

A front-face 'S_Ni synthase' engineered from a retaining 'double-S_N2' hydrolase

Javier Iglesias-Fernández^{1,2,8,10}, Susan M Hancock^{3,10}, Seung Seo Lee^{3,4,10}, Maola Khan³, Jo Kirkpatrick^{3,8}, Neil J Oldham^{3,8}, Katherine McAuley⁵, Anthony Fordham-Skelton^{6,9}, Carme Rovira^{1,2,7*} & Benjamin G Davis^{3*}

S_Ni-like mechanisms, which involve 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-like, S_N1 and S_N2 substitution pathways can be energetically comparable, engineered switching could be feasible. Here, engineering of *Sulfolobus solfataricus* β -glycosidase, which originally catalyzed double S_N2 substitution, changed its mode to S_Ni-like. Destruction of the first S_N2 nucleophile through E387Y mutation created a β -stereoselective catalyst for glycoside synthesis from activated substrates, despite lacking a nucleophile. The 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 that is similar to those of natural, S_Ni-like glycosyltransferases. This observation of a front-face mechanism in a β -glycosyltransferase enzyme highlights that S_Ni-like pathways may be engineered in catalysts with suitable environments and suggests that 'S_Ni' mechanisms may be feasible for natural glycosyltransfer enzymes.

Since the seminal demonstration of front-side (same face) nucleophilic attack in chemical, α -glycosyl transfer substitution¹, the possibility of a 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 mechanism⁵ that is found in glycoside hydrolases (GHs)⁴. Although typically observed chemical nucleophilic substitution involves likely intermediacy of solvent-exposed and accessible reaction centers, even for such reactions, S_Ni-like mechanisms are observed, facilitated by assisted delivery of the nucleophile to the electrophile^{6,7}. In proteins, more constrained environments (and possible alternative pathways) exist. Structures of several retaining GTs^{8–11} show the 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 simulations¹⁴. These investigations were followed by 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 mechanism⁷; phosphorylation of α -glucosyl fluoride in mutant phosphorylases is also suggested to follow a similar path¹⁵. Subsequent QM/MM studies on the retaining GTs lipopolysaccharyl α -galactosyltransferase C (LgtC)¹⁶, α -1,2-mannosyltransferase Kre2p/Mnt1p¹⁷, 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 GTs⁴. More recently, the functionally essential Notch-modifying xylosyltransferase was proposed to follow this S_Ni pathway¹¹. Together, these studies suggest that the unusual front-face mechanism may, in fact, play an important and potentially widespread role in nature, considering the importance and ubiquity of glycosyltransferases. Thus far, no β -selective front-face-retaining reaction has been observed. One apparently crucial feature of the α -selective mechanism suggested in these studies (Fig. 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 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 data 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 engineered¹⁵ rather than simply observed.

Here we demonstrate that the front-face reaction not only can operate in retaining GTs but also can be created in engineered GHs through the exploitation of such features. Selection of a suitable, robust GH scaffold created an enzyme with 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 activity arises from front-face nucleophilic substitution, similar to that proposed for retaining

¹Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica), Universitat de Barcelona, Barcelona, Spain. ²Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Barcelona, Spain. ³Department of Chemistry, University of Oxford, Oxford, UK. ⁴School of Chemistry, University of Southampton, Southampton, UK. ⁵Diamond Light Source, Didcot, UK. ⁶CLRC, Daresbury Laboratory, Warrington, UK. ⁷Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. ⁸Present addresses: Institut de Química Computacional i Catalisi and Departament de Química, Universitat de Girona, Girona, Spain (J.I.-F.), School of Chemistry, University of Nottingham, Nottingham, UK (N.J.O.) and Leibniz Institute on Aging-Fritz Lipmann Institute (FLI), Jena, Germany (J.K.). ⁹Deceased. ¹⁰These authors contributed equally to this work. *e-mail: c.rovira@ub.edu or ben.davis@chem.ox.ac.uk

GTs. To the best of our knowledge, this is the first description of a front-face mechanism for 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* (SSG) shows stability to mutation^{21,22}, solvents²³ and even typically denaturing conditions^{24,25}. Prior nucleophile-free mutants bearing smaller residues than the natural Glu387 (for example, Gly387; ref. 26) act as classical, inverting glycosynthases²⁷ with suitable (α -glycosyl fluoride) substrates²⁶. By contrast, our initial modeling suggested that to ensure sufficient protection and putative stabilizing interactions and yet be small enough to be accommodated, only certain residues (for example, Tyr and Phe) would prove suitable. Tyr387 was therefore chosen, and site-directed mutagenesis of SSG-WT yielded stable, folded, soluble protein S β SG-E387Y, C-terminally His-tagged to allow for exhaustive nickel-affinity chromatography (**Supplementary Results, Supplementary Fig. 1**), giving good protein yields of ~28 mg/L of growth. N-terminal sequencing, LC-ESI-MS (found 57,450 Da; expected 57,447 Da) (**Supplementary Table 1**) and circular dichroism (CD) analysis (**Supplementary Fig. 2**) confirmed identity and unaffected secondary structure, respectively.

Glu387Tyr nucleophile-mutant displays altered activity

By design, we chose *para*-nitrophenoxide ($pK_{aH} \sim 7$), which has a similar pK_a to those of UDP ($pK_{aH1} \sim 7$ and $pK_{aH2} \sim 9$), as a suitable leaving group for our putative 'activated' substrates. Determination of the hydrolytic kinetic parameters (**Supplementary Table 2**) of S β SG-E387Y toward *p*-nitrophenyl β -D-glycosides and comparison with S β SG-WT revealed reduced but clear activity toward *p*NP β Gal (**1**) and *p*NP β Glc (**2**) substrates. Consistent with the loss of the nucleophilic Glu387 residue in S β SG-WT, 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 S β SG-WT to 3:1 in S β SG-E387Y—a ratio that may more closely reflected the inherent chemical reactivity of Gal as compared to Glc²⁸. Interestingly, tyrosine residues are observed in positions similar to that of Tyr387 in glycosidase enzymes that exploit substrate-assisted catalysis, such as the hexosaminidases²⁹. These are thought to stabilize the formation of corresponding oxazolium ion intermediates. However, S β SG-E387Y displayed no hexosaminidase activity toward either *p*NP β GlcNAc (**3**) or corresponding activated oxazoline substrates (2-methyl-(1,2-dideoxy- α -D-glucopyranosyl) [2,1-d]- Δ^2 -oxazoline, **4**). Consistent with the designed requirement for a suitable activated leaving group, S β SG-E387Y failed to hydrolyze either methyl β -D-galactopyranoside (MeGal, **5**) or *p*-nitrophenyl 6-O-(β -D-galactopyranosyl)- β -D-galactopyranoside (*p*NPGal1,6Gal, **6**).

Incubation with mechanism-based inhibitor³⁰ 2,4-dinitrophenyl 2-deoxy-2-fluoro- β -D-glucopyranoside (**7**) had no substantial effect, thereby suggesting that the observed activity did not arise from S β SG-WT (wild type) or other (for example, endogenously expressed host *Escherichia coli*) glycosidases that use nucleophilic catalysis. It also intriguingly suggested that this altered catalytic activity of S β SG-E387Y was no longer nucleophile dependent (as discussed below). When S β SG-E387Y was thermally denatured (16–20 h at 45 °C) all activity was lost, implying that native protein conformation was required for catalytic activity.

S β SG-E387Y is a 'synthase'

Given this striking selectivity for activated substrates, with negligible activity in the hydrolysis of unactivated glycosides (and hence potential products), S β SG-E387Y was a potentially useful catalyst for glycosidic bond formation from activated

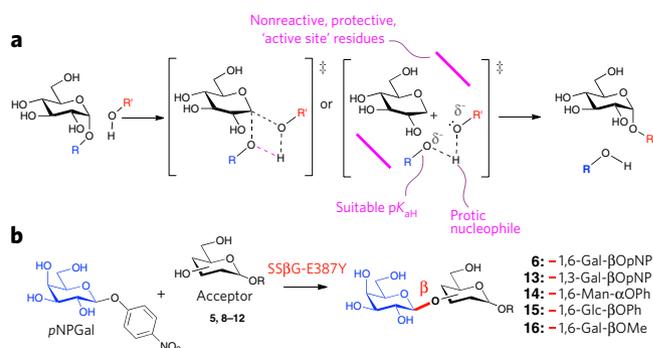


Figure 1 | Comparisons of front-face glycosyl transfer. (a) The front-face reaction mechanism of known α -selective retaining glycosyltransferases. (b) The transglycosylation reactions catalyzed by the β -selective 'front-face' synthase described here. Reactions are detailed in **Table 1**.

*p*NP substrates. We surveyed a small set of representative monosaccharides as nucleophilic acceptors under different conditions (**Fig. 1b** and **Table 1**).

S β SG-E387Y did not process nonaromatic sugar acceptors to any substantial extent, resulting in reactions that instead primarily gave Gal β -*p*NPGal disaccharidic products **6** and **13** (**Table 1**), suggesting a strong preference for using *p*NPGal **1** as an acceptor. This observed preference for aromatic sugar acceptors was consistent with aromatic stacking interactions in the +1 or +2 acceptor pockets used by the GH naturally for binding oligosaccharide substrates^{31,32}. Indeed, aromatic Gal β , Glc β and Man α glycosides all proved to be suitable nucleophile substrates (**Table 1**). Unlike several other synthases, under these conditions, trisaccharides and higher or branched oligosaccharides (from uncontrolled 'self-condensation') were not observed in measurable amounts; these are isolated in reactions catalyzed by classical glycosynthases, notably including a variant derived from S β SG²⁶. Only under more extreme conditions were small amounts of trisaccharides observed (see below). In all reactions, either exclusive 1,6- or 1,6-/1,3-linked regioselectivity was observed³³; in contrast to the behavior of other S β SG-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 Ph β Glc **11** (**Supplementary Table 3**). Strikingly, variable conditions allowed the improvement of the synthesis:hydrolysis (S:H) ratio to over 99:1. Under these conditions, the enzyme was both selective and, essentially, exclusively synthetic, yielding **15** as the predominant product in >70% isolable yield, with only the formation of smaller amounts of trisaccharides as side products (**Table 1**). In control experiments, under essentially identical conditions, S β SG-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: S β SG-E387Y did not process donor substrates with α -anomeric configuration, thereby confirming that S β SG-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 S β SG-E387Y (several > 80%) were high, and only rivaled by some of the more potent glycosynthases³⁶. It should be noted, however, that estimated transglycosylation rates ($k_{cat}/K_M \sim 0.0052 - 0.025 \text{ min}^{-1}\text{mM}^{-1}$) were ~2,000-fold lower compared to those of classical glycosynthases (see below for further details).

Table 1 | SsβG-E387Y catalyzes transglycosylation

Acceptor	Temp / °C	Product and yield / % ^a							Total	S/H	Conversion ^d / %
		13	6	14	15	16	H ^b	S ^c			
MeβGal 5	45	18	24	-	-	2	37	44	81	1.2	92
MeβGal 5	80	51	36	-	-	1	<1	88	88	>88	78
Cellobiose 8	45	14	15	-	-	-	44	29	73	0.7	100
Cellobiose 8	80	22	27	-	-	-	6	49	55	8.2	79
Lactose 9	45	21	29	-	-	-	33	50	75	2.0	80
Lactose 9	80	30	54	-	-	-	16	84	100	5.3	91
MeβMan 10	45	16	38	-	-	-	46	54	100	1.2	100
MeβMan 10	80	39	46	-	-	-	15	85	100	5.7	92
PhβGlc 11	45	9	46	-	26	-	17	81	98	4.8	97
PhβGlc 11	80	0	28	-	12	-	37	-	-	-	100
PhαMan 12	45	0	3	12	-	-	85	15	100	0.2	100
PhαMan 12	80	1	10	25	-	-	64	36	100	0.6	100
PhβGlc ^e 11	45	5	0	-	72	-	<1	>99	100	>100	- ^e

^aYields 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 -15 h or longer. ^bH, total yield of hydrolysis products. ^cS, total yield of glycosides/synthesis products. ^dBased on the consumption of starting material. ^eAfter optimization for yield, including additional production of trisaccharide as mass balance (see **Supplementary Table 3**).

Disaccharides synthesized from pNPGal **1** as a glycosyl donor (see **Fig. 1b** for relevant reaction).

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 transglycosidases. The *trans*-sialidase from *Trypanosoma cruzi* of GH family 33 utilizes a tyrosine residue as a nucleophile³⁷, and, although modeling and design (as discussed above) 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 inactivators³⁰. Thus, SsβG-E387Y was incubated with DNP-2FGlc³⁰ **7** and analyzed by LC-MS (**Fig. 2** and **Supplementary Fig. 3**). Over 6 h, no change in the hydrolytic activity of SsβG-E387Y was observed. Concomitant monitoring of DNP release (absorbance at 405 nm) revealed no acceleration compared to uncatalyzed chemical DNP-2FGlc hydrolysis. *Agrobacterium faecalis* β-glucosidase can form a stable α-D-glucopyranosyl tyrosine product at nonrelevant Y298 upon mutation of the active site nucleophile³⁸; peptide 'mapping' of SsβG-E387Y did not show trapping of Tyr387. Neither proteolytic (trypsin, pepsin, thermolysin or clostripain) MS/MS nor CNBr-cleavage-MS/MS (including neutral loss analysis of the 2FGlc moiety) indicated peptides with attached 2FGlc moieties (**Supplementary Figs. 4–6**), even though the coverage of this 'mapping' successfully included peptides containing Tyr387 (and Glu206) as putative trapping sites. In control experiments, under essentially similar conditions, SsβG-WT was successfully labeled (**Supplementary Figs. 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 nonspecifically at low abundance over multiple nonspecific locations that could not be detected by proteolytic cleavage-MS/MS 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 Fig. 2**).

None of these mutations caused a dramatic loss of function; indeed, the similar activities of SsβG-E387Y, SsβG-E387F, SsβG-E206A-E387Y and SsβG-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 (see above), the additional mutation of the acid/base residue (Glu206) along with that of the 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 the design (**Fig. 1a**). This was also in agreement with the observation that the basic limb

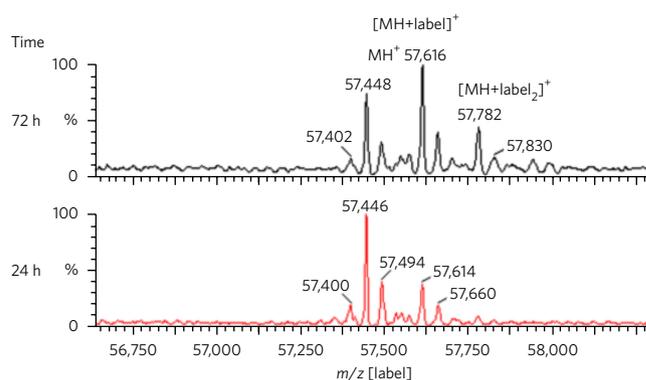


Figure 2 | Mass spectrometric analysis of incubation of SsβG-E387Y with covalent inhibitor DNP-2FGlc. The reaction with 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (**7**) was monitored over time by ESI-MS. A slow reaction and the emergence of additional peaks ($2 \times +165 \pm 3$ Da, etc.) after extended incubation and with an apparent statistical distribution suggest nonspecific chemical modification; incubation with 2FGlc did not cause direct glycation (**Supplementary Fig. 3**). This nonspecific, non-'activity-based' cause is also consistent with the thermal denaturation of SsβG-E387Y at 45 °C for 16 h (as discussed above). Data are from a single experiment representative of five repeated 'trapping' experiments.

of the pH profile of SsβG-E387Y was also shifted ~0.6 pK_a units to a value similar to that for *para*-nitrophenol (Supplementary Fig. 10).

Finally, transglycosylation kinetics were determined for SsβG-E387Y using a range of substrates (Supplementary Table 5 and Supplementary Fig. 11). Notably, both activity (as judged by $k_{\text{cat}}/K_{\text{M}}$) and regioselectivity (1,6 versus 1,3; Supplementary Fig. 11b) varied with leaving group; tentative linear free-energy analysis (Supplementary Fig. 12) revealed a small β-value (-0.049), consistent with computational analysis results suggesting a stepwise mechanism with a higher barrier for the collapse of an oxocarbenium-ion intermediate than that for leaving group departure (see below).

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 (Fig. 3a; Supplementary Fig. 13; Online Methods; Supplementary Table 6) and compared to the previously reported SsβG-WT structure²⁵. Despite the mutation, the structures superimposed with very little divergence (r.m.s. deviation of 0.26 Å calculated using 486 Cα positions). In the active site, only two amino acids shifted substantially as a result of the mutation: Tyr322 and His342 (Fig. 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-WT³⁹ complexed with D-galactohydroximolactam (PDB ID: 1UWT) (Supplementary Fig. 14). The SsβG-E387Y active site was very similar to that of SsβG-WT (Fig. 3a), consistent with the similar K_{M} values obtained for *p*NPβGal and *p*NPβGlc substrates for SsβG-WT and SsβG-E387Y SsβG (Supplementary Table 2).

A combination of classical molecular dynamics (MD) and metadynamics techniques were used to model a ternary Michaelis complex of SsβG-E387Y with two molecules of *p*NPβGal as putative acceptor and donor substrates corresponding to one of the observed synthase activities (discussed above). In the first step, the two molecules were manually placed at the entrance of the enzyme catalytic groove (see Online Methods; Supplementary Figs. 15 and 16). After 200 ns of MD simulation, one of the molecules partially entered the catalytic site, sitting at ~8 Å from the catalytic residues, whereas the other remained at the entrance (Supplementary Fig. 16a). Further MD simulation did not lead to substantial change, indicating that complete entrance of the two molecules is associated with a certain free-energy barrier. Therefore, the ligand-binding process was activated using an enhanced-sampling technique (metadynamics)⁴⁰. Two collective variables were chosen to drive the binding of the two *p*NPβGal molecules to the active site of SsβG-E387Y. The first (CV₁; Supplementary Fig. 15) measures the degree of penetration of the first *p*NPβGal molecule (as the donor) into the active site; the second (CV₂) accounts for the formation of a O1-H' interaction, providing a measure of distance between donor and acceptor.

The free-energy landscape (FEL) of ligand binding obtained from the classical metadynamics simulation (Supplementary Fig. 16c) showed an energy minimum (the global one) in which the two *p*NPβGal molecules were inside the enzyme active site (the ternary complex, shown in Fig. 3b,c). Analysis of the water content around the active site showed that a number of water molecules were displaced during binding (13 ± 4 from a region of ≤5 Å from Y387 and Y322). Among the remaining water molecules, two were located within 5 Å of the donor anomeric carbon. Although these water molecules were not well positioned for catalysis, they could account for the observed residual hydrolysis. Close examination of the orientation of the two sugar 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*NP group (i.e., the leaving group). This was an optimum topology for a front-face mechanism, which could ultimately lead to a transglycosylation product

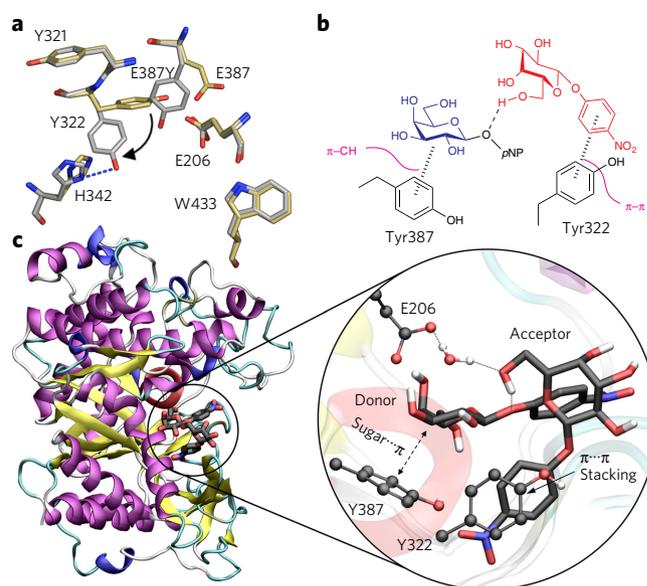


Figure 3 | Structural analysis of SsβG-E387Y. (a) The X-ray structure of apo-SsβG-E387Y (determined in this work; PDB ID: 5I3D, silver) superimposed on SsβG-WT (PDB ID: 1GOW, gold) shows the highly localized rearrangement (indicated by curled black arrow) of residues Y322 and H342 to accommodate the 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 interactions⁴¹ (sugar hydrogen atoms point toward 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 *p*NPβGal 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 ID: 5I3D, silver) shown in a. The inset shows an expanded view of the active site.

with net retention of configuration. The hydrogen atom of acceptor OH-6 was directed toward O-1 of the donor molecule, favoring the formation of a 1,6-glycosidic linkage, consistent with the observed regiochemical preferences of SsβG-E387Y. This hydrogen bonding interaction may guide the nucleophile to the same face as the leaving group, akin to interactions observed in retaining 'S_Ni-like' GTs^{14,16}. Furthermore, this result 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 (Fig. 1a). It was also consistent with the removal of the general acid/base residue (Glu206) in SsβG-E387Y-E206A having a nondetrimental effect on activity; in SsβG-E387Y with *p*NPGal, the phenolic base appeared to be sufficient to deprotonate the incoming hydroxyl nucleophile.

There were crucial substrate-protein interactions (Fig. 3b) that contributed to the stability of the above 'front-face arrangement'. First of all, Tyr387 formed stabilizing donor sugar-π interactions⁴¹ (sugar hydrogen atoms pointed toward the center of the Y387 phenol ring, with distances <3 Å; Fig. 3c), consistent with the overlay of the starting apo SsβG-E387Y X-ray crystal structure with the SsβG-WT-inhibitor complex (Supplementary Fig. 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

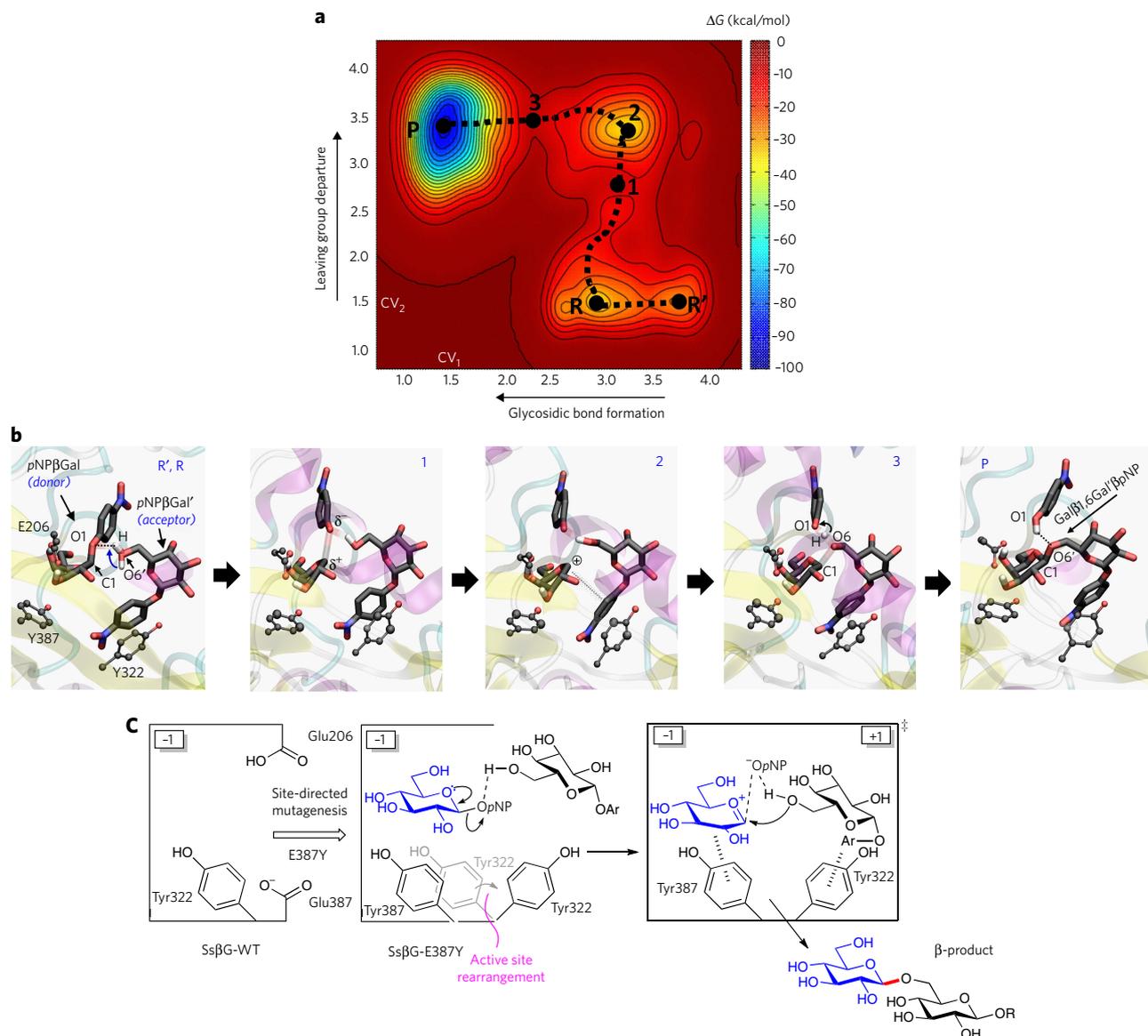


Figure 4 | Analysis of the S_Ni reaction pathway. (a) Free-energy landscape (FEL) reconstructed from the metadynamics simulation of the transglycosylation reaction (projection on two collective variables CV_1 and CV_2). Contour lines are at 5 kcal/mol. The second transition state (labeled as state 3) is higher in energy with respect to the first one (labeled as state 1) by 3 kcal/mol. (b) Representative active site structures for the most relevant states along the reaction pathway. Hydrogen atoms have been omitted for clarity, except the one being transferred from the sugar acceptor to the pNP leaving group of the donor molecule and the hydroxyl hydrogen atoms of the Gal donor that interact with E206. Bonds being broken/formed are represented by a transparent bond (in states 1 and 3), whereas dotted lines indicate hydrogen bonding interactions. (c) Proposed front-face substitution mechanism of SsβG-E387Y.

attack the anomeric carbon of the sugar donor. These π - π stacking interactions explained why pNPβGal and other aromatic glycosides were preferred substrates for the synthase activity of SsβG-E387Y (as discussed above). Essentially identical analysis of a possible O3-regioselective pathway also generated an appropriate Michaelis complex (Supplementary Fig. 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 reaction, 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 in atomic detail and to obtain the FEL from which, in turn, reaction coordinates were defined. From the ternary complex determined above (Fig. 3b,c), three collective variables, corresponding to the main bonds undergoing breaking or formation, were used (Supplementary Fig. 18 and Online Methods). 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 self-selected any specific reaction pathway. The FEL for the transglycosylation reaction, reconstructed from the QM/MM metadynamics simulation

(Fig. 4a and Supplementary Video 1) 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 \cdot mol $^{-1}$, similar to the value obtained for the OtsA glycosyltransferase with essentially similar computational methodology¹⁴.

The structure of the reactants complex (R in Fig. 4b) was very similar to the one obtained from classical (i.e., force-field based) metadynamics simulation (Fig. 3c), except that the donor galactosyl ring was distorted into a 1S_3 conformation in the QM/MM structure as opposed to a relaxed 4C_1 . This was not surprising in view of the known limitations of force fields in describing 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, which has previously been observed on the basis of QM/MM calculations for GTs^{14,16,18,19} (the hydrogen bond forms either in the reactant complex or in the early stages of the reaction), is known as a common feature of enzymes operating via a front-face mechanism and is part of the design invoked for Ss β G-E387Y (Fig. 1 and above).

The reaction pathway (Fig. 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 state 1; Supplementary Fig. 19 and Supplementary Table 7). This bond was completely broken at intermediate state 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 center atom was further supported by a shift toward trigonal geometry, which was also associated with changes in the conformation of the pyranose ring along the reaction (see Supplementary Fig. 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 center (by 0.30 e $^-$ when going from R to state 2).

The oxocarbenium ion pair corresponded to a minimum along the reaction pathway. It was stabilized by the O6'–H–O1 hydrogen bond (state 2 in Fig. 4b), which also played a role in orienting the acceptor for nucleophilic attack. Afterward, a slight displacement of the hydroxymethyl moiety coupled with proton transfer (from hydroxymethyl to *p*NP-oxygen) formed the new glycosidic bond (state 3 \rightarrow P in Fig. 4a). Notably, the observation of a slightly higher barrier of ~ 3 kcal/mol for collapse of the oxocarbenium ion was consistent not only with prior observations in GTs^{14,18} but also with the low Brønsted coefficient (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 state 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 lifetime >15 ps. This again indicated that the ion-pair species was a minimum of the FEL. 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 to 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 residue⁴⁶ was also considered and tested (Online Methods and Supplementary Fig. 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 -*p*NP bond and formation of the Gal β 1–6Gal bond were entirely asynchronous and followed a front-side stepwise mechanism.

The donor conformational itinerary observed in Ss β G-E387Y during transglycosylation (Fig. 4b) was as follows: 1S_3 (reactants) \rightarrow ${}^4H_3/E_3$ (reaction intermediate) \rightarrow 4C_1 (products). This pathway was the same as that delineated experimentally^{44,45,47} and theoretically⁴⁸ for retaining β -D-gluco-active glycoside hydrolases such as Ss β G-WT. Remarkably, therefore, despite the very different mechanism, the engineered $S_{\text{N}}\text{i}$ synthase Ss β G-E387Y appears to synthesize 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, independently 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 $S_{\text{N}}\text{i}$ -like mechanisms are only implied in retaining α -glycosyltransferases; the engineered system we present here appears to constitute an example of a retaining glycosyltransferase-like enzyme with β -glycosidic bond selectivity. Structural and computational analyses suggest a critical role for the installed Tyr387 through sugar- π and π - π interactions in recruiting to the Michaelis complex (Fig. 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 \rightarrow Phe mutation in Ss β G-E387F did not affect activity, 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- to 5-fold decrease in k_{cat} , with minimal effect on K_{M} ; this suggests that a polarizable π -aromatic ring system might have the capacity for transition state stabilization⁴⁹. FEL 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 of ~ 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 Ss β G-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 the donor) that was devoid of any catalytic residue but that nonetheless provided stabilization and specified that reactants can only form β -products. This reactivity and selectivity was provided, at least in part, by the box's favoring of particular conformers along the corresponding itinerary (Fig. 4b). Such a box is highly reminiscent of the catalytic activity proposed for serine protease mutants that lack their entire catalytic triad but nonetheless show rate accelerations of $\sim 10^3$ -fold over background⁵⁰. Notably, the box provided by catalytic antibodies acting as glycosidases⁵¹ that also lack participating residues are similarly highly hydrophobic and, indeed, less efficient (rate accelerations of $\sim 10^3$ -fold over background; $k_{\text{cat}} = 0.007$ min $^{-1}$ and $K_{\text{M}} = 0.53$ mM) than the designed $S_{\text{N}}\text{i}$ synthase that we have created here (rate accelerations of $\sim 10^5$ -fold over background; $k_{\text{cat}} = 0.48$ min $^{-1}$ and $K_{\text{M}} = 0.17$ mM). It should be noted that our $S_{\text{N}}\text{i}$ synthase was, in turn, a similar magnitude less active than prior $S_{\text{N}}2$ synthases. Further future activity optimization might be

considered through forced evolution strategies, for example. Given the previously suggested⁵² conceptual kinship of some glycosyl units and terpenes, it is interesting to note that our initial inspection of known structures of terpene cyclase structures suggests prominently placed aromatic side chains, akin to the role that is suggested for Y387 here. Altogether, although there may be other causes for the activities that we observed here, our combined results suggest that the once seemingly improbable and rare same-face nucleophilic substitution is a more broadly viable mechanistic possibility and can be considered an accessible mechanism in the design of catalysts for substitution⁵³.

Received 22 April 2016; accepted 8 March 2017;
published online 12 June 2017

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

References

- Sinnott, M.L. & Jencks, W.P. Solvolysis of D-Glucopyranosyl derivatives in mixtures of ethanol and 2,2,2-trifluoroethanol. *J. Am. Chem. Soc.* **102**, 2026–2032 (1980).
- Persson, K. *et al.* Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat. Struct. Biol.* **8**, 166–175 (2001).
- 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. *Chem. Biol.* **9**, 1337–1346 (2002).
- Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
- Koshland, D.E. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Camb. Philos. Soc.* **28**, 416–436 (1953).
- Lewis, E.S. & Boozer, C.E. The kinetics and stereochemistry of the decomposition of secondary alkyl chlorosulfites¹. *J. Am. Chem. Soc.* **74**, 308–311 (1952).
- Chan, J., Tang, A. & Bennet, A.J. A stepwise solvent-promoted S_Ni reaction of α-D-glucopyranosyl fluoride: mechanistic implications for retaining glycosyltransferases. *J. Am. Chem. Soc.* **134**, 1212–1220 (2012).
- Vetting, M.W., Frantom, P.A. & Blanchard, J.S. Structural and enzymatic analysis of MshA from *Corynebacterium glutamicum*: substrate-assisted catalysis. *J. Biol. Chem.* **283**, 15834–15844 (2008).
- Batt, S.M. *et al.* Acceptor substrate discrimination in phosphatidylmyo-inositol mannoside synthesis: structural and mutational analysis of mannosyltransferase *Corynebacterium glutamicum* PimB'. *J. Biol. Chem.* **285**, 37741–37752 (2010).
- Chaikwad, A. *et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proc. Natl. Acad. Sci. USA* **108**, 21028–21033 (2011).
- Yu, H. *et al.* Notch-modifying xylosyltransferase structures support an S_Ni-like retaining mechanism. *Nat. Chem. Biol.* **11**, 847–854 (2015).
- Errey, J.C. *et al.* Mechanistic insight into enzymatic glycosyl transfer with retention of configuration through analysis of glycomimetic inhibitors. *Angew. Chem. Int. Ed. Engl.* **49**, 1234–1237 (2010).
- Lee, S.S. *et al.* Mechanistic evidence for a front-side, S_Ni-type reaction in a retaining glycosyltransferase. *Nat. Chem. Biol.* **7**, 631–638 (2011).
- Ardévol, 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 (2011).
- Goedl, C. & Nidetzky, B. Sucrose phosphorylase harbouring a redesigned, glycosyltransferase-like active site exhibits retaining glycosyl transfer in the absence of a covalent intermediate. *ChemBioChem* **10**, 2333–2337 (2009).
- Gómez, H., Polyak, I., Thiel, W., Lluç, J.M. & Masgrau, L. Retaining glycosyltransferase mechanism studied by QM/MM methods: lipopolysaccharyl-α-1,4-galactosyltransferase C transfers α-galactose via an oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **134**, 4743–4752 (2012).
- Bobovská, A., Tvaroška, I. & Kóňa, J. A theoretical study on the catalytic mechanism of the retaining α-1,2-mannosyltransferase Kre2p/Mnt1p: the impact of different metal ions on catalysis. *Org. Biomol. Chem.* **12**, 4201–4210 (2014).
- Lira-Navarrete, E. *et al.* Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase 2. *Angew. Chem. Int. Ed. Engl.* **53**, 8206–8210 (2014).
- Gómez, H. *et al.* A computational and experimental study of O-glycosylation. Catalysis by human UDP-GalNAc polypeptide:GalNAc transferase-T2. *Org. Biomol. Chem.* **12**, 2645–2655 (2014).
- Albesa-Jové, D. *et al.* A native ternary complex trapped in a crystal reveals the catalytic mechanism of a retaining glycosyltransferase. *Angew. Chem. Int. Ed. Engl.* **54**, 9898–9902 (2015).
- Corbett, K., Fordham-Skelton, A.P., Gatehouse, J.A. & Davis, B.G. Tailoring the substrate specificity of the beta-glycosidase from the thermophilic archaeon *Sulfolobus solfataricus*. *FEBS Lett.* **509**, 355–360 (2001).
- Hancock, S.M., Corbett, K., Fordham-Skelton, A.P., Gatehouse, J.A. & Davis, B.G. Developing promiscuous glycosidases for glycoside synthesis: residues W433 and E432 in *Sulfolobus solfataricus* beta-glycosidase are important glucoside- and galactoside-specificity determinants. *ChemBioChem* **6**, 866–875 (2005).
- Trincone, A., Improta, R. & Gambacorta, G. Enzymatic synthesis of polyol and masked polyol glucosides using β-glycosidase of *Sulfolobus solfataricus*. *Biotransformation* **12**, 77–88 (1995).
- Trincone, A. *et al.* Enzyme catalyzed synthesis of alkyl beta-D-glycosides with crude homogenate of *Sulfolobus solfataricus*. *Biotechnol. Lett.* **13**, 235–240 (1991).
- Aguilar, C.F. *et al.* Crystal structure of the beta-glycosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J. Mol. Biol.* **271**, 789–802 (1997).
- Trincone, A., Perugini, G., Rossi, M. & Moracci, M. A novel thermophilic glycosynthase that effects branching glycosylation. *Bioorg. Med. Chem. Lett.* **10**, 365–368 (2000).
- Mackenzie, L.F., Wang, Q., Warren, R.A.J. & Withers, S.G. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.* **120**, 5583–5584 (1998).
- Zhang, Z. *et al.* Programmable one-pot oligosaccharide synthesis. *J. Am. Chem. Soc.* **121**, 734–753 (1999).
- Williams, S.J., Mark, B.L., Vocadlo, D.J., James, M.N.G. & Withers, S.G. Aspartate 313 in the *Streptomyces plicatus* hexosaminidase plays a critical role in substrate-assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state. *J. Biol. Chem.* **277**, 40055–40065 (2002).
- Withers, S.G., Street, I.P., Bird, P. & Dolphin, D.H. 2-deoxy-2-fluoroglucosides: a novel class of mechanism-based glucosidase inhibitors. *J. Am. Chem. Soc.* **109**, 7530–7531 (1987).
- Lopez, R. & Fernandez-Mayoralas, A. Enzymatic β-galactosidation of modified monosaccharides: study of the enzyme selectivity for the acceptor and its application to the synthesis of disaccharides. *J. Org. Chem.* **59**, 737–745 (1994).
- Yamamoto, K. & Davis, B.G. Creation of an α-mannosynthase from a broad glycosidase scaffold. *Angew. Chem. Int. Ed. Engl.* **51**, 7449–7453 (2012).
- Petzelbauer, I., Reiter, A., Splachtna, B., Kosma, P. & Nidetzky, B. Transgalactosylation by thermostable beta-glycosidases from *Pyrococcus furiosus* and *Sulfolobus solfataricus*. Binding interactions of nucleophiles with the galactosylated enzyme intermediate make major contributions to the formation of new beta-glycosides during lactose conversion. *Eur. J. Biochem.* **267**, 5055–5066 (2000).
- Reuter, S., Rusborg Nygaard, A. & Zimmermann, W. beta-Galactooligosaccharide synthesis with beta-galactosidases from *Sulfolobus solfataricus*, *Aspergillus oryzae*, and *Escherichia coli*. *Enzyme Microb. Technol.* **25**, 509–516 (1999).
- Crout, D.H.G. & Vic, G. Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis. *Curr. Opin. Chem. Biol.* **2**, 98–111 (1998).
- Shim, J.-H., Chen, H.M., Rich, J.R., Goddard-Borger, E.D. & Withers, S.G. Directed evolution of a β-glycosidase from *Agrobacterium* sp. to enhance its glycosynthase activity toward C3-modified donor sugars. *Protein Eng. Des. Sel.* **25**, 465–472 (2012).
- Watts, A.G. *et al.* *Trypanosoma cruzi* trans-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. *J. Am. Chem. Soc.* **125**, 7532–7533 (2003).
- Lawson, S.L., Warren, R.A.J. & Withers, S.G. Mechanistic consequences of replacing the active-site nucleophile Glu-358 in *Agrobacterium* sp. beta-glycosidase with a cysteine residue. *Biochem. J.* **330**, 203–209 (1998).
- Gloster, T.M. *et al.* Structural studies of the beta-glycosidase from *Sulfolobus solfataricus* in complex with covalently and noncovalently bound inhibitors. *Biochemistry* **43**, 6101–6109 (2004).
- Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **99**, 12562–12566 (2002).
- Asensio, J.L., Ardá, A., Cañada, F.J. & Jiménez-Barbero, J. Carbohydrate-aromatic interactions. *Acc. Chem. Res.* **46**, 946–954 (2013).
- 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. *J. Biol. Chem.* **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. *J. Am. Chem. Soc.* **137**, 7528–7547 (2015).
44. Davies, G.J., Planas, A. & Rovira, C. Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* **45**, 308–316 (2012).
45. Speciale, G., Thompson, A.J., Davies, G.J. & Williams, S.J. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* **28**, 1–13 (2014).
46. Bottoni, A., Miscione, G.P. & De Vivo, 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. *Curr. Opin. Chem. Biol.* **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. *J. Am. Chem. Soc.* **133**, 20301–20309 (2011).
49. Berrin, J.-G. Substrate (aglycone) specificity of human cytosolic β -glucosidase. *Biochem. J.* **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. *Nat. Chem. Biol.* **3**, 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. *Org. Biomol. Chem.* **12**, 2993–3003 (2014).

Acknowledgments

We thank the Engineering and Physical Sciences Research Council (EPSRC) and High Force Research (S.M.H.), the BBSRC (S.S.L., BB/E004350/1), MINECO (grant CTQ2014-55174-P to C.R.) and AGAUR (grant and 2014SGR-987 to C.R.) for funding. B.G.D. was a Royal Society Wolfson Research Merit Award recipient during the course of this work. The authors gratefully acknowledge the computer resources at *MareNostrum* and the technical support provided by BSC-CNS (RES-QCM-2013-3-0011). We would like to thank the referee who suggested possibly similar roles of aromatic side chains 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

J.I.-F. designed and performed calculations. S.M.H., S.S.L., M.K. performed biochemical experiments. S.M.H., M.K., J.K., N.J.O. performed mass spectrometric experiments. S.M.H., K.M., A.F.-S. determined X-ray structures. All authors analyzed results. C.R., S.S.L., B.G.D. wrote the manuscript. All authors except A.F.-S. read and commented on the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.R. or B.J.D.

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 purchased from Life Technologies (Paisley, UK). Bradford reagent concentrate was purchased from Bio-Rad (Hemel, Hempstead, UK). Clostripain was purchased from Worthington Biochemicals Corporation (Lakewood, NJ, USA). LB medium, kanamycin, and IPTG were purchased from Melford (Ipswich, UK). Methyl β -D-galactopyranoside and 2,3,4,6-tetra-O-acetyl- α -D-galactopyranosyl fluoride were 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 K.P. Corbett and A.P. Fordham-Skelton (University of Durham). All chemical and biochemical reagents were purchased from Sigma-Aldrich Chemical Co. unless otherwise noted.

Biological methods. H₂O was purified using a Milli-Q Synthesis system (Millipore), and was heat-sterilized before DNA applications. Agar plates were poured and streaked in a MDH Microflow laminar flow cabinet using media and equipment that had been sterilized in an autoclave at 121 °C for 20 min. Solutions of IPTG (0.1 M) and kanamycin (50 mg•ml⁻¹) were filter-sterilized through 0.2 μ m filters and stored at -20 °C. The final concentrations of kanamycin and IPTG in growth media, unless otherwise indicated, were 50 μ gml⁻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 of Oxford, Department of Biochemistry DNA sequencing service on a ABI 377XL Prism DNA sequencer. S β G was sequenced using primers to the T7 promoter and terminator sequences and the following internal primers:

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

Internal reverse: 5'-GGA ATG AGC TAT TAG C

Cloning the S β G gene into pET28a(+). The gene of S β G was originally inserted in pET24d and was subcloned into pET28a(+). PCR reaction was conducted using the following primers.

Forward: 5'-GGTGGTCATATGTCATTTCCAAATAGC

Reverse: 5'-GGTGGTCTCGAGTTAGTGCC

The amplified WT S β G insert was cleaned using a QIAquick gel extraction kit (Qiagen, Manchester, UK) and digested to generate sticky ends using XhoI and NdeI (both from Promega, Southampton, UK). The restriction digest was carried out for 3 h at 37 °C.

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

Site-directed mutagenesis. Site-directed mutagenesis of S β G WT and further mutagenesis of E387Y S β G 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'-CTATATGTACGTTACTTACAAATGGTATTGCGGATGATGCC Reverse: 5'-GATAATCGGCATCATCCGCAATACCATTGTAAGTACACGTAC

E206A:

Forward: 5'-CAATGAATGCACCTAACGTGGTGG Reverse: 5'-CAACGTTAGGTGCATTTCATTTGTTG

E387F:

Forward: 5'-CTATATGTACGTTACTTTTCAATGGTATTGCGGATGATGCC

Reverse: 5'-GATAATCGGCATCATCCGCAATACCATTGAAAGTAAACGTAC

Y322F:

Forward: 5'-GGAGTTAATTATTTCACTAGGACTGTTGTG Reverse: 5'-CAGTCTAGTGAAAATAATTAACCCAATCC

E387A:

Forward: 5'-CTATATGTACGTTACTGCAAATGGTATTGCGGATGATGCC

Reverse: 5'-GATAATCGGCATCATCCGCAATACCATTTGCAGTAAACGTAC

E. coli TOP10 cells (25 μ l) were transformed with DpnI digested PCR products (1 μ l) 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 S β G 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 medium containing 85 μ g/ml kanamycin. 2 L flasks containing LB medium (500 ml) and kanamycin were prewarmed 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 r.p.m. until an optical density (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 9,000 r.p.m. 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 \times 15 amplitude micron bursts of 30 s separated by 1 min intervals. The lysed cells were centrifuged at 10,000 r.p.m. in a JA20 Beckman rotor for 30 min at 4 °C. Lysates were filtered through a Nalgene 0.2 μ m filter before protein purification.

The filtered supernatant was applied to a pre-equilibrated (binding buffer) GE Healthcare 5 ml 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 binding buffer. Protein was detected with an on-line detector monitoring A280 and column fractions were collected and analyzed by SDS-PAGE. Fractions containing the ca. 57 kDa protein were pooled. The protein was further purified by gel filtration. Using an Äkta FPLC system, S β G (~54 mg, 3 ml in 100 mM Tris-HCl, pH 7.5) was loaded onto a Superdex 75 (320 ml) gel filtration column and eluted with Tris-HCl buffer (100 mM, pH 7.5). Buffers were filtered and degassed before use. Fractions were checked with SDS-PAGE and those containing desired protein pooled. Pooled protein samples were dialyzed into the appropriate buffer (2 L) overnight with two changes. The protein was concentrated using Vivaspin centrifugal concentrator (Sartorius, Goettingen, Germany). Protein concentration was determined by Coomassie Brilliant Blue binding via absorption at 595 nm⁵⁴. N-terminal sequencing was performed by the University of Oxford, Department of Biochemistry Immunochemistry protein characterization service, using an Applied Biosystems Procise 494A protein sequencer employing Edman degradation.

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

Protein mass spectrometry. S β G and its mutants were analyzed by ESI-MS in positive ion mode on a Micromass LCT mass spectrometer interfaced with a Waters 2790 Alliance HT separations module using a Jupiter C5 5 μ m 300 Å 150 \times 2 mm i.d. column (Phenomenex, Macclesfield, UK). Before injection, protein samples (20 μ l, ~60 μ M) were washed with H₂O using spin concentrators. Proteins were eluted at 0.2 ml/min by a gradient from Buffer A (0.1% formic acid in 95% water, 5% acetonitrile) to Buffer B (0.1% formic acid in 95% acetonitrile, 5% water) as follows: 5% B for 3 min, increase to 100% B over 13 min, hold at 100% B for 2 min, decrease to 5% B over 1 min, hold at 5% B for 6 min. The eluent was split 1:1 waste:mass spectrometer. The following MS parameters were used: capillary voltage, 3,000 V; sample cone, 35 V; desolvation temperature, 200 °C; source temperature, 80 °C; desolvation flow (N₂), 425 Lh⁻¹; no cone flow; pusher cycle time, 94; and ion energy, 34 V; *m/z* scan range 200 to 2,100; scan time, 1 s; interscan time, 0.1 s. The electrospray mass spectra were processed using the Maximum Entropy method (MaxEnt1).

Proteolysis reactions. Cyanogen bromide cleavage was conducted as follows. Protein (0.5 ml, 1.69 mg·ml⁻¹, 15 nmol; total methionine = 150 nmol) was concentrated to 25 µl and washed with H₂O (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 equivalent) was added and the reaction mixture incubated in the dark at room temperature for 24 h. The mixture was diluted with water (2 × 2 ml), freeze-dried, and resuspended 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 H₂O, and was either (a) Incubated at 95 °C for 5 min with 2% w/v RapiGest SF in NH₄HCO₃ 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 NH₄HCO₃ (180 µl, 20 mM, pH 8.0). Once cooled, sequencing-grade modified trypsin (20 µg) in NH₄HCO₃ (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.

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

Semi-preparative scale HPLC purification of CNBr-generated peptides. CNBr-generated peptides (100 µl, 116 µM) were injected onto a Jupiter C4 5 µm 300 Å column (150 × 4.6 mm i.d.) connected to a Waters 2790 Alliance HT separations module, directly interfaced to a Micromass LCT mass spectrometer fitted with an ESI source. The peptides were eluted at 1 ml/min by a gradient from Buffer A (0.1% formic acid in 95% water, 5% acetonitrile) to Buffer B (0.1% formic acid in 95% acetonitrile, 5% water): 5.6% B for 5 min, increase to 42.2% B over 25 min, increase to 100% B over 20 min, hold at 100% B for 10 min, decrease to 5.6% B over 10 min, hold at 5.6% B for 10 min. A post-column splitter was used to direct 85% of flow to the fraction collector, while 15% entered the mass spectrometer for analysis. Peptide elution was followed by ESI-MS, in positive ion mode, using the same MS parameters as reported for protein MS, except for *m/z* scan range, which was 200–2,500. Fractions (1 ml) were collected and freeze dried for further digestion/analysis.

Peptide analysis by MS and MS/MS. Protein digestion mixtures were injected onto a Jupiter C4 5 µm 300 Å capillary column (150 × 0.5 mm i.d.) connected to an Agilent 1100 series capillary HPLC. The peptides were eluted at 15 µl/min by a gradient from Buffer A (0.1% formic acid in water) to Buffer B (0.1% formic acid in acetonitrile): 10% B for 5 min, increase to 43% B over 25 min, increase to 95% B over 20 min, hold at 95% B for 10 min, decrease to 10% B over 10 min, hold at 10% B for 10 min. UV absorption at 210 and 280 nm and positive ion mode ESI-MS (Q-TOF micro, Micromass) followed peptide elution using the following parameters: capillary voltage, 3,000 V; sample cone, 35 V; extraction cone, 6 V; desolvation temperature, 150 °C; source temperature, 80 °C; desolvation flow (N₂), 150 Lh⁻¹; *m/z* scan range, 400–2,000; step size, 1 s; interscan time, 0.1 s. MS/MS data was obtained by selectively introducing the precursor ion (mass triggered) into the collision cell and fragmentation was induced by collision with Ar using charge state recognition, calculating the collision energy from the standard profile within MassLynx 4. The resulting product ions produced were scanned over the range 100 to 3,000, scan

time 1 s, and with 0.1 s interscan time. Product ion spectra were deconvoluted using MaxEnt3 and *de novo* sequenced using PepSeq within MassLynx 4.

2,4-Dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside labeling. To label E387Y, E387F, E387Y:E206A and E387Y:Y322F SsβG, 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (1.0 mg, 2.9 µmol, 1,000 eq.) was incubated with enzyme (0.20 mg, 3.5 nmol) in sodium phosphate buffer (200 µl, 50 mM, pH 6.5) for 3 d at 45 °C.

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

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

Solvolysis of 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside with or without SsβG-E387Y was performed as follows. 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside **16** (1 mg, 2.9 µmol, 1,000 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 96-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 (1 mg, 2.9 µmol, 1,000 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 *p*NP and DNP were determined by measuring the absorbance at various 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 *p*NP and DNP were 3,212 M⁻¹·cm⁻¹ and 4,230 M⁻¹·cm⁻¹, respectively.

p-Nitrophenyl glycosides were hydrolyzed 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 concentration range of substrate), 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, nonlinear 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 s.d. of the rates and protein concentration.

2,4-Dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside labeling 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, 1,000 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 to 0.1 M. Data were interpreted using Dixon plot analysis⁵⁵.

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 D₂O. 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)-Δ²-oxazoline was examined as follows. 2-Methyl-4,5-(2-deoxy-α-D-glucopyrano)-Δ²-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-galactopyranoside (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 Na₂CO₃ (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 equation⁵⁶:

$$\left(\frac{k_{\text{cat}}}{K_{\text{M}}}\right)_{\text{obs}} = \left(\frac{k_{\text{cat}}}{K_{\text{M}}}\right)_{\text{max}} \left(\frac{1}{1 + \frac{10^{-\text{pH}}}{10^{-\text{p}K_{\text{a}1}}} + \frac{10^{-\text{p}K_{\text{a}2}}}{10^{-\text{pH}}}} \right)$$

HPLC analyses. For initial hydrolysis assays, the reaction mixture containing *p*-nitrophenyl β-D-galactoside (10 mM) and SsβG-E387Y (0.93 μM) in sodium phosphate buffer (50 mM, pH 6.5) was incubated at 45 °C. 40 μl aliquots were withdrawn at every 1 h interval and aliquots were immediately filtered using Vivaspin centrifugal filter (Sartorius, MWCO 10,000). Each filter was washed with 20 μl of distilled water. Volume was increased to 60 μl. A 5 μl portion of the sample was injected into Phenomenex Luna NH₂ HPLC column on Dionex UltiMate 3000 system (Dionex, Hemel Hempstead, UK). Analysis was performed by eluting isocratic 70/30 acetonitrile/water at 1 ml/min flow rate.

Transglycosylation timecourse. To determine the time course of the reaction, 2 μM SsβG-E387Y (in 50 mM sodium phosphate buffer, pH 6.5) was incubated with 10 mM 4-nitrophenyl β-galactoside at 45 °C. 60 μl aliquots were withdrawn at 15 min, 30 min, 45 min, 1 h, 3 h, 5 h and 6 h. Samples were filtered using Vivaspin centrifugal filter (Sartorius, MWCO 10,000). A 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.

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 6 h 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 ~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-visible absorbance as measured by area under the curve from chromatograms at fixed time points 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 (>5 μM), 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; 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, before 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_α 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 MOLREP⁵⁷ using the SsβG-WT crystal structure (PDB ID: 1GOW)²⁵ as a search model. Molecular replacement was followed by rigid body refinement in REFMAC5 (ref. 58), and the sequence was corrected by manual building in XtalView⁵⁹. Structures were refined using REFMAC5, including TLS refinement. 5% of the reflections were excluded for calculation of *R_{free}*. Strong noncrystallographic symmetry restraints were imposed for the four or eight molecules within the asymmetric unit. Water molecules were added automatically using ARP/wARP⁶⁰ and checked manually. The models were validated by PROCHECK⁶¹ and WHAT_CHECK⁶². 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 data sets, collected at a wavelength of 0.87 Å at SRS Daresbury beamline 9.6, in space group *P*2₁2₁ were reprocessed using XIA2 (ref. 63) to generate an apo E387Y data set with a resolution of 2.16 Å. The apo E387Y structure was solved by molecular replacement with Phaser⁶⁴ using the same search model as above. This model was optimized by iterative cycles of manual building using the graphics program Coot⁶⁵ and refinement as implemented in phenix.refine⁶⁶. MolProbity⁶⁷ was used to validate the final model; the all-atom clashscore is 0.84 and the percentage of residues in the favored 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 with the program Coot⁶⁵ using the default parameters and figures were

prepared using PyMOL. First, a ligand was modeled into the active site of SsβG-E387Y by structural alignment of the protein with that of the SsβG-WT³⁹ complexed with D-galactohydroximolactam (PDB ID: 1UWT). Although this modeled 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 pNPGal 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-galactohydroximolactam inhibitor in the SsβG-E387Y structure (Supplementary Fig. 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.

Computational studies. The initial protein structure for the simulation was taken from the X-ray crystal structure of the SsβG-E387Y mutant (PDB ID: 513D). Two molecules of pNPGal were manually placed at the entrance of the enzyme catalytic groove, at distances of 12 and 20 Å, respectively, from residue Y387. The protonation states and hydrogen atom positions of all ionizable amino acids residues were selected based on their hydrogen bond environment. Eleven histidine residues were modeled in their neutral states, while the rest were modeled in their protonated states. All the crystallographic water molecules were retained and extra water molecules were added to form a 10 Å water box around the protein surface. Three sodium ions were added to neutralize the enzyme charge.

Molecular dynamics (MD) simulations of the enzyme (E387Y mutant) in complex with two molecules of pNPGal were performed with the Amber10 software package⁶⁸. The protein was modeled with the FF99SB force field⁶⁹, whereas all carbohydrate molecules were modeled with the GLYCAM06 force field⁷⁰. The partial charges for the aglycon leaving groups were obtained using a RESP fit to a HF/6-31G* calculation, which was performed with the Gaussian 03 software package. The MD simulations were carried out in several steps. First, the system was minimized, maintaining the protein and both substrate molecules fixed. In a second step, the entire system was allowed to relax. Weak spatial constraints were initially added to the protein and substrates to gradually reach the desired temperature of 300 K, while the rest of the system was allowed to move freely. The constraints were subsequently removed and the system was subjected to 100 ps of constant pressure MD simulation to adjust the density of the water environment. After the equilibration process, 200 ns of constant volume MD simulation was performed. During this time, one of the two pNPGal molecules partially entered the catalytic groove, whereas the other one remained at its entrance. In particular, one pNPGal molecule accommodates near the –1 subsite (at ~8 Å from the catalytic residues) and the second one remains far from it (more than 20 Å). As complete entrance of the two molecules in the active site did not occur during the 200 ns timescale window, it was subsequently activated using an enhanced sampling methodology (metadynamics^{40,71}).

A snapshot of the equilibrium MD simulation was taken for the metadynamics simulations, which were performed with the NAMD2.9 software⁷². Two collective variables were chosen (Supplementary Fig. 15). The first one (CV_1) was taken as the distance between the center of mass of the Gal ring of the inner pNPGal molecule and the center of mass of a few residues defining the –1 enzyme subsite (Tyr387, Trp425 and His150, as identified in structural studies of WT SsβG in complex with inhibitors and PDB ID: 1UWT). Thus, this collective variable measures the degree of penetration of the first pNPGal molecule into the active site. This pNPGal molecule will act as the donor of the transglycosylation reaction. The second collective variable (CV_2) was taken as the distance between the glycosidic oxygen (O1) of the donor and the terminal hydrogen atom of the hydroxymethyl substituent of the outer pNPGal molecule (the acceptor of the transglycosylation reaction). Therefore, CV_2 accounts for the formation of a O1–H interaction. In other words, it measures the distance between the donor and acceptor molecules. The values of the height and width of the

Gaussian-like biasing potential were selected as 1.0 kcal/mol and 0.15 Å, respectively. A new Gaussian-like potential was added every 2 ps. Reflective walls were placed at proper distances (15 Å in CV_1 and 25 Å in CV_2) to avoid sampling of the ligands in a fully solvated environment with loss of the substrate–protein interactions. The simulation was continued until the system completely explored the free energy surface (Supplementary Fig. 16), which, in terms of the simulation time, corresponds to approximately 100 ns.

An additional metadynamics simulation in which CV_2 accounts for the formation of the 1,3 bond (by smoothly biasing the distance between O1 and H3') led to a ternary complex that can be considered as precursor of the 1,3-linkage product. Interestingly, the acceptor pNPGal enters the binding site in a different orientation with respect to the previous simulation, resulting in a different pose in the active site (Supplementary Fig. 17). Nevertheless, the hydroxyl group of the acceptor is quite fixed in the active site in both cases and the donor sugar is stabilized by CH–π interactions engendered by Y387. In the case of the 1,6 precursor (Supplementary Fig. 17b), the acceptor aglycon makes a π–π interaction with Y322, whereas in the case of the 1,3 pathway (Supplementary Fig. 17a), the acceptor places the aglycon away from Y322, which enables the formation of sugar–π (CH–π) interactions between the acceptor sugar and the donor aglycon. Thus, in both cases the sugar makes use of π–π and sugar–π interactions to stabilize the substrates in the optimum orientation for catalysis.

Quantum mechanics/molecular mechanics (QM/MM)-based metadynamics were used to model a transglycosylation reaction leading to the major 1,6-linked product using a QM/MM method⁷³ that combined *ab initio* (Car–Parrinello) molecular dynamics⁷⁴, based on density functional theory (DFT), with classical molecular dynamics, which is based on the use of a force field, as implemented in the CPMD code (<http://www.cpmo.org>). The QM region was chosen as to include both *p*-nitrophenyl glycosides (72 atoms) and was enclosed in an isolated 18 × 18.5 × 18.5 Å³ supercell. Kohn–Sham orbitals were expanded in a plane-wave basis set with a kinetic energy cutoff of 70 Ry. Troullier–Martins *ab initio* pseudopotentials with dispersion-corrected atom-centered potentials (DCACPs)⁷⁵ were used for all elements. The PBE functional in the generalized gradient-corrected approximation of DFT was used. Structural optimizations were done by MD with annealing of the ionic velocities until the maximal component of the nuclear gradient was $<5 \times 10^{-4}$ a.u. A constant temperature of 300 K was reached by coupling the system to a Nosé–Hoover thermostat⁷⁶. A time step of 0.12 fs and a fictitious electron mass of 650 a.u. for the Car–Parrinello Lagrangian were used. The metadynamics algorithm^{40,71} was used to model the enzymatic reaction and to reconstruct the free energy landscape of the transglycosylation. First of all, the structure of the products complex was obtained by steered MD, dragging the system toward a stable configuration by pulling four collective variables (C1–OpNP, C1–O6, OpNP–H and O6–H). The products structure (E387Y SsβG in complex with Gaβ1,6GaβpNP and pNP) was further equilibrated for 10 ps of QM/MM MD simulation, in which no substantial changes of the active site or the protein structure were observed. Subsequently, a metadynamics simulation to obtain the reaction pathway was initiated. Three collective variables containing the main bonds undergoing breaking or formation were used (Supplementary Fig. 18). The first collective variable (CV_1) was taken as the distance between the donor anomeric carbon (C1) and the acceptor O6 atom. Thus, this variable measures the degree of formation of the new glycosidic bond. The second collective variable (CV_2) was taken as the distance between C1 and the leaving group oxygen (O1), thus this CV accounts for the departure of the leaving group. Finally, the third collective variable (CV_3), taken as the difference in coordination number (CN) of the O6–H and O1–H bonds, measures the extent of proton transfer between the acceptor and the pNP moiety. The following definition of CN is used:

$$CN_{ij} = \frac{1 - \left(\frac{d_{ij}}{d_0}\right)^p}{1 - \left(\frac{d_{ij}}{d_0}\right)^{p+q}}$$

where d_{ij} is the internuclear distance of the atoms involved, d_0 is the threshold distance for bonding, and p and q are exponents that determine the steepness of CN_{ij} decay with respect to d_{ij} . CN values range from 0 (no bond) to 1 (a bond). The values for the CN parameters were selected as $d_0 = 2.50$ a.u., $p = 12$, and $q = 4$.

An extended Lagrangian version of the metadynamics method was used⁷⁷. The selected mass values of the fictitious particles were 2 a.m.u. (CV_1 and CV_2) and 1 a.m.u. (CV_3). A force constant value of 20 a.u. was taken for all CVs. The height of the Gaussian terms was set to 2.0 kcal•mol⁻¹ and the width was set to 0.3 a.u., 0.4 a.u. and 0.1 arbitrary units for CV_1 , CV_2 and CV_3 , respectively. A new Gaussian-like potential was added every 200 MD steps and the simulation was stopped after one recrossing of the TS. Approximately 940 Gaussian hills were added during the metadynamics simulation. To confirm the stability of the oxocarbenium ion pair, two snapshots corresponding to this intermediate were optimized by annealing of the atomic velocities, using a threshold for nuclear gradients of 5×10^{-4} a.u. (maximum component). The oxocarbenium ion-pair state was stable in all optimizations, featuring a planar geometry at the anomeric carbon (the C5-O-C1-C2 dihedral angle is close to zero, characteristic of a sp² hybridization). Afterwards, ten independent QM/MM MD simulations of 15 ps starting with random velocities were performed. Of these, two evolved to the reactants state and one evolved toward products at ≈ 4 ps. The oxocarbenium ion-pair state was stable during the whole simulation time in the remaining seven simulations. These results confirm that the oxocarbenium ion-pair state is an intermediate of the reaction, with a lifetime > 15 ps.

To further validate our model, an additional metadynamics simulation was designed considering an alternative mechanism in which E206 acts as a nucleophile, forming a glycosyl-enzyme intermediate. Both pNPGal molecules and the E206 residue (up to the C β atom) were included in the QM region. As in the previous case, three collective variables were used. The first collective variable (CV_1) was taken as the distance between the donor anomeric carbon (C1) and one oxygen atom of the nucleophile E206 residue. Thus, this variable measures the degree of formation of the new glycosidic bond. The second and third collective variables were the same as the ones described above, as well as the metadynamics setup (hills width, height and deposition time). The metadynamics simulation was run for ≈ 10 ps (416 deposited Gaussian functions). Although a glycosyl-enzyme covalent intermediate was sampled (Supplementary Fig. 21), it does not correspond to a minimum of the FES and it is very high in energy (32 kcal/mol, involving an energy barriers of the same magnitude along the reaction pathway). Therefore, our results show that E206 cannot act as a nucleophile in a double-displacement reaction.

Statistical methods. Kinetic parameters were analyzed through either nonlinear regression using the Michaelis–Menten equation or linear regression using the Lineweaver–Burk equation. Apparent pK_a 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.).

Data availability. The data that support the findings of this study are available in the Supplementary Information 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.

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 (1953).
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. & Teplyaev, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
58. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
59. McRee, D.E. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
60. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein models. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 129–147 (1993).
61. Laskowski, 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 Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
64. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
65. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
66. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
67. Davis, I.W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
68. Pearlman, D.A. *et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1–41 (1995).
69. Cornell, W.D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
70. Kirschner, K.N. *et al.* GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655 (2008).
71. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 826–843 (2011).
72. Phillips, J.C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
73. Laio, A., VandeVondele, J. & Rothlisberger, U. A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.* **116**, 6941–6947 (2002).
74. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
75. Lin, I.-C. Library of dispersion-corrected atom-centered potentials for generalized gradient approximation functionals: Elements H, C, N, O, He, Ne, Ar, and Kr. *Phys. Rev. B* **75**, 205131 (2007).
76. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
77. Iannuzzi, M., Laio, A. & Parrinello, M. Efficient exploration of reactive potential energy surfaces using Car–Parrinello molecular dynamics. *Phys. Rev. Lett.* **90**, 238302 (2003).