Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

Supplementary Information

The role of catalytic residue pK_a on the hydrolysis/transglycosylation partition in Family 3 β-glucosidases

Inacrist Geronimo,[a] Christina M. Payne[b] and Mats Sandgren[a]

[a] Department of Molecular Sciences, Swedish University of Agricultural Sciences, Box 7015, 750 07 Uppsala, Sweden

[b] Department of Chemical and Materials Engineering, University of Kentucky, Lexington, Kentucky 40506-0046, USA

I. Computational Method

A. System preparation

The initial structure of *Hj*Cel3A-GEI was built using crystal structures PDB IDs 3ZYZ (*Hj*Cel3A with glucose)[1] and 1IEW (barley β-D-glucan glucohydrolase isoenzyme Exo1 with covalently bound 2-deoxy-2-fluoro-α-D-glucopyranose).[2] Coordinates of the glycosyl moiety and side chains of D236, E441, D61, R125, K158, H159, Y204, and E128 were copied from the latter. Initial coordinates of glucose in the +1 site were obtained by superimposing the resulting structure with PDB ID 1IEX (Exo1 with thiocellobiose).[2] W286 of Exo1 (equivalent to W237 in *Hj*Cel3A) forms a stacking interaction with a glucose unit in PDB ID 1IEX. However, the original coordinates of W237, in which the side chain is swung inwards to the -1 site, were retained because using the W286 coordinates leads to a clash with the W37 side chain. Moreover, the study of Seidle et al. on the *Aspergillus niger* Family 3 β-glucosidase indicated that the tryptophan residue located next to the nucleophile is important to the binding stability of a sugar at the -1 site but has an insignificant effect on binding at positive sites.[3] The non-reducing end of cellobiose was positioned similarly at the +1 site as the glucose acceptor; doing so placed the reducing end in a stacking interaction with F260, though too far from D370 to form a hydrogen bond.

Each system was minimized for 100 steps using the steepest descent (SD) method followed by 1000 steps using the adopted basis Newton-Raphson (ABNR) method. The system was then solvated in a truncated octahedral box of TIP3P water, with a buffer distance of 12 Å between each wall and the closest atom in each direction. Fifty-three sodium and 54 chloride ions were added to make a final solution with an ionic strength of 0.1 M. The system size is approximately 98,000 atoms.

B. Equilibration

SD minimization of the system was performed for 1000 steps with the protein and sugar (glucosyl moiety and glucose or cellobiose acceptor) fixed. Residues within 5 Å of the sugar were released, while the sugar and remainder of the protein were restrained with a harmonic force constant of 10.0 kcal/mol/Å, and 100 steps of SD minimization were performed. This was followed by an additional 100 minimization steps, with only the sugar restrained. Finally, the entire system was minimized for 1000 steps using the SD method, followed by 1000 steps using the ABNR method.

The protein and sugar were again restrained with a harmonic force constant of 10.0 kcal/mol/Å, and the system was gradually heated for 20 ps from 100 to 300 K in steps of 5 K every 0.5 ps. The system was equilibrated at the optimum pH of the enzyme (pH 5.0) for 100 ps without restraints using constant pH molecular dynamics. Conformational dynamics was performed in the *NPT* ensemble, using a modified Hoover thermostat[4] to maintain the temperature at 300 K and Langevin piston[5] to keep the pressure at 1 atm. Long-range electrostatics were calculated using the Particle Mesh Ewald method[6] with a 6th order b-spline, Gaussian parameter width of 0.340 Å, and mesh size of 120 x 120 x 120 Å (subsequently reduced to 108 Å after density equilibration). Van der Waals interactions were switched off between 10 and 12 Å. Bonds involving hydrogen were constrained using the SHAKE algorithm[7] and a 2-fs time step was used.

II. Constant pH Molecular Dynamics Results

The temperature, energy, and density plots for the independent simulations demonstrate the stability of the three systems (Fig. S1). In the *Hj*Cel3A-GEI and *Hj*Cel3A-GEI-glucose models, loop A (residues 251–267) at the active site entrance and loop B (residues 353–393) near the active site loop (loop C, residues 438–460) exhibited the highest flexibility. Residues have comparatively lower RMSF values in the *Hj*Cel3A-GEI-cellobiose model (Fig. S2).

A. *Hj*Cel3A-GEI

The increase in the backbone RMSD over time (Fig. S3) is caused by the high flexibility of domains 2 (residue 317–522) and 3 (residues 580–713) (Fig. S2). However, domain 1 (residues 7–300, excluding loop A), which contains most of the active site residues, and loop C stabilized after about 4 ns (Fig. S3). The backbone RMSD of the residues lining the -1 and +1 sites, including W37, V43, D61, D67,

L110, R125, E128, K158, H159, E166, R169, M201, Y204, D236, W237, S384, E441, and Y443, were also calculated. Fig. S4 shows that local conformational change in the substrate pocket is caused by the high mobility of W37 at the entrance (black line) and/or breaking of the hydrogen bond network around E441 involving R125, E128, E166, R169, Y204, S384, and Y443 (red line). In simulation 1, the E441–R169 interaction broke within 1 ns, with Y204 moving away from the -1 site along with R169 to form a cation-π interaction. The breaking of the E441 hydrogen bond network occurred at ~5 ns, after which the E441 side chain moved away from the glycosyl moiety (Fig. S8). In simulation 2, the C–C_α–C_β–C_γ dihedral of Y204 began to rotate at ~6 ns from 160° to -100°, bringing the side chain closer to R169 and E441 (Fig. S14A). The C_α–C_β–C_γ–C_δ dihedral of E441 subsequently rotated from -170° to 70° so that the side chain shifted away from the glycosyl moiety and interacted with the E128 side chain (Fig. S14B). These changes account for the larger backbone RMSDs observed for this simulation (Figs. S3 and S4). In simulation 3, R169 and Y204 moved away from the +1 and -1 sites at ~4 ns, with the latter's C–C_α–C_β–C_γ dihedral also changing to -90°. E441 moved away from the glycosyl moiety at ~7 ns (Fig. S8). The plot of deprotonated fraction for each simulation (Figs. S9-S11) shows that sampling of the protonated state by E441 coincides with the breaking of its hydrogen bond network and movement of its side chain away from the glycosyl moiety (e.g., ~6 ns onwards in simulation 2). At the -1 site, D61 is predominantly deprotonated while H159 is predominantly doubly protonated. At the putative +2 site, D370 is predominantly deprotonated in two of the three simulations and formed hydrogen bonds with R67 and Y443, both located at the +1 site. In simulation 1, the residue was initially deprotonated as it was intermittently hydrogen bonded only to R67. As for the two glutamate residues close to E441, E166 is predominantly deprotonated while E128 sampled both states. The E128 protonated state occurred when its interaction with R125 is broken and/or it interacts directly with E441.

B. *Hj*Cel3A-GEI with glucose acceptor

The backbone RMSD plots (Figs. S3 and S4) show that structural stability was attained only for simulation 2, and a conformational change at the last half of the other two simulations occurred in domain 1. In simulations 1 and 3, Y204 side chain eventually rotated (as described above) to form a cation-π interaction with R169, breaking its hydrogen bond with the W237 backbone O atom. This also disrupted the hydrogen bond network around E441, evidenced by the increase in backbone RMSD at ~8 ns and ~5 ns for simulations 1 and 3 (Fig. S4). In simulation 2, both R169 and the glucose acceptor moved away from the +1 site at ~6 ns, while Y204 remained. Thus, the active site conformation did not change significantly during the simulation (Fig. S4). The glucose acceptor was also very mobile in simulation 3. After its hydrogen bond with E441 was broken at ~5 ns, glucose left the +1 site but remained in the substrate pocket through hydrogen bonding with either D370, R67, or the glycosyl moiety. E441 was predominantly deprotonated (Figs. S17 and S18) except in simulation 3, where its hydrogen bond with the glucose acceptor was broken early and those with R125 and S384 were frequently disrupted (Fig. S19). At the -1 site, D61 is predominantly deprotonated while H159 is predominantly doubly protonated, as in the absence of glucose. In simulation 2, H159 was briefly deprotonated (2–3 ns, Fig. S18) and became the hydrogen bond acceptor to the glycosyl moiety. At the putative +2 site, D370 is also predominantly deprotonated, except for a brief period at ~2 ns of simulation 2 (Fig. S18), when both R67 and Y443 were temporarily replaced by Y68 as the hydrogen bond donor. The protonation states of E128 and E166 are discussed in the main text.

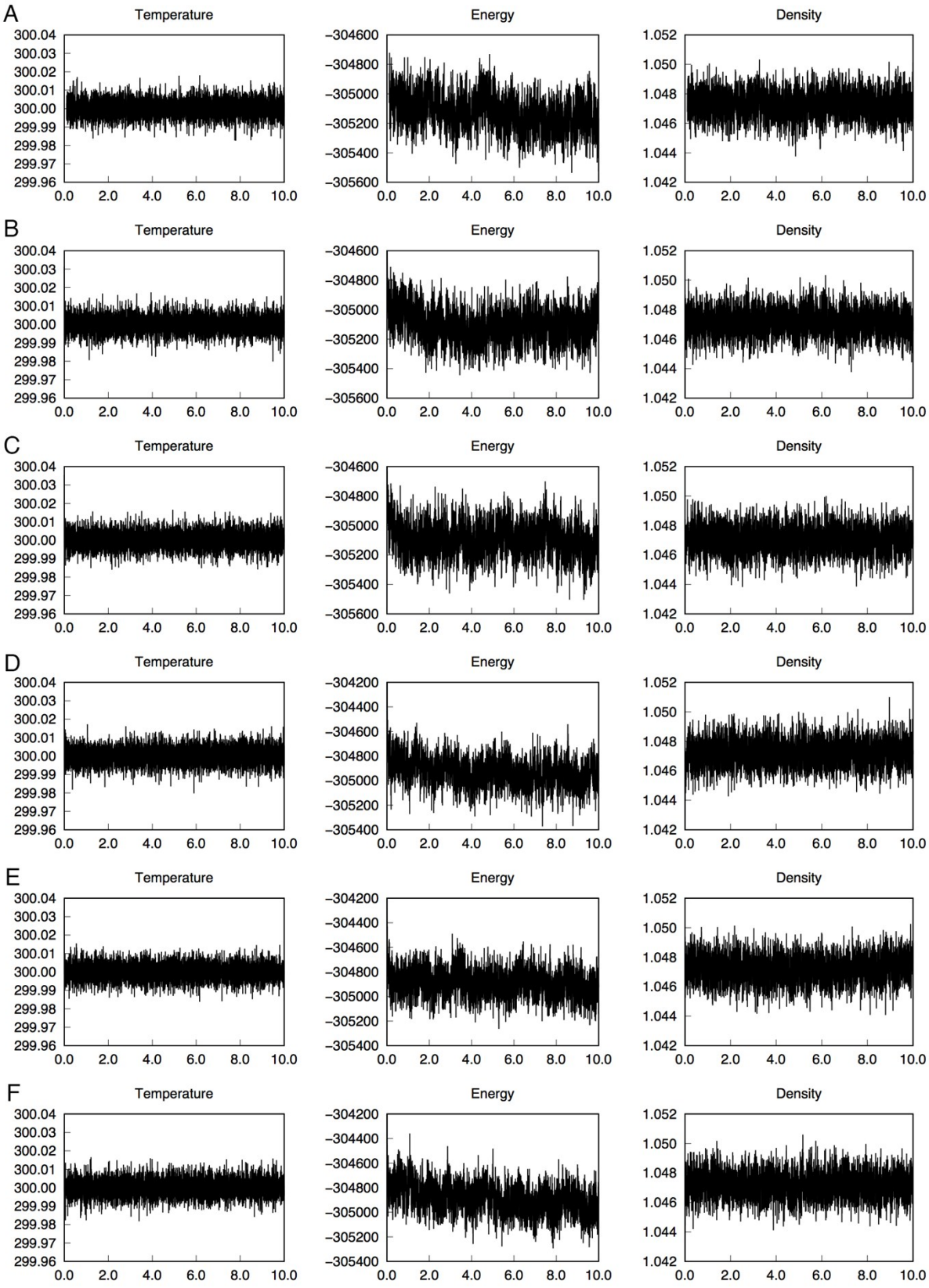
C. *Hj*Cel3A-GEI with cellobiose acceptor

Residues were less flexible, likely due to the presence of another glucose unit in the +2 site. The protein structure was consequently more stable for a longer simulation period compared to *Hj*Cel3A-GEI with a glucose acceptor (Figs. S3 and S4). Unbinding of cellobiose from the +1/+2 sites was observed only at ~9 ns of simulation 3. As in the other transglycosylation model, the R169 and Y204 side chains were at their positions at the -1/+1 sites in simulation 1 and most part of simulation 3 (R169 eventually moved away while the Y204 C–C_α–C_β–C_γ dihedral rotated at ~8 ns). In simulation 2, however, the two residues moved away to form a cation-π interaction. Nevertheless, the non-reducing end of cellobiose remained in the +1 site because of hydrogen bond interactions with R67 and the glycosyl moiety. The reducing end of cellobiose at the +2 site interacted mainly with F260; however, it also formed a hydrogen bond with N238 in simulation 1 and D370 in simulation 2. E441

was predominantly deprotonated, except at ~7 ns of simulation 3, when its hydrogen bonds with R125 and/or S384 were temporarily broken. At the -1 site, D61 is predominantly deprotonated while H159 also sampled the singly protonated state in simulations 1 and 2 upon the breaking of its interaction with the glycosyl moiety. At the putative +2 site, D370 is also predominantly deprotonated, except at ~3 ns of simulation 1, when it did not form a hydrogen bond with any of its neighboring residues. E166 was predominantly deprotonated as in the previous models (Figs. S20-S22). The protonation state of E128 is discussed in the main text.

Movie S1. Simulation of *Hj*Cel3A-GEI showing dynamic protonation of E441, E128, E166, and D61. Snapshots where both carboxylate oxygen atoms are protonated represent the mixed tautomeric state, a consequence of using a continuous coordinate to switch between protonated and deprotonated states. Since the mixed state is unphysical, it is disregarded in the deprotonated fraction calculation. Water molecules within the first solvation shell of E441 are also shown. The simulation shows R169 and Y204 moving away from their original positions near E441 to form a cation-π interaction.

Movie S2. Simulation of *Hj*Cel3A-GEI with a glucose acceptor showing dynamic protonation of E441, E128, E166, and D61. Snapshots where both carboxylate oxygen atoms are protonated represent the mixed tautomeric state, a consequence of using a continuous coordinate to switch between protonated and deprotonated states. Since the mixed state is unphysical, it is disregarded in the deprotonated fraction calculation. Y204 did not form a cation-π interaction with R169 because of its hydrogen bond with W237. Several snapshots show direct interaction between E441 and protonated E128, due to the breaking of the latter's hydrogen bond with R125. Glucose is held by hydrogen bonds to E441 and R169 and positioned for deprotonation at the HO6 hydroxyl group. The glucose left the +1 site toward the end of each of the three 10-ns simulations performed.


A
Temperature
300.04
300.03
300.02
300.01
300.00
299.99
299.98
299.97
299.96
0.0 2.0 4.0 6.0 8.0 10.0
Energy
−304600
−304800
−305000
−305200
−305400
−305600
0.0 2.0 4.0 6.0 8.0 10.0
Density
1.052
1.050
1.048
1.046
1.044
1.042
0.0 2.0 4.0 6.0 8.0 10.0
B
Temperature
300.04
300.03
300.02
300.01
300.00
299.99
299.98
299.97
299.96
0.0 2.0 4.0 6.0 8.0 10.0
Energy
−304600
−304800
−305000
−305200
−305400
−305600
0.0 2.0 4.0 6.0 8.0 10.0
Density
1.052
1.050
1.048
1.046
1.044
1.042
0.0 2.0 4.0 6.0 8.0 10.0
C
Temperature
300.04
300.03
300.02
300.01
300.00
299.99
299.98
299.97
299.96
0.0 2.0 4.0 6.0 8.0 10.0
Energy
−304600
−304800
−305000
−305200
−305400
−305600
0.0 2.0 4.0 6.0 8.0 10.0
Density
1.052
1.050
1.048
1.046
1.044
1.042
0.0 2.0 4.0 6.0 8.0 10.0
D
Temperature
300.04
300.03
300.02
300.01
300.00
299.99
299.98
299.97
299.96
0.0 2.0 4.0 6.0 8.0 10.0
Energy
−304200
−304400
−304600
−304800
−305000
−305200
−305400
0.0 2.0 4.0 6.0 8.0 10.0
Density
1.052
1.050
1.048
1.046
1.044
1.042
0.0 2.0 4.0 6.0 8.0 10.0
E
Temperature
300.04
300.03
300.02
300.01
300.00
299.99
299.98
299.97
299.96
0.0 2.0 4.0 6.0 8.0 10.0
Energy
−304200
−304400
−304600
−304800
−305000
−305200
−305400
0.0 2.0 4.0 6.0 8.0 10.0
Density
1.052
1.050
1.048
1.046
1.044
1.042
0.0 2.0 4.0 6.0 8.0 10.0
F
Temperature
300.04
300.03
300.02
300.01
300.00
299.99
299.98
299.97
299.96
0.0 2.0 4.0 6.0 8.0 10.0
Energy
−304200
−304400
−304600
−304800
−305000
−305200
−305400
0.0 2.0 4.0 6.0 8.0 10.0
Density
1.052
1.050
1.048
1.046
1.044
1.042
0.0 2.0 4.0 6.0 8.0 10.0

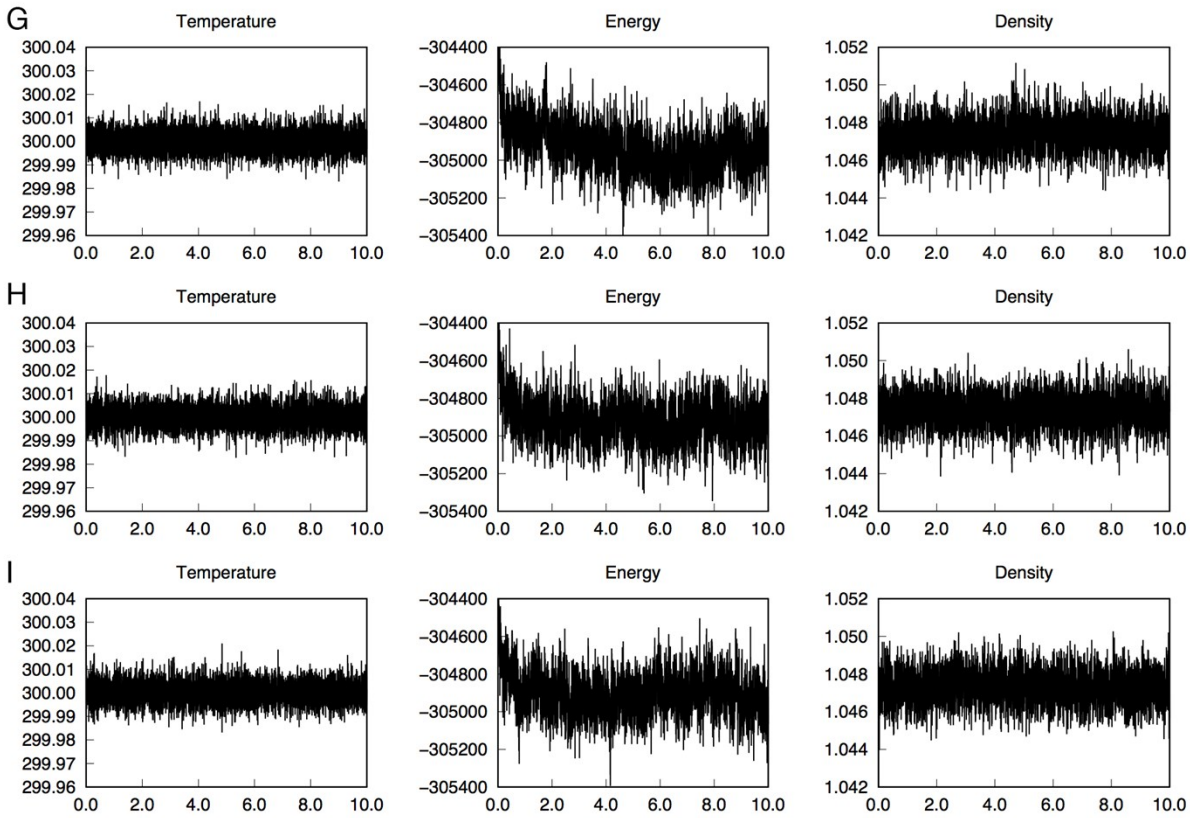


Fig. S1. Temperature (K), energy (kcal/mol), and density (g/cm^3) during the 10 ns-simulations. A–C. *Hj*Cel3A-GEI. D–F. *Hj*Cel3A-GEI with glucose acceptor. G–I. *Hj*Cel3A-GEI with cellobiose acceptor.

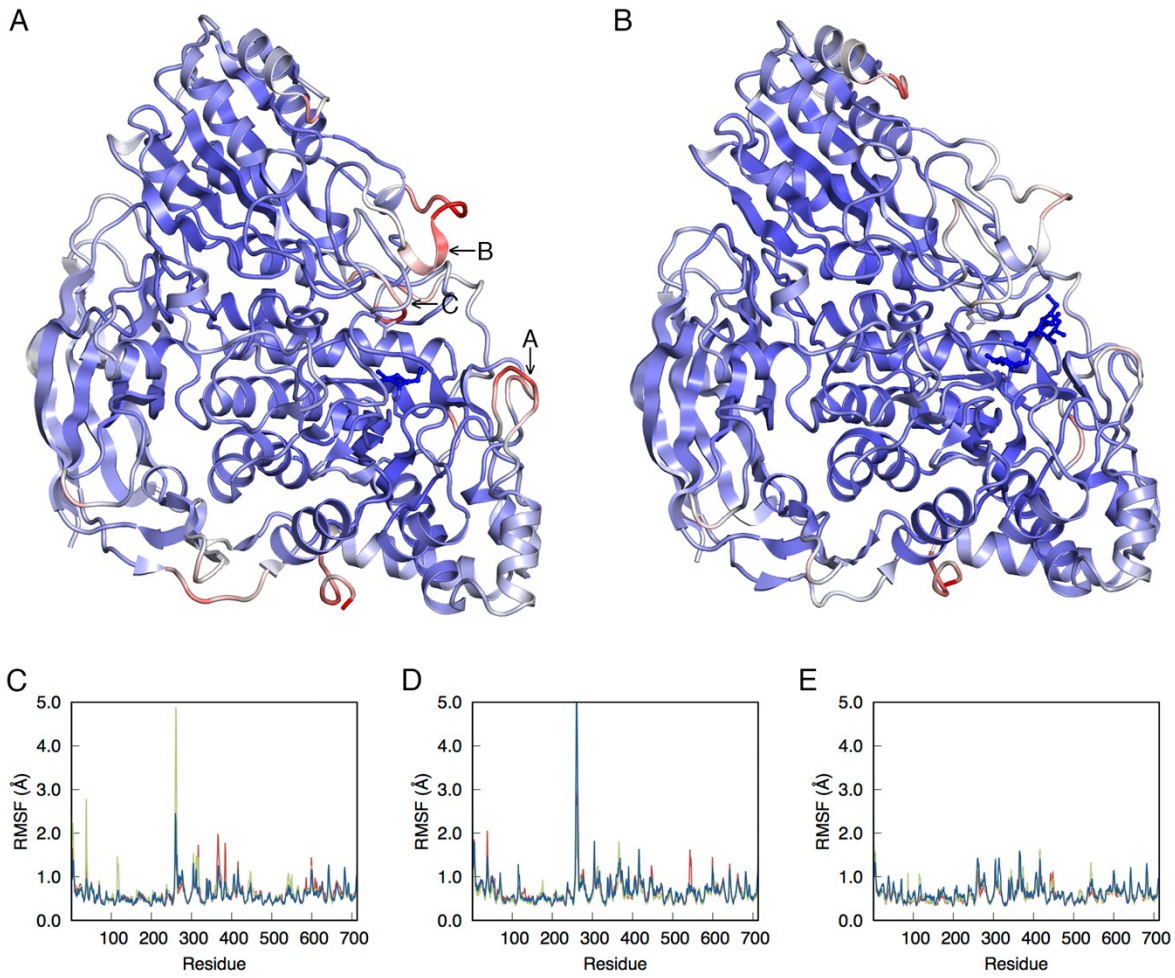


Fig. S2. Structures of A. *Hj*Cel3A-GEI and B. *Hj*Cel3A-GEI with cellobiose acceptor, with residues colored using a blue-white-red palette based on RMSF (lowest and highest values corresponding to blue and red, respectively). The corresponding figure for *Hj*Cel3A-GEI-glucose is not shown because it is similar to *Hj*Cel3A-GEI. The mobile loops around the active site, 251–267, 353–393, and 438–460, are labeled A, B, and C, respectively. C–E. Plots of residue RMSF values for *Hj*Cel3A-GEI, *Hj*Cel3A-GEI-glucose, and *Hj*Cel3A-GEI-cellobiose models, respectively.

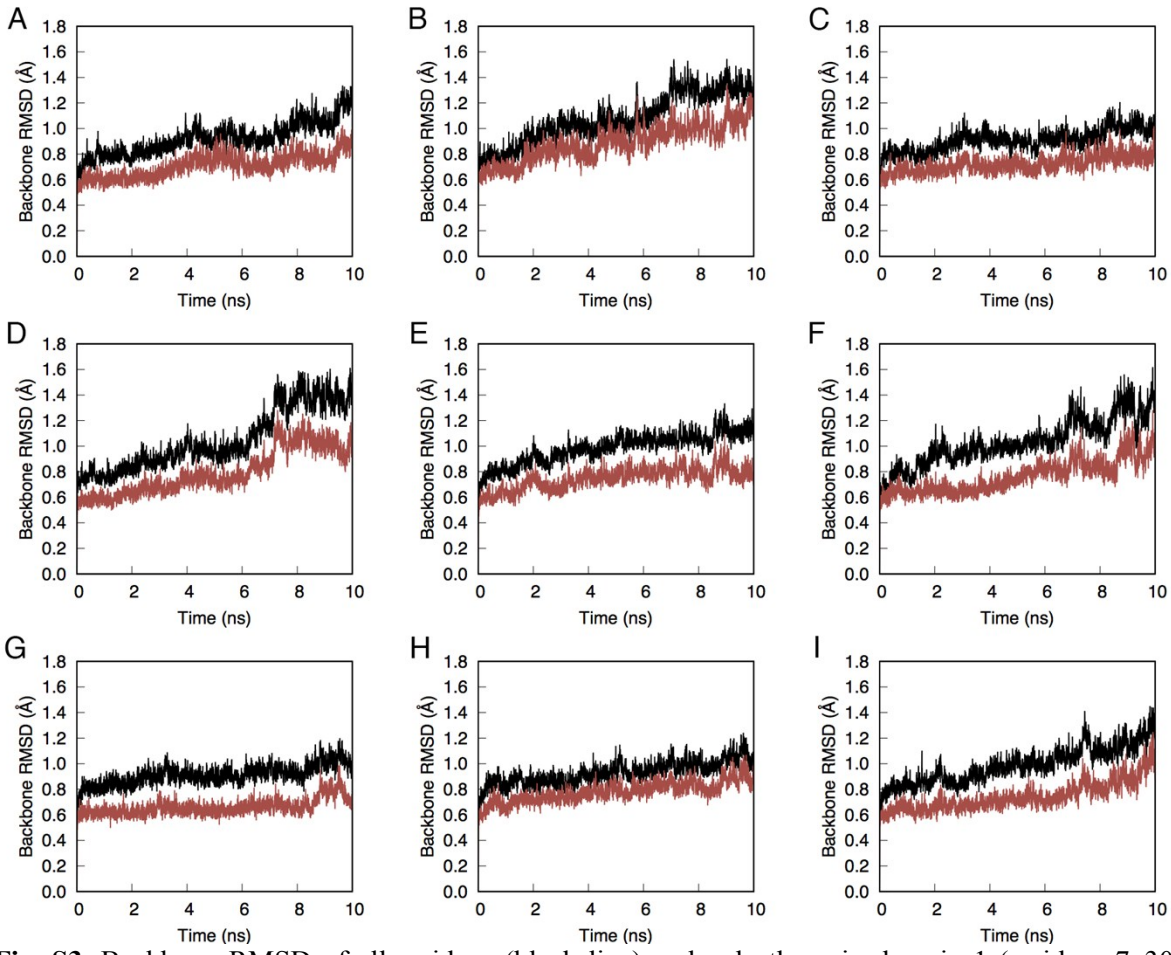


Fig. S3. Backbone RMSD of all residues (black line) and only those in domain 1 (residues 7–300, excluding loop A) and loop C (red line). A–C. *Hj*Cel3A-GEI. D–F. *Hj*Cel3A-GEI with glucose acceptor. G–I. *Hj*Cel3A-GEI with cellobiose acceptor.

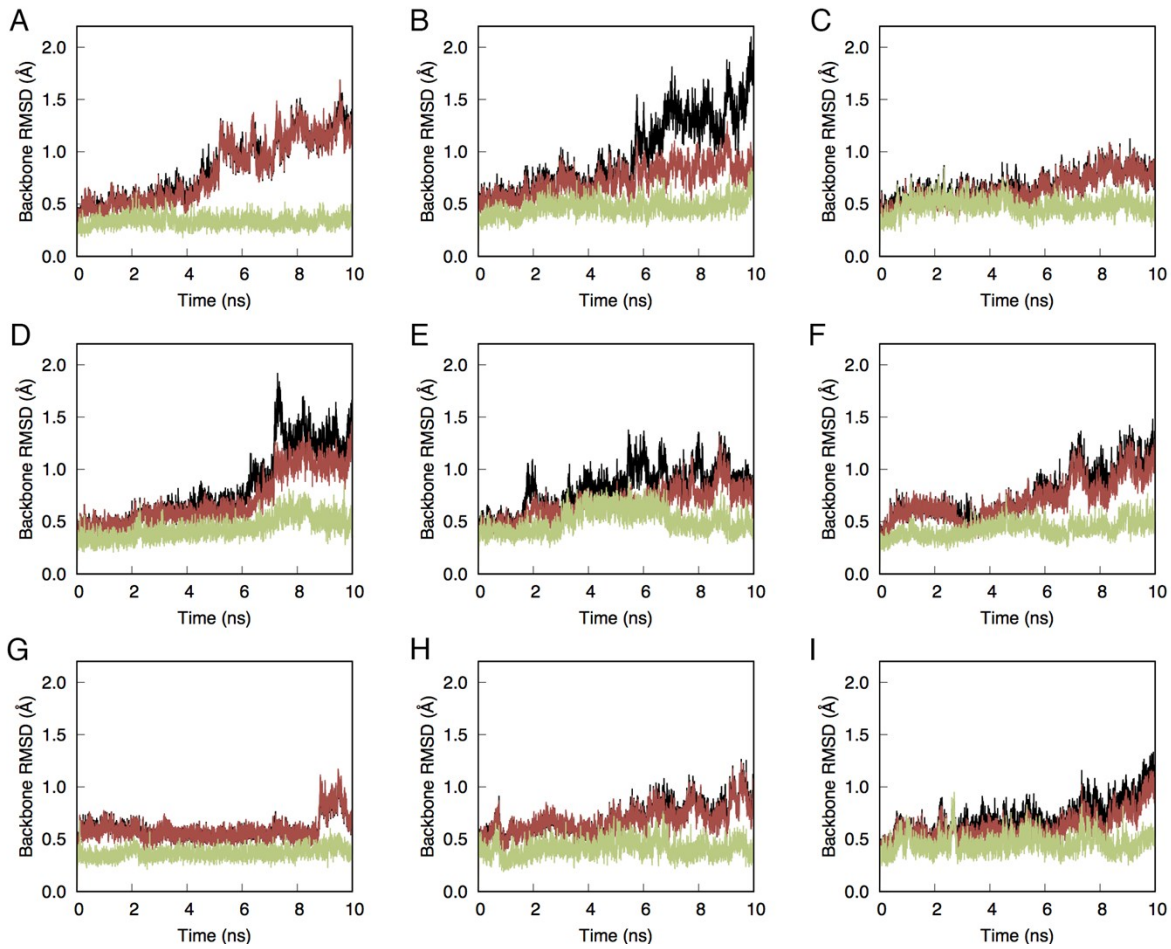


Fig. S4. Backbone RMSD of the substrate pocket residues W37, V43, D61, D67, L110, R125, E128, K158, H159, E166, R169, M201, Y204, D236, W237, S384, E441, and Y443 (black line). RMSDs were also calculated without W37 (red line) and without the residues participating in the E441 hydrogen bond network (green line). A–C. *Hj*Cel3A-GEI. D–F. *Hj*Cel3A-GEI with glucose acceptor. G–I. *Hj*Cel3A-GEI with cellobiose acceptor.

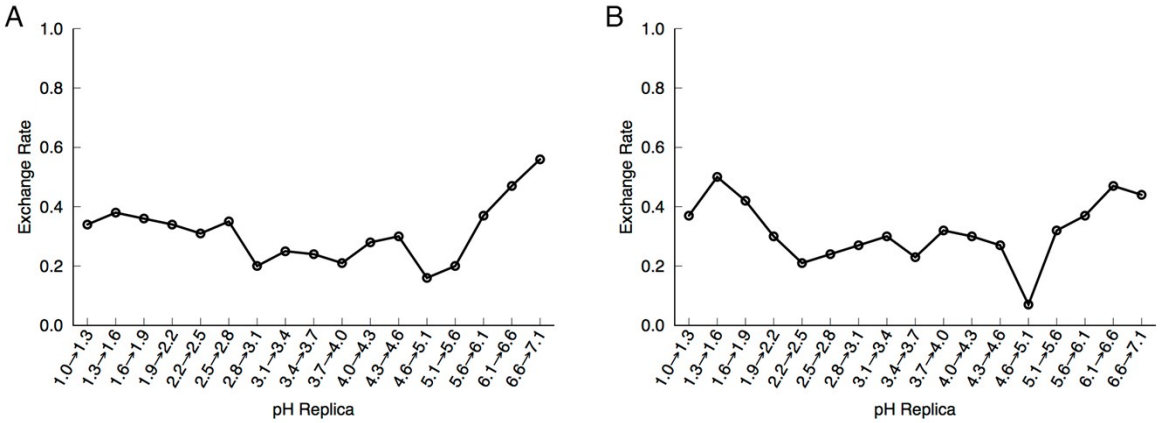


Fig. S5. Replica exchange rates of the pH-REX simulation. A. *Hj*Cel3A-GEI. B. *Hj*Cel3A-GEI with glucose acceptor.

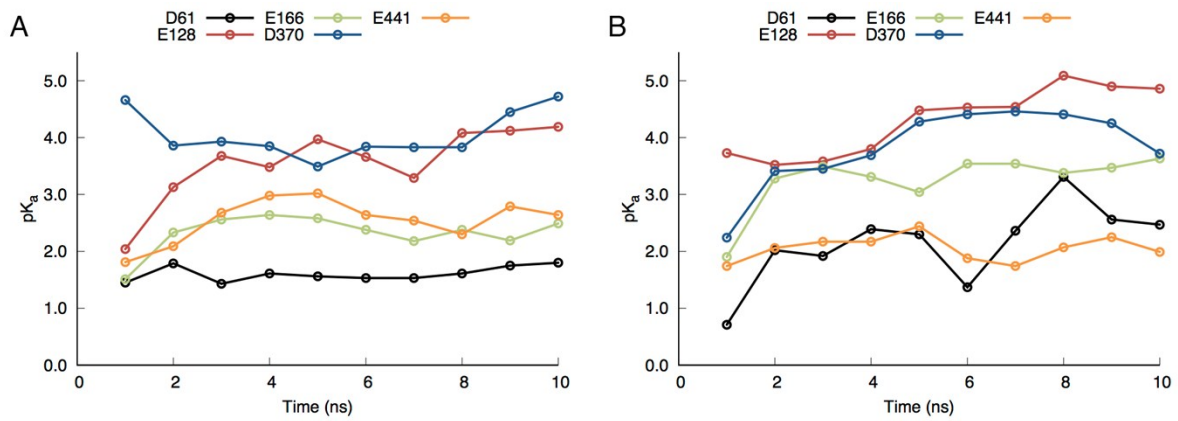


Fig. S6. pK_a of ionizable active site residues at each 1-ns block of pH-REX simulation. A. *Hj*Cel3A-GEI. B. *Hj*Cel3A-GEI with glucose acceptor.

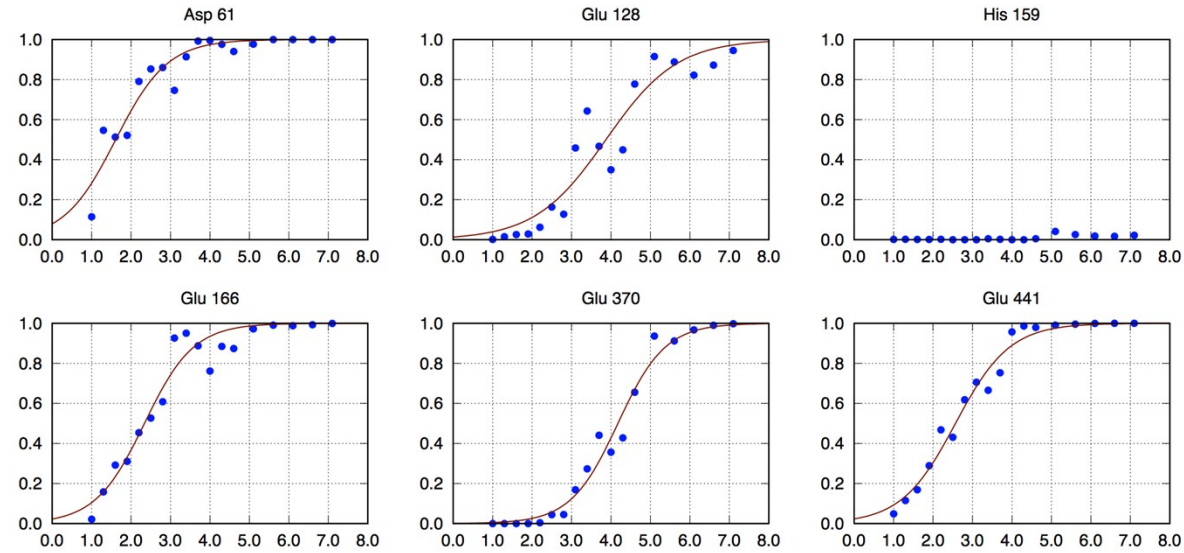


Fig. S7. Titration curves of ionizable active site residues in *Hj*Cel3A-GEI. The deprotonated fraction *S* (y-axis) at different pH (x-axis) (blue points) is fit to the generalized Henderson-Hasselbalch equation (red line). Each data point is the average from the last 5 ns of simulation.

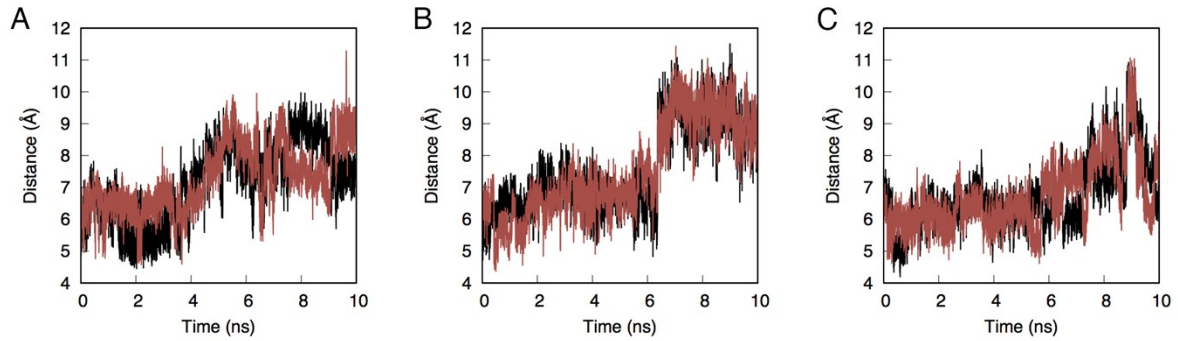


Fig. S8. Plots of E441:OE1–glucose:C1 (black line) and E441:OE2–glucose:C1 (red line) distances from the three *Hj*Cel3A-GEI simulations.

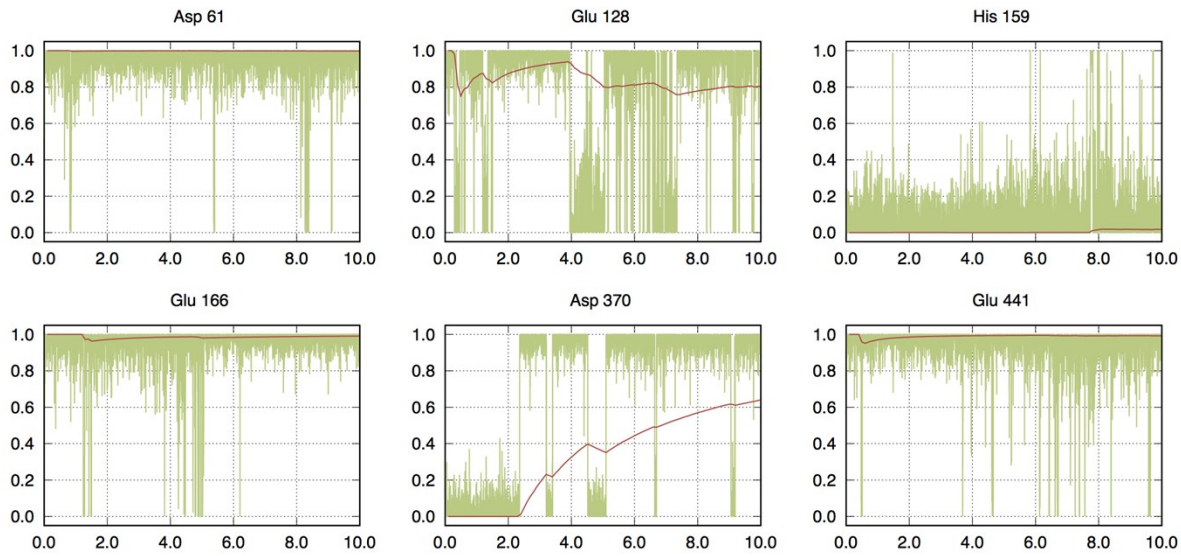


Fig. S9. Titration data from simulation 1 of *Hj*Cel3A-GEI. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

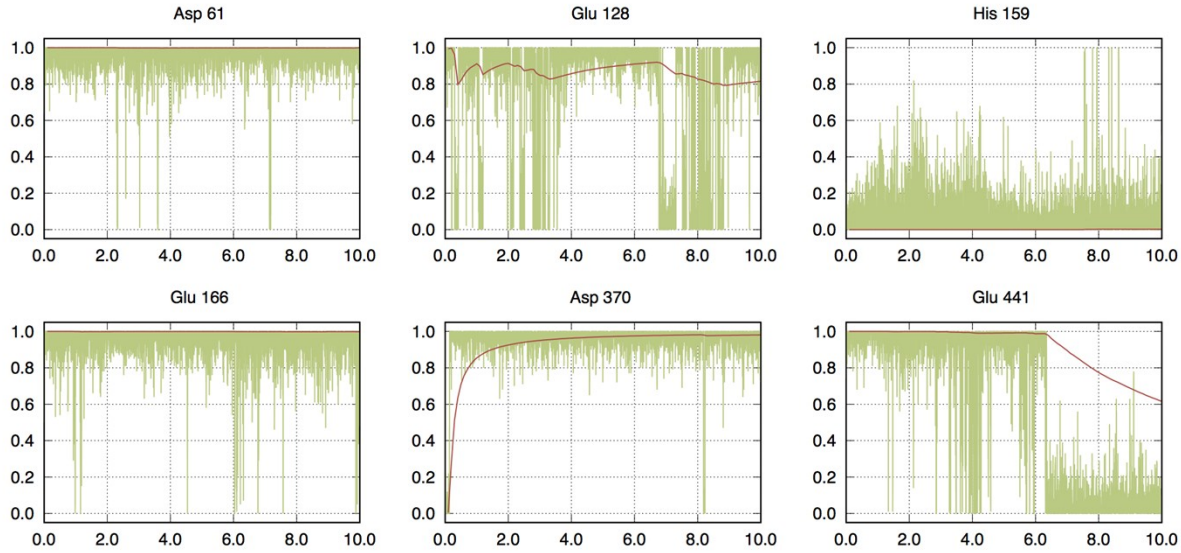


Fig. S10. Titration data from simulation 2 of *Hj*Cel3A-GEI. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

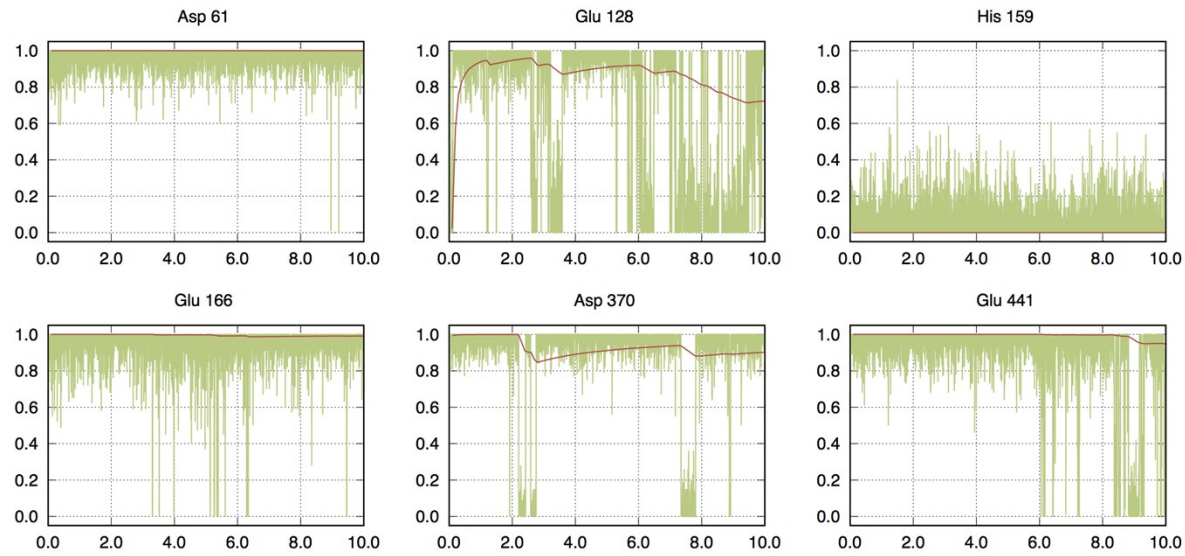


Fig. S11. Titration data from simulation 3 of *Hj*Cel3A-GEI. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

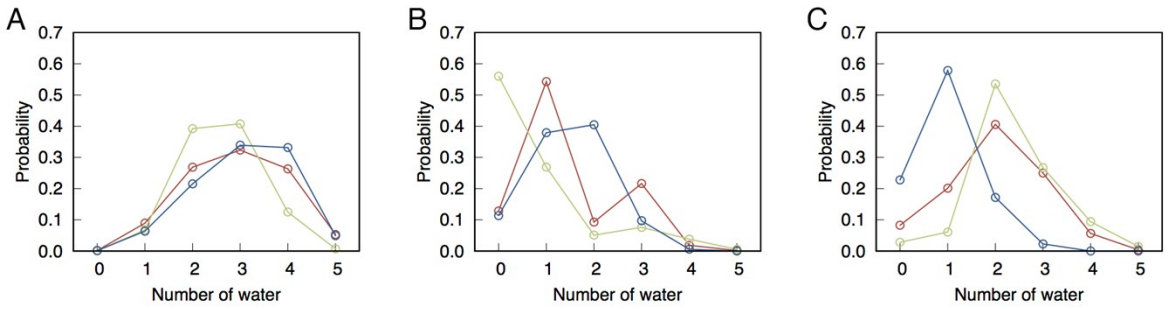


Fig. S12. Histogram of the number of water molecules within the first solvation shell of E441 from three simulations. A. *Hj*Cel3A-GEI. B. *Hj*Cel3A-GEI with glucose acceptor. C. *Hj*Cel3A-GEI with cellobiose acceptor.

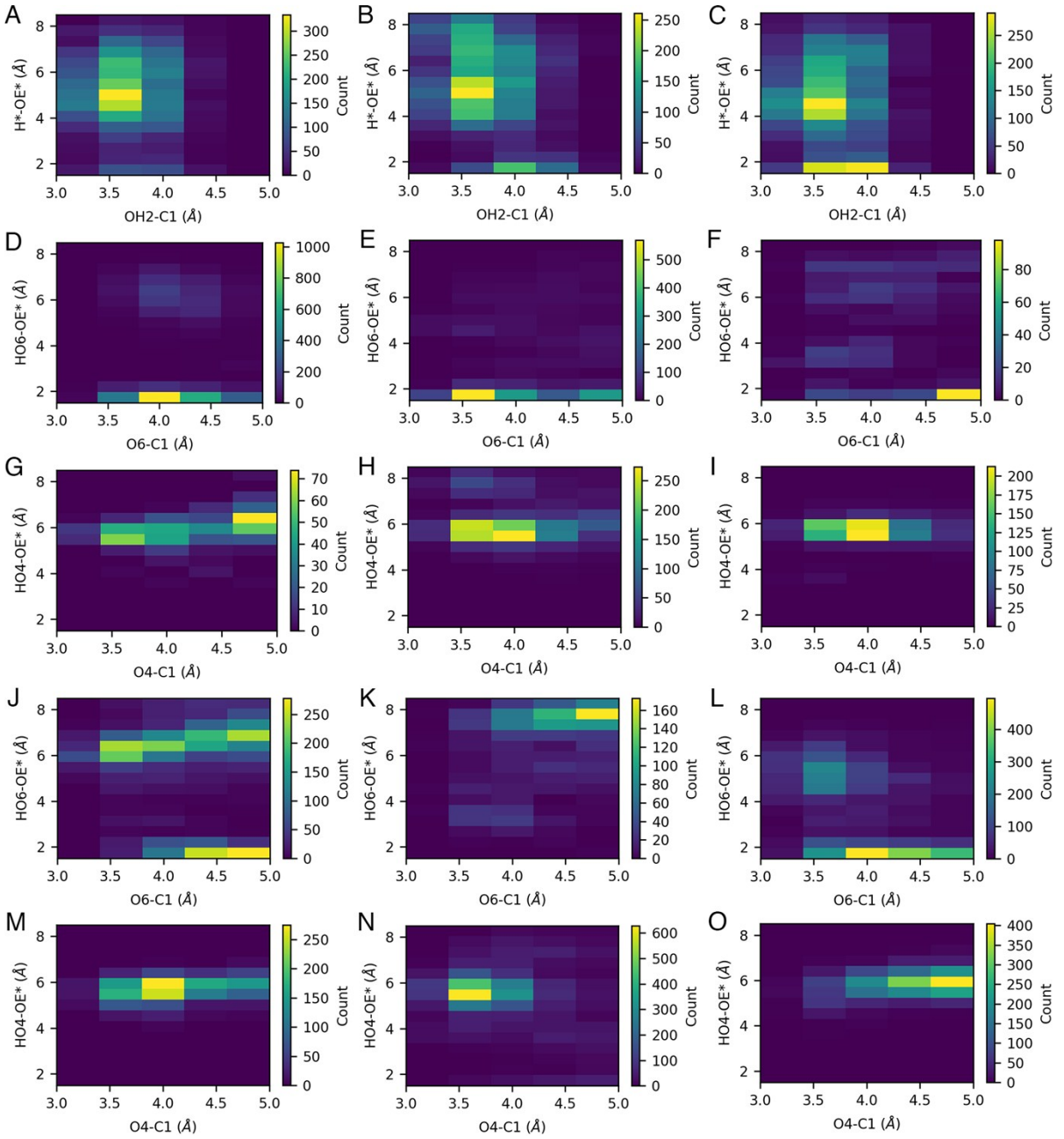


Fig. S13. 2D-Histograms of acceptor oxygen-anomeric carbon (C1) and acceptor hydrogen-E441 carboxylate oxygen (OE*) distances. Data collected from 10 ns of simulation (5000 trajectories). Only the shortest hydrogen-oxygen distance in each trajectory was tallied. A–C. *Hj*Cel3A-GEI. D–I. *Hj*Cel3A-GEI with glucose acceptor. J–O. *Hj*Cel3A-GEI with cellobiose acceptor.

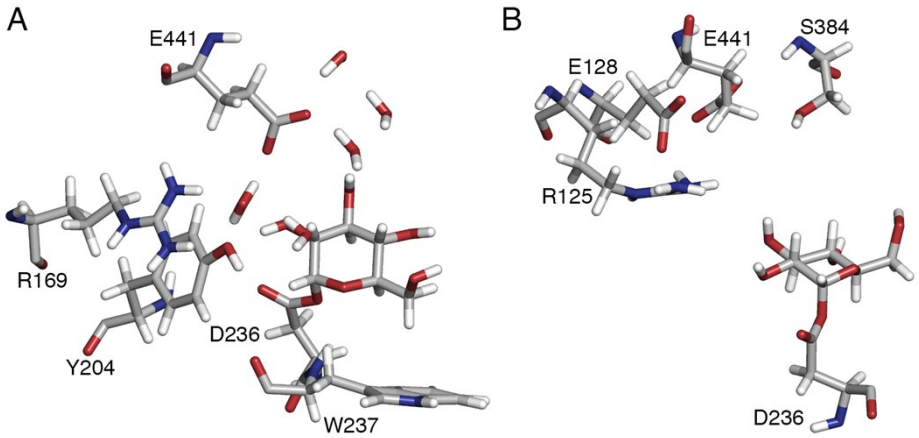


Fig. S14. Snapshots from the *Hj*Cel3A-GEI simulations. A. R169 and Y204 moved away from the +1/-1 sites to form a cation-π interaction. Water molecules in the first solvation shell of E441 are also shown. B. E441 side chain rotated to interact directly with E128.

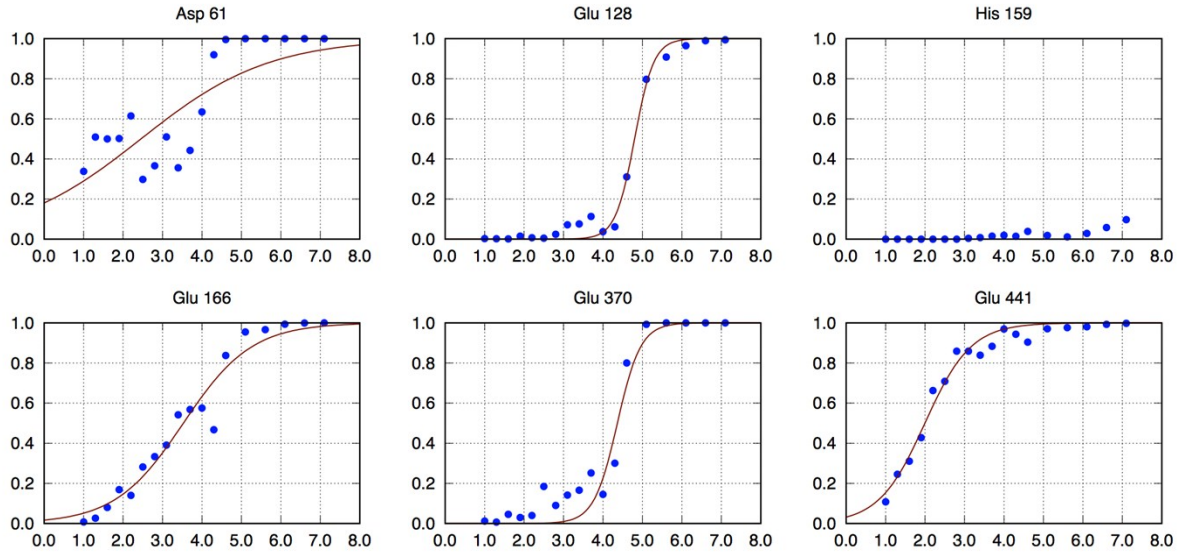


Fig. S15. Titration curves of ionizable active site residues in *Hj*Cel3A-GEI with glucose acceptor. The deprotonated fraction *S* (y-axis) at different pH (x-axis) (blue points) is fit to the generalized Henderson-Hasselbalch equation (red line). Each data point is the average from the last 5 ns of simulation.

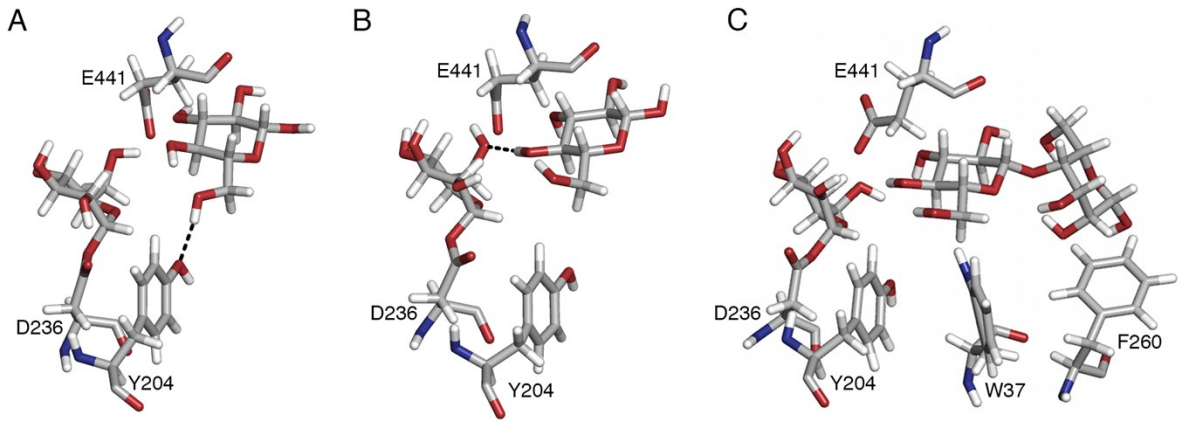


Fig. S16. A. Initial and B. equilibrated structures of *Hj*Cel3A-GEI with glucose acceptor. C. Equilibrated structure of *Hj*Cel3A-GEI with cellobiose acceptor. In A, the glucose O4 atom is positioned closer to the anomeric carbon (cellobiose product), while in B and C, it is the glucose O6 atom (gentiobiose/6-*O*-β-D-glucopyranosyl-4-*O*-β-D-glucopyranosyl-D-glucopyranose product).

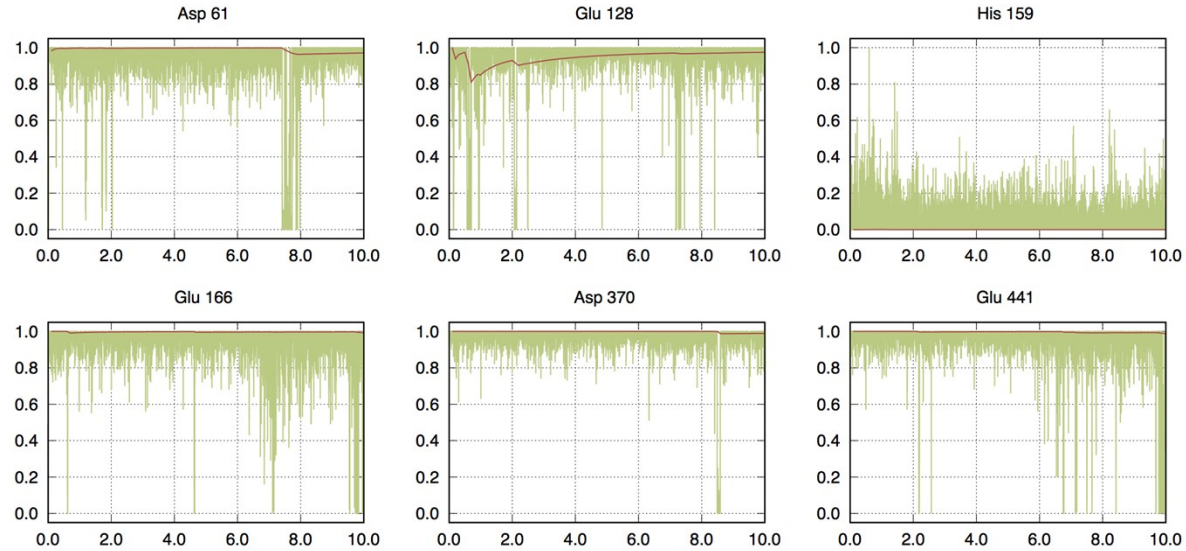


Fig. S17. Titration data from simulation 1 of *Hj*Cel3A-GEI with glucose acceptor. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

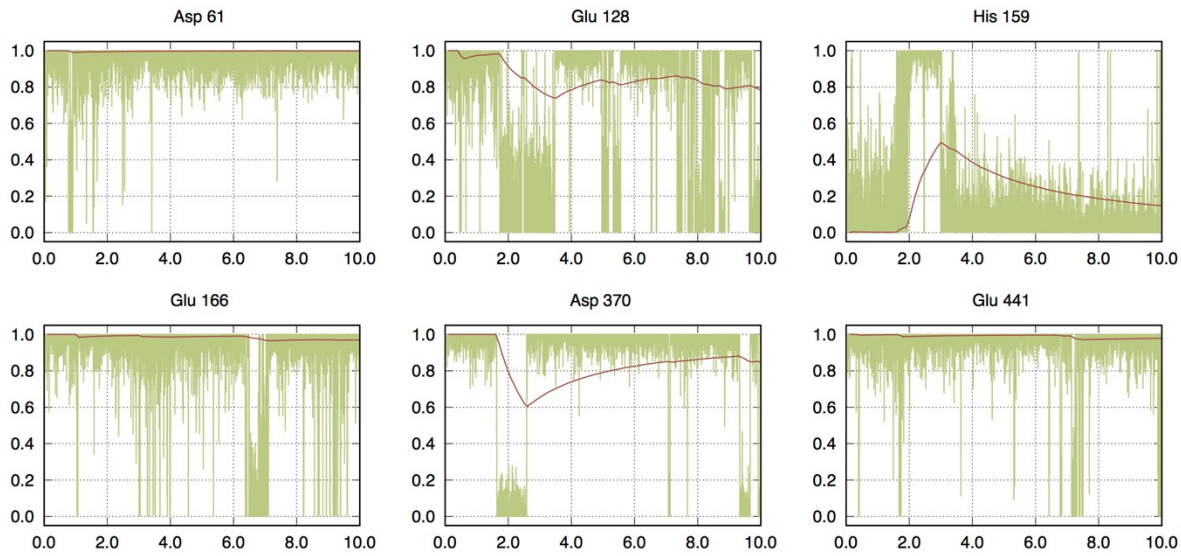


Fig. S18. Titration data from simulation 2 of *Hj*Cel3A-GEI with glucose acceptor. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

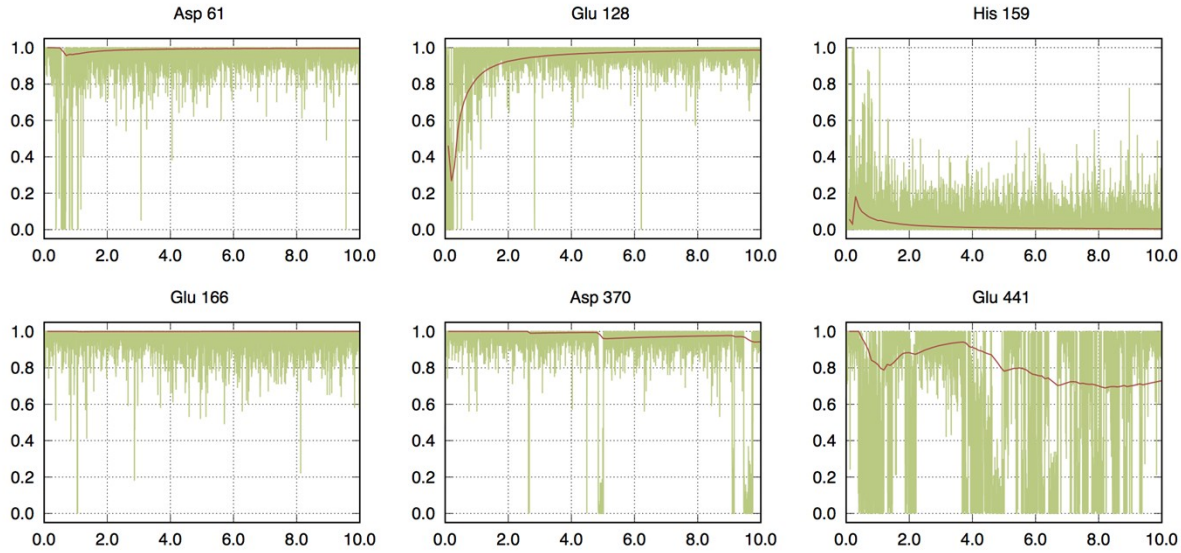


Fig. S19. Titration data from simulation 3 of *Hj*Cel3A-GEI with glucose acceptor. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

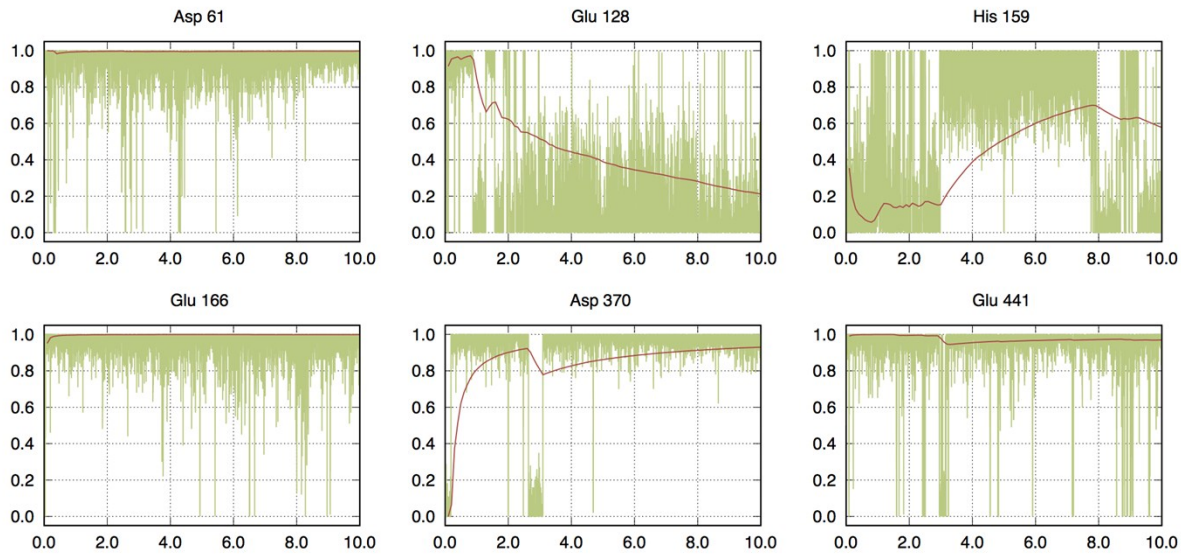


Fig. S20. Titration data from simulation 1 of *Hj*Cel3A-GEI with cellobiose acceptor. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

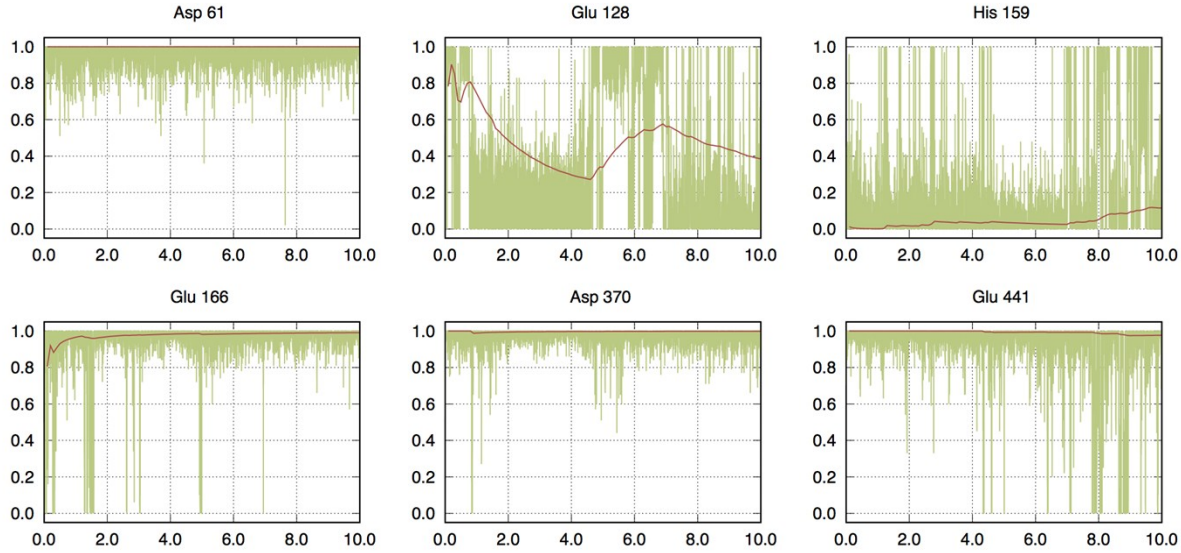


Fig. S21. Titration data from simulation 2 of *Hj*Cel3A-GEI with cellobiose acceptor. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

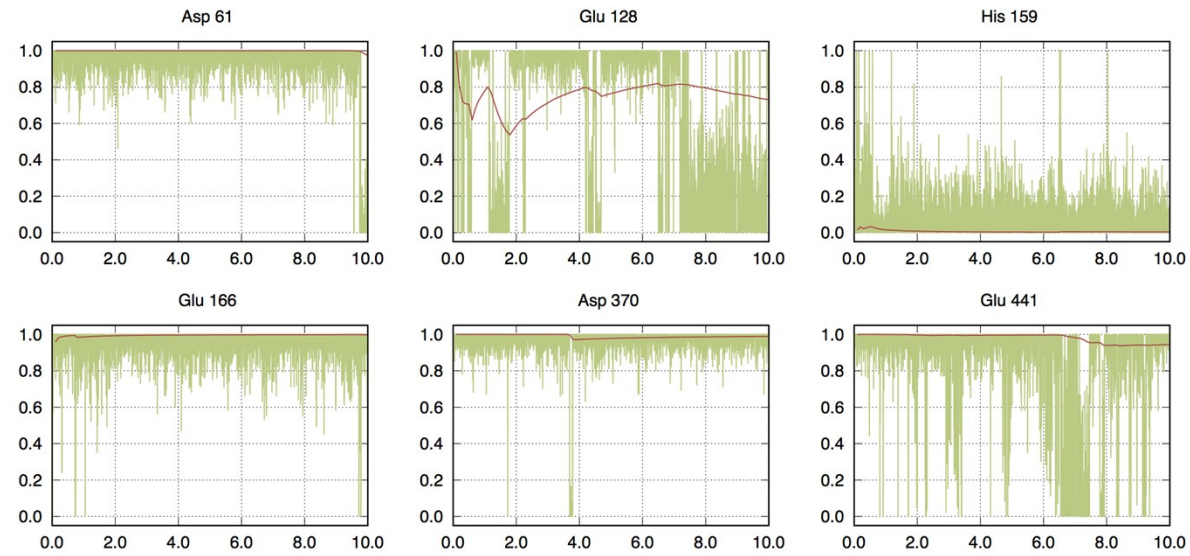


Fig. S22. Titration data from simulation 3 of *Hj*Cel3A-GEI with cellobiose acceptor. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) at pH 5 plotted as a function of simulation time in ns.

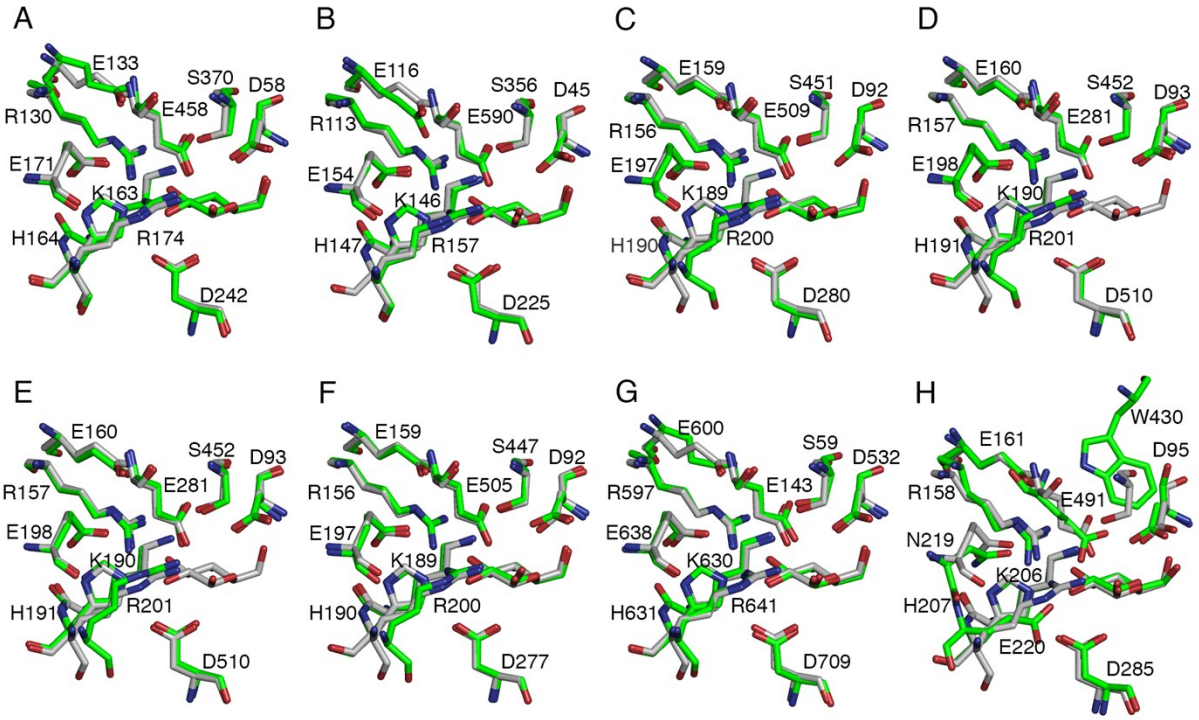


Fig. S23. Comparison of the active site of other GH3 βgls (green) with that of *Hj*Cel3A (PDB ID 3ZYZ,[1] gray). A. *Thermatoga neapolitana* (PDB ID 2X41[8]). B. *Kluyveromyces marxianus* (PDB ID 3AC0[9]). C. *Aspergillus aculeatus* (PDB ID 4IIG[10]). D. *A. fumigatus* (PDB ID 5FJI[11]). E. *A. oryzae* (PDB ID 5FJJ[11]). F. *Rasamsonia emersonii* (PDB ID 5JU6[12]). G. cow rumen metagenome (PDB ID 5K6M[13]). H. *Hordeum vulgare* (PDB ID 1IEQ[2]). Unlike the other GH3 βgls, the *H. vulgare* βgl has N219 and E220 at positions corresponding to E166 and R169 in *Hj*Cel3A. The position of residues corresponding to R67 and Y204 in *Hj*Cel3A were also similar in the other GH3 βgls but were not shown here for clarity.

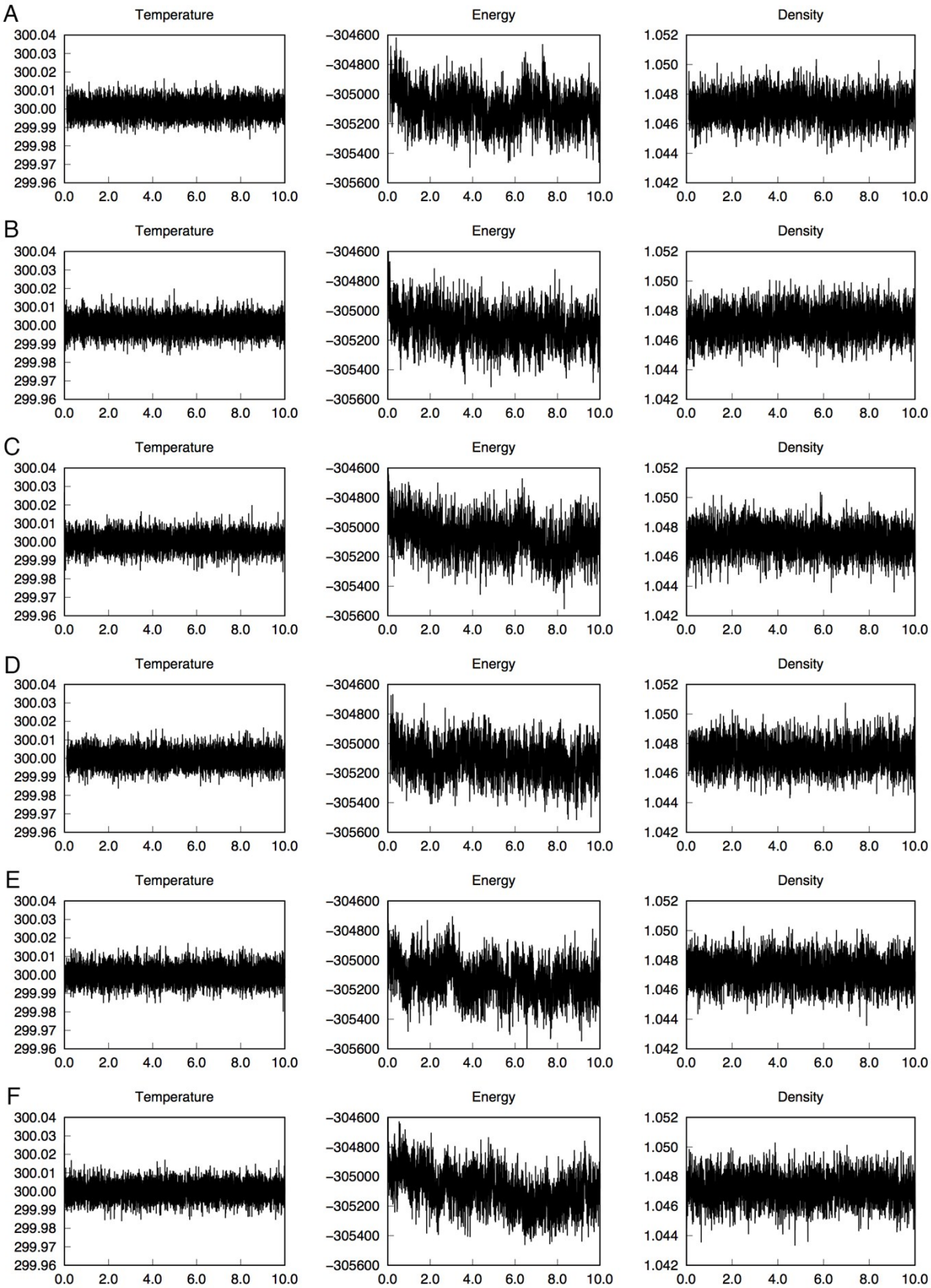


Fig. S24. Temperature (K), energy (kcal/mol), and density (g/cm^3) during the 10 ns-simulations. A–C. E128A-GEI. D–F. E128Q-GEI.

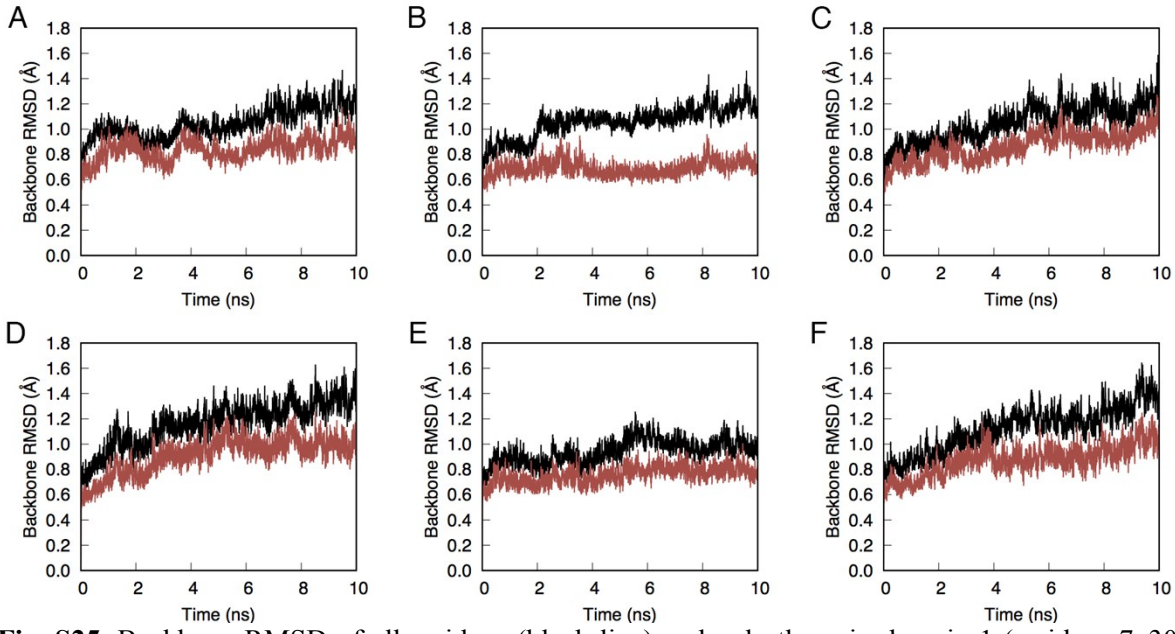


Fig. S25. Backbone RMSD of all residues (black line) and only those in domain 1 (residues 7–300, excluding loop A) and loop C (red line). A–C. E128A-GEI. D–F. E128Q-GEI.

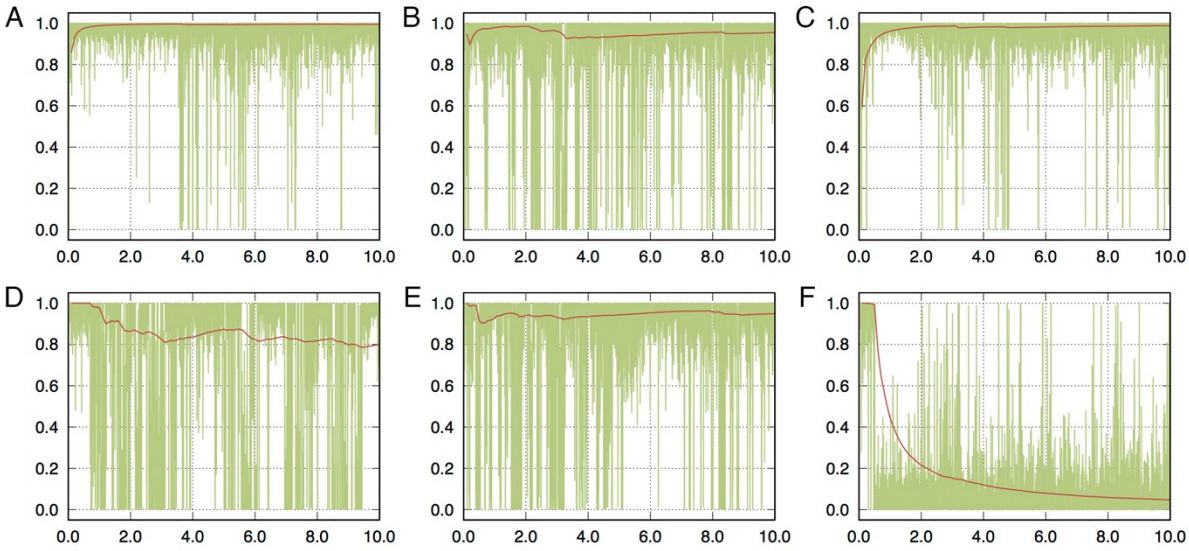


Fig. S26. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) of E441 at pH 5 plotted as a function of simulation time in ns. A–C. E128A-GEI. D–F. E128Q-GEI. In F, the predominance of the protonated state (0) is caused by rotation of the E441 side chain away from the glycosyl moiety and S384.

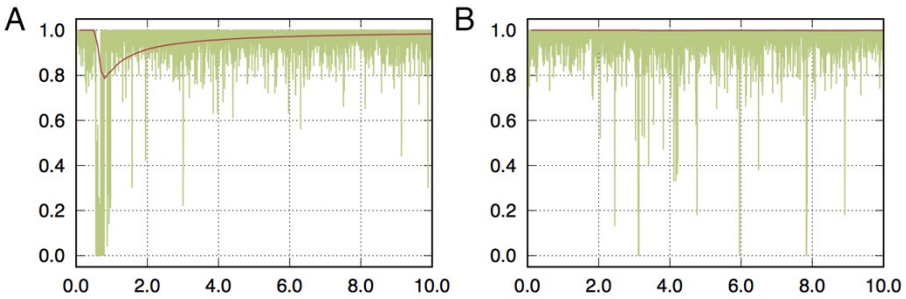


Fig. S27. Titration coordinate λ (green line) and cumulative average (0.1 ns-window) of the deprotonated fraction S (red line) of E441 at pH 5 plotted as a function of simulation time in ns. A. E166A-GEI. B. E166Q-GEI.

Table S1. Pearson correlation coefficient of ionizable active site residues in *Hj*Cel3A-GEI.[a,b]

	D61	E128	H159	E166	D370	E441
D61		-0.07 (-0.01) [-0.02]	-0.03 (-0.01) [0.00]	-0.02 (-0.02) [0.00]	0.00 (-0.02) [-0.02]	-0.02 (-0.01) [-0.03]
E128	-0.07 (-0.01) [-0.02]		0.03 (-0.06) [-0.03]	0.19 (0.04) [0.07]	0.10 (-0.05) [0.30]	0.02 (0.00) [0.16]
H159	-0.03 (-0.01) [0.00]	0.03 (-0.06) [-0.03]		0.02 (0.01) [-0.02]	0.11 (0.04) [-0.01]	-0.01 (-0.03) [-0.01]
E166	-0.02 (-0.02) [0.00]	0.19 (0.04) [0.07]	0.02 (0.01) [-0.02]		0.17 (-0.01) [-0.06]	-0.03 (-0.01) [0.00]
D370	0.00 (-0.02) [-0.02]	0.10 (-0.05) [0.30]	0.11 (0.04) [-0.01]	0.17 (-0.01) [-0.06]		-0.05 (-0.04) [-0.03]
E441	-0.02 (-0.01) [-0.03]	0.02 (0.00) [0.16]	-0.01 (-0.03) [-0.01]	-0.03 (-0.01) [0.00]	-0.05 (-0.04) [-0.03]	

[a] A positive value indicates that the two residues have the same protonation state. A negative value indicates that they have the opposite protonation state.
[b] Values from each of the three simulations are reported. Trajectories wherein the E441 side chain has rotated away from the glycosyl moiety to interact with E128 were not included in the analysis.

Table S2. Hydrogen bond interactions of E441 and the glycosyl group (BGLC) in *Hj*Cel3A-GEI.[a,b]

Acceptor	Donor	Occupancy (%)	Average lifetime (ps)
E441:OE1	S384:HN	17 ± 4	7.2 ± 0.4
	S384:HG1	26 ± 6	10 ± 1
	R125:HH21	12 ± 3	6.3 ± 0.9
E441:OE2	S384:HN	29 ± 6	8.6 ± 0.6
	S384:HG1	28 ± 12	11 ± 1
	R125:HH21	10 ± 3	7.0 ± 0.8
D61:OD1	BGLC:HO4	14 ± 3	8 ± 3
D61:OD2	BGLC:HO4	38 ± 10	9 ± 3
	BGLC:HO6	46 ± 21	14 ± 3
S384:OG	BGLC:HO3	30 ± 12	8.9 ± 0.7
BGLC:O3	H159:HE2	13 ± 1	4.7 ± 0.2
BGLC:O6	D61:HD1	17 ± 9	5.9 ± 0.7

[a] Only hydrogen bonds with occupancy >10% are listed. Those of D236 have occupancy <10%.
[b] Mean and standard deviation values were calculated from three independent simulations. Trajectories wherein the E441 side chain has rotated away from the glycosyl moiety to interact with E128 were not included in the analysis.

Table S3. Pearson correlation coefficient of ionizable active site residues in *Hj*Cel3A-GEI with glucose acceptor.[a,b]

	D61	E128	H159	E166	D370	E441
D61		-0.02 (-0.10) [-0.01]	0.03 (0.02) [-0.09]	0.00 (-0.05) [0.03]	-0.02 (-0.02) [-0.05]	-0.03 (0.00) [0.10]
E128	-0.02 (-0.10) [-0.01]		-0.10 (-0.54) [-0.43]	0.01 (-0.11) [-0.02]	-0.01 (0.36) [-0.05]	-0.02 (-0.04) [-0.14]
H159	0.03 (0.02) [-0.09]	-0.10 (-0.54) [-0.43]		0.01 (0.13) [-0.06]	0.01 (-0.60) [0.05]	0.00 (0.06) [0.01]
E166	0.00 (-0.05) [0.03]	0.01 (-0.11) [-0.02]	0.01 (0.13) [-0.06]		0.00 (-0.11) [0.05]	0.22 (-0.02) [0.05]
D370	-0.02 (-0.02) [-0.05]	-0.01 (0.36) [-0.05]	0.01 (-0.60) [0.05]	0.00 (-0.11) [0.05]		0.01 (0.00) [0.21]
E441	-0.03 (0.00) [0.10]	-0.02 (-0.04) [-0.14]	0.00 (0.06) [0.01]	0.22 (-0.02) [0.05]	0.01 (0.00) [0.21]	

[a] A positive value indicates that the two residues have the same protonation state. A negative value indicates that they have the opposite protonation state.
[b] Values from each of the three simulations are reported. Trajectories wherein glucose has left the +1 site were not included in the analysis.

Table S4. Hydrogen bond interactions of E441, glycosyl group [BGLC(1)], and glucose acceptor [BGLC(2)] in *Hj*Cel3A-GEI.[a,b]

Acceptor	Donor	Occupancy (%)	Average lifetime (ps)
E441:OE1	S384:HG1	30 ± 15	13 ± 4
	S384:HN	24 ± 12	9 ± 2
	R125:HH21	11.0 ± 0.3	6 ± 1
	BGLC(2):HO6	16 ± 6	7.7 ± 0.2
E441:OE2	S384:HG1	32 ± 8	14 ± 3
	S384:HN	19 ± 9	8 ± 1
	R125:HH11	21 ± 11	9 ± 2
	R125:HH21	12 ± 2	7.4 ± 0.4
	BGLC(2):HO6	24 ± 4	9 ± 2
D61:OD1	BGLC(1):HO4	41 ± 23	11 ± 2
D61:OD2	BGLC(1):HO4	17 ± 5	5.8 ± 0.1
	BGLC(1):HO6	43 ± 26	11 ± 3
S384:OG	BGLC(1):HO3	14.3 ± 0.7	8 ± 1
	E441:HE2	11.0 ± 0.4	5.6 ± 0.1
BGLC(1):O3	H159:HE2	17 ± 6	5.2 ± 0.6
BGLC(1):O4	D61:HD1	14 ± 3	5.3 ± 0.2
BGLC(1):O6	D61:HD2	17 ± 3	6.8 ± 0.3
BGLC(2):O6	R169:HH11	30 ± 16	8.5 ± 0.9

[a] Only hydrogen bonds with occupancy >10% are listed. Those of D236 have occupancy <10%.
[b] Mean and standard deviation values are calculated from three independent simulations. Trajectories wherein glucose has left the +1 site were not included in the analysis.

Table S5. Pearson correlation coefficient of ionizable active site residues in *Hj*Cel3A-GEI with cellobiose acceptor.[a,b]

	D61	E128	H159	E166	D370	E441
D61		-0.11 (-0.01) [0.01]	0.03 (0.01) [-0.01]	-0.01 (0.03) [0.02]	0.05 (0.00) [-0.01]	-0.02 (-0.01) [-0.02]
E128	-0.11 (-0.01) [0.01]		-0.37 (-0.19) [-0.11]	0.01 (0.03) [-0.03]	-0.10 (-0.03) [-0.07]	0.02 (0.08) [0.09]
H159	0.03 (0.01) [-0.01]	-0.37 (-0.19) [-0.11]		0.00 (0.04) [0.01]	0.18 (0.02) [0.02]	-0.10 (-0.09) [0.03]
E166	-0.01 (0.03) [0.02]	0.01 (0.03) [-0.03]	0.00 (0.04) [0.01]		0.05 (0.02) [-0.01]	-0.04 (-0.05) [0.01]
D370	0.05 (0.00) [-0.01]	-0.10 (-0.03) [-0.07]	0.18 (0.02) [0.02]	0.05 (0.02) [-0.01]		0.19 (0.01) [-0.05]
E441	-0.02 (-0.01) [-0.02]	0.02 (0.08) [0.09]	-0.10 (-0.09) [0.03]	-0.04 (-0.05) [0.01]	0.19 (0.01) [-0.05]	

[a] A positive value indicates that the two residues have the same protonation state. A negative value indicates that they have the opposite protonation state.
[b] Values from each of the three simulations are reported. Trajectories wherein cellobiose has left the +1/+2 sites were not included in the analysis.

Table S6. Hydrogen bond interactions of E441, glycosyl group [BGLC(1)], and non-reducing end of cellobiose acceptor [BGLC(2)] in *Hj*Cel3A-GEI.[a,b]

Acceptor	Donor	Occupancy (%)	Average lifetime (ps)
E441:OE1	S384:HG1	46 ± 1	17 ± 5
	S384:HN	31 ± 1	8 ± 1
	R125:HH21	17 ± 9	5.8 ± 0.8
E441:OE2	S384:HG1	36 ± 27	16 ± 3
	S384:HN	24 ± 19	7.8 ± 0.3
D61:OD1	BGLC(1):HO4	30 ± 27	8 ± 4
D61:OD2	BGLC(1):HO4	30 ± 1	7.0 ± 0.4
	BGLC(1):HO6	61 ± 13	13 ± 4
S384:OG	BGLC(1):HO3	33 ± 12	8 ± 1
	E441:HE2	14 ± 5	6 ± 1
BGLC(1):O2	H159:HE2	12 ± 3	6 ± 0
BGLC(1):O4	D61:HD1	21 ± 3	5.3 ± 0.1
BGLC(1):O6	BGLC(2):HO4	20 ± 10	7 ± 1
BGLC(2):O6	R169:HH21	28 ± 10	9.7 ± 0.2

[a] Only hydrogen bonds with occupancy >10% are listed. Those of D236 and reducing end of cellobiose have occupancy <10%.
[b] Mean and standard deviation values are calculated from three independent simulations. Trajectories wherein cellobiose has left the +1/+2 sites were not included in the analysis.

Table S7. Hydrogen bond interactions of E441 and glycosyl group (BGLC) in E128A-GEI.[a,b]

Acceptor	Donor	Occupancy (%)	Average lifetime (ps)
E441:OE1	S384:HG1	39 ± 16	12 ± 2
	S384:HN	29 ± 7	8 ± 1
E441:OE2	S384:HG1	23 ± 1	10.3 ± 0.4
	S384:HN	14 ± 3	8 ± 2
S384:OG	E441:HE1	22 ± 5	6.3 ± 0.3
	E441:HE2	17 ± 2	6.9 ± 0.3
D61:OD2	BGLC(1):HO4	33 ± 20	9 ± 2
	BGLC(1):HO6	24 ± 8	13 ± 1

[a] Only hydrogen bonds with occupancy >10% are listed. Those of D236 have occupancy <10%.
[b] Mean and standard deviation values were calculated from three independent simulations.

Table S8. Hydrogen bond interactions of E441 and glycosyl group (BGLC) in E128Q-GEI.[a,b]

Acceptor	Donor	Occupancy (%)	Average lifetime (ps)
E441:OE1	S384:HG1	14 ± 2	11 ± 2
	S384:HN	14 ± 5	7.2 ± 0.3
S384:OG	E441:HE1	10.4 ± 0.3	6.0 ± 0.1
D61:OD1	BGLC(1):HO4	18 ± 5	8 ± 2
D61:OD2	BGLC(1):HO4	32 ± 7	8 ± 1
	BGLC(1):HO6	42 ± 7	13 ± 1

[a] Only hydrogen bonds with occupancy >10% are listed. Those of D236 have occupancy <10%.
[b] Mean and standard deviation values were calculated from two independent simulations. The third simulation was not included because the E441 side chain has rotated away from the glycosyl moiety.

References

1. S. Karkehabadi, K. E. Helmich, T. Kaper, H. Hansson, N.-E. Mikkelsen, M. Gudmundsson, K. Piens, M. Fujdala, G. Banerjee, J. S. Scott-Craig, J. D. Walton, G. N. Phillips and M. Sandgren, *J. Biol. Chem.*, 2014, **289**, 31624-31637.
2. M. Hrmova, J. N. Varghese, R. De Gori, B. J. Smith, H. Driguez and G. B. Fincher, *Structure.*, 2001, **9**, 1005-1016.
3. H. F. Seidle, K. McKenzie, I. Marten, O. Shoseyov and R. E. Huber, *Arch. Biochem. Biophys.*, 2005, **444**, 66-75.
4. W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695-1697.
5. S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**, 4613-4621.
6. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
7. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
8. T. Pozzo, J. L. Pasten, E. N. Karlsson and D. T. Logan, *J. Mol. Biol.*, 2010, **397**, 724-739.
9. E. Yoshida, M. Hidaka, S. Fushinobu, T. Koyanagi, H. Minami, H. Tamaki, M. Kitaoka, T. Katayama and H. Kumagai, *Biochem. J.*, 2010, **431**, 39.
10. K. Suzuki, J. Sumitani, Y. W. Nam, T. Nishimaki, S. Tani, T. Wakagi, T. Kawaguchi and S. Fushinobu, *Biochem. J.*, 2013, **452**, 211-221.
11. J. Agirre, A. Ariza, W. A. Offen, J. P. Turkenburg, S. M. Roberts, S. McNicholas, P. V. Harris, B. McBrayer, J. Dohnalek, K. D. Cowtan, G. J. Davies and K. S. Wilson, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2016, **72**, 254-265.
12. M. Gudmundsson, H. Hansson, S. Karkehabadi, A. Larsson, I. Stals, S. Kim, S. Sunux, M. Fujdala, E. Larenas, T. Kaper and M. Sandgren, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2016, **72**, 860-870.
13. M. Ramírez-Escudero, M. V. del Pozo, J. Marín-Navarro, B. González, P. N. Golyshin, J. Polaina, M. Ferrer and J. Sanz-Aparicio, *J. Biol. Chem.*, 2016, **291**, 24200-24214.