

**Mechanism of Oligosaccharide Synthesis via a Mutant GH29
Fucosidase**

Journal:	<i>Reaction Chemistry & Engineering</i>
Manuscript ID	RE-ART-10-2018-000240.R1
Article Type:	Paper
Date Submitted by the Author:	30-Nov-2018
Complete List of Authors:	Burgin, Tucker; University of Michigan, Department of Chemical Engineering Mayes, Heather; University of Michigan,



Cite this: DOI: 10.1039/xxxxxxxxxx

Mechanism of Oligosaccharide Synthesis via a Mutant GH29 Fucosidase[†]

Tucker Burgin^a and Heather Mayes^{a‡}Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Techniques for synthesis of bespoke oligosaccharides currently lag behind those for other biopolymers such as polypeptides and polynucleotides, in part because of the absence of satisfactory enzymatic tools to perform the synthetic reactions. One promising avenue of development for this problem is glycoside hydrolase enzymes with mutated nucleophile residues (called glycosynthases), which retain some elements of their native specificity and work with cheaply available substrates. However, the mechanistic underpinnings of this class of enzymes are not yet well-understood, and what few atomistic studies have been conducted have found different reaction pathways. In this paper, we describe the first unbiased computational study of the mechanism of a GH29 glycosynthase enzyme, *Thermotoga maritima* α -L-fucosidase (*TmAfc*) D224G. We find a single-step endothermic reaction step with an oxocarbenium-like transition state, demonstrating how stabilization of this transition state structure (which is common to many retaining glycoside hydrolases) can be repurposed in mutant enzymes to perform synthesis instead of hydrolysis. Our results are consistent with previous experimental observations and help both to clarify the mechanism of the existing single-mutant and to provide directions for further engineering of this and other glycosynthases.

1 Introduction

Oligosaccharides have long been known to play a wide variety of important roles in biology,¹ from structural support to signaling cascades and mediation of cell-cell interactions — as one 1993 review paper put it: “all of the theories are correct.”² Further understanding of the properties and roles of particular oligosaccharides requires synthesis of homogeneous samples of sufficient quantities for research studies. Unfortunately, techniques for synthesizing oligosaccharides have lagged significantly behind those for other biological polymers, owing in part to the complexities of regio- and stereochemistry.^{3,4} Though significant progress has been made since the 1990s, the variety of methods that have been developed are narrow in their applicability, usually taking place over many successive steps with a loss of conversion at each step, and require meticulous control over the reaction conditions to minimize competing off-pathway reactions.^{5–7}

The natural alternative to arduous organic synthesis routes is the use of enzymes to catalyze highly specific glycosynthetic reactions. Enzymes remove the need for careful protection and deprotection steps or the tuning of highly sensitive reaction parameters.

Unfortunately, the enzymes evolved in nature to perform these reactions, glycotransferases, are largely not amenable to biotechnological applications because of their low stability outside the cell and reliance upon expensive nucleotide-sugar substrates.⁸ Although strategies to circumvent these issues are in development, such as the recycling of reacted nucleotides or directed evolution of the enzymes to accept more readily available precursors, this has proven to be a non-trivial problem.^{9,10}

Other enzymatic alternatives are mutant glycosidases, dubbed “glycosynthases.”^{4,12} To modify a two-step retaining bond cleavage mechanism, the nucleophilic residue responsible for forming the stable intermediate is mutated into a non-reactive residue in order to preclude the forward reaction, leaving the catalytic base (Glu-266) intact and able to aid synthesis of a glycosidic bond between suitable glycosyl donor and acceptor molecules. Such enzymes provide a powerful framework for building oligosaccharides from readily available substrates (employing much simpler leaving groups than nucleotides, such as azide groups or fluorine atoms).¹³ Glycoside hydrolases are more stable and soluble than glycosyltransferases and thus more amenable to *in vitro* and industrial conditions. However, because they are evolved for a different reaction pathway, they lack the high specificity and efficiency characteristic of most wild-type enzymes. Furthermore, much of the work to date on the discovery of new glycosynthases has been a series of shots in the dark: it is not well-understood

^a University of Michigan Department of Chemical Engineering, 2800 Plymouth Ave., Ann Arbor, MI USA.

[‡] E-mail: hbmeyes@umich.edu

[†] Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

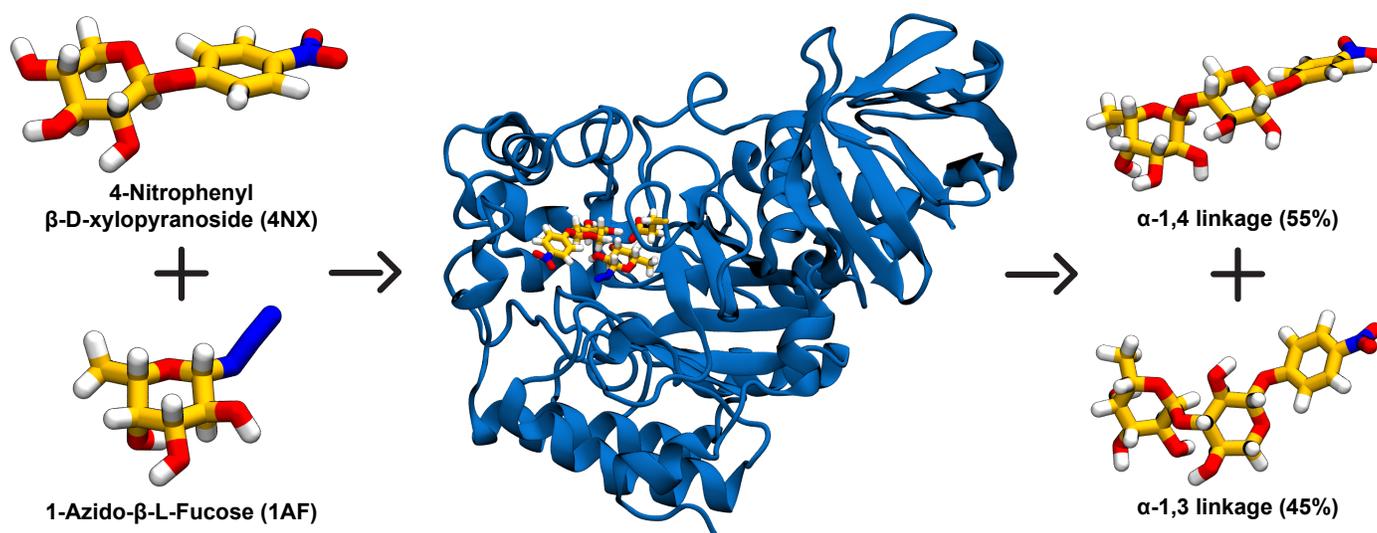


Fig. 1 The *TmAfc* D224G reaction of interest in this work. This reaction was experimentally studied by Cobucci-Ponzano *et al.* in 2009.¹¹ As reported there, this reaction produces α -1,4 and α -1,3 products in nearly equal amounts, and has an overall specificity of 91% for transfer of the donor (4NX) to the shown acceptor (1AF) versus water.

which nucleophile mutations apply to which glycoside hydrolases will produce an active glycosynthase or why, and the most successful work for obtaining new or improved glycosynthases has relied on random, semi-random, or otherwise exploratory approaches.^{11,14,15}

These shortcomings motivate attempts to rationally engineer glycosynthases to produce a given oligosaccharide with high specificity and efficiency. This endeavor will require a clear understanding of the reaction mechanism at the atomic level. However, although glycosynthases have been present in the literature for over 20 years,¹² investigations into the atomistic underpinnings of these mutants' reactions have been scant. One 2013 study by Zhang *et al.* described a metadynamics study on the *Humicola insolens* Cel7B cellulase E197S mutant in its glycosylation reaction between α -lactosyl fluoride and the flavonoid luteolin.¹⁶ In that same year, Wang *et al.* performed partitioned-rational function optimization calculations to find energy minima and maxima in the cellulose synthase reaction of rice BGlul1 β -glucosidase mutants E386G, -S and -A.¹⁷ Although both papers help to clarify the mechanisms of their respective reactions, taken together they underscore how individual studies do not capture the complete picture of how this class of enzyme functions. For example, although the transition state search algorithm employed by Wang *et al.* identified only a single step in the BGlul1 mutant reaction mechanisms, Zhang *et al.* describe a three-step reaction with stable intermediates in *HiCel7B* E197S. Furthermore, although in Zhang *et al.* the mutant serine participates in the reaction by stabilizing the leaving fluorine, the corresponding interaction in BGlul1 E386S was shown to be less favorable compared to the lack of interaction of E386G. Because of these stark disagreements, it is clear that further study is required in order to gain further understanding of these promising enzymes.

Herein we present a validated reaction mechanism for the transglycosylation reaction between 1-azido- β -L-fucose (1AF)

and 4-nitrophenyl β -D-xylopyranoside (4NX) catalyzed by *Thermotoga maritima* α -L-fucosidase (*TmAfc*) D224G, as diagrammed in Figure 1.¹¹ Significantly, we present the first study of the complete reaction pathway of a glycosynthase performed strictly using methods that do not bias the Hamiltonian by introducing non-physical energy terms into the model. Because the only decision made in our study that biases our results is the choice of reactant and product definitions (as discussed in the Computational Methods section), this result represents the most rigorous computational study of a glycosynthase reaction to date. *TmAfc* D224G differs from the enzymes in previous computational studies (and most studies of glycosynthases in general) in that it uses α -fucosyl donors (compared to the α -lactosyl and α -glucosyl donors of Zhang *et al.* and Wang *et al.* respectively) with azide leaving groups instead of fluorine atoms. Despite the underrepresentation of fucosylsynthases in the literature, fucosyl oligosaccharides are of particular interest in biomedical applications, including as anti-cancer and anti-inflammatory drugs.¹⁸ They are also major constituents in human milk oligosaccharides present in natural breast milk and implicated in healthy gut microbiota development, but typically not included in infant formula.^{19–21}

The results of our molecular models are in good agreement with the experimental observations in Cobucci-Ponzano *et al.* with respect to the relative abundances of the two isomeric products shown in Figure 1.¹¹ They reveal a one-step, endothermic reaction mechanism, wherein the dissociation of the leaving azide from the electron donor occurs in concert with the transfer of the hydrogen from the acceptor to the catalytic residue, as well as with the bond formation between the donor and acceptor, through an oxocarbenium-like transition state. Our results explain the experimental observation that, unlike in the D224G mutant, fucosidase activity in the *TmD224S* mutant cannot be rescued by the addition of free azide,¹¹ and also provide clues for rational engineering of this and similar enzymes in the future.

2 Computational Methods

2.1 Model building

The *TmAfc* model was based on the crystal structure PDB ID: 2ZXD,²² chosen over the two earlier crystal structures published (PDB IDs: 1ODU²³ and 2WSP¹¹) because they both contain incomplete loops near the active site, and because the resolution of 2ZXD is considerably better than the next most recent (2.15 Å vs. 2.65 Å). Although the 2ZXD structure is complexed with an inhibitor molecule, the protein backbone overlays very closely with that of 1ODU and 2WSP (complexed with fucose and α -L-Fucose-(1-2)- β -L-Fucosyl-Azide, respectively), indicating that complexing with the inhibitor does not result in any major conformational changes.

The substrate models had to be custom-built for this study, as parameters for neither 1AF nor 4NX were available. In both cases, parameterization began with the Generalized Amber Force Field (GAFF),²⁴ to which appropriate GLYCAM06 parameters²⁵ and other custom parameters were added as needed to obtain qualitatively reasonable agreement between minimized structures obtained using the custom force field and quantum mechanical SCC-DFTB calculations.^{26,27} For 1AF, the additional parameters required were those for the azide group and its connection to the sugar, and were taken from Carvalho *et al.*²⁸ and Weiner *et al.*²⁹, respectively. For 4NX, one parameter was calculated directly using Gaussian 16 Rev. A.03,³⁰ using the B3LYP quantum mechanics model^{31–34} with the 6-311+G(d,p) basis set.^{35,36} Comparisons between the parameterized molecular mechanics models and the quantum mechanical models, as well as documentation of all the parameters added to GAFF to build the custom force fields, are available in the ESI.[†]

To insert the substrates into the active site of the enzyme, first the donor 1AF was overlaid atop the inhibitor present in the crystal structure using the RMSD Visualizer tool in VMD³⁷, taking advantage of the six-membered ring structure shared between the two. The acceptor 4NX was inserted manually into the open cleft nearby the donor in such a way as to place its sugar's O4 close to the donor's C1 atom (in anticipation of the bond that forms between them). This structure was solvated in a box of TIP3P water molecules³⁸ such that there was everywhere at least 10 Å between the solute and the edge of the box. The model was minimized with Amber 16³⁹ over 2500 steps and then heated from 100 K to 300 K over 10,000 steps (followed by 1000 additional steps at 300 K) using the Andersen thermostat⁴⁰ to randomize velocities every 1000 steps in an NVT ensemble. Finally, the structure was equilibrated with molecular mechanics (MM) for 10,000 steps with isotropic pressure scaling turned on (NPT ensemble) and velocity randomization every 100 steps. A step size of 2 fs and a cutoff distance of 8.0 Å were used throughout, and the SHAKE algorithm⁴¹ was applied to restrict the covalent bond lengths of hydrogen atoms during heating and equilibration.

2.2 Transition path sampling

Our transition path sampling (TPS) methodology is divided into several steps: transition state hypothesizing, aimless shooting, likelihood maximization,⁴² committer analysis,⁴³ and equilib-

rium path sampling.⁴⁴ Taken together, they represent a method of sampling the transition state ensemble without biasing the Hamiltonian, obtaining a reaction coordinate from that sample, verifying the transition state described by the resulting reaction coordinate, and then measuring the free energy surface along that reaction coordinate.

In the remainder of this section, we will detail our methodology for building the models and performing the aimless shooting step, which is responsible for producing the data that the following steps (likelihood maximization, committer analysis, and equilibrium path sampling) were used to analyze. The methodologies for those analysis steps can be found in the ESI.[†]

2.2.1 Transition state hypothesizing.

Aimless shooting requires at least one (and preferably more) initial structure(s) close to the separatrix (the surface in phase space along which any trajectory with randomly selected velocities for all atoms will have an equal chance of collapsing to the product state or to the reactant state.) Because we don't have an *a priori* definition of the separatrix, hypothesized transition states are created by changing the distances between the atoms involved in either formation or cleavage of bonds during the reaction of interest, to a range of distances between those observed in the reactants and products. In the case of the reaction at hand, these bond lengths (and in brackets the corresponding distances tested in Å) were those between: (1) the 4NX O4 hydrogen and closest oxygen of Glu-266 (the catalytic base) [1.1, 1.2, 1.3, 1.4]; (2) that same hydrogen and the 4NX O4 itself [1.1, 1.2, 1.3, 1.4]; (3) the 4NX O4 and the 1AF C1 atoms [2.1, 2.2, 2.3, 2.4]; and (4) the 1AF C1 atom and the primary nitrogen of the azide group [2.5, 2.6, 2.7, 2.8]. Structures with each combination of the given bond distances were built using restraints in combined quantum mechanics and molecular mechanics (QM/MM) simulations with SCC-DFTB.^{26,27} This semi-empirical quantum mechanical model was selected for its good compromise between speed and accuracy, and its history of successful application to systems similar to our own.^{45,46} Details of these simulations are available in the ESI.[†] Combinations with more than one extreme value (defined as either the largest or smallest allowable value for a given bond length) were omitted to reduce computational expense, resulting in a total of 80 starting conformations.

2.2.2 Aimless shooting.

Each of the 80 initial conformations were used to seed two "threads" of aimless shooting using the flexible length shooting algorithm of Mullen *et al.*⁴⁷ and the SCC-DFTB quantum mechanical model,^{26,27} and threads were canceled if they were rejected 10 times in a row to prevent excessive sampling of regions far from the transition state. This procedure was followed to yield 2305 unbiased shooting moves, of which 2069 committed to either the reactant or product basin from the "forward" trajectory. Of those, 275 showed the "backward" trajectory committed to the opposite basin than from the "forward" trajectory, and thus were "accepted" as points along the ensemble of pathways connecting the reactant and product basins. New shooting points were generated after an accepted shooting move by randomly choos-

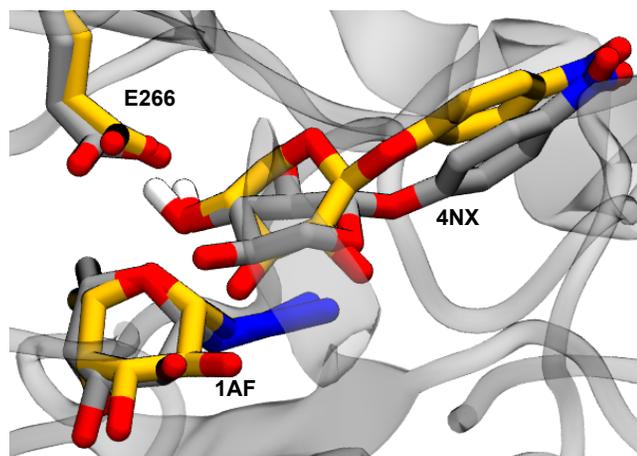


Fig. 2 Snapshots from the ensemble of reactant state structures for the α -1,4 (gold) and α -1,3 (silver) reactions. The key features – namely, the positioning of the acceptor’s active oxygen and hydrogen relative to the catalytic residue and the donor – are conserved with the acceptor rotated 180° in place, motivating the hypothesis that the reaction mechanisms (and reaction coordinates) are homologous. The protein structures are fitted onto one another, though for clarity only that of the α -1,3 product is shown here. Non-reactive hydrogens are also omitted for clarity.

ing (with equal probability) the configuration from the first 50 1-fs frames of either the forward or reverse trajectory, also chosen randomly with equal probability. The basin definitions were: for the products, the 4NX O4 and the 1AF C1 atoms closer than 1.60 Å and the 1AF C1 atom and the primary nitrogen of the azide group further than 2.75 Å; and for the reactants, the former distance further than 2.75 Å and the latter closer than 2.00 Å. These conservative basin definitions were chosen to avoid errors due to potential recrossing. The full set of collective variables that were included in each observation are listed in the ESI.[†]

2.3 Gaussian energy calculations

We calculated the overall energy of each reaction using Gaussian 16, Rev. B.01⁴⁸ using the B3LYP quantum mechanics model^{31–34} with the 6-311+G(d,p) basis set.^{35,36} These calculations were performed on the donor, acceptor, azide, and α -1,3 and α -1,4 product structures solvated in implicit water using the polarizable continuum model.^{49–51} The overall reaction energy was calculated from the Gibbs free energies of the constituent molecules as:

$$\Delta G_{rxn} = G_{product} + G_{azide} - G_{donor} - G_{acceptor}. \quad (1)$$

3 Results and Discussion

3.1 Results

Analysis of how 4NX binds in the active site revealed two clear modes, shown in Figure 2. In the mode shown in gold in the figure, the C4 of the 4NX fucosyl group is closest to the donor C1, which we hypothesized leads to the α -1,4 product, while the mode shown in silver has the C3 of the 4NX fucosyl group closest to the donor C1, corresponding to a α -1,3 product. Binding energy measurements for the α -1,3 and -1,4 reactant states showed

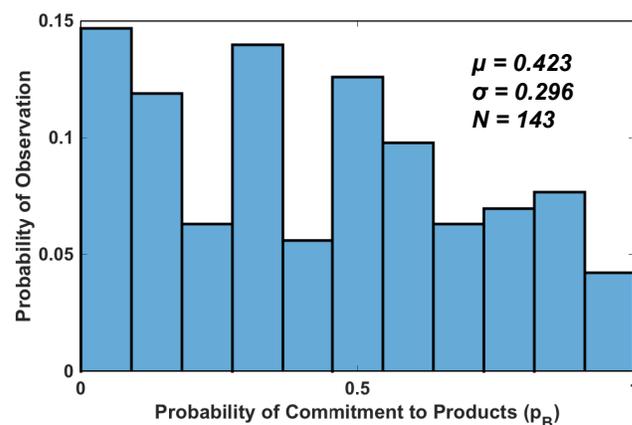


Fig. 3 Committor distribution for the α -1,4 reaction, using the reaction coordinate definition in Equation 2. Each of the 143 shooting points had an RC value with absolute value less than 0.1 and was simulated 10 times in order to obtain a p_B value. Although the committor distribution is not of the ideal, sharply peaked shape, its average is appropriate and its standard deviation is reasonable, suggesting a decent fit between the reaction coordinate and the underlying committor surface.

no significant differences in the protein’s affinity for binding either mode. For this reason, and based on the highly similar reactant state binding modes hypothesized to account for the two products, we studied only the reaction coordinate for the formation of the α -1,4 product. We propose that the reaction coordinate for the α -1,3 would be analogous to that for the α -1,4 reaction, with the identities of the donor O4 and its hydrogen changed to the O3 and its hydrogen in the definitions of the relevant CVs.

Likelihood maximization was performed for the α -1,4 reaction on a set of 54 candidate CVs. The top three CVs whose values and rates of change were most predictive of commitment to the product and reactant basins were: the distance between the acceptor O4 and the donor C1 (CV_3); the distance between the donor C1 and the primary azide nitrogen (CV_4); and the difference of the distances between the transferred hydrogen and the glycosidic and Glu-266 carboxyl oxygens, respectively (CV_{21}). See the ESI for a complete list of CVs.[†] The reaction coordinate constructed only from the configurational parts of these CVs was:

$$RC = -1.35 - 3.66 \text{ \AA}^{-1} CV_3 + 3.83 \text{ \AA}^{-1} CV_4 - 1.69 \text{ \AA}^{-1} CV_{21} \quad (2)$$

where $RC = 0$ represents the transition state and the RC is dimensionless. Because all three of these CVs represent a different bond breaking and/or forming, their importance in describing the progress of this reaction is unsurprising. Validation of the reaction coordinate was performed using committor analysis, and the results are shown in Figure 3, and the same RC is used in analysis of the α -1,4 and α -1,3 reaction paths.

The energy profiles along this RC were obtained via EPS runs and are shown in Figure 4. The activation energy for the α -1,4 reaction is slightly lower (6.4 kcal/mol α -1,4 vs. 7.1 kcal/mol α -1,3), and the overall ΔG of this reaction step is marginally more favorable for the α -1,3 product (3.2 kcal/mol α -1,4 vs. 1.8 kcal/mol α -1,3). Notably, the reaction step is endothermic overall for both products, as was the case in the results of Wang *et al.*

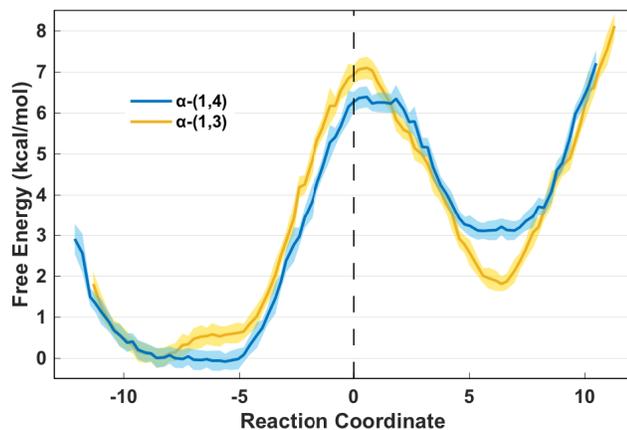


Fig. 4 Energy profiles for both reactions obtained via equilibrium path sampling along the unitless reaction coordinate (RC). The negative RC values represent the side of the separatrix including the reactant basin while positive values represent the side including the product basin. The energy profiles are normalized to the minimum along the α -1,3 reaction, for clarity. The shaded regions represent one standard deviation of the mean values from 1000 iterations of bootstrapping with a sample size of 25 conformations per EPS window.

for BGlu1.¹⁷ It is unlikely that this is the rate determining step in the overall reaction from the reactants in solution to products in solution, which also including binding and unbinding steps; an estimation of the rate coefficient for bond rearrangement is on the order of $1E8$ 1/s for both reactions, using the Eyring equation with $\kappa = 1$ at the experimental reaction temperature of 70°C . If this were the rate determining step, the reaction would proceed in less than the order of seconds and reach an equilibrium of approximately 13:87 (α -1,4: α -1,3) (with low yield due to the endothermic reaction), rather than the observed ratio of 55:45.¹¹

In addition to the energies of activation and ΔG values for the reaction step, we calculated the overall reaction ΔG using Gaussian⁴⁸ as described in the Computational Methods section. The overall reaction energy (the difference in energy between the free product and free reactant states) was slightly exothermic at -1.27 kcal/mol for α -1,4 and -1.24 kcal/mol for α -1,3. This difference in overall ΔG values are consistent with the ratios of products reported by Cobucci-Ponzano *et al.*¹¹ of 55:45 (α -1,4 : α -1,3), giving a ratio of 51:49.

The experimentally observed equilibrium constants (K_{eq}) were very low ($6.6E-3$ and $5.4E-3$ for the α -1,4 and -1,3 reactions, respectively; personal communication, see ESI for details[†]). Based on the conditions reported in that study, if thermodynamics of the overall (slightly exothermic) reaction dominated, a much higher K_{eq} (approximately 6) would be expected. It is possible that the reactant binding and/or product unbinding steps have significant barriers and limited the rate of conversion. Studying these steps would thus be of interest for future work.

3.2 Discussion

A close inspection of the α -1,4 energy profile (Figure 4) provides a potential explanation for the leftward-bias observed in the committor distribution (Figure 3): although the transition state is very

close to the $RC = 0$ value, in fact the energy maximum is slightly to the right of this threshold, owing to minor errors in the likelihood maximization step. Furthermore, the region directly surrounding the transition state is flat towards the products, but falls off quickly towards the reactant basin on the other side. This explanation helps ameliorate concerns that the $RC = 0$ surface does not approximate the true separatrix along the relevant region of phase space; the observed error is instead attributable to a mild bias in the selected shooting points, sampling error, and recrossing (particularly back towards the reactants).

Snapshots corresponding to the reactants, products, and transition state of the α -1,4 reaction are shown in Figure 5. The reaction was observed to proceed via an oxocarbenium-like transition state; unsurprising, given that this is the same transition state structure typical of wild-type glycoside hydrolases,⁵² although in this case a nitrogen atom is substituted for the more typical oxygen atom. The reaction mechanism takes place in a concerted manner, with none of the relevant bond lengths changing significantly earlier than the others in either the forward or reverse directions, and a single energy barrier is observed.

One potential explanation for the poor reaction efficiency observed for this enzyme is that there are no nearby residues or water molecules to stabilize the departure of the azide group. Instead, the product state azide ion is left unbound in the active site cleft, presumably until it is able to diffuse into the bulk solvent (although this was not observed during our simulations, which had limited sampling time and a limited QM region that could prevent such an observation). In this light, it is not surprising that the rescue of fucosidase activity with the addition of free azide was not observed in the *TmAfc* D224S mutant,¹¹ as there is already limited space for the azide ion with a glycine at position 224; the relatively bulky serine side chain would restrict the azide group's access to the substrate. Based on this reasoning, a possible target for engineering this enzyme is via mutation of Met-225, as this is a bulky, hydrophobic side chain in prime position for interacting with the azide and potentially stabilizing its departure with a positive charge.

4 Conclusions

We investigated the mechanism of oligosaccharide synthesis in the reaction between 1-azido- β -L-Fucose (1AF) and 4-nitrophenyl β -D-xylopyranoside (4NX) via the glycosynthase *T. maritima* α -L-fucosidase (*TmAfc*) D224G. We propose that the reaction proceeds in a single endothermic step via an oxocarbenium-like transition state, wherein the role of the mutant residue Gly-224 is solely to provide room for the leaving azide group. Experimental results indicated that the α -1,3 product is produced in a slightly lower quantity compared to the α -1,4 product, and that fucosidase activity in the D224S mutant could not be rescued with the addition of free azide. Our results explain both of these observations, and provide new information for use in designing and engineering *TmAfc* and other glycoside hydrolases for improved glycosynthetic activity.

Conflicts of interest

There are no conflicts to declare.

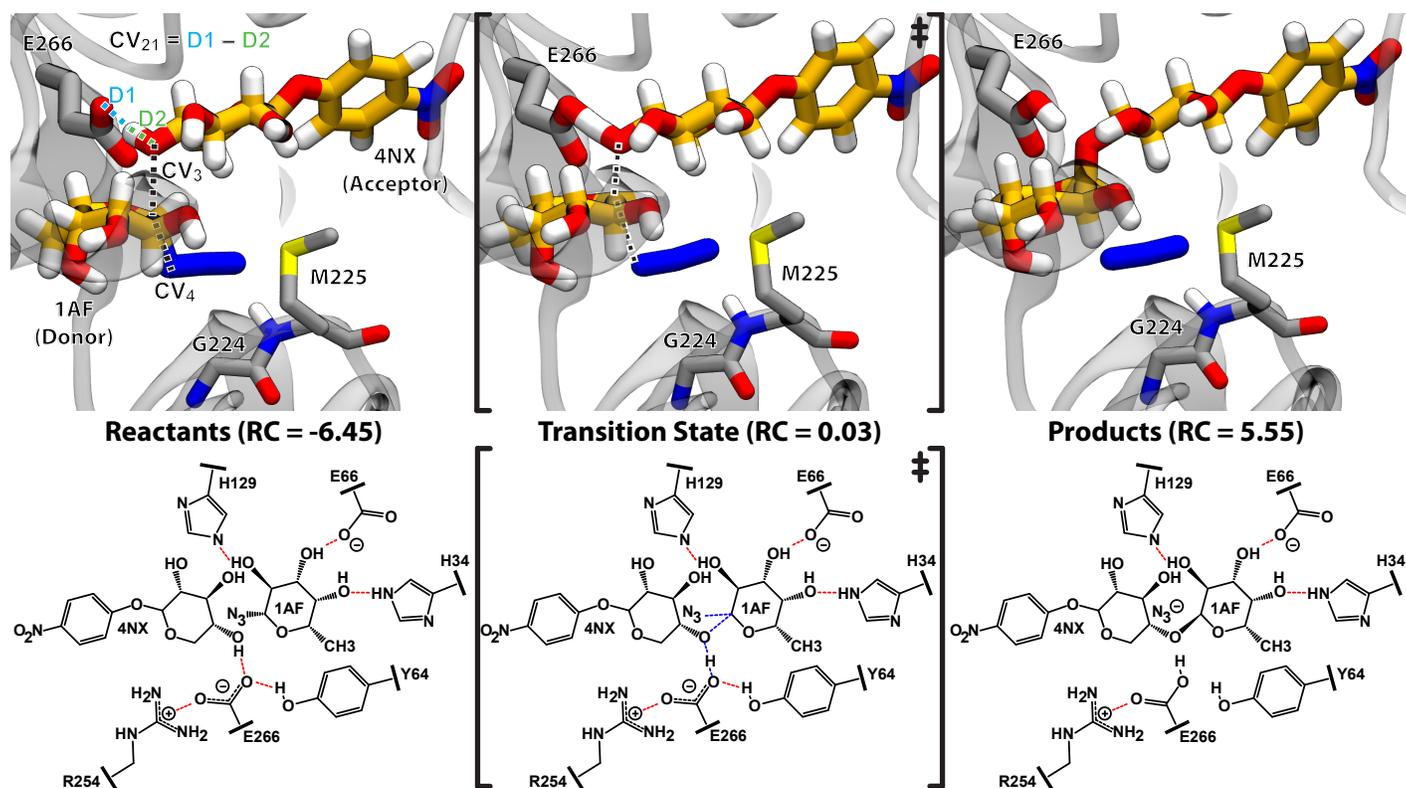


Fig. 5 Snapshots and schematics of a representative transition path for the α -1,4 reaction. For this figure, a transition pathway from late in the aimless shooting procedure and with a relatively high acceptance ratio preceding it was chosen to ensure maximum decorrelation from the initial configuration. Red dashes in the schematic representations indicate hydrogen bonding interactions. Dashed lines in the reactant state snapshot (at top-left) indicate the CVs constituting the reaction coordinate obtained with likelihood maximization. Dashed lines in the transition state snapshot (top-center) help delineate the oxocarbenium-like transition state structure, where the catalytic hydrogen is caught between the two oxygen atoms and the resulting partial charge on the donor O4 is compensated by an elongated bond between the anomeric carbon and the azide group. These intermediate bonds are represented with blue dashes in the corresponding schematic. Finally, in the product state the azide is completely dissociated from the fucose and the new glycosidic bond is formed as the donor O4 bonds fully with the catalytic Glu-266 residue. Residues Met-225 and Gly-224 are also shown in the snapshots to illustrate the molecular context around the azide group, whereas various hydrogen bonding residues are shown in the schematics to depict the stabilization of the donor and the catalytic residue.

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE),⁵³ which is supported by National Science Foundation grant number ACI-1548562. The authors thank Dr. Shishir Chundawat, Rutgers, The State University of New Jersey, for helpful discussions, as well as Dr. Beatrice Cobucci-Ponzano, National Research Council of Italy.

Notes and references

- 1 A. Varki, R. D. Cummings, J. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger, *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 3rd edn, 2017.
- 2 A. Varki, *Glycobiology*, 1993, **3**, 97–130.
- 3 P. H. Seeberger, *Chem. Soc. Rev.*, 2008, **37**, 19–28.
- 4 P. M. Danby and S. G. Withers, *ACS Chem. Biol.*, 2016, **11**, 1784–1794.
- 5 L.-X. Wang and B. G. Davis, *Chem. Sci.*, 2013, **4**, 3381–3394.
- 6 D. Lloyd and C. S. Bennett, *Chem. - Eur. J.*, 2018, **24**, 7610–7614.
- 7 M. Panza, S. G. Pistorio, K. J. Stine and A. V. Demchenko, *Chem. Rev.*, 2018, **118**, 8105–8150.
- 8 Z. Armstrong and S. G. Withers, *Biopolymers*, 2013, **99**, 666–674.
- 9 S. M. Hancock, M. D. Vaughan and S. G. Withers, *Curr. Opin. Chem. Biol.*, 2006, **10**, 509–519.
- 10 K. Schmörlzer, M. Lemmerer and B. Nidetzky, *Biotechnol. Bioeng.*, 2018, **115**, 545–556.
- 11 B. Cobucci-Ponzano, F. Conte, E. Bedini, M. M. Corsaro, M. Parrilli, G. Sulzenbacher, A. Lipski, F. Dal Piaz, L. Lepore, M. Rossi and M. Moracci, *Chem. Biol.*, 2009, **16**, 1097–1108.
- 12 L. F. Mackenzie, Q. Wang, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 1998, **120**, 5583–5584.
- 13 B. Cobucci-Ponzano, A. Strazzulli, M. Rossi and M. Moracci, *Adv. Synth. Catal.*, 2011, **353**, 2284–2300.
- 14 W. Huang, J. Giddens, S.-Q. Fan, C. Toonstra and L.-X. Wang, *J. Am. Chem. Soc.*, 2012, **134**, 12308–12318.
- 15 J.-H. Shim, H.-M. Chen, J. R. Rich, E. D. Goddard-Borger and S. G. Withers, *Protein Eng., Des. Sel.*, 2012, **25**, 465–472.
- 16 Y. Zhang, S. Yan and L. Yao, *Theor. Chem. Acc.*, 2013, **132**,

- 1–10.
- 17 J. Wang, S. Pengthaisong, J. R. K. Cairns and Y. Liu, *Biochim. Biophys. Acta, Proteins Proteomics*, 2013, **1834**, 536–545.
- 18 P. T. Vanhooren and E. J. Vandamme, *J. Chem. Technol. Biotechnol.*, 1999, **74**, 479–497.
- 19 P. A. Prieto, *Adv. Nutr.*, 2012, **3**, 456S–464S.
- 20 J. T. Smilowitz, C. B. Lebrilla, D. A. Mills, J. B. German and S. L. Freeman, *Annu. Rev. Nutr.*, 2014, **34**, 143–169.
- 21 D. L. Ackerman, R. S. Doster, J.-H. Weitkamp, D. M. Arono, J. A. Gaddy and S. D. Townsend, *ACS Infect. Dis.*, 2017, **3**, 595–605.
- 22 H.-J. Wu, C.-W. Ho, T.-P. Ko, S. D. Popat, C.-H. Lin and A. H.-J. Wang, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 337–340.
- 23 G. Sulzenbacher, C. Bignon, T. Nishimura, C. A. Tarling, S. G. Withers, B. Henrissat and Y. Bourne, *J. Biol. Chem.*, 2004, **279**, 13119–13128.
- 24 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comp. Chem.*, 2004, **25**, 1157–1174.
- 25 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, C. R. Daniels, B. L. Foley and R. J. J. Woods, *J. Comp. Chem.*, 2008, **29**, 622–655.
- 26 R. Rüger, A. Yakovlev, P. Philipsen, S. Borini, P. Melix, A. F. Oliveira, M. Franchini, T. Soini, M. de Reus, M. G. Asl, D. McCormack, S. Patchkovskii and T. Heine, *ADF DFTB 2017*, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, <http://www.scm.com>.
- 27 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B*, 1998, **58**, 7260–7268.
- 28 A. T. P. Carvalho, P. A. Fernandes and M. J. Ramos, *Int. J. Quantum Chem.*, 2007, **107**, 292–298.
- 29 S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comp. Chem.*, 1986, **7**, 230–252.
- 30 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16, Revision A.03*, Gaussian, Inc., Wallingford CT, 2016.
- 31 F. J. Devlin, J. W. Finley, P. J. Stephens and M. J. Frisch, *J. Phys. Chem.*, 1995, **99**, 16883–16902.
- 32 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 33 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
- 34 S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.
- 35 A. D. McLean and G. S. Chandler, *J. Chem. Phys.*, 1980, **72**, 5639–5648.
- 36 R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650–654.
- 37 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.
- 38 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 39 D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, *Amber 16*, University of California, San Francisco, 2016.
- 40 T. A. Andrea, W. C. Swope and H. C. Andersen, *J. Chem. Phys.*, 1983, **79**, 4576–4584.
- 41 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comp. Phys.*, 1977, **23**, 327–341.
- 42 B. Peters, *Chem. Phys. Lett.*, 2012, **554**, 248–253.
- 43 G. T. Beckham and B. Peters, *Computational Modeling in Lignocellulosic Biofuel Production*, American Chemical Society, Washington, D.C., 2010, ch. 13, pp. 299–332.
- 44 B. Peters, N. E. R. Zimmermann, G. T. Beckham, J. W. Tester and B. L. Trout, *J. Am. Chem. Soc.*, 2008, **130**, 17342–17350.
- 45 H. B. Mayes, B. C. Knott, M. F. Crowley, L. J. Broadbelt, J. Ståhlberg and G. T. Beckham, *Chemical Science*, 2016, **7**, 5955–5968.
- 46 M. Elstner, *Theor. Chem. Acc.*, 2006, **116**, 316–325.
- 47 R. G. Mullen, J. E. Shea and B. Peters, *J. Chem. Theory Comput.*, 2015, **11**, 2421–2428.
- 48 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision B.01*, Gaussian, Inc., Wallingford

- CT, 2016.
- 49 S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
- 50 S. Miertuš and J. Tomasi, *Chem. Phys.*, 1982, **65**, 239–245.
- 51 J. L. Pascual-ahuir, E. Silla and I. Tuñon, *J. Comp. Chem.*, 1994, **15**, 1127–1138.
- 52 T. M. Gloster and G. J. Davies, *Org. Biomol. Chem.*, 2010, **8**, 305–320.
- 53 J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott and N. Wilkens-Diehr, *Comput. Sci. Eng.*, 2014, **16**, 62–74.