A Front-Face 'S_Ni Synthase' Engineered from a Retaining 'Double-S_N2' Hydrolase

Javier Iglesias-Fernández[1,2]ψΦ, Susan M. Hancock[3]ψ, Seung Seo Lee[3,7]ψ, Maola Khan[3], Jo Kirkpatrick[3]ς, Neil J. Oldham[3]‡, Katherine McAuley[4], Anthony Fordham-Skelton[5]†, Carme Rovira[1,2,6] and Benjamin G. Davis[3]**

[1] Departament de Química Orgànica, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

[2] Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

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

[4] Diamond Light Source, Diamond House, Harwell Science & Innovation Campus, Didcot, Oxfordshire, OX11 0DE, UK

[5] CLRC, Daresbury Laboratory, Warrington, Cheshire, UK

[6] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.

[7] School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

* c.rovira@ub.edu; ben.davis@chem.ox.ac.uk

ψ These authors contributed equally.

† Deceased.

Φ Current address: Department of Chemistry, King's College London, London SE1 1DB, UK

ǂ Department of Chemistry, University Park, Nottingham, NG7 2RD, UK

ς Current address: Leibniz Institute on Aging - Fritz Lipmann Institute (FLI), Beutenbergstraße 11, 07745 Jena, Germany

Abstract

S_Ni-like mechanisms with 'front-face' leaving group departure and nucleophile approach have been observed experimentally and computationally in chemical and enzymatic substitution at α-glycosyl electrophiles. Since S_Ni-vs-S_N1-vs-S_N2-substitution pathways can be energetically comparable, engineered switching could be feasible. Here, engineering of a protein that catalyzes "double S_N2"-substitution, *Sulfolobus solfataricus* β-glycosidase apparently changes the mode to "S_Ni". Destruction of "first S_N2" nucleophile, through Glu387Tyr mutation, created a β-stereoselective catalyst for glycoside synthesis from activated substrates despite lacking a nucleophile. pH-profile, kinetic and mutational analyses; mechanism-based inactivators; x-ray structure and subsequent metadynamics simulations together suggest recruitment of substrates by π-sugar interaction and reveal a quantum mechanics/molecular mechanics (QM/MM) free energy landscape for the substitution reaction similar to those of natural, S_Ni-like glycosyltransferases. This observation of a front-face mechanism in a β-glycosyltransfer enzyme highlights that S_Ni-like pathways may be engineered in catalysts with suitable environments; and suggests that 'β-S_Ni' may be feasible for natural glycosyltransfer enzymes with more widespread existence of S_Ni-like mechanisms in nature.

Introduction

Since the seminal demonstration of front-side (same face) nucleophilic attack in chemical, α-glycosyl transfer substitution,[1] the possibility of the wider existence of such an unusual mechanism has been rarely but carefully considered.[2,3] Such a front-side mechanism is invoked to explain the seemingly unusual behavior of retaining glycosyltransferases (GTs).[4] Most retaining GTs do not contain obvious, conserved, functional nucleophiles and/or acid/base residues required to operate the double-displacement mechanism[5] that is found in glycoside hydrolases (GHs).[4] Whilst typically-observed 'chemical' nucleophilic substitution involves likely intermediacy of solvent exposed and accessible reactions centres, even for such reactions, S_Ni-like mechanisms, facilitated by assisted delivery of the nucleophile to the electrophile, are observed.[6,7] In proteins, more constrained environments (and possible alternative pathways) exist. Structures of several retaining GTs[8-11] show positioning of substrates, leaving group and nucleophile in positions suitable for front-face mechanisms.[2,12]

Recently, we have provided experimental evidence that supports the operation of a front-face mechanism in the retaining GT trehalose-6-phosphate synthase (OtsA)[13] consistent with detailed computational QM/MM metadynamics simulations.[14] These were followed by an experimental and computational studies of glycosyl transfer in solution chemistry, indicating that the solvolysis of α-glucosyl fluoride in hexafluoro-2-propanol, a non-nucleophilic environment, also follows a front-face mechanism;[7] phosphorolysis of α-glucosyl fluoride in mutant phosphorylases is also suggested to follow a similar path.[15] Subsequent QM/MM studies on the retaining GTs lipopolysaccharyl α-galactosyltransferase C (LgtC),[16] α-1,2-mannosyltransferase Kre2p/Mnt1p,[17] polypeptide GalNAc-transferase T2 (GalNAc-T2)[18,19] and glucosyl-3-phosphoglycerate synthase (GpgS)[20] further disentangle the molecular details of

the front-face mechanism for these α-selective retaining GTs.[4] More recently, the functionally essential Notch-modifying xylosyltransferase is proposed to follow this S_Ni-pathway.[11] Together these studies suggest that the unusual, front-face mechanism may, in fact, play an important and potentially widespread role in nature, when considering the importance and ubiquity of glycosyltransferases. Thus far, no β-selective retaining reaction has been observed. One apparent crucial feature of the α-selective mechanism suggested in these studies (**Figure 1a**) is the role of an asymmetric and shielding environment (the active site) as a 'reaction compartment' with sufficient space to not only accommodate the nucleophile and the leaving group on the same face but to do so in a protective manner that allows sufficient lifetime for oxocarbenium ion-like intermediates. In essence, the active site provides a 'protective box' that allows the acceptor nucleophile to separate the ion-pair that is generated from the donor electrophile. Together these suggest common features (suitable shielding by active site moieties to exclude solvent; no competing protein nucleophile; reduced requirement for protein general acid/base; and suitable leaving group pK_a) that, in principle, could be engineered[15] rather than simply observed.

Here we demonstrate that the front-face reaction is operative not only in retaining GTs but can also be created in engineered GHs through the exploitation of such features. Selection of a suitable, robust GH scaffold created an enzyme with highly specific transglycosylation activity capable of stereospecific creation of β-glycosidic linkages from activated β-donors such as *p*-nitrophenyl glycosides, and incapable of hydrolyzing the unactivated glycosidic linkages in the product. Mechanistic investigations (including kinetic, biochemical, mutagenic, structural and computational studies) suggested that this novel, unnatural 'synthase' utilizes front-face nucleophilic substitution, similar to that proposed for retaining GTs. To the best of our knowledge, this is the first description of a frontal face mechanism of a β-retaining enzyme.

Results

Design and Creation of a Nucleophile-free GH.

We chose the robust and representative GH family 1 scaffold as a protein platform for design. The retaining β-glycosidase from *Sulfolobus solfataricus* (SSβG) shows stability to mutation,[21,22] solvents[23] and even typically denaturing conditions.[24,25] Prior nucleophile-free mutants bearing smaller residues than the natural Glu387 (e.g., Gly387[26]) act as classical, *inverting* glycosynthases[27] with suitable (α-glycosyl fluoride) substrates.[26] In contrast, our initial modeling suggested that to ensure sufficient protection and putative stabilizing interactions and yet small enough to be accommodated, only certain bulkier residues (e.g. Tyr, Phe) would prove suitable. Tyr387 was therefore chosen and site-directed mutagenesis of SSβG-WT, yielded stable, folded, soluble protein SSβG-E387Y, *C*-terminally-His-tagged to allow exhaustive nickel-affinity chromatography (**Supplementary Results, Supplementary Figure 1**) giving good protein yields of ~28 mg per L of growth. *N*-terminal sequencing, LC-mass spectrometry (ESI-MS, found 57,450; expected 57,447 Da) (**Supplementary Table 1**) and circular dichroism (CD) analysis (**Supplementary Figure 2**) confirmed identity and unaffected secondary structure, respectively.

Glu387Tyr Nucleophile-Mutant Displays Altered Activity

By design, we chose *para*-nitrophenoxide (pKa_H ~7) with a similar pKa to those of UDP (pKa_{H1} ~7, pKa_{H2} ~9) as a suitable leaving group for our putative 'activated' substrates. Determination of the kinetic parameters (**Supplementary Table 2**) of SSβG-E387Y towards *p*-nitrophenyl β-D-glycosides and comparison with SSβG-WT revealed reduced but clear activity towards pNPβGlc and pNPβGal substrates. Consistent with the loss of SSβG-WT's

nucleophilic Glu387 residue, the decrease in activity was manifested exclusively in k_{cat}. Notably, substrate selectivity (as judged by k_{cat}/K_M) was reversed from Gal:Glc = 1:1.6 in SSβG-WT to 3:1 in SsβG-E387Y, a ratio that more closely reflected inherent, chemical reactivity of Gal *vs* Glc.[28] Interestingly, tyrosinyl residues are observed in similar positions to Tyr387 in glycosidase enzymes that exploit substrate-assisted catalysis, such as the hexosaminidases.[29] These are thought to stabilize the formation of corresponding oxazolinium ion intermediates. However, SsβG-E387Y displayed no hexosaminidase activity either towards pNPβGlcNAc or even corresponding activated oxazoline substrates (2-methyl-(1,2-dideoxy-α-D-glucopyrano)[2,1-d]-Δ^2-oxazoline) (**Supplementary Figure 3**). Consistent with the designed requirement for a suitable activated leaving group, SsβG-E387Y failed to hydrolyze either methyl β-D-galactopyranoside (MeGal) or *p*-nitrophenyl 6-*O*-(β-D-galactopyranosyl)-β-D-galactopyranoside (pNPGal1,6Gal).

Incubation with mechanism-based inhibitor[30] 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside with no significant effect discounted the possibility of activity arising from SsβG-WT or other (e.g. endogenous expression host *E. coli*) glycosidases that use nucleophilic catalysis. It also intriguingly suggested that this altered catalytic activity of SsβG-E387Y was no longer nucleophile-dependent (*vide infra*). When SsβG-E387Y was thermally denatured (16-20h at 45°C) all activity was lost, implying that native protein conformation was required for catalytic activity.

SSβG-E387Y is a 'Synthase'

Given this striking selectivity for activated substrates, with negligible activity towards the hydrolysis of unactivated glycosides (and hence potential products), SSβG-E387Y suggested

itself as a potentially useful catalyst for glycosidic bond formation from activated *p*NP substrates. We surveyed a small set of representative monosaccharides as nucleophilic acceptors under different conditions (**Figure 1b, Table 1** and **Supplementary Notes**).

SsβG-E387Y did not process non-aromatic sugar acceptors to any significant extent, resulting in reactions that instead primarily gave GalβGal*p*NP disaccharidic products **4** and **12** (**Table 1**, entries i-viii). This suggested a strong preference for utilizing GalpNP **1** as an acceptor. This observed preference for aromatic sugar acceptors was consistent with aromatic stacking interactions in the + 1 or + 2 acceptor pockets that are used by the GH naturally for binding oligosaccharide substrates.[31,32] Indeed, aromatic Galβ, Glcβ and Manα glycosides all proved to be suitable nucleophile substrates (**Table 1**, entries ix-xii). Unlike several other synthases, under these conditions trisaccharides and higher or branched oligosaccharides (from uncontrolled 'self condensation') were not synthesized in measurable amounts; these are normally isolated in reactions catalyzed by classical glycosynthases including, notably, a variant derived from SsβG.[26] Only under more extreme conditions were small amounts of trisaccharides observed (see below and **Supplementary Notes**). In all reactions, either exclusive 1,6- or 1,6-/1,3-linked regioselectivity was observed;[33] in contrast to the behavior of other SsβG-related catalysts,[26,34] no 1,4-linked disaccharides were isolated. Notably, *all* transglycosylation reactions displayed exclusive, retentive β-stereoselectivity.

Having demonstrated initial synthetic potential, the synthetic application was explored in a model reaction of donor *p*NPGal **1** with acceptor PhβGlc **10** (**Supplementary Table 3**). Strikingly, variation of conditions allowed the improvement of the synthesis(S):hydrolysis(H) ratio to up to >99. Under these conditions, the enzyme was both selective and essentially, exclusively synthetic, yielding **14** as the predominant product in >70% isolable yield with only the formation of smaller amounts of trisaccharides as side products (**Table 1,** entry xiii). In

control experiments under essentially identical conditions, SsβG-WT simply hydrolyzed the donor sugar and gave none of the desired synthetic product. No transglycosylation activity was observed using α-D-galactopyranosyl fluoride donor and representative acceptors: SsβG-E387Y did not process donor substrates with α-anomeric configuration, thereby confirming that SsβG-E387Y did not act as a classical glycosynthase. Notably, in comparison to reactions that are catalyzed by glycosidases, which typically give transglycosylation yields from 20-40%,[35] the general yields of transglycosylation products synthesized with SsβG-E387Y (several > 80%) were high and only rivaled by some of the more potent glycosynthases.[36] Although it should be noted that estimated transglycosylation rates (k_{cat}/K_M ~ 0.0052 – 0.025 $min^{-1}mM^{-1}$) were ~ 2,000-fold lower compared to classical glycosynthases (see below for further details).

SsβG-E387Y requires no nucleophile nor general acid/base

This useful transglycosylation / 'synthase' activity again highlighted the differing mechanism of SsβG-E387Y and suggested comparison with natural, trans-glycosidases. The trans-sialidase from *Trypanosoma cruzi* of GH family 33 utilizes a tyrosine residue as a nucleophile,[37] and although modeling and design (*vide supra*) had suggested incompatible geometries for Tyr387 in SsβG-E387Y to play this role, we attempted to clarify this aspect of its mechanism. First, to test Tyr387 as a catalytic nucleophile, trapping experiments were designed that were intended to yield a covalent intermediate from mechanism-based fluorosugar inactivators.[30] Thus, SsβG-E387Y was incubated with DNP-2FGlc[30] **16** (1000 equivalents, 45°C, pH 6.5 50 mM sodium phosphate buffer) and analyzed by LC-MS (**Figure 2 and Supplementary Figure 3**). Over 6h, no change in SsβG-E387Y's hydrolytic activity was observed. Concomitant monitoring of DNP release (absorbance at 405 nm) revealed no

acceleration over uncatalyzed chemical DNP-2FGlc hydrolysis. *Agrobacterium faecalis* β-glucosidase can form a stable α-D-glucopyranosyl tyrosine product at non-relevant Y298 upon mutation of the active site nucleophile;[38] peptide 'mapping' of SsβG-E387Y did not show trapping of Tyr387. Neither proteolytic (trypsin, pepsin, thermolysin, clostripain)-MSMS and/or CNBr-cleavage-MSMS (including neutral loss analysis of the 2FGlc moiety) indicated peptides with attached 2FGlc moieties (**Supplementary Figures 4-6**), even though the coverage of this 'mapping' successfully included peptides containing Y387 (and E206) as putative trapping sites. In control experiments, under essentially similar conditions, SsβG-WT was successfully labeled (**Supplementary Figures 7-9**). Together these results suggested that Tyr387 (or even Glu206) was not acting as a catalytically nucleophilic residue in SsβG-E387Y (and that observed mass changes in the total protein MS were distributed non-specifically at low abundance over multiple non-specific locations that could not be detected by proteolytic-cleavage-MSMS analyses).

Next, to further probe the mechanism of SsβG-E387Y, and prompted by this apparent lack of any functioning nucleophilic catalytic residue, a range of representative mutants of SsβG were constructed (**Supplementary Table 4**). Their identities (primary and secondary structure) were confirmed by ESI-MS (**Supplementary Table 1**) and CD analysis (**Supplementary Figure 2**).

None of these mutations caused a dramatic loss of function; indeed, the similar activities of SsβG-E387Y, -E387F, -E206A:E387Y, and -Y322F:E387Y suggested that none of these residues were necessary for the observed catalytic mechanism, i.e. none play a required role as a nucleophile or a general acid/base in their catalytic mechanisms. It is particularly notable that, consistent with the designed mechanism (*vide supra*) the additional mutation of the acid/base residue (Glu206) along with that of nucleophile (Glu387) to give SsβG-E206A:E387Y had no detrimental effect on activity; in the catalytic mechanism a general

acid/base catalyst was also apparently not required, consistent with design (**Figure 1a**). This was also consistent with the observation that the basic limb of the pH profile of SsβG-E387Y was also shifted ~0.6 pKa units to a value similar to that for *para*-nitrophenol (**Supplementary Figure 10**).

Finally, transglycosylation kinetics were determined for SsβG-E387Y with a range of substrates (**Supplementary Table 5 and Supplementary Figure 11**). Notably, both activity (as judged by k_{cat}/K_M) and regioselectivity (1,6 vs 1,3, **Supplementary Figure 11b**) varied with leaving group; tentative linear free energy analysis (**Supplementary Figure 12**) revealed a small β value (-0.049), consistent with computational analysis suggesting a step-wise mechanism with a higher barrier for the collapse of oxocarbenium-ion intermediate than that for leaving group departure (*vide infra*).

Structural Determinants of Catalysis in SsβG-E387Y

To further probe the mechanism of SsβG-E387Y, the *apo* x-ray crystal structure of SsβG-E387Y was successfully determined (**Figure 3a** and **Supplementary Figure 13, Online Methods and Supplementary Table 6**) and compared to the previously reported SsβG-WT structure.[25] Despite the mutation, the structures were superimposed with very little divergence; the r.m.s. deviation is 0.26 Å was calculated using 486 Cα positions. Essentially in the active site, only 2 amino acids shifted significantly as a result of the mutation *i.e.* Tyr322 and His342 (**Figure 3a**). Attempts to generate *holo* structures in complex with either substrate or inhibitor were unsuccessful. Therefore, the structures of appropriate ternary complexes were modeled informed by both the *apo* SsβG-E387Y structure and structural alignments with SsβG-WT[39] complexed with D-galactohydroximolactam (pdb: 1uwt) (**Supplementary Fig. 14**). The SsβG-E387Y active site was very similar to that of SsβG-WT (**Figure 3a**), consistent with the similar

236 K_M values obtained for pNPβGal and pNPβGlc substrates for SsβG-WT and SsβG-E387Y
237 SsβG (**Supplementary Table 2**).

238 A combination of classical molecular dynamics and metadynamics techniques were used to
239 model a ternary Michaelis complex of SsβG-E387Y with two molecules of pNPβGal, as
240 putative acceptor and donor substrates corresponding to one of the observed synthase
241 activities (*vide supra*). In a first step, the two molecules were manually placed at the entrance
242 of the enzyme catalytic groove (see **Online Methods** and **Supplementary Figures 15, 16**).
243 After 200 nanoseconds of molecular dynamics (MD) simulation, one of the molecules partially
244 entered the catalytic site, sitting at ~8 Å from the catalytic residues, whereas the other
245 remained at the entrance (**Supplementary Figure 16a**). Further MD simulation did not lead to
246 significant change, indicating that complete entrance of the two molecules is associated with a
247 certain free energy barrier. Therefore, the ligand binding process was activated using an
248 enhanced-sampling technique (metadynamics).[40] Two collective variables were chosen to
249 drive the binding of the two pNPβGal molecules to the active site of SsβG-E387Y. The first
250 (CV_1, **Supplementary Figure 15**) measures the degree of penetration of the first pNPβGal
251 molecule (as the *donor*) into the active site; the second (CV_2) accounts for the formation of a
252 O1···H' interaction, providing a measure of distance between *donor* and *acceptor*.

253 The free energy landscape (FEL) of ligand binding that was obtained from the classical
254 metadynamics simulation (**Supplementary Figure 16c**) showed an energy minimum (the
255 global one) in which the two pNPβGal molecules were inside the enzyme active site (the
256 ternary complex, shown in **Figure 3c**). Analysis of the water content around the active site
257 showed that a number of water molecules were displaced during binding (13 ± 4 from a region
258 of ≤ 5 Å from Y387 and Y322). Among the remaining water molecules, there were two that are
259 located within 5 Å of the donor anomeric carbon. Although these water molecules were not

well oriented for catalysis, they could account for the observed residual hydrolysis. Close examination of the orientation of the two molecules in the active site revealed that the hydroxymethyl group of the acceptor molecule is located on the same face of the donor sugar as the *p*-nitrophenyl group (i.e. the leaving group) of the donor molecule. This was an optimum topology for a front-face mechanism, which could ultimately lead to a transglycosylation product with net retention of configuration. The hydrogen atom of acceptor OH-6 was directed towards O-1 of the donor molecule, favouring the formation of a 1,6-glycosidic linkage, consistent with the observed regiochemical preferences of SsβG-E387Y. This hydrogen bonding interaction may provide a guide for the nucleophile to the same face as the leaving group, akin to interactions observed in retaining "S_Ni-like" GTs.[14,16] Furthermore, this was consistent with the intended, designed role of the leaving group glycosidic oxygen as a general base that deprotonates the incoming protic OH-6-hydroxyl (**Figure 1a**). It was also consistent with the non-detrimental effect on activity of the removal of the general acid/base residue (Glu206) in SsβG-E387Y:E206A; in SsβG-E387Y with *p*NPGal the phenolic base appeared sufficient to deprotonate the incoming hydroxyl nucleophile.

There were crucial substrate-protein interactions (**Figure 3b**) that contributed to the stability of the above "front-face arrangement". First of all, Tyr387 formed stabilizing *donor* sugar···π interactions[41] (sugar hydrogen atoms pointed towards the center of the Y387 phenol ring, with distances < 3 Å, **Figure 3c**), consistent with the overlay of the starting *apo* SsβG-E387Y x-ray crystal structure with the SsβG-WT•inhibitor complex (**Supplementary Figure 14**). Second, Tyr322 swung to form π···π stacking interactions with the *acceptor p*NPGal moiety (the distance between carbon atoms of both six-membered rings amounts to ~ 3.5 Å). This, in turn, appeared to position the OH-6-hydroxyl group in an optimum orientation to attack the anomeric carbon of the sugar donor. These π···π stacking interactions explained why

*p*NPβGal and other aromatic glycosides were preferred substrates for the synthase activity of SsβG-E387Y (*vide supra*). Essentially identical analysis of a possible O3-regioselective pathway also generated an appropriate Michaelis complex (**Supplementary Figure 17**). The binding modes corresponding to the 1,6- or 1,3-reaction were quite different, especially for the acceptor molecule. However, notably, in *both* cases (1,6 and 1,3), the *donor* sugar was stabilized by CH···π interactions engendered by Y387. In the corresponding 1,3- pathway the major difference was that in the *acceptor* the aglycon is oriented away from Y322 enabling sugar-CH···π interactions between acceptor and donor (c.f. *acceptor* aglycon π···π interactions with Y322 for the 1,6-, see above). Thus, in both cases π···π and sugar···π interactions stabilized the substrates in optimum orientation for catalysis. Together these structural analyses (x-ray structure and metadynamics simulations of ligand binding) suggested clearly that the donor anomeric carbon was spatially accessible to the acceptor OH-6 or OH-3 hydroxyl groups from the 'front face'.

QM/MM Analysis of Mechanism and Reaction Landscape

QM/MM simulations, using the metadynamics approach, were performed to elucidate precise details of this unusual glycosyl transfer reaction at atomic detail and to obtain the free energy landscape from which, in turn, reaction coordinates were defined. From the ternary complex determined above (**Figure 3b,c**) three collective variables, corresponding to the main bonds undergoing breaking or formation, were used (**Supplementary Fig. 18** and **Supplementary Discussion**). As a test of one of the critical design elements in this "S_Ni-synthase", it is important to note that none of the CVs used 'self-select' any specific reaction pathway. The free energy landscape for the transglycosylation reaction, reconstructed from the QM/MM metadynamics simulation (**Figure 4a**) showed three main minima and two

transition states (TS). The free energy difference between the reactants state and the highest TS amounted to ~ 25 kcal.mol^{-1}, similar to the value that is obtained for the OtsA glycosyltransferase with essentially similar computational methodology.[14]

The structure of the reactants complex (**R** in **Figure 4b**) was very similar to the one from classical (i.e. force-field based) metadynamics simulation (**Figure 3c**), except that the donor galactosyl ring was distorted into a $^{1}S_{3}$ conformation in the QM/MM structure as opposed to a relaxed $^{4}C_{1}$. This was not surprising in view of the known limitations of force-fields to describe the precise conformation of the sugar ring in glycoside hydrolases.[42,43] The more detailed QM/MM metadynamics simulations instead supported a distorted conformation for the saccharide ring at the -1 donor enzyme subsite, essentially similar to that expected for a β-glucoside hydrolase mechanism.[44,45] Of particular interest was the hydrogen bond between the hydroxymethyl group of the acceptor molecule and the leaving group (*p*NP) of the donor molecule in the reactants complex. This type of interaction, is observed on the basis of QM/MM calculations for GTs[14,16,18,19] (the hydrogen bond forms either at the reactants complex or in the early stages of the reaction), is a common feature of enzymes operating via a front-face mechanism and was part of the design invoked for SsβG-E387Y (**Figure 1** and *vide supra*).

The reaction pathway (**Figure 4**) started with the elongation of the C1-O1 bond of the donor molecule (the C-O distance increases more than 1 Å when going from **R** to **1**, Supplementary **Figure 19** and **Supplementary Table 7**). This bond was completely broken by intermediate **2** (C1-O1 = 3.4 Å). At this stage of the reaction, the distance between donor and acceptor (C1⋯O6') was still long (~3 Å), indicating formation of an oxocarbenium–phenoxide ion pair. Such a change in electronics at the anomeric centre atom was further supported by a shift towards trigonal geometry, which was also associated with changes in the conformation of the

pyranose ring along the reaction (see **Supplementary Figure 20** and discussion below). This change coincided with a decrease in the C1-O5 bond length (from 1.41 Å to 1.27 Å, **Supplementary Table 7**) and an increase in the charge carried by the anomeric centre (by 0.30 e^- when going from **R** to **2**).

The oxocarbenium ion-pair corresponded to a minimum along the reaction pathway. It was stabilized by the O6'-H···O1 hydrogen bond (**2** in **Figure 4b**), which also played a role in orienting the acceptor for nucleophilic attack. Afterwards, a slight displacement of the hydroxymethyl moiety coupled with proton transfer (from hydroxymethyl to *p*NP-oxygen) formed the new glycosidic bond (**3** → **P** in **Figure 4a**). Notably, the observation of a slightly higher barrier ~3 kcal/mol for collapse of the oxocarbenium ion was not only consistent with prior observations in GTs[14,18] but also with the low β_{lg} determined experimentally (see above). As a further characterization of this species, we extracted two snapshots of the metadynamics simulation that correspond to minimum **2** and performed geometry optimizations and subsequent QM/MM MD simulations (see **Online Methods**). The ion-pair species was stable under optimization and MD simulation with a life-time > 15 ps. This again indicated that the ion-pair species was a minimum of the free energy landscape. Interestingly, *in silico* mutation of Y387 to F387 generated an oxocarbenium-ion species that was still a stable minimum, with a slightly longer distance between the aryl ring and the sugar donor anomeric carbon compared with the E387Y variant. This was consistent with the experimental findings that the E387F variant still exhibits clear activity (**Supplementary Table 4**). An alternative mechanism in which the oxocarbenium ion collapsed with the E206 acid base residue,[46] was also considered and tested (**Online Methods** and **Supplementary Figure 21**). However, this mechanism was discarded in view of the high-energy barrier obtained and the low stability of such an intermediate. Therefore, the simulation showed that cleavage of the donor Gal-β-pNP

bond and formation of the Galβ1,6Gal bond were entirely asynchronous and followed a front-side stepwise mechanism.

The donor conformational itinerary observed in SsβG-E387Y during transglycosylation (**Figure 4b**) was: $^{1}S_{3}$ (reactants) – $^{4}H_{3}/E_{3}$ (reaction intermediate) – $^{4}C_{1}$ (products). This pathway was the same that is delineated experimentally[44,45,47] and theoretically[48] for retaining β-D-gluco-active glycoside hydrolases such as SsβG-WT. Remarkably, therefore, despite the very different mechanism, the engineered 'S_{N}i-synthase' SsβG-E387Y synthesized glycosidic bonds by exploiting essentially the same conformational itinerary (and associated distortional strategies to guide catalysis) used by the WT enzyme for hydrolysis. This suggested that, independent of the type of reaction catalyzed by the enzyme, the active site served as a 'box' for the donor to accommodate a given reduced set of pyranose ring conformers.

Discussion

Until now, frontal face or S_Ni-like mechanisms are only implied in retaining α-glycosyltransferases; the engineered system we present here constitutes an example of a retaining glycosyltransferase-like enzyme with β-glycosidic bond selectivity. Structural and computational analyses supported a critical role for the installed Tyr387 through sugar-π and π-π interactions in recruiting to the Michaelis complex (**Figure 3c**) and in stabilizing the reaction pathway through the formation of a hydrogen bond between the acceptor OH and the donor glycosidic oxygen. Given that the dehydroxylating Tyr→Phe mutation in SsβG-E387F did not affect activity, it suggested that any such stabilization might not be (entirely) *via* interactions with the hydroxyl group and/or was not dramatically altered by the change in π-density that this would also cause; this slight effect was supported by computation. Mutagenesis of an analogous tyrosine to phenylalanine in human cytosolic β-glucosidase, causes only a 2-5 fold decrease in k_{cat}, with minimal effect on K_M; this suggests that a polarisable π-aromatic ring system might have the capacity for transition state stabilization.[49] Free energy landscape analyses showed some shortening of the sugar-phenol distances ~0.5 Å at the point of ion pair formation, consistent with π-cation stabilization, albeit at a distance ~5-6 Å. Consistent with this reasoning, the aromatic residues (Tyr or Phe) at position 387 were found to be essential for activity: removal of the aromatic group by mutagenesis to Ala in SsβG-E387A resulted in a protein with no activity (**Supplementary Table 4**).

The front-face mechanism therefore appeared to proceed *via* an oxocarbenium ion-pair intermediate that, due to the greater steric bulk of the active site upon tyrosine introduction, was largely prevented from reacting with water to give the hydrolysis product. Instead, an acceptor bound in the +1 subsite, preferentially stabilized by the relocated Tyr322 residue, attacked the carbocation. The enzyme scaffold provided a shaped 'protein box' (primarily for

393 the donor) devoid of any catalytic residue but that nonetheless provided stabilization and
394 specified that reactants can only form β-products. This reactivity and selectivity was provided
395 (at least in part) by the box's favoring of particular conformers along the corresponding
396 itinerary (**Figure 4b**). Such a 'box' was highly reminiscent of the catalytic activity proposed for
397 serine protease mutants that, although lacking their entire catalytic triad, nonetheless show
398 rate accelerations of ~10^3-fold over background.[50] Notably the 'box' that is provided by
399 catalytic antibodies that act as glycosidases[51] that also lack participating residues are similarly
400 highly hydrophobic and, indeed, less efficient (rate accelerations of ~10^3-fold over background;
401 k_{cat} 0.007 min^{-1}, K_M 0.53 mM) than the designed 'S_Ni synthase' that we have created here (rate
402 accelerations of ~10^5-fold over background; k_{cat} 0.48 min^{-1}, K_M 0.17 mM). It should be noted
403 that our 'S_Ni synthase' was, in turn, a similar magnitude less active than prior 'S_N2 synthases'.
404 Further future activity optimization might be considered, through forced evolution strategies,
405 for example. Given the previously suggested[52] 'conceptual kinship' of some glycosyl units and
406 terpenes it is interesting to note that our initial inspection of known structures of terpene
407 cyclase structures suggests prominently placed aromatic sidechains, akin to the Y387 that we
408 have discovered here. Altogether, these results suggested that the, once seemingly
409 improbable and rare, same-face nucleophilic substitution is a viable mechanistic possibility in
410 many respects in nature and can be considered an accessible mechanism in the design of
411 catalysts for substitution.[53]
412
413

Acknowledgements. We thank the EPSRC and High Force Research (SMH), the BBSRC (SSL, BB/E004350/1), MINECO (grant CTQ2014-55174 to CR) and AGAUR (grant and 2014SGR-987 to CR) for funding. BGD was a Royal Society Wolfson Research Merit Award recipient during the course of this work. We acknowledge the computer support provided by the Barcelona Supercomputing Center (BSC−CNS). We would like to thank the referee who suggested possibly similar roles of aromatic sidechains in glycosyl- and terpenyl- processing enzymes that we note in the discussion. This paper is dedicated to the memory of Tony Fordham-Skelton, a friend, mentor and comrade who is still very much missed.

Author Contributions. JIF designed and performed calculations. SMH, SSL, MK performed the biochemical experiments. SMH, KM, AF-S determined x-ray structures. All authors analyzed results. CR, SSL, BGD wrote the manuscript. All authors except AF-S read and commented on the manuscript.

Supporting Information Available: Supporting Figures, Tables, Notes and Movie. This material is available free of charge via the Internet.

Competing Financial Interests: The authors declare no competing financial interests.

Data Availability Statement: The data that support the findings of this study are available in the SI and from the corresponding author upon reasonable request. X-ray crystallographic data that support the findings of this study have been deposited in the Protein Data Bank with the accession codes 5i3d.

437

References

1. Sinnott, M.L. & Jencks, W.P. Solvolysis of D-Glucopyranosyl Derivatives in Mixtures of Ethanol and 2,2,2-Trifluoroethanol. *Journal of the American Chemical Society* **102**, 2026-2032 (1980).
2. Persson, K.*, et al.* Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor analogs. *Nature Structural Biology* **8**, 166-175 (2001).
3. Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R. & Davies, G.J. Insights into Trehalose Synthesis Provided by the Structure of the Retaining Glucosyltransferase OtsA. *Chemistry & Biology* **9**, 1337-1346 (2002).
4. Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. Glycosyltransferases: Structures, Functions and Mechanisms. *Annual Review of Biochemistry* **77**, 521-555 (2008).
5. Koshland, D.E. Stereochemistry and the Mechanism of Enzymatic Reactions. *Biological Reviews* **28**, 416-436 (1953).
6. Lewis, E.S. & Boozer, C.E. The Kinetics and Stereochemistry of the Decomposition of Secondary Alkyl Chlorosulfites1. *Journal of the American Chemical Society* **74**, 308-311 (1952).
7. Chan, J., Tang, A. & Bennet, A.J. A Stepwise Solvent-Promoted S_Ni Reaction of a-D-Glucopyranosyl Fluoride: Mechanistic Implications for Retaining Glycosyltransferases. *Journal of the American Chemical Society* **134**, 1212-1220 (2012).
8. Vetting, M.W., Frantom, P.A. & Blanchard, J.S. Structural and Enzymatic Analysis of MshA from Corynebacterium glutamicum. *Journal of Biological Chemistry* **283**, 15834-15844 (2008).
9. Batt, S.M.*, et al.* Acceptor substrate discrimination in phosphatidyl-myo-inositol mannoside synthesis: structural and mutational analysis of mannosyltransferase Corynebacterium glutamicum PimB'. *Journal of Biological Chemistry* **285**, 37741-37752 (2010).
10. Chaikuad, A.*, et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proceedings of National Academy of Sciences USA* **108**, 21028-21033 (2011).
11. Yu, H.*, et al.* Notch-modifying xylosyltransferase structures support an SNi-like retaining mechanism. *Nature Chemical Biology* **11**, 847-854 (2015).
12. Errey, J.C.*, et al.* Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors. *Angewandte Chemie International Edition* **49**, 1234-1237 (2010).
13. Lee, S.S.*, et al.* Mechanistic evidence for a front-side, S_Ni-type reaction in a retaining glycosyltransferase. *Nature Chemical Biology* **7**, 631-638 (2011).
14. Ardevol, A. & Rovira, C. The Molecular Mechanism of Enzymatic Glycosyl Transfer with Retention of Configuration: Evidence for a Short-Lived Oxocarbenium-Like Species. *Angewandte Chemie International Edition* **50**, 10897-10901 (2011).
15. Goedl, C. & Nidetzky, B. Sucrose Phosphorylase Harbouring a Redesigned, Glycosyltransferase-Like Active Site Exhibits Retaining Glucosyl Transfer in the Absence of a Covalent Intermediate. *ChemBioChem* **10**, 2333-2337 (2009).
16. Gomez, H., Polyak, I., Thiel, W., Lluch, J.M. & Masgrau, L. Retaining Glycosyltransferase Mechanism Studied by QM/MM Methods: Lipopolysaccharyl-alpha-

1,4-galactosyltransferase C Transfers alpha-Galactose via an Oxocarbenium Ion-like Transition State. *Journal of the American Chemical Society* **134**, 4743-4752 (2012).
17. Bobovska, A., Tvaroska, I. & Kona, J. A theoretical study on the catalytic mechanism of the retaining [small alpha]-1,2-mannosyltransferase Kre2p/Mnt1p: the impact of different metal ions on catalysis. *Organic & Biomolecular Chemistry* **12**, 4201-4210 (2014).
18. Lira-Navarrete, E.*, et al.* Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide N-Acetylgalactosaminyltransferase 2. *Angewandte Chemie International Edition* **53**, 8206-8210 (2014).
19. Gomez, H.*, et al.* A computational and experimental study of O-glycosylation. Catalysis by human UDP-GalNAc polypeptide:GalNAc transferase-T2. *Organic & Biomolecular Chemistry* **12**, 2645-2655 (2014).
20. Albesa-Jove, D.*, et al.* A Native Ternary Complex Trapped in a Crystal Reveals the Catalytic Mechanism of a Retaining Glycosyltransferase. *Angewandte Chemie International Edition* **54**, 9898-9902 (2015).
21. Corbett, K., Fordham-Skelton, A.P., Gatehouse, J.A. & Davis, B.G. Tailoring the substrate specificity of the b-glycosidase from the thermophilic archeon Sulfolobus solfataricus. *FEBS Letters* **509**, 355-360 (2001).
22. Hancock, S.M., Corbett, K., Fordham-Skelton, A.P., Gatehouse, J.A. & Davis, B.G. Developing promiscous glycosidases for glycoside synthesis: residues W433 and E432 in Sulfolobus solfataricus b-glycosidase are important glucoside and galactoside specificity determinants. *ChemBioChem* **6**, 866-875 (2004).
23. Trincone, A., Improta, R. & Gambacorta, G. Enzymatic synthesis of polyol and masked polyol glucosides using b-glycosidase of Sulolobus solfataricus. *Biocatalysis and Biotransformations* **12**, 77-88 (1995).
24. Trincone, A.*, et al.* Enzyme catalysed syntheis of alkyl beta-D-glycosides with crude homogenate of Sulfolobus solfataricus. *Biotechnology Letters* **13**, 235-240 (1991).
25. Aguilar, C.F.*, et al.* Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *Journal of Molecular Biology* **271**, 789-802 (1997).
26. Trincone, A., Perugino, G., Rossi, M. & Moracci, M. A novel thermophilic glycosynthase that effects branching glycosylation. *Bioorganic and Medcinal Chemistry Letters* **10**, 365-368 (2000).
27. Mackenzie, L.F., Wang, Q., Warren, R.A.J. & Withers, S.G. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *Journal of the American Chemical Society* **120**, 5583-5584 (1998).
28. Zhang, Z.*, et al.* Programmable One-Pot Oligosaccharide Synthesis. *Journal of the American Chemical Society* **121**, 734-753 (1999).
29. Williams, S.J., Mark, B.L., Vocadlo, D.J., James, M.N.G. & Withers, S.G. Aspartate 313 in the Streptomyces plicatusHexosaminidase Plays a Critical Role in Substrate-assisted Catalysis by Orienting the 2-Acetamido Group and Stabilizing the Transition State. *Journal of Biological Chemistry* **277**, 40055-40065 (2002).
30. Withers, S.G., Street, I.P., Bird, P. & Dolphin, D.H. 2-deoxy-2-fluoroglucosides: a novel class of mechanism based glucosidase inhibitors. *Journal of the American Chemical Society* **109**, 7530-7531 (1987).

31. Lopez, R. & Fernandez-Mayoralas, A. Enzymatic galactosidation of modified monosaccharides:study of the enzyme selectivity for the acceptor and its application to the synthesis of disaccharides. *Journal of Organic Chemistry* **59**, 737-745 (1994).
32. Yamamoto, K. & Davis, B.G. Creation of an α-Mannosynthase from a Broad Glycosidase Scaffold. *Angewandte Chemie International Edition* **51**, 7449-7453 (2012).
33. Petzelbauer, I., Reiter, A., Splechtna, B., Kosma, P. & Nidetzky, B. Transgalactosylation by thermostable beta-glycosidases from Pyrococcus furiosus and Sulfolobus solfataricus. *European Journal of Biochemistry* **267**, 5055-5066 (2000).
34. Reuter, S., Rusborg Nygaard, A. & Zimmermann, W. beta-Galactooligosaccharide synthesis with beta-galactosidases from *Sulfolobus solfataricus*, *Aspergillus oryzae*, and *Escherichia coli*. *Enzyme and Microbial Technology* **25**, 509-516 (1999).
35. Crout, D.H.G. & Vic, G. Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis. *Current Opinion in Chemical Biology* **2**, 98-111 (1998).
36. Shim, J.-H., Chen, H., Rich, J.R., Goddard-Borger, E.D. & Withers, S.G. Directed evolution of a β-glucosidase from Agrobacterium sp. enhances its glycosynthase catalytic activity toward C3-OH modified donor sugar. *PEDS* **25**, 465-472 (2012).
37. Watts, A.G.*, et al. Trypanosoma cruzi trans*-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. *Journal of the American Chemical Society* **125**, 7532-7533 (2003).
38. Lawson, S.L., Warren, R.A.J. & Withers, S.G. Mechanistic consequences of replacing the active site nucleophile Glu-358 in *Agrobacterium sp.* beta-glucosidase with a cysteine residue. *Biochemical Journal* **330**, 203-209 (1998).
39. Gloster, T.M.*, et al.* Structural studies of the beta-glycosidase from Sulfolobus solfataricus in complex with covalently and non-covalently bound inhibitors. *Biochemistry* **43**, 6101-6109 (2004).
40. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **99**, 12562-12566 (2002).
41. Asensio, J.L., Ardá, A., Cañada, F.J. & Jiménez-Barbero, J. Carbohydrate–Aromatic Interactions. *Accounts of Chemical Research* **46**, 946-954 (2013).
42. Biarnés, X., Nieto, J., Planas, A. & Rovira, C. Substrate Distortion in the Michaelis Complex of Bacillus 1,3–1,4-β-Glucanase: Insight From First Principles Molecular Dynamics Simulations. *Journal of Biological Chemistry* **281**, 1432-1441 (2006).
43. Ardèvol, A. & Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *Journal of the American Chemical Society* **137**, 7528-7547 (2015).
44. Davies, G.J., Planas, A. & Rovira, C. Conformational Analyses of the Reaction Coordinate of Glycosidases. *Accounts of Chemical Research* **45**, 308-316 (2012).
45. Speciale, G., Thompson, A.J., Davies, G.J. & Williams, S.J. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Current Opinion in Structural Biology* **28**, 1-13 (2014).
46. Bottoni, A., Miscione, G.P. & DeVivo, M. A theoretical DFT investigation of the lysozyme mechanism: computational evidence for a covalent intermediate pathway. *Proteins* **59**, 118-130 (2005).
47. Vocadlo, D.J. & Davies, G.J. Mechanistic insights into glycosidase chemistry. *Current Opinion in Chemical Biology* **12**, 539-555 (2008).
48. Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A. & Rovira, C. Catalytic Itinerary in 1,3-1,4-β-Glucanase Unraveled by QM/MM Metadynamics. Charge Is Not

Yet Fully Developed at the Oxocarbenium Ion-like Transition State. *Journal of the American Chemical Society* **133**, 20301-20309 (2011).
49. Berrin, J.-G.*, et al.* Substrate (aglycone) specificity of human cytosolic -glucosidase. *Biochemical Journal* **373**, 41-48 (2003).
50. Carter, P. & Wells, J.A. Dissecting the catalytic triad of a serine protease. *Nature* **332**, 564-568 (1988).
51. Janda, K.D.*, et al.* Chemical selection for catalysis in combinatorial antibody libraries. *Science* **275**, 945-948 (1997).
52. Schreiber, S.L. Rethinking relationships between natural products. *Nature Chemical Biology* **3**, 352-352 (2007).
53. An, J., Denton, R.M., Lambert, T.H. & Nacsa, E.D. The development of catalytic nucleophilic substitution reactions: challenges, progress and future directions. *Organic & Biomolecular Chemistry* **12**, 2993-3003 (2014).

Figures and Legends

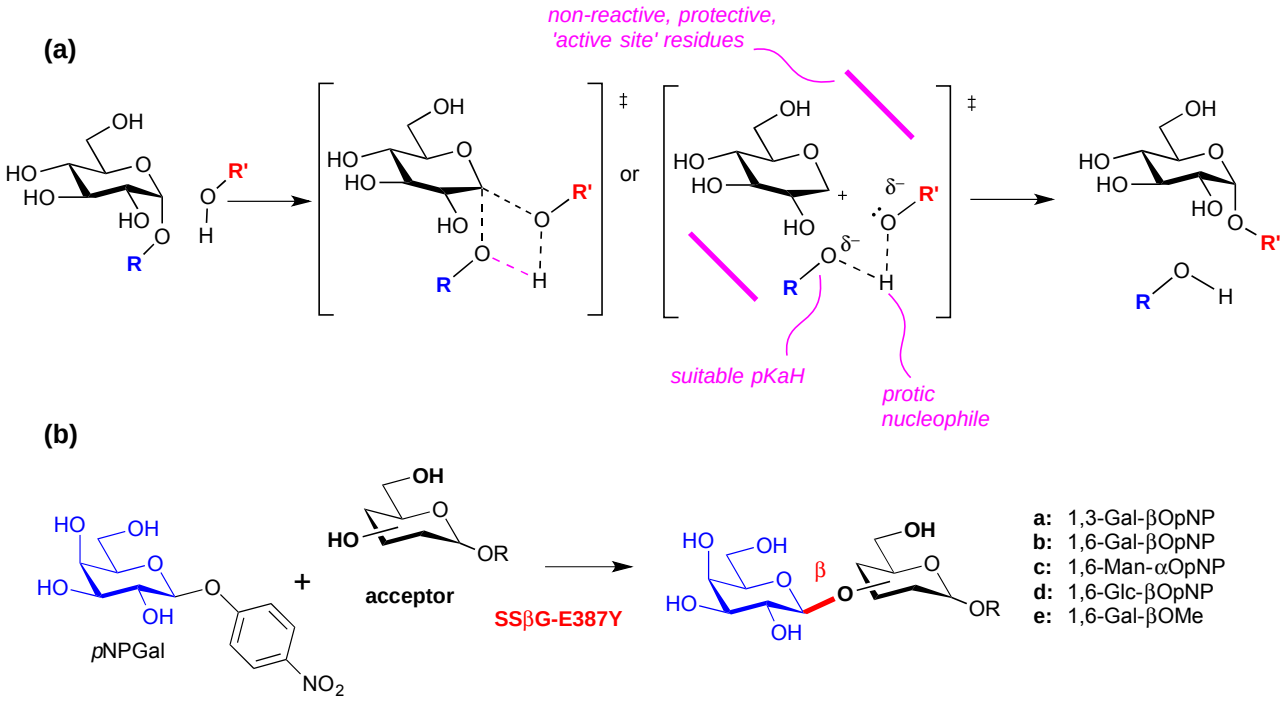


Figure 1. Comparisons of 'Front-face' Glycosyl Transfer **(a)** Front-face reaction mechanism of known α-selective retaining glycosyltransferases. **(b)** Transglycosylation reactions catalysed by the β-selective 'front-face' synthase described here, reactions detailed in Table 1

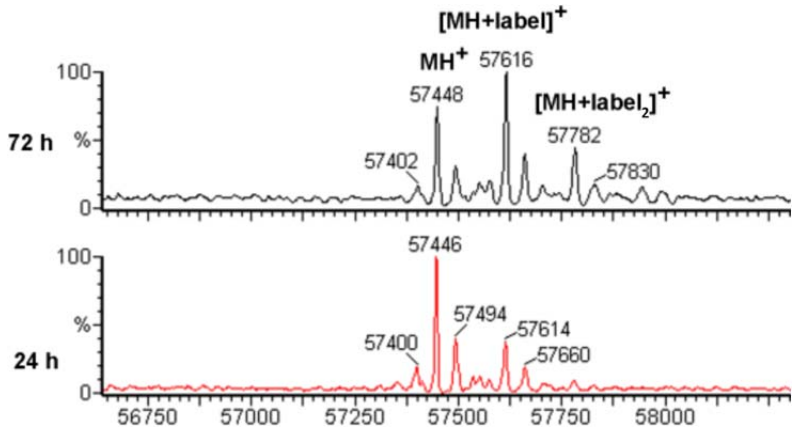


Figure 2. Mass Spectrometric Analysis of Incubation of SsβG-E387Y with Covalent Inhibitor DNP-2FGlc. Reaction with 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (**16**) was monitored over time by ESI-MS. Slow reaction and emergence of additional peaks (2 × +165 ±3 Da etc) after extended incubation and with an apparent statistical distribution suggest non-specific chemical modification; incubation with 2FGlc did not cause direct glycation (**Supplementary Figure 3**). This non-specific, non-'activity-based' cause is also consistent with the thermal denaturation of SsβG-E387Y at 45°C >16h (*vide supra*).

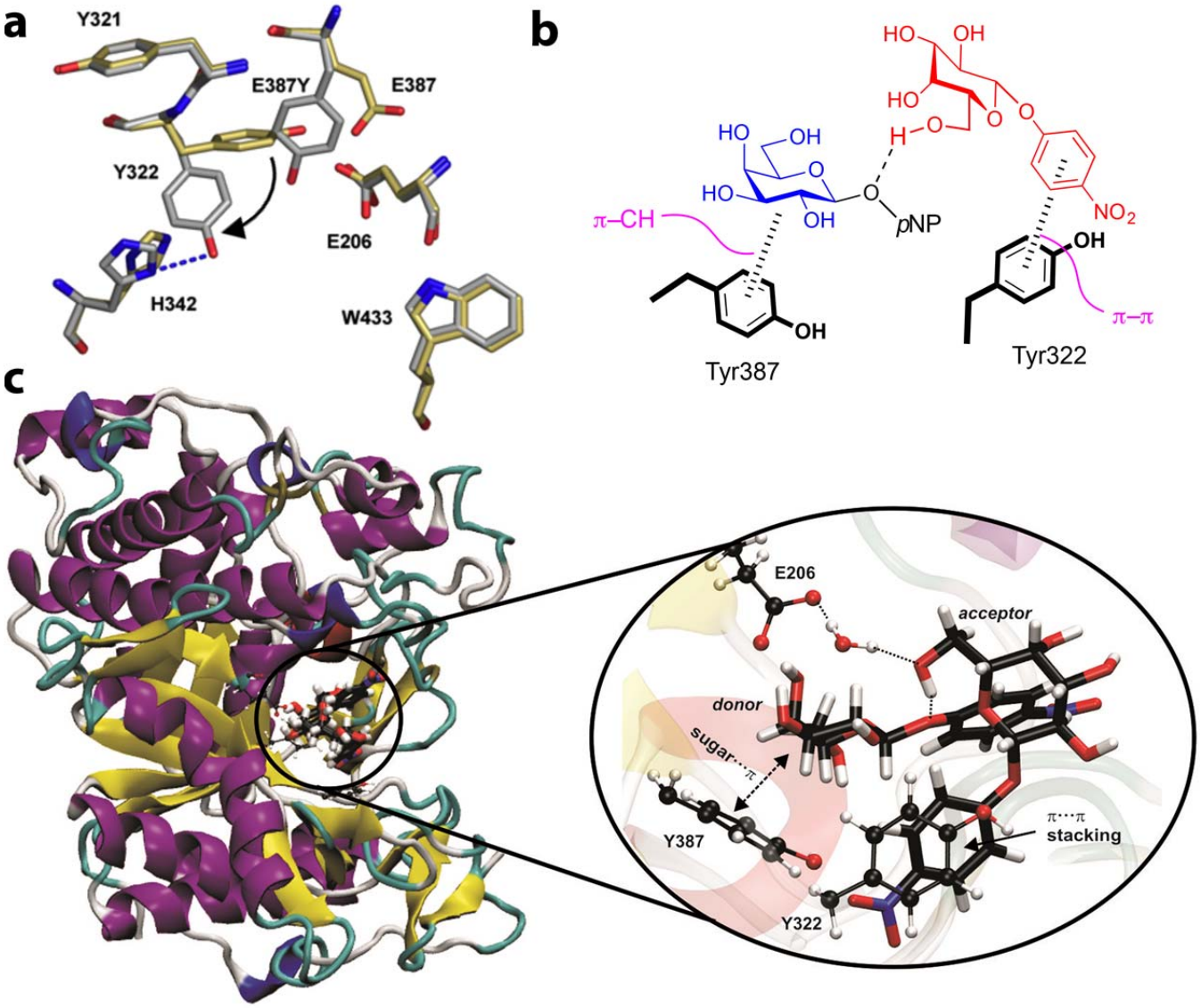
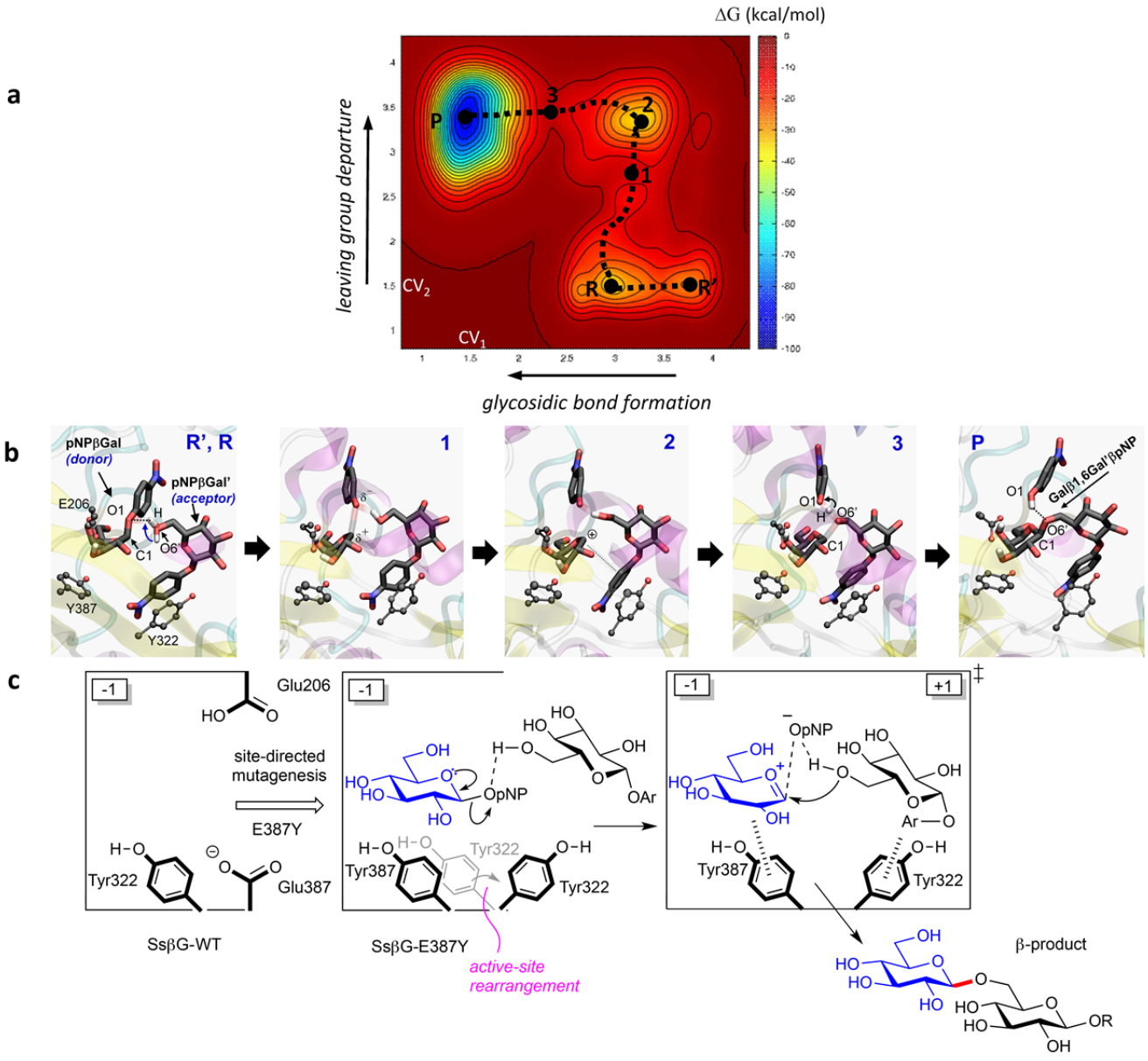


Figure 3. **Structural Analysis of SSβG-E387Y. (a)** The X-ray structure of *apo* SSβG-E387Y (determined in this work: pdb 5i3d, silver) superimposed on SSβG-WT (pdb: 1gow, gold) shows the highly localized rearrangement (indicated by curled black arrow) of residues Y322 and H342 to accommodate the changed residue at 388 (E387Y). The hydroxyl of Y322 is within ~3.1 Å of the Nδ1 of H342, suggesting that a hydrogen bond stabilizes this amino acid side chain migration (blue dashes). Essentially negligible alterations are observed in the rest of the structure. **(b)** Schematic interaction diagram of proposed substrate-protein interactions based on (a) and (c): Y387 forms stabilizing *donor* sugar···π interactions[41] (sugar hydrogen atoms point towards the center of the Y387 phenol ring, with distances < 3 Å, see (c)); the

localized Y322 rearrangement creates π···π stacking interactions with the *acceptor p*NPGal moiety. This, in turn, positions the acceptor OH-6 in an orientation to attack the anomeric carbon of the sugar donor. **(c)** Structure of SsβG-E387Y in complex with two pNPβGal molecules. This Michaelis complex was obtained from classical metadynamics simulations (see **Online Methods**) based upon the determined *apo* x-ray structure (determined in this work: pdb 5i3d, silver) shown in (a). The inlay shows an expanded view of the active site.



634

635 **Figure 4. Analysis of the 'S$_N$i' reaction pathway. (a)** Free energy landscape (FEL)
636 reconstructed from the metadynamics simulation of the transglycosylation reaction (projection
637 on two collective variables CV_1 and CV_2). Contour lines are at 5 kcal/mol. The second
638 transition state (labelled as **3** on the reaction pathway in (b)) is above in energy with respect to
639 the first one (labelled as **1**) by 3 kcal/mol. **(b)** Hydrogen atoms have been omitted for clarity,
640 except the one being transferred from the sugar acceptor to the *p*NP leaving group of the
641 donor molecule and the hydroxyl hydrogen atoms of the Gal donor that interact with E206.

642 Bonds being broken/formed are represented by a transparent bond (in **1** and **3**), whereas
643 dotted lines indicate hydrogen-bonding interactions. **(c)** Proposed 'front-face' substitution
644 mechanism of SsβG-E387Y.

645

646

647

Tables

Table 1. SsβG-E387Y catalyzes transglycosylation. Disaccharides synthesised from *p*NPGal **1** as a glycosyl donor (see **Figure 1b** for relevant reaction).

	Acceptor	Temp / °C	Product and Yield / %[a]								S / H	Conversion[d] / %
			12	4	13	14	15	H[b]	S[c]	Total		
i	MeβGal **6**	45	18	24	-	-	2	37	44	81	1.2	92
ii	MeβGal **6**	80	51	36	-	-	1	<1	88	88	>88	78
iii	cellobiose **7**	45	14	15	-	-	-	44	29	73	0.7	100
iv	cellobiose **7**	80	22	27	-	-	-	6	49	55	8.2	79
v	lactose **8**	45	21	29	-	-	-	33	50	75	2.0	80
vi	lactose **8**	80	30	54	-	-	-	16	84	100	5.3	91
vii	MeβMan **9**	45	16	38	-	-	-	46	54	100	1.2	100
viii	MeβMan **9**	80	39	46	-	-	-	15	85	100	5.7	92
ix	PhβGlc **10**	45	9	46	-	26	-	17	81	98	4.8	97
x	PhβGlc **10**	80	0	28	-	12	-	37	-	-	-	100
xi	PhαMan **11**	45	0	3	12	-	-	85	15	100	0.2	100
xii	PhαMan **11**	80	1	10	25	-	-	64	36	100	0.6	100
xiii	PhβGlc[e] **9**	45	5	0	-	72	-	<1	>99	100	>100	-[e]

[a] Yields were determined by NMR analysis of the per-acetylated reaction mixture, separated by flash chromatography and based on the recovery of starting material. Reaction times were determined by period of catalytic activity i.e. until no further progression ~15h or longer. [b] Total yield of **H**ydrolysis products. [c] Total yield of glycosides/**S**ynthesis products. [d] based on the consumption of starting material [e] After optimization for yield, including additional production of trisaccharide as mass balance – see **Supplementary Table 3**.

660

Online Methods

Materials

BL21(DE3) competent *E. coli* cells and pET28a(+) vector were purchased from Merck Bioscience (Nottingham, UK). TOP10 competent E. coli cells were from Life Technologies (Paisley, UK). Bradford reagent concentrate was purchased from Bio-Rad (Hemel Hempsted, UK). Clostripain was from Worthington Biochemicals Corporation (Lakewood, NJ USA). LB broth, kanamycin, and IPTG were from Melford (Ipswich, UK). Methyl β-D-galactopyranoside and 2,3,4,6-tetra-*O*-acetyl-α-d-galactopyranosyl fluoride was purchased from Carbosynth (Compton, UK). RapiGest™ SF reagent was from Waters, Ltd (Elstree, UK). Recombinant *Sulfolobus solfataricus* β-glycosidase and mutants in pET24d in *E. coli* strain BL21(DE3) were kindly provided by Mr. K. P. Corbett and Dr. A. P. Fordham-Skelton (University of Durham). All chemical and biochemical reagents were purchased from Sigma-Aldrich Chemical Co. unless otherwise noted.

Biological Methods

H_2O was purified using a Milli-Q Synthesis system (Millipore), and was heat-sterilised before DNA applications. Agar plates were poured and streaked in a MDH Microflow laminar flow cabinet using media and equipment that had been sterilised in an autoclave at 121°C for 20 min. Solutions of IPTG (0.1 M) and kanamycin (50 $mgmL^{-1}$) were filter-sterilised through 0.2 μm filters and stored at -20°C. The final concentrations of kanamycin and IPTG in growth media, unless otherwise indicated, were 50 μgmL^{-1} and 0.1 mM, respectively. All growth media were autoclaved at 121°C for 20 min before use. Sequencing was carried out by the University

683 of Oxford, Department of Biochemistry DNA sequencing service on a ABI 377XL Prism DNA
684 sequencer. SsβG was sequenced using primers to the T7 promoter and terminator sequences
685 and the following internal primers:

686 Internal forward: 5′-CGT AGG CAT ATG TAT AAC ATC

687 Internal reverse: 5′-GGA ATG AGC TAT TAG C.

688

689 *Cloning the SsβG gene into pET28a(+)*

690 The gene of SsβG was originally inserted in pET24d and was subcloned into pET28a(+). PCR
691 reaction was conducted using the following primers.

692 Forward: 5′-GGTGGTCATATGTCATTTCCAAATAGC

693 Reverse: 5′-GGTGGTCTCGAGTTAGTGCC

694 The amplified WT SsβG insert was cleaned using a Qiaquick gel extraction kit (Qiagen,
695 Manchester, UK) and digested to generated sticky ends using XhoI and NdeI (both from
696 Promega, Southampton, UK). The restriction digest was carried out for 3 h at 37°C.

697 The reaction mixture was electrophoresed on a agarose gel and the digested DNA fragments
698 were extracted. The purified insert was ligated to pET28a(+) vector that was already digested
699 with XhoI and NdeI. The ligation reaction was effected by T4 DNA ligase (Promega) and
700 conducted at 4°C overnight. TOP10 cells (25 μL) were transformed with ligation mixture (5 μL)
701 by the manufacturer's standard protocol.

702

Site-directed mutagenesis

Site-directed mutagenesis of SsβG WT and further mutagenesis of E387Y SsβG in pET28a(+) was conducted using QuickChange II Site-Directed Mutagenesis Kit (Agilent, Stockport, UK) following the manufacturer's protocol. Used primers were as follows (Mutagenic codons are underlined):

E387Y

Forward: 5′-CTATATGTACGTTACT**TAC**AATGGTATTGCGGATGATGCC

Reverse: 5′-GATAATCGGCATCATCCGCAATACCATT**GTA**AGTAACGTAC

E206A

Forward: 5′-CAATGAAT**GCA**CCTAACGTGGTGG

Reverse: 5′-CAACGTTAGG**TGC**ATTCATTGTTG

E387F

Forward: 5′-CTATATGTACGTTACT**TTC**AATGGTATTGCGGATGATGCC

Reverse: 5′-GATAATCGGCATCATCCGCAATACCATT**GAA**AGTAACGTAC

Y322F

Forward: 5′-GGAGTTAATTAT**TTC**ACTAGGACTGTTGTG

Reverse: 5′-CAGTCCTAGT**GAA**ATAATTAACTCCAATCC

E387A

Forward: 5′-CTATATGTACGTTACT**GCA**AATGGTATTGCGGATGATGCC

Reverse: 5′-GATAATCGGCATCATCCGCAATACCATT**TGC**AGTAACGTAC

E. coli TOP10 cells (25 μL) were transformed with Dpn1 digested PCR products (1 μL) and the miniprep of mutated DNAs was conducted by the manufacturers' protocol. Resulting DNAs were sequenced and the mutagenesis was confirmed.

Protein expression and purification

The plasmids coding for E387Y SsβG and its further mutants inserted in pET28a(+) were used to transform *E. coli* BL21 DE3 cells. Transformed cells were grown overnight in 50 mL of LB broth containing 85 μg/mL kanamycin. 2 L flasks containing LB broth (500 mL) and kanamycin were pre-warmed at 37°C for 0.5 h and inoculated with 25 mL of an overnight culture. The culture was grown at 37°C at 200 rpm until an OD of 0.6-0.9 (typically, 1.5-2 h). Expression was induced by the addition of IPTG (0.5 mL, 0.1 M) and grown for a further 5-6 h. Cells were harvested by spinning at 9000 rpm in the JA10 Beckman rotor for 20 min at 4°C. Pellets were stored at -20°C. The frozen cell pellets were resuspended in binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris pH 7.8). Cell suspensions were sonicated on ice. This consisted of 3 × 15 amplitude micron bursts of 30 s separated by 1 min intervals. The lysed cells were centrifuged at 10,000 rpm in a JA20 Beckman rotor for 30 min at 4°C. Lysates were filtered through a Nalgene 0.2 μm filter prior to protein purification.

The filtered supernatant was applied to a pre-equilibrated (binding buffer) GE Healthcare 5mL HisTrap Ni-NTA column on an Äkta FPLC system (GE Healthcare, Bucks, UK). The column was washed at 1 mL/min with 20 column volumes of the same buffer and then eluted with a linear gradient of imidazole (10 mM to 500 mM over 25 column volumes) in Buffer A. Protein

744 was detected with an on-line detector monitoring A280 and column fractions were collected
745 and analyzed by SDS-PAGE. Fractions containing the ca. 57 kDa protein were pooled. The
746 protein was further purified by the gel filtration. Using an Äkta FPLC system, SsβG (~54 mg, 3
747 mL in 100 mM Tris-HCl, pH 7.5) was loaded onto a Superdex 75 (320 mL) gel filtration
748 column, and eluted with Tris-HCl buffer (100 mM, pH 7.5). Buffers were filtered and degassed
749 before use. Fractions were checked with SDS PAGE and those containing desired protein
750 pooled. Pooled protein samples were dialyzed into the appropriate buffer (2 L) overnight with
751 two changes. The protein was concentrated using Vivaspin centrifugal concentrator (Sartorius,
752 Goettingen, Germany). Protein concentration was determined by Coomassie Brilliant Blue
753 binding via absorption at 595 nm.[54] *N*-terminal sequencing was performed by the University
754 of Oxford, Department of Biochemistry, Immunochemistry protein characterisation service,
755 using an Applied Biosystems Procise 494A protein sequencer employing Edman degradation.

756

757 *Circular dichroism*

758 CD spectra were recorded from 190-250 nm on a Jasco J-720 spectropolarimeter using a
759 1 mm quartz cuvette. Samples contained 50 mM (or lower) sodium phosphate buffer, pH 6.5
760 and were approximately 0.2 mg/mL. The spectral background was subtracted from the data.

761

762 *Protein mass spectrometry*

763 SsβG and its mutants were analyzed by electrospray ionization-mass spectrometry (ESI-MS)
764 in positive ion mode on a Micromass LCT mass spectrometer interfaced with a Waters 2790
765 Alliance HT separations module using a Jupiter C5 5 μm 300 Å 150 × 2 mm i.d. column

766 (Phenomenex, Macclesfield, UK). Before injection, protein samples (20 μL, ~60 μM) were
767 washed with H_2O using spin concentrators. Proteins were eluted at 0.2 mL/min by a gradient
768 from Buffer A (0.1% formic acid in 95% water, 5% acetonitrile) to Buffer B (0.1% formic acid in
769 95% acetonitrile, 5% water) as follows:

Time / min	**Buffer B / %**	**Curve**
0	5	1
3	5	1
16	100	4
18	100	1
19	5	10
25	5	1

770 The eluent was split 1:1 waste:mass spectrometer. The following MS parameters were used:
771 capillary voltage, 3000 V; sample cone, 35 V; desolvation temperature, 200°C; source
772 temperature, 80°C; desolvation flow (N_2), 425 Lh^{-1}; no cone flow; pusher cycle time, 94; and
773 ion energy, 34 V; *m/z* scan range 200 to 2100; scan time, 1 s; interscan time, 0.1 s. The
774 electrospray mass spectra were processed using the Maximum Entropy method (MaxEnt1).

775

776 *Proteolysis reactions*

Cyanogen bromide cleavage was conducted as follows. Protein (0.5 mL, 1.69 $mgmL^{-1}$, 15 nmol (total methionine = 150 nmol)) was concentrated to 25 μL and washed with H_2O (3 × 300 μL) in a 500 μL spin concentrator (MWCO 10,000). HCl (100 μL, 0.1 M, 10 μmol) was added to the protein and nitrogen gas was bubbled through the solution for 2 min. CNBr (2.9 μL, 5 M in MeCN, 15 μmol, 100 eq) was added and the reaction mixture incubated in the dark at rt for 24 h. The mixture was diluted with water (2 × 2 mL), freeze-dried, and re-suspended in water (125 μL) for MS analysis. For larger scale digestions, CNBr-digested peptides were passed through a 20 mL G10 Sephadex column (VL11 × 250 mm) eluting with phosphate buffer (pH 6.5) at 2-3 mL/min. A Bradford test determined which fractions were collected and freeze dried.

Trypsin digests were conducted as follows. Protein (0.3 mg) was concentrated, washed with H_2O, and was either: a) Incubated at 95°C for 5 min with 2% w/v RapiGest™ SF in NH_4HCO_3 buffer (20 mM, pH 8.0) in a total volume of 65 μL; b) Incubated at 95°C for 20 min in a total volume of 65 μL; c) Incubated at 95°C for 15 min in Tris-HCl (50 mM, pH 8.0) containing guanidine-HCl (6 M) and β-mercaptoethanol (4 mM) in a total volume of 40 μL. Following heating, the solution was diluted with NH_4HCO_3 (180 μL, 20 mM, pH 8.0). Once cooled, sequencing-grade modified trypsin (20 μg) in NH_4HCO_3 (25 μL, 20 mM, pH 8.0) was added and incubated at 37°C overnight. The RapiGest™ SF detergent was removed by adding HCl (10 μL, 0.5 M) and incubating for 0.5 h with at 37°C. The resulting precipitant was removed by centrifugation and the supernatant was injected directly into the LC-MS.

Clostripain digests were performed as follows. HPLC purified CNBr-generated peptides were re-suspended after lyophilization in incubation buffer (85 μL). To this was added clostripain (0.1 mg) in re-suspension buffer (5 μL) and activation solution (10 μL) and the mixture was incubated at 37°C overnight. Peptides were injected directly into the LC-MS.

800 Pepsin digests were performed as follows. WT SsβG (0.1 mg, 50 μL) was digested by
801 incubating with pepsin (10 μg) in sodium phosphate (20 μL, 2 M, pH 2) for 1 h at rt. The
802 solution was slowly neutralized with NaOH (1 M) before ESI-MS analysis.

803

804 *Semi-preparative scale HPLC purification of CNBr generated peptides*

Time / min.	**Buffer B / %**	**Curve**
0	5.6	1
5	5.6	1
30	42.2	6
50	100	6
60	100	1
70	5.6	6
80	5.6	1

805 CNBr-generated peptides (100 μL, 116 μM) were injected onto a Jupiter C4 5 μm 300 Å
806 column (150 × 4.6 mm i.d.) connected to a Waters 2790 Alliance HT separations module,
807 directly interfaced to a Micromass LCT mass spectrometer fitted with an ESI source. The
808 peptides were eluted at 1 mL/min by a gradient from Buffer A (0.1% formic acid in 95% water,
809 5% acetonitrile) to Buffer B (0.1% formic acid in 95% acetonitrile, 5% water) as shown in the
810 table above. A post-column splitter was used to direct 85% of flow to the fraction collector,

811 whilst 15% entered the mass spectrometer for analysis. Peptide elution was followed by ESI-
812 MS, in positive ion mode, using the same MS parameters as reported for protein mass
813 spectrometry, except for *m/z* scan range, which was 200-2500. Fractions (1 mL) were
814 collected and freeze dried for further digestion/analysis.

815

816 *Peptide analysis by MS and MS/MS*

Time / min	**Buffer B / %**	**Curve**
0	10	1
5	10	1
30	43	6
50	95	6
60	95	6
70	10	6
80	10	1

817

818 Protein digestion mixtures were injected onto a Jupiter C4 5 μm 300 Å capillary column (150 ×
819 0.5 mm i.d.) connected to an Agilent 1100 series capillary HPLC. The peptides were eluted at
820 15 μL/min by a gradient from Buffer A (0.1% formic acid in water) to Buffer B (0.1% formic acid
821 in acetonitrile) as shown in the table above. UV absorption at 210 and 280 nm and positive ion

822 mode ESI-MS (Q-TOF micro, Micromass) followed peptide elution using the following
823 parameters: capillary voltage, 3000 V; sample cone, 35 V; extraction cone, 6 V; desolvation
824 temperature, 150°C; source temperature, 80°C; desolvation flow (N_2), 150 Lh^{-1}; *m/z* scan
825 range, 400–2000; step size, 1 s; interscan time, 0.1 s. MS/MS data was obtained by
826 selectively introducing the precursor ion (mass triggered) into the collision cell and
827 fragmentation was induced by collision with Ar using charge state recognition, calculating the
828 collision energy from the standard profile within MassLynx 4. The resulting product ions
829 produced were scanned over the range 100 to 3000, scan time 1 s, and with 0.1 s interscan
830 time. Product ion spectra were deconvoluted using MaxEnt3 and *de novo* sequenced using
831 Pepseq within Masslynx 4.

832

833 *2,4-Dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside labelling*

834 To label E387Y, E387F, E387Y:E206A and E387Y:Y322F SsβG, 2,4-dinitrophenyl 2-deoxy-2-
835 fluoro-β-d-glucopyranoside (1.0 mg, 2.9 μmol, 1000 eq) was incubated with enzyme (0.20 mg,
836 3.5 nmol) in sodium phosphate buffer (200 μL, 50 mM, pH 6.5) for 3 d at 45°C.

837 To label WT SsβG, 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-d-glucopyranoside (3.0 mg, 8.6 μmol,
838 40 eq) was incubated with enzyme (10 mg, 0.2 μmol) in sodium phosphate buffer (7.0 mL,
839 50 mM, pH 6.5). Complete conversion had occurred after 24 h at 45°C.

840 To determine whether E387Y SsβG activity was due to a contaminant, the enzyme was
841 incubated with 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-d-glucopyranoside (0.25-0.35 eq) for 16 h
842 at 45°C and then measured for their hydrolytic activity with *p*-nitrophenyl β-d-
843 galactopyranoside.

Solvolysis of 2,4 dinitrophenyl 2-deoxy-2-fluro-β-D-glucopyranoside with or without SSβG-E387Y was performed as follows. 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside **16** (1mg, 2.9 μmol, 1000 eq) was incubated with E387Y-SSβG (0.2 mg, 3.5 nmol in 50 mM sodium phosphate, pH 6.5) or buffer (50 mM sodium phosphate, pH 6.5) at 45°C. The UV absorption of DNP release was monitored at 405 nm on a 97-well SpectraMax Plus plate reader (Molecular Devices, New Milton, UK). Graphs were drawn using Graphpad prism software.

Test of glycation of SSβG-E387Y after incubating with 2-deoxy-2-fluoro-D-glucose was performed as follows. 2-deoxy-2-fluoro-D-glucose (1mg, 2.9 μmol, 1000 eq) was incubated with SSβG-E387Y (0.2 mg, 3.5 nmol) in 50 mM sodium phosphate, pH 6.5 at 45°C. Aliquots of samples (20 μl) were collected at different time points and analyzed by ES-MS.

Kinetic analyses

Kinetic parameters for glycoside hydrolysis were determined as follows. The extinction coefficients of pNP and DNP were determined by measuring the absorbance at various of concentrations (200 μL, 0.125 - 1 mM) of each compound in 50 mM phosphate buffer at pH 6.5 at 405 nm. Substrate concentration was plotted against absorbance and according to the Beer-Lambert law, the gradients were equal to the extinction coefficient. R^2 values were always greater than 0.99. The extinction coefficients of pNP and DNP were 3212 $M^{-1}cm^{-1}$ and 4230 $M^{-1}cm^{-1}$, respectively.

Hydrolyses of *p*-nitrophenyl glycosides were performed at 45°C in sodium phosphate buffer (50 mM, pH 6.5). Assays were initiated by adding enzyme (10 μL, final 0.079 μM for WT, 0.93

μM for SsβG) to substrate (190 μL, 0.0125 - 10 mM) and *p*-nitrophenol release was monitored at 405 nm on a 96-well SpectraMax Plus plate reader (Molecular Devices, New Milton, UK). The initial rates were used and K_M and V_{max} were determined from curve fitting, non-linear regression using GraFit 4 (Erithacus Software). Errors in kinetic parameters were calculated from the standard error of curve fitting.

In cases where the high K_M and substrate solubility prevented determination of the Michaelis-Menten parameters, approximate k_{cat}/K_M was determined using the limiting case of the Michaelis-Menten equation at low substrate concentration. Errors were determined from the standard deviation of the rates and protein concentration.

2,4-Dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside labelling kinetics were performed as follows. SSβG-E387Y (0.2 mg, 3.5 nmol) were incubated with 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (1 mg, 2.9 μmol, 1000 eq) in sodium phosphate buffer (200 μL, 50 mM, pH 6.5) at 45°C. The UV absorption of the DNP released was measured, and aliquots (10 μL) were added to *p*-nitrophenyl β-D-galactopyranoside (190 μL, 0.1 mM) in sodium phosphate buffer (50 mM, pH 6.5) to determine hydrolytic activity over time.

Inhibition assays were performed as follows. Enzyme/inhibitor solution (15 μL) was added to *p*-nitrophenyl glycopyranoside (185 μL, 2 - 0.05 mM) in sodium phosphate buffer (50 mM, pH 6.5) at 45°C and *p*-nitrophenol release was monitored at 405 nm. Inhibitor concentrations were varied from 0-0.1 M. Data were interpreted using Dixon plot analysis.[55]

NMR kinetics were performed as follows. Deuterated sodium phosphate buffer (50 mM, pH 6.5) and enzyme solutions were prepared by lyophilizing and resuspending in D_2O. The pH refers to that of the non-deuterated solution from which the buffer was prepared.

SSβG-E387Y hydrolysis of 2-methyl-4,5-(2-deoxy-α-d-glucopyrano)-Δ^2-oxazoline was examined as follows. 2-Methyl-4,5-(2-deoxy-α-d-glucopyrano)-Δ^2-oxazoline (final concentration 400 – 2.0 mM), was dissolved in SSβG-E387Y solution (500 μL) in deuterated sodium phosphate buffer (50 mM, pH 6.5). NMR spectra were collected at 45°C on a Bruker Avance 500 spectrometer at 5 min intervals, with referencing to an internal drop of dioxane. Integration of the anomeric peaks of the oxazoline and *N*-acetylglucosamine allowed the initial rates of hydrolysis to be calculated. Background degradation was measured using the same protocol, dissolving the oxazoline in deuterated buffer (500 μL).

NMR substrate hydrolysis kinetics for SSβG-E387Y were performed as follows. *p*-Nitrophenyl 6-*O*-(β-D-galactopyranosyl)-β-D-galactopyranoside (0.1 – 1 mM) in sodium phosphate buffer (50 mM, pH 6.5) was incubated with WT (23 μg) or SSβG-E387Y (60 μg) at 45°C for 5 and 25 min, respectively. Protein was removed by a spin concentrator (MWCO 10 000), and samples were freeze dried for NMR analysis.

pH dependence of enzyme activity

Enzyme solution (10 μL) was added to *p*-nitrophenyl β-D-glycopyranoside (190 μL, 0.25/0.1 mM) at 45°C in the appropriate buffer for the given pH range (pH 2-5, 20 mM succinate; pH 5-7, 20 mM MES; pH 7-8, 20 mM HEPES; pH 8-11, 20 mM CHES). Reactions were incubated for an appropriate time for the enzyme activity (SSβG-E387Y, 15 min; SSβG-WT, 5 min) and stopped by the addition of Na_2CO_3 (50 μL, 1 M, pH 11.0). Initial rates were calculated from the release of *p*-nitrophenol (405 nm). Apparent pK_a values were determined by fitting the k_{cat}/K_M data as a function of pH using the following equation[56]:

$$\left(\frac{k_{cat}}{K_M}\right)_{obs} = \left(\frac{k_{cat}}{K_M}\right)_{max}\left(\frac{1}{1 + \frac{10^{-pH}}{10^{-pK_{a\,1}}} + \frac{10^{-pK_{a\,2}}}{10^{-pH}}}\right)$$

910

911

912 *HPLC Analyses*

913 For initial hydrolysis assays, the reaction mixture containing p-nitrophenyl β-D-galactoside (10
914 mM) and SsβG-E387Y (0.93 μM) in sodium phosphate buffer (50 mM, pH 6.5) was incubated
915 at 45°C. 40 μL aliquots were withdrawn at every 1 hour interval and aliquots were immediately
916 filtered using viva spin centrifugal filter (Sartorius, mwco 10,000). Each filter was washed with
917 20 μL of distilled water. Volume was increased to 60 μL. 5 μL portion of the sample was
918 injected into Phenomenex Luna NH_2 HPLC column on Dionex UltiMate 3000 system (Dionex,
919 Hemel Hempstead, UK). Analysis was performed by eluting isocratic 70/30 acetonitrile/water
920 at 1 mL/min flow rate.

921

922 *Transglycosylation timecourse.*

923 To determine the time course of the reaction, 2 μM SsβG-E387Y (in 50 mM sodium phosphate
924 buffer, pH 6.5) was incubated with 10 mM 4-nitrophenyl β-galactoside at 45°C. 60 μl aliquots
925 were withdrawn at 15 min, 30 min, 45 min, 1h, 3h, 5h and 6h. Samples were filtered using
926 vivaspin centrifugal filter (Sartorius, MWCO 10,000). 5 μl portion of the sample was injected
927 into Phenomenex Luna NH2 column on Shimadzu HPLC system. Samples were eluted in
928 isocratic 80/20 acetonitrile/water solution at 1 ml/min and monitored at 305 nm.

929

Transglycosylation kinetic parameter determination.

To determine the kinetic parameters of SsβG-E387Y-mediated transglycosylation, 2 μM SsβG-E387Y (28 μl in 50 mM sodium phosphate buffer, pH 6.5) was incubated with substrate at different concentrations for 6h at 45°C. 50 mM sodium phosphate buffer, pH 6.5 was added to make the final reaction mixture of 100 μl. Samples were filtered using Vivaspin centrifugal filter (Sartorius, MWCO 10,000). 5 μl portion of the sample was injected into Phenomenex Luna NH2 column on Shimadzu HPLC system. Samples were eluted in isocratic 80/20 acetonitrile/water solution at 1 ml/min and monitored at 305 nm. 4-nitrophenyl-β-galactoside and 4-nitrophenyl-β-lactoside were used as standards to estimate approximate retention times for mono- and di-saccharides (eluting at ~ 4.2 and 6.3 min, respectively).

From the reaction mixture samples two major peaks were observed at retention time ~ around 6 min corresponding to 1,3 and 1,6 products and the concentrations of the products were determined using calibration curves for 4-nitrophenyl-β-lactoside and UV-vis absorbance as measured by area under the curve from chromatograms at fixed timepoints within the linear range allowing direct estimate of k_{cat}/K_M using the low substrate approximation. Under similar conditions but with higher extremes of enzyme concentration (> 5μM) small amounts of trisaccharide (typically < 5%) were also observed. In all cases, in the absence of enzyme, no products were seen.

Protein crystallography

Commercial crystals screens (Crystal Screens 1 and 2, and Additive screens, Hampton Research, Aliso Viejo, CA, USA) and the crystal growth conditions for the previously determined crystal structure of WT-SsβG (without heptahistidine tag) were used as starting

points for crystallization screening. Crystallization screens were made by combining stock solutions of salts (1 M), buffers (1 M) and precipitant (50% w/v). All buffers were filtered before use and stored in the dark.

Protein solution (2 μL) and reservoir buffer (2 μL) were mixed as a drop on a cover slip which was sealed with vacuum grease over wells containing reservoir solution (200 μL). Wells were checked daily for the first week and every 3 d thereafter, using a microscope connected to a digital camera.

To determine whether crystals were protein or salt, trace amounts of Izit dye (Hampton Research) were added to crystal drops or crystals and were probed with an acupuncture needle.

SSβG-E387Y rod-shaped crystals were produced after 2 weeks using sodium acetate buffer (0.1 M, pH 3.25), ammonium acetate (0.2 M) and 20% (w/v) PEG 4000 and 10 mg/mL E387Y SsβG in Tris (10 mM, pH 7.5).

Substrate soaking was carried out by transferring crystals to substrate solution (40 or 10 mM in reservoir buffer), adding tiny amounts of solid substrate to the crystal drop, or by co-crystallization with substrate solutions made up in reservoir buffer.

Crystals were cryo-cooled by plunging into liquid nitrogen, and X-ray data were collected at 100 K using a nitrogen stream. Cryo-protection was accomplished by transferring crystals to 30% glycerol, 30% (w/v) glucose or 30% (w/v) gentiobiose in reservoir buffer, prior to flash freezing. The “in-house” machine refers to a MAR Research image plate detector (345 mm) mounted on a Rigaku RU200 rotating anode generator operating at 3.9 kW with Cu $K_{\square}$ radiation and equipped with Osmic mirrors, in the Laboratory of Molecular Biophysics, University of Oxford. Data were collected at 100 K on beamlines 9.6 and 10.1 at the

Synchrotron Radiation Source Daresbury, Warrington using ADSC Quantum-4 CCD and MarCCD 165 detectors, and at EMBL beamline X11 at the DORIS storage ring, Hamburg using a MarCCD 165 detector.

The structures were solved by molecular replacement with MOLREP[57] using the SsβG-WT crystal structure (pdb: 1gow)[25] as a search model. Molecular replacement was followed by rigid body refinement in REFMAC5,[58] and the sequence was corrected by manual building in XtalView.[59] Structures were refined using REFMAC5, including TLS refinement. 5% of the reflections were excluded for calculation of R_{free}. Strong non-crystallographic symmetry restraints were imposed for the 4 or 8 molecules within the asymmetric unit. Water molecules were added automatically using ARP/wARP[60] and checked manually. The models were validated by PROCHECK[61] and WHAT_CHECK.[62] No substrate or inhibitor molecules were observed in the electron density maps for these structures. Data collection parameters and refinement statistics are given in **Supplementary Table 6**.

Three of the highest resolution datasets, collected at a wavelength of 0.87 Å at SRS Daresbury beamline 9.6, in space group $P2_1\ 2_1\ 2_1$ were reprocessed using XIA2[63] to generate an *apo* E387Y dataset with a resolution of 2.16 A. The apo E387Y structure was solved by molecular replacement with Phaser[64] using the same search model as above. This model was optimized by iterative cycles of manual building using the graphics program Coot[65] and refinement as implemented in phenix.refine.[66] MolProbity[67] was used to validate the final model; the all-atom clashscore is 0.84 and the percentage of residues in the favoured regions of the Ramachandran plot is 97.3%. The crystallographic data have been deposited in the Protein Data Bank as entry 5I3D.

Structural alignment

Structural comparisons and superimpositions were made using the program Coot[65] using the default parameters, and figures were prepared using Pymol.[68] First of all, a ligand was modeled into the active site of SsβG-E387Y by structural alignment of the protein with that of the SsβG-WT[39] complexed with D-galactohydroximolactam (pdb: 1uwt). Although this modelled structure differed in the absolute location of the sugar atoms in the crystal structure, the space occupied by the substrate was identical. Therefore, the E387Y SsβG active site still has sufficient space and binding interactions to interact with a sugar substrate in exactly the same manner as WT SsβG. This is consistent with the very similar K_M values for pNPβGal and pNPβGlc substrates for WT and SsβG-E387Y (**Supplementary Table 2**). These results implied that it might be informative to overlay the structure of SsβG-WT in complex with d-galactoximolactam and SsβG-E387Y, as the substrate location in SsβG-E387Y was likely to be similar. The overlay of the D-galactohydroimolactam inhibitor in the SsβG-E387Y structure (**Supplementary Figure 10**) confirmed multiple potential contacts between the active site and substrate. Many residues are within 3 Å, permitting useful binding interactions and allowing most of the interactions observed in SsβG-WT•substrate complex to persist.

Statistical Methods

Kinetic parameters were analyzed through either nonlinear regression using the Michaelis-Menten equation or linear regression using the Lineaver-Burke equation. Apparent pKa values were analyzed through nonlinear regression using the pH-dependent k_{cat}/K_M equation. All data fitting were carried out using GraFit 7.0 (Erithacus Software). Data were typically collected

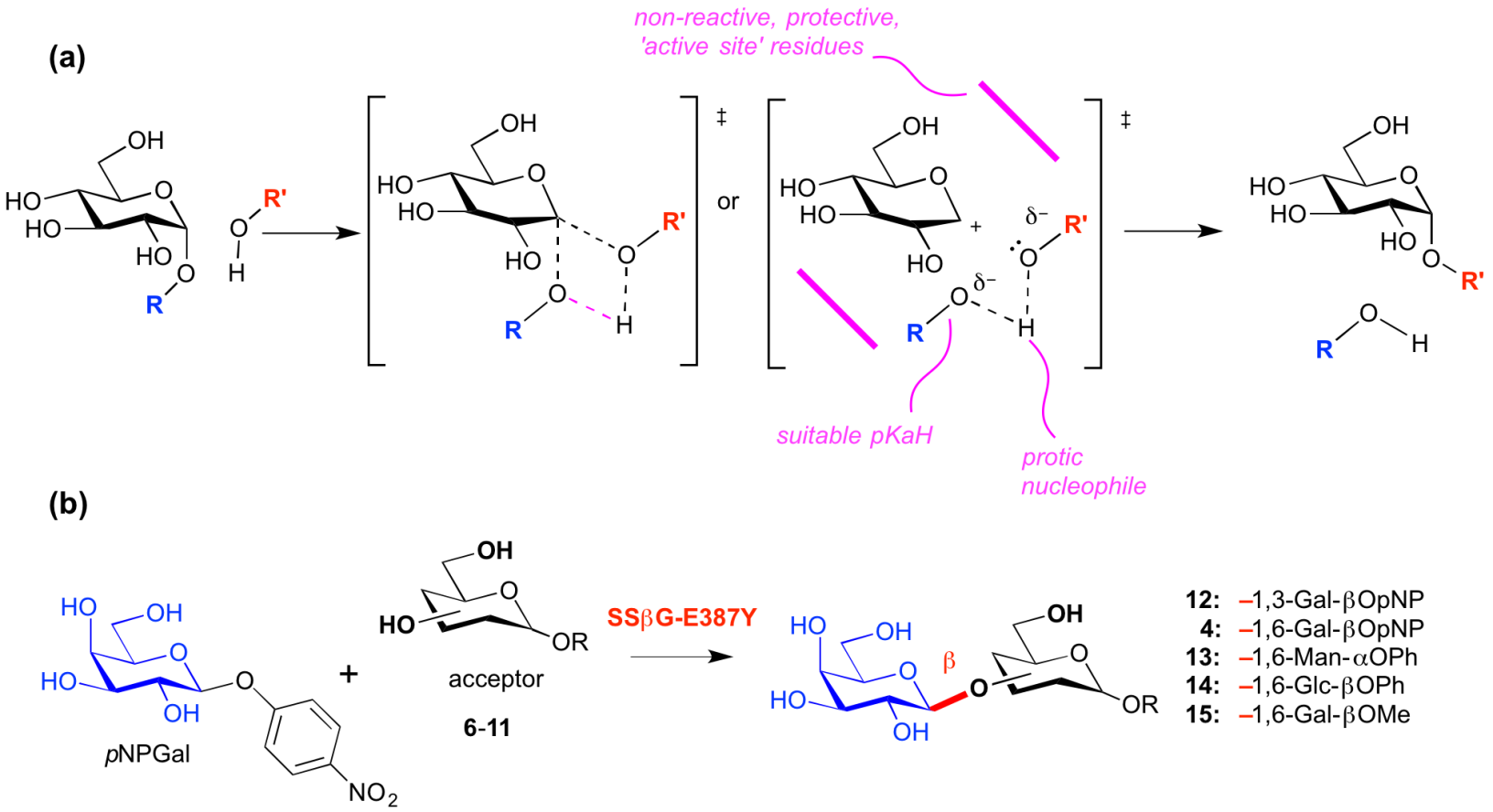
from two or three individual experiments, and all regressions generated standard errors of means (s.e.m.).

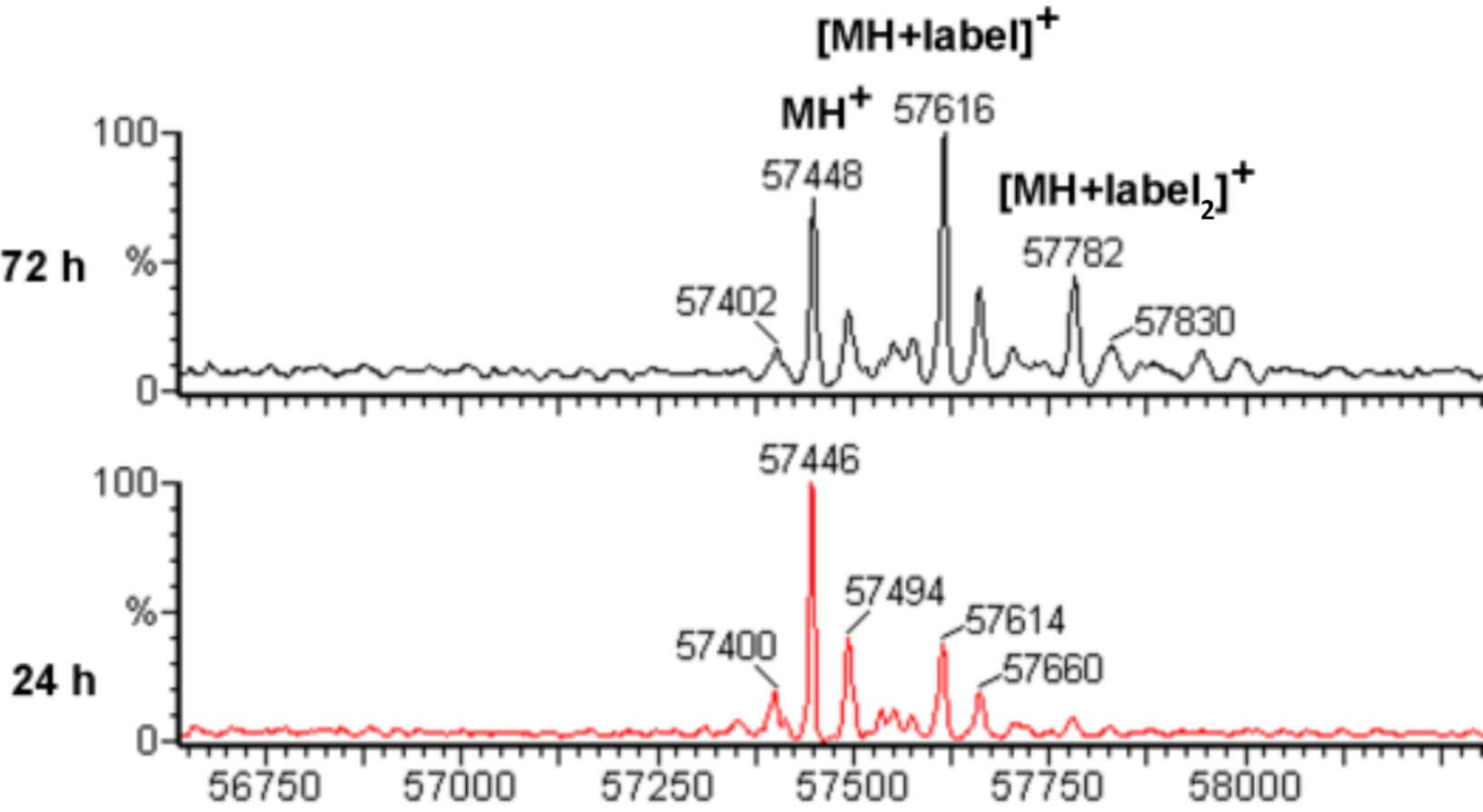
Online Databases

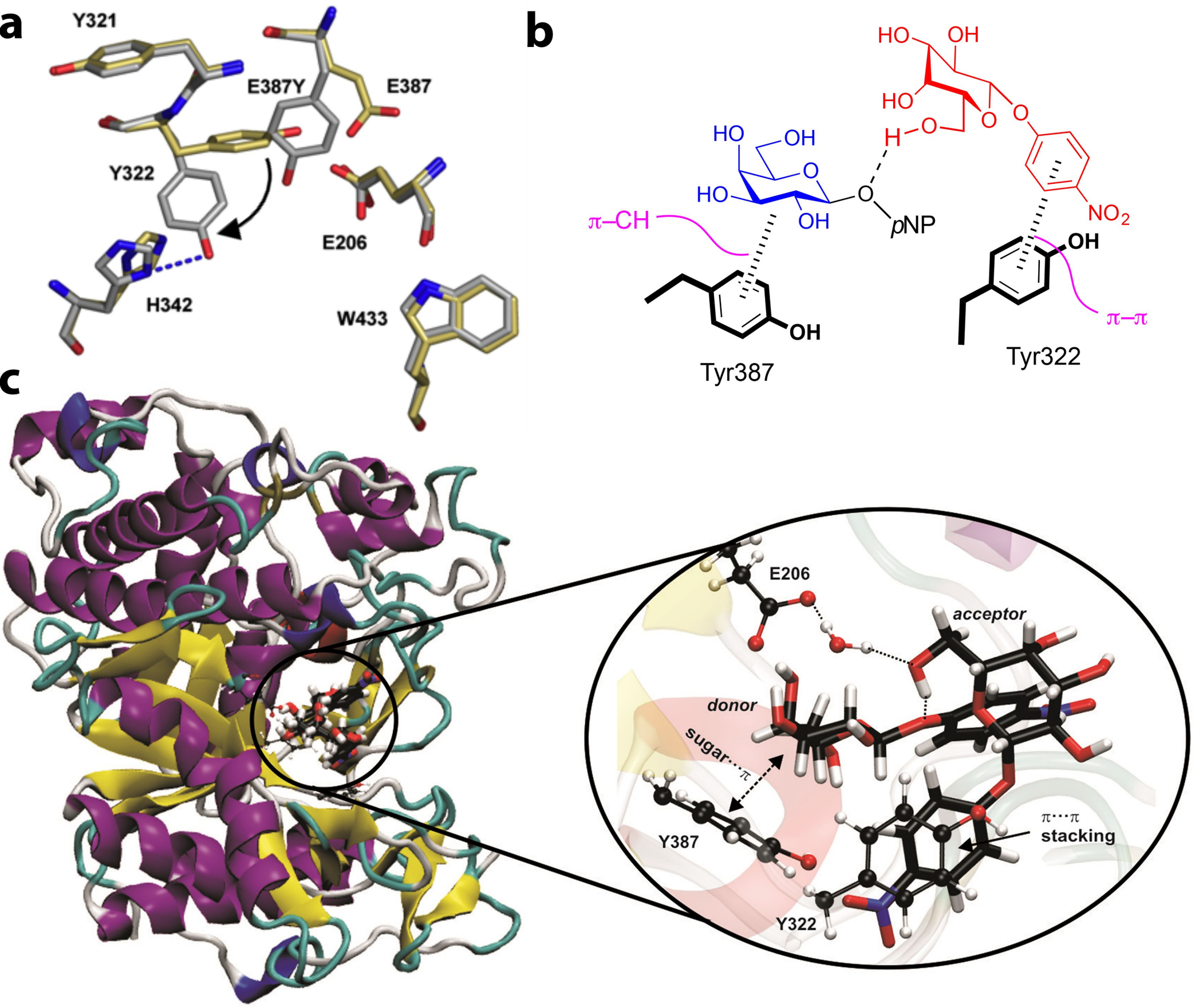
The crystallographic data have been deposited in the Protein Data Bank as entry 5I3D.

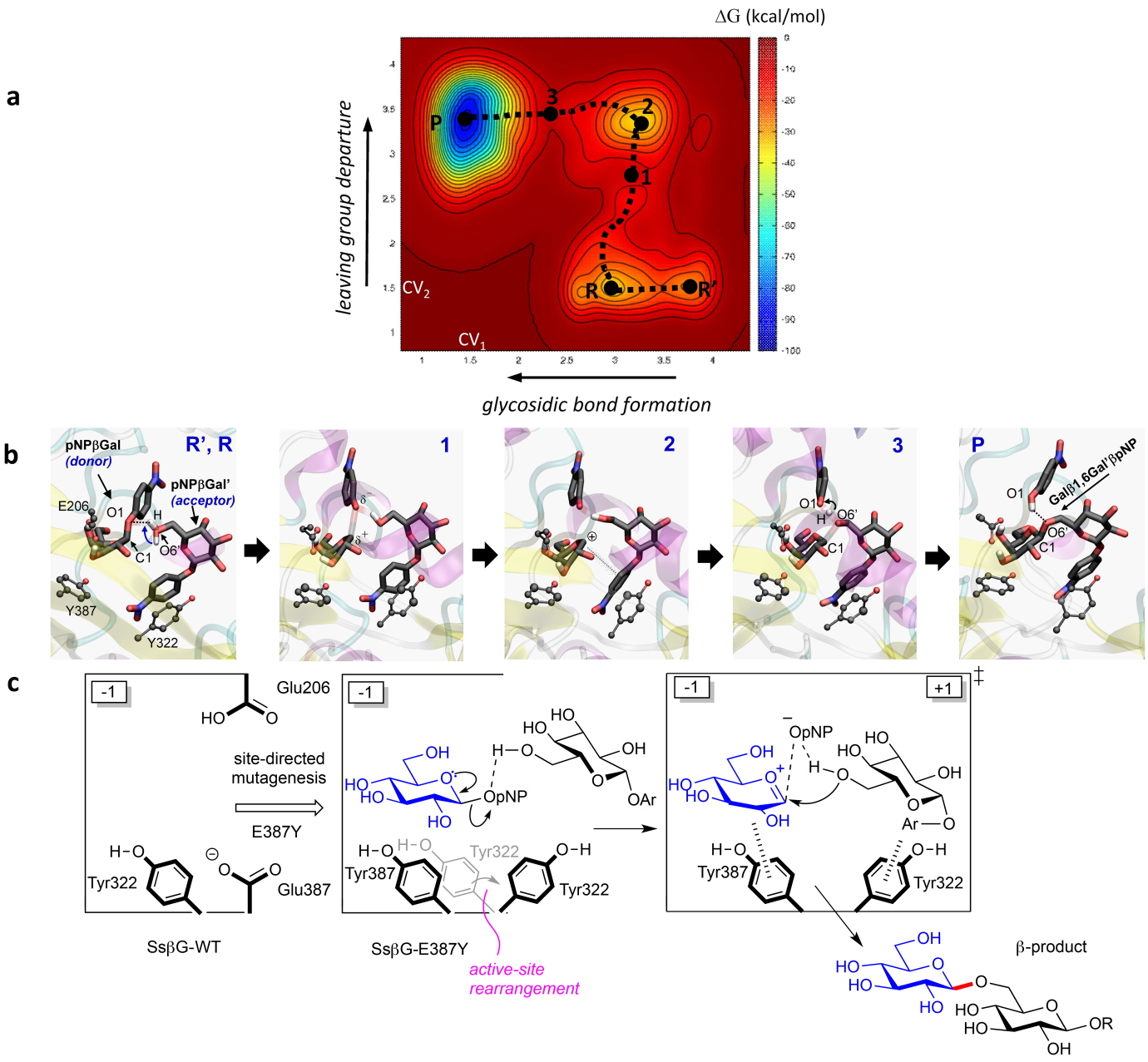
References for Online Methods

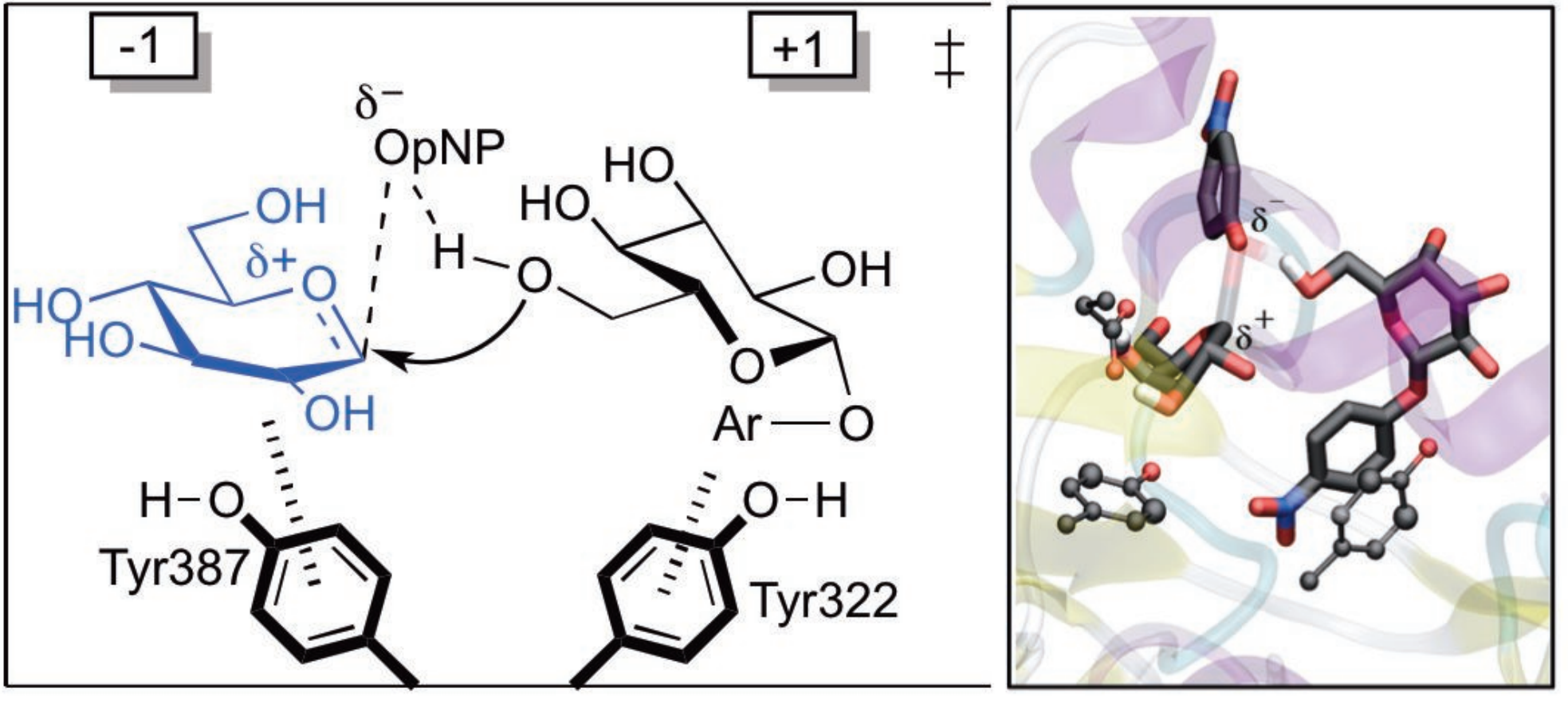
54. Bradford, M.M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding *Anal. Biochem.* **72**, 248-254 (1976).
55. Dixon, M. The determination of enzyme inhibitor constants. *Biochem. J.* **55**, 170-171 (1952).
56. Joshi, M.D.*, et al.* Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. *J. Mol. Biol.* **299**, 255-279 (2000).
57. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J Appl. Crystallogr.* **30**, 1022-1025 (1997).
58. Murshudov, G.N., Vagin, A.A. & Dobson, E.J. Refinement of macromolecular systems by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240-255 (1997).
59. McRee, D.E. XtalView/Xfit - a versatile program for manipulating atomic cooridnates and electron density. *J. Struct. Biol.* **125**, 156-165 (1999).
60. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein models. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **49**, 129-147 (1993).
61. Lakowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291 (1993).
62. Hooft, R.W.W., Vriend, G., Sander, C. & Abola, E.E. Errors in protein structures. *Nature* **381**, 272 (1996).
63. Evans, P. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography* **62**, 72-82 (2006).
64. McCoy, A.J.*, et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674 (2007).
65. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501 (2010).
66. Afonine, P.V.*, et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* **68**, 352-367 (2012).
67. Davis, I.W.*, et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research* **35**, W375-383 (2007).
68. DeLano, W.L. The Pymol Molecular Graphics System. (DeLano Scientific, San Carlos, 2002).


(a)
non-reactive, protective, 'active site' residues
suitable pKaH
protic nucleophile
or
(b)
pNPGal
acceptor
6-11
SSβG-E387Y
12: –1,3-Gal-βOpNP
4: –1,6-Gal-βOpNP
13: –1,6-Man-αOPh
14: –1,6-Glc-βOPh
15: –1,6-Gal-βOMe


[MH+label]+
MH+
57616
57448
[MH+label2]+
57782
57402
57830
72 h
100
%
0
57446
57494
57614
57400
57660
24 h
100
%
0
56750
57000
57250
57500
57750
58000


a
Y321
E387Y
E387
Y322
E206
H342
W433
b
HO
OH
HO
O
H
O
O
HO
OH
HO
O
OH
O
pNP
π–CH
NO2
OH
π–π
OH
Tyr387
Tyr322
c
E206
acceptor
donor
sugar···π
Y387
π···π
stacking
Y322


a
ΔG (kcal/mol)
leaving group departure
glycosidic bond formation
CV1
CV2
P
R
R'
1
2
3
b
pNPβGal
(donor)
pNPβGal'
(acceptor)
R', R
E206
O1
H
C1
O6'
Y387
Y322
1
2
3
P
Galβ1,6Gal'βpNP
c
-1
+1
Glu206
site-directed
mutagenesis
E387Y
Tyr322
Glu387
SsβG-WT
OpNP
OAr
Tyr387
SsβG-E387Y
active-site
rearrangement
β-product


-1
+1
‡
δ−
OpNP
OH
δ+
O
HO
HO
OH
H
O
HO
HO
OH
O
Ar — O
H−O
Tyr387
O−H
Tyr322
δ−
δ+