).


GYG-wt
(GYG-Tyr195)
GYG-pIPhe195
INACTIVE
sugar-mimic
template
Mimics of defined
GYG glycoforms
ACTIVE ?
"OH → I"
Amber codon
suppression
Pd
C(sp²)–C(sp²)
cross coupling
angle °
C-C bond
formation
length
195
HO
B(OH)2

1-Glc
1-GlcGlc
1-Glc6
HO
OH
O


a
= Glc
Pd-mediated
enzyme activation
1-Glc
Pd(L2)2(OAc)2
pH 8.0, 37 °C, 3 h
ligand L2
GYG-Glc
29786
Calc: 29784
Obs: 29786
30000
MnCl2
UDP-Glucose
pH 7.5, 37 °C
3.5 min
UDP
n = 5
30596
n = 4
30434
n = 6
30757
n = 7
30919
n = 3
30272
Calc: 29784 + 162n
Obs: 29786 + 162n
25000
30000
35000
b
1-GlcGlc
Pd(L2)2(OAc)2
pH 8.0
37 °C, 3 h
Calc: 29946
Obs: 29946
29946
29565
Calc: 29946 + 162n
Obs: 29946 + 162n
30594
29568
30756
30918
c
Average Glucoses Attached (per enzyme)
Time (s)
GYG-Glc
GYG-GlcGlc
d
% Cross-coupled Enzyme Remaining
Time (s)
GYG-Glc
GYG-GlcGlc


a
GYG-Glc
UDP
MnCl2
pH 7.5, 37 °C
or
= Glc
= Gal
GYG-GlcGlc
30110
30434
30596
30000
Calc: 29784 + 162n
Obs: 29786 + 162n
29567
30108
Calc: 29946 + 162n
Obs: 29945 + 162n
Average sugars attached (per enzyme)
Time (min)
GYG-Glc
GYG-GlcGlc
% Cross-coupled enzyme remaining
b
Pd(L2)2(OAc)2
pH 8.0
37 °C, 3 h
1-Mann
GYG-Man
= Man
1-Gal
GYG-Gal
29786
Calc: 29784
Obs: 29786
29783
Calc: 29784
Obs: 29786
c
GYG-Man
29570
31246
31408
Calc: 29784 + 162n
Obs: 29788 + 162n
29566
29783
31243
31404
Calc: 29784 + 162n
Obs: 29783 + 162n
Time (s)


a
UDP
Tyr195
n
Intra
UDP-Glc
n=3
n=2
n=1
n=0
(not stable)
Inter
UDP-Glc
n=2
n=0
n=1
n=3
Glucosylation Pathway
Galactosylation Pathway
Multimer Mode
b
GYG-Glc
GYG-Gal
Gal-Glc transfer
UDP
No Gal-Gal transfer
Glc-Gal transfer
PHASE 1: "priming"
Slow
Low plasticity:
Limited tolerance of
non-glucose transfer
GYG-GlcGlc
PHASE 2: "extension"
Rapid
High plasticity:
tolerant of non-glucose
transfer throughout
n = 2-3
(4-5 sugars total)
PHASE 3: "refining"
Slow
Low plasticity:
limited tolerance of
non-glucose transfer
LIMITED, CURTAILED
GAL-GAL TRANSFER
n = at most 6
(8 total)
- Up to 13 sugars attached (n = 11)
- No more than 8 sugars attached in total (n = 6)
- Predominantly < 6 sugars
Legend:
= Glc
= Gal
= Glc or Gal