Computational insights into active site shaping for substrate and regiospecificity in the EXTL2 retaining glycosyltransferase

Fernanda Mendoza,[1] José M. Lluch [2,3] and Laura Masgrau [2,4,*]

[1] Departamento de Ciencias Químicas, Facultad de Ciencias Exactas, Universidad Andres Bello, Sede Concepción, Talcahuano, 4260000, Chile.
[2] Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain.
[3] Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain.
[4] Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain.
Corresponding author: Laura.Masgrau@uab.cat

Supporting Information

1. Results for the six E:S complexes initially used in the reactivity studies (starting reaction paths).

Table S1. QM/MM highest energy point ($\Delta E^{\ddagger}$) and reaction energy (ΔE) along the starting front-side reaction pathway for the six different E:S complexes investigated in the reactivity studies. Energies are in kcal/mol.

	QM(BP86/TZVP)		QM(M05-2X/TZVP//BP86/TZVP)	
Frame	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE
R-175	6.64	-13.79	13.23	-14.39
R-208	18.60	-18.86	25.18	-18.12
R-488	18.20	9.15	24.90	9.12
R-495	11.77	-2.52	17.48	-3.78
R-350	18.52	2.72	23.78	5.14
R-609	21.85	9.67	30.15	11.02

Figure S1. QM(BP86/TZVP)/MM optimized reactant structures for the six different E:S complexes investigated in the reactivity studies. R-175 (orange), R-208 (red), R-488 (indigo), R-495 (green), R-350 (light blue), R-609 (purple).

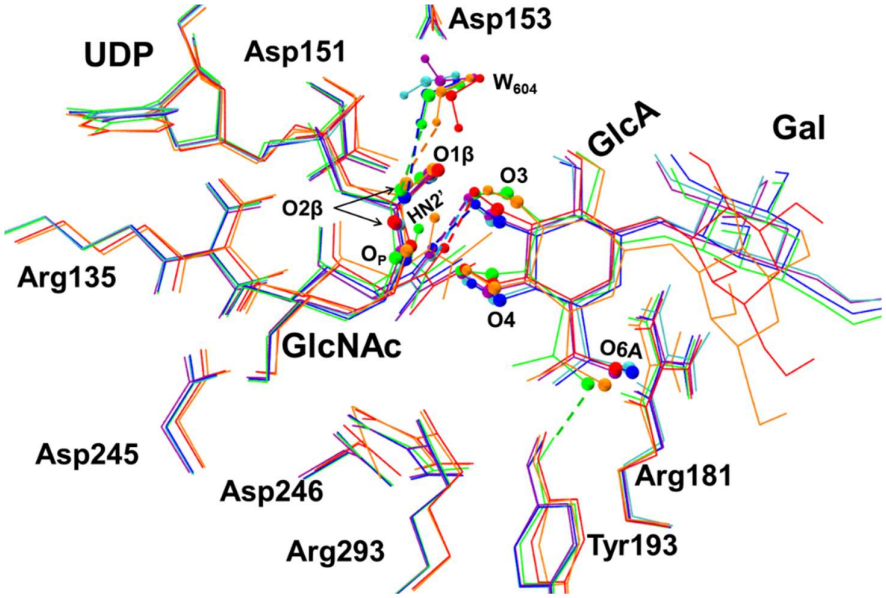
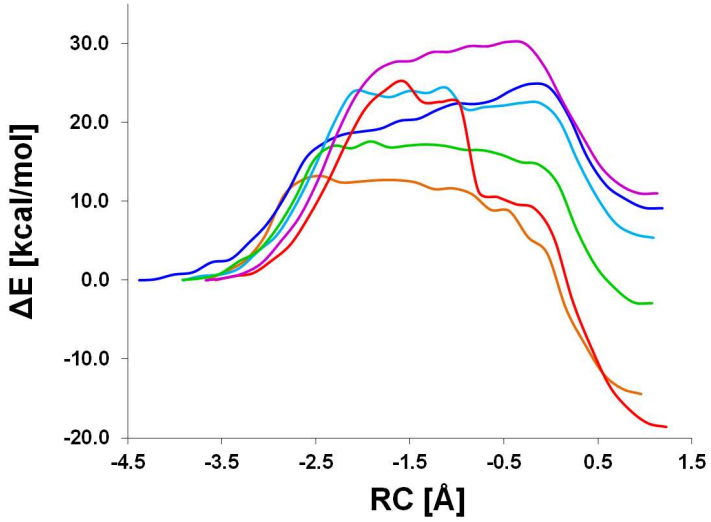


Figure S2. QM/MM [QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM] potential energy profiles for the starting front-side attack mechanism of the six different E:S complexes investigated in the initial reactivity studies. The color code corresponds to that of Figure S1 (R-175 (orange), R-208 (red), R-488 (indigo), R-495 (green), R-350 (light blue), R-609 (purple)).



As can be seen in Table S1 and Figure S2, the estimated energy barriers obtained for the different frames are diverse. The exponential averaged value is of 14.4 kcal/mol. The different barrier heights can be attributed to small variations in the interactions between the two substrates and between the enzyme and the substrates. Among them: $HN2'_{GlcNAc}$-$O3_{GlcA}$, $O1\beta$-$HO3_{GlcA}$ and O_P-$HO4_{GlcA;}$ HH_{Tyr193}-$O6A_{GlcA}$, $NE/NH2_{Arg293}$-$O6B_{GlcA}$, $OD1/OD2_{Asp246}$-$C1_{GlcNAc}$ and W_{604}-$O2\beta$.

The interaction W_{604}-$O2\beta$ is formed along the reaction path for frames starting at reactant complexes R-208 and R-350, whereas it is present from the beginning in the case of R-175, R-488 and R-495. This correlates quite well with the barrier heights obtained. For R-609, which present the highest energy barrier, this interaction is not established at any moment. In the case of R-488, where the interaction with W604 can stabilize UDP dissociation, the higher barrier height is associated with the second part of the path (thus in GlcA attack rather than ion-pair formation). These results suggest that this interaction has an important role in stabilizing glycosidic bond breakage in the donor substrate.

Other important interactions are the ones involving $HN2'_{GlcNAc}$, which can interact with $O1\beta_{UDP}$ or with $O3_{GlcA}$. This latter hydrogen bond is present in R-208, R-488 and R-609 reactants, and for the others it is formed during the reaction. This, such interaction may favor a correct orientation for catalysis.

Finally, between the two substrates a hydrogen bond between O_P-$HO4_{GlcA}$ is present in all cases, which is an interaction known to be very important for these reactions.

On the other side, the enzyme-substrate interaction HH_{Tyr193}-$O6A_{GlcA}$ seems to participate to the correct positioning of GlcA and has a bigger effect on the part of the reaction path corresponding to GlcA attack. In R-350, R-488 and R-609 reactants this interaction is missing and the barrier heights are over 20 kcal/mol. As seen in Figure S1, this interaction seems to correlate with a positioning of GlcA in which the attacking oxygen ($O4_{GlcA}$) will approach the anomeric carbon from a more above orientation.

2. Results for optimized reaction paths for the R-495 reported in the main text and two other catalytically competent E:S complexes (R-175, R-350).

Apart from the E:S complex selected to be reported in the main text (R-495), the stationary points and electrostatic contributions analysis were performed for two more complexes (R-175 and R-350). These results are presented here.

Figure S3. QM/MM energies for the stationary points along the optimized front-side attack paths for R-495 (green), R-175 (orange) and R-350 (blue). OC is used to name oxocarbenium species optimized as TS but for which an imaginary frequency has not been obtained due to the flatness of the potential energy surface. Moreover, due to the use of a dual-level approach, the [QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM] energy of the IP is often higher than that of the OC to which it is connected (something that does not happen at the QM(BP86/TZVP)/MM level). Thus, these energies are only to be taken as orientative.

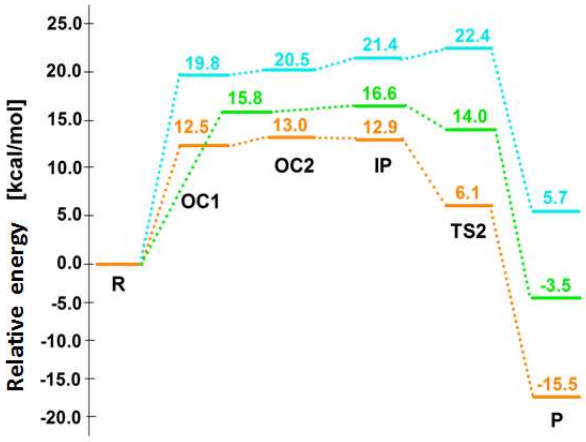


Figure S4. QM/MM optimized reactant (R) and IP for frames R-175 and R-350. Some relevant distances are shown (in Å). The substrates and some protein residues are represented in sticks, colored by element, with pink carbon atoms for the substrates and cyan ones for the enzyme. The Mg^{2+} is depicted as a green sphere.

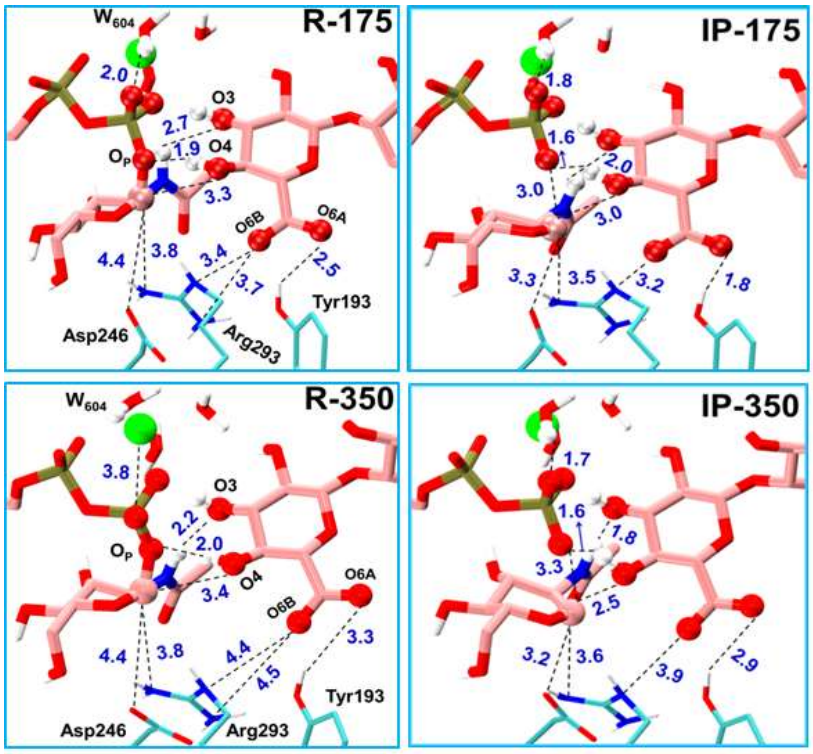


Table S2. For frame R-495, distances (in Å), NPA charges (Δq, in a.u.), QM/MM energies (ΔE, in kcal/mol) and frequency (in cm^{-1}) for the optimized structures of the front-side attack mechanism.

	R	OC	IP	OC2	P
d($C1_{GlcNAc}$-O_P)	**1.52**	**2.56**	**3.25**	**3.38**	**3.48**
d($C1_{GlcNAc}$-$O4_{GlcA}$)	**3.45**	**3.11**	**2.49**	**2.11**	**1.52**
d($OD1_{D246}$-$C1_{GlcNAc}$)	**4.62**	**3.59**	**3.38**	**3.37**	**3.98**
d($C1_{GlcNAc}$-$O5_{GlcNAc}$)	1.38	1.27	1.28	1.29	1.39
d($HO4_{GlcA}$-O_P)	1.92	1.66	1.56	1.39	1.01
d($HO4_{GlcA}$-$O4_{GlcA}$)	0.98	1.01	1.03	1.09	1.69
d(WAT604-$O2B_{UDP}$)	2.01	1.94	1.77	1.76	1.82
d($HN2'_{GlcNAc}$-$O1B_{UDP}$)	2.32	3.15	3.45	3.63	3.86
d($HN2'_{GlcNAc}$-$O3_{GlcA}$)	2.92	2.45	2.20	2.05	2.00
d($O1B_{GlcNAc}$-$HO3_{GlcA}$)	1.71	1.55	1.53	1.57	1.75
d($NH1_{R293}$-$O6_{GlcNAc}$)	3.29	3.03	2.96	2.93	2.86
d($NH1_{R293}$-$C1_{GlcNAc}$)	**3.89**	**3.72**	**3.68**	**3.73**	**4.05**
d($NH1_{R293}$-$O5_{GlcNAc}$)	**3.23**	**3.03**	**2.99**	**2.98**	**3.09**
d($HH22_{R293}$- $OD2_{D246}$)	1.65	1.67	1.66	1.66	1.68
d($HH12_{R293}$- $OD1_{D246}$)	1.78	1.83	1.83	1.82	1.80
d(NE_{R293}-$O6B_{GlcA}$)	3.90	3.83	3.70	3.55	3.29
d($NH2_{R293}$-$O6B_{GlcA}$)	3.80	3.55	3.41	3.32	3.30
d($HH21_{R293}$-OH_{Y193})	2.03	1.99	1.98	1.98	1.97
d(HH_{Y193}-$O6A_{GlcA}$)	1.88	1.84	1.81	1.78	1.76
d(HH_{Y193}-$O6B_{GlcA}$)	2.41	2.35	2.34	2.37	2.44
d(HE_{R181}-$O6A_{GlcA}$)	1.73	1.73	1.74	1.73	1.73
d($HH21_{R181}$-$O5_{GlcA}$)	1.66	1.66	1.68	1.69	1.71
$\Delta q(C1+H1+O5)_{GlcNAc}$	0.00	0.37	0.36	0.29	-0.02
ΔE	0.00	15.84	16.58	14.03	-3.46
ν (cm^{-1})[1]		70.61[a]	83.82[b]	73.71[c]	

[a] Frequency associated with the O_P-$C1_{GlcNAc}$ vibration.

[b] Lowest frequency for the IP. It describes a translation of the GlcNAc ring, which has $C1_{GlcNAc}$ in the plane of $(-C3-C2-O5-C5-)_{GlcNAc}$.

[c] Frequency related to new glycosidic bond formation and $HO4_{GlcNAc}$ transfer.

Table S3. For frame R-175, distances (in Å), NPA charges (Δq, in a.u.), QM/MM energies (ΔE, in kcal/mol) and frequency (in cm^{-1}) for the optimized structures of the front-side attack mechanism.

	R	OC1	OC2	IP	OC3	P
d($C1_{GlcNAc}$-O_P)	**1.54**	**2.14**	**2.35**	**2.99**	**3.55**	**3.56**
d($C1_{GlcNAc}$-$O4_{GlcA}$)	**3.29**	**3.25**	**3.26**	**3.04**	**2.17**	**1.50**
d($OD1_{D246}$-$C1_{GlcNAc}$)	**4.43**	**3.80**	**3.63**	**3.30**	**3.08**	**3.63**
d($C1_{GlcNAc}$-$O5_{GlcNAc}$)	1.37	1.29	1.28	1.27	1.29	1.40
d($HO4_{GlcA}$-O_P)	1.86	1.65	1.63	1.60	1.43	1.01
d($HO4_{GlcA}$-$O4_{GlcA}$)	0.98	1.00	1.01	1.02	1.08	1.69
d(WAT604-$O2B_{UDP}$)	1.99	1.93	1.88	1.79	1.74	1.81
d($HN2'_{GlcNAc}$-$O1B_{UDP}$)	2.33	3.04	3.19	3.47	3.76	3.98
d($HN2'_{GlcNAc}$-$O3_{GlcA}$)	2.71	2.14	2.11	2.07	2.19	2.14
d($O1B_{UDP}$-$HO3_{GlcA}$)	1.72	1.57	1.57	1.54	1.60	1.78
d($NH1_{R293}$-$O6_{GlcNAc}$)	3.10	3.12	3.09	3.02	2.96	2.80
d($NH1_{R293}$-$C1_{GlcNAc}$)	**3.80**	**3.58**	**3.55**	**3.48**	**3.52**	**3.87**
d($NH1_{R293}$-$O5_{GlcNAc}$)	**3.10**	**3.00**	**2.98**	**2.97**	**2.90**	**2.99**
d($HH12_{R293}$- $OD1_{D246}$)	1.84	1.85	1.87	1.91	1.90	1.83
d($HH22_{R293}$- $OD2_{D246}$)	1.66	1.73	1.74	1.73	1.71	1.70
d(NE_{R293}-$O6B_{GlcA}$)	3.39	3.14	3.14	3.16	2.99	2.94
d($NH2_{R293}$-$O6B_{GlcA}$)	3.71	3.25	3.25	3.16	3.14	3.14
d($HH21_{R293}$-OH_{Y193})	1.89	1.96	1.95	1.96	1.93	1.91
d(HH_{Y193}-$O6B_{GlcA}$)	1.77	2.22	2.23	2.23	2.30	1.91
d(HH_{Y193}-$O6A_{GlcA}$)	2.51	1.84	1.82	1.80	1.77	1.74
d(HE_{R181}-$O6A_{GlcA}$)	1.71	1.71	1.71	1.72	1.73	1.74
d($HH21_{R181}$-$O5_{GlcA}$)	1.79	1.73	1.73	1.73	1.76	1.75
$q(C1+H1+O5)_{GlcNAc}$	0.00	0.31	0.35	0.36	0.31	-0.03
ΔE	0.00	12.48	12.99	12.88	6.14	-15.52
v (cm^{-1})		63.35[a]	81.59[a]	78.41[b]	87.55[c]	

[a] Frequency associated with the O_P-$C1_{GlcNAc}$ vibration.

[b] Lowest frequency for the IP. It describes a translation of the GlcNAc ring, which has $C1_{GlcNAc}$ in the plane of $(-C3-C2-O5-C5-)_{GlcNAc}$.

[c] Frequency related to new glycosidic bond formation and $HO4_{GlcNAc}$ transfer.

Table S4. For frame R-350, distances (in Å), NPA charges (Δq, in a.u.), QM/MM energies (ΔE, in kcal/mol) and frequency (in cm^{-1}) for the optimized structures of the front-side attack mechanism.

	R	OC1	OC2	IP	OC3	P
d($C1_{GlcNAc}$-O_P)	**1.49**	**2.28**	**2.41**	**3.28**	**3.34**	**3.49**
d($C1_{GlcNAc}$-$O4_{GlcA}$)	**3.35**	**3.03**	**3.04**	**2.52**	**2.02**	**1.51**
d($OD2_{D246}$-$C1_{GlcNAc}$)	**4.39**	**3.60**	**3.54**	**3.18**	**3.46**	**4.01**
d($C1_{GlcNAc}$-$O5_{GlcNAc}$)	1.38	1.28	1.28	1.27	1.30	1.39
d($HO4_{GlcA}$-O_P)	2.00	1.79	1.80	1.62	1.38	1.01
d($HO4_{GlcA}$-$O4_{GlcA}$)	0.98	1.00	1.00	1.02	1.10	1.66
d(WAT604-$O2B_{UDP}$)	3.76	1.88	1.86	1.73	1.73	1.76
d($HN2'_{GlcNAc}$-$O1B_{UDP}$)	3.50	3.27	3.30	3.61	3.73	3.97
d($HN2'_{GlcNAc}$-$O3_{GlcA}$)	2.24	1.85	1.85	1.79	1.79	1.83
d($O1B_{GlcNAc}$-$HO3_{GlcA}$)	1.78	1.55	1.55	1.53	1.57	1.77
d($NH1_{R293}$-$O6_{GlcNAc}$)	3.20	3.08	3.09	2.95	2.86	2.80
d($NH1_{R293}$-$C1_{GlcNAc}$)	**3.84**	**3.66**	**3.64**	**3.62**	**3.76**	**4.02**
d($NH1_{R293}$-$O5_{GlcNAc}$)	**3.31**	**3.01**	**2.99**	**2.95**	**2.96**	**3.04**
d($HH12_{R293}$- $OD2_{D246}$)	1.78	1.83	1.84	1.85	1.83	1.80
d($HH22_{R293}$- $OD1_{D246}$)	1.62	1.62	1.62	1.64	1.64	1.65
d(NE_{R293}-$O6B_{GlcA}$)	4.42	4.28	4.32	3.93	3.80	3.67
d($NH2_{R293}$-$O6B_{GlcA}$)	4.49	4.28	4.31	3.92	3.78	3.67
d($HH21_{R293}$-OH_{Y193})	2.00	1.95	1.95	1.94	1.94	1.96
d(HH_{Y193}-$O6A_{GlcA}$)	3.28	3.18	3.20	2.86	2.70	2.56
d(HH_{Y193}-$O6B_{GlcA}$)	1.73	1.71	1.72	1.74	1.80	1.86
d(HE_{R181}-$O6A_{GlcA}$)	1.70	1.70	1.70	1.70	1.72	1.73
d($HH21_{R181}$-$O5_{GlcA}$)	1.75	1.73	1.73	1.71	1.71	1.71
$q(C1+H1+O5)_{GlcNAc}$	0.00	0.36	0.38	0.37	0.25	-0.01
ΔE	0.00	19.81	20.47	21.41	22.42	5.67
ν (cm^{-1})[1]		63.49[a]	68.46[a]	86.74[b]	21.29[c]	

[a] Frequency associated with the O_P-$C1_{GlcNAc}$ vibration.
[b] Lowest frequency for the IP. It describes a translation of the GlcNAc ring, which has $C1_{GlcNAc}$ in the plane of $(-C3-C2-O5-C5-)_{GlcNAc}$.
[c] Frequency related to new glycosidic bond formation and $HO4_{GlcNAc}$ transfer.

Table S5. NPA charges (in a.u.) for Asp246 and Arg293 residues in WT, revealing that there is some charge transfer within the QM atoms that make their charge slightly different from the formal -1/+1 a.u. (Asp/Arg), with no significant change between the reactants (R) and the IP. The charges for Asp246 in the Arg293Ala mutant are also given to illustrate the effect of deleting Arg293.

	R	IP
WT		
Asp246	- 0.763	- 0.774
Arg293	0.889	0.894
Arg293Ala		
Asp246	- 0.997	- 0.908

Figure S5. QM(SCC-DFTB)/MM potential (black) and free energy (blue) profiles for the front-side attack mechanism. Energies are in kcal/mol.

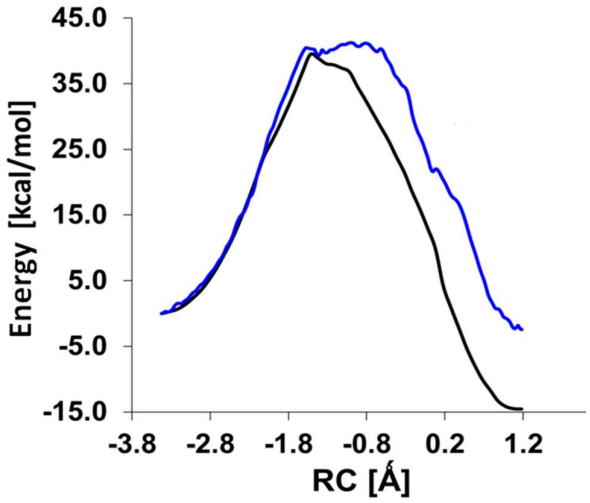
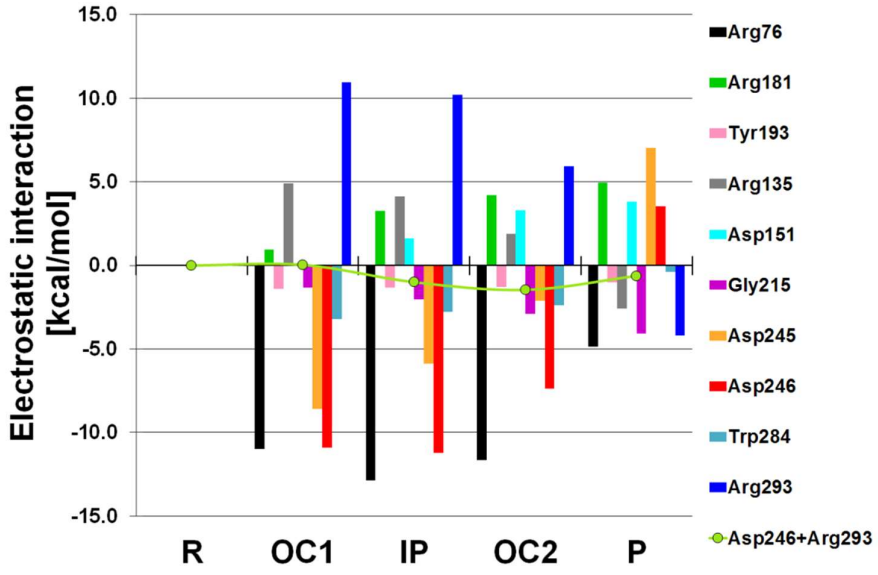
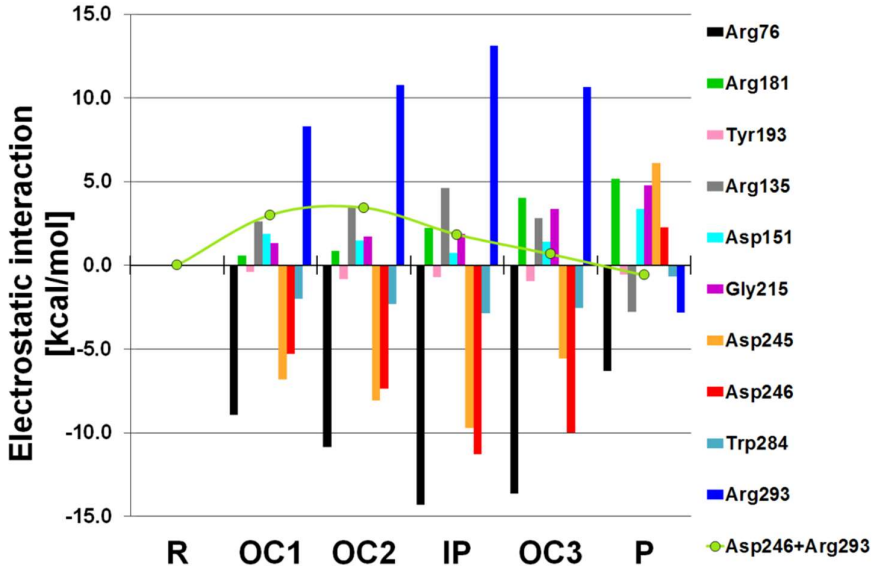


Figure S6. Electrostatic interaction (in kcal/mol) between residues and the QM zone for the stationary points along the front-side attack optimized reaction path for (A) R-495, (B) R-175 and (C) R-350. Values are given relative to the interaction in the reactants. A positive value indicates less stabilization in the corresponding point than in the reactants. Asn243, not included in the graph, has a practically negligible contribution of -0.91 kcal/mol at the IP.

A)



B)



C)

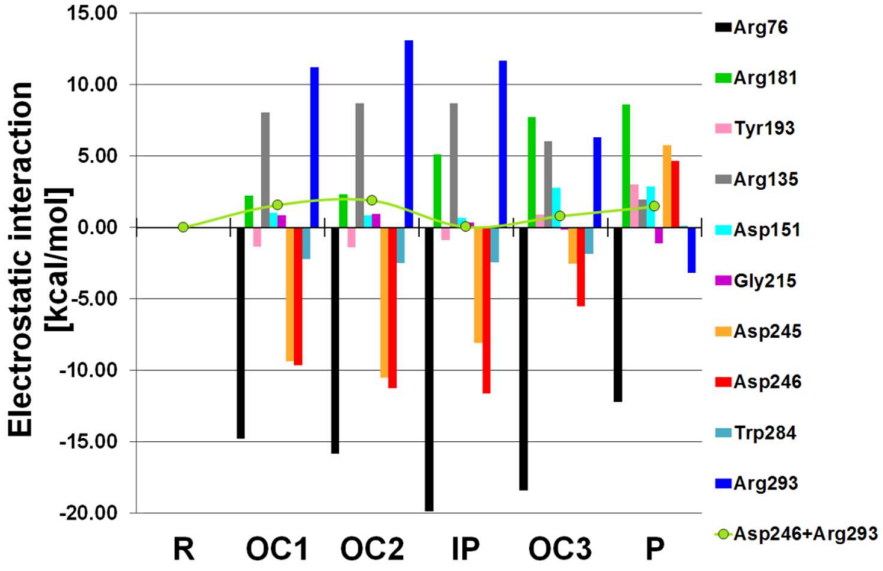


Figure S7. QM/MM [QM(M05-2X/TZVP)/MM//QM(BP86/TZVP)/MM] potential energy profiles (left-hand axis, in kcal/mol) for the Asp246 attack mechanism of three different E:S complexes (R-495 (green, reported in main text), R-175 (orange) and R-350 (blue). Relevant distances (in Å) along the path are shown in the secondary vertical axis. IP and CGE optimized points are marked with a circle. The triangles show the location of optimized TSs or OCs. For R-175 (orange) the CGE is not formed but the attack by GlcA takes place.

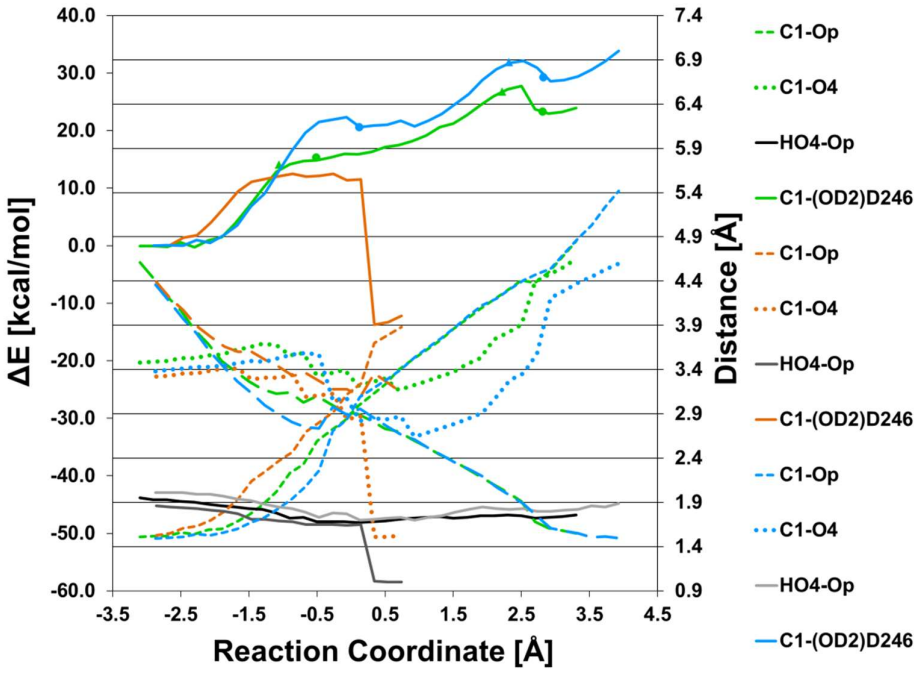


Table S6. For frame R-495, distances (in Å), NPA charges (Δq, in a.u.), QM/MM energies (ΔE, in kcal/mol) and frequency (in cm^{-1}) for the optimized structures of the Asp246 attack mechanism.

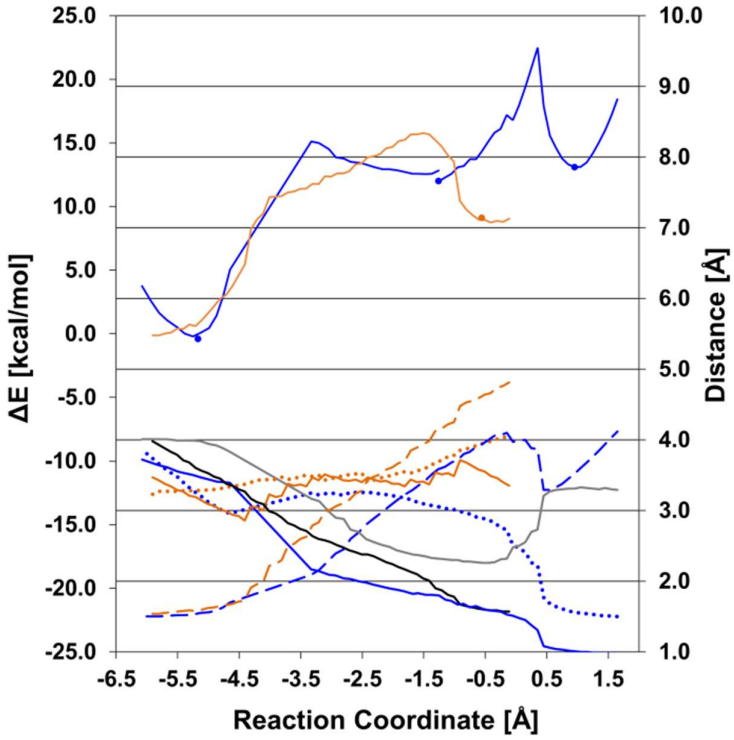
	R-495	OC1-495	IP-495	OC2-495	CGE-495
d($C1_{GlcNAc}$-O_P)	**1.52**	**2.26**	**2.73**	**4.23**	**4.45**
d($C1_{GlcNAc}$-$O4_{GlcA}$)	**3.45**	**3.32**	**3.06**	**3.93**	**4.51**
d($OD1_{Asp246}$-$C1_{GlcNAc}$)	**4.62**	**3.53**	**3.24**	**2.01**	**1.64**
d($C1_{GlcNAc}$-$O5_{GlcNAc}$)	1.38	1.28	1.27	1.30	1.35
d($HO4_{GlcA}$-O_P)	1.92	1.68	1.65	1.71	1.73
d($HO4_{GlcA}$-$O4_{GlcA}$)	0.98	1.00	1.01	1.01	1.01
d(WAT604-$O2B_{UDP}$)	2.01	2.03	1.88	1.80	1.84
d($HN2'_{GlcNAc}$-$O1B_{UDP}$)	2.32	2.95	3.23	3.69	3.56
d($HN2'_{GlcNAc}$-$O3_{GlcA}$)	2.92	2.63	2.44	2.43	2.60
d($O1B_{GlcNAc}$-$HO3_{GlcA}$)	1.71	1.56	1.53	1.52	1.50
d($NH1_{R293}$-$O6_{GlcNAc}$)	3.29	3.11	2.99	2.78	2.89
d($NH1_{R293}$-$C1_{GlcNAc}$)	**3.89**	**3.71**	**3.67**	**4.00**	**4.00**
d($NH1_{R293}$-$O5_{GlcNAc}$)	**3.23**	**2.98**	**2.97**	**3.06**	**2.96**
d($HH22_{R293}$- $OD2_{D246}$)	1.65	1.67	1.67	1.86	1.94
d($HH12_{R293}$- $OD1_{D246}$)	1.78	1.86	1.93	3.57	3.65
d(NE_{R293}-$O6B_{GlcA}$)	3.90	3.87	3.81	3.76	3.68
d($NH2_{R293}$-$O6B_{GlcA}$)	3.80	3.60	3.51	3.21	2.95
d($HH21_{R293}$-OH_{Y193})	2.03	1.97	1.95	2.01	2.22
d(HH_{Y193}-$O6A_{GlcA}$)	1.88	1.81	1.80	1.78	1.77
d(HH_{Y193}-$O6B_{GlcA}$)	2.41	2.41	2.39	2.33	2.34
d(HE_{R181}-$O6A_{GlcA}$)	1.73	1.74	1.73	1.73	1.73
d($HH21_{R181}$-$O5_{GlcA}$)	1.66	1.65	1.66	1.66	1.66
$\Delta q(C1+H1+O5)_{GlcNAc}$	0.00	0.35	0.38	0.24	0.00
ΔE	0.00	14.12	15.33	26.80	23.34
ν (cm^{-1})		43.74	107.94	7.00	

Table S7. For frame R-350, distances (in Å), NPA charges (Δq, in a.u.), QM/MM energies (ΔE, in kcal/mol) and frequency (in cm^{-1}) for the optimized structures of the Asp246 attack mechanism.

	R-350	OC-350	IP-350	TS2-350	CGE-350
d($C1_{GlcNAc}$-O_P)	**1.49**	**2.70**	**3.08**	**4.24**	**4.48**
d($C1_{GlcNAc}$-$O4_{GlcA}$)	**3.35**	**3.05**	**2.82**	**3.68**	**4.17**
d($OD2_{D246}$-$C1_{GlcNAc}$)	**4.39**	**2.99**	**2.96**	**1.93**	**1.66**
d($C1_{GlcNAc}$-$O5_{GlcNAc}$)	**1.38**	**1.27**	**1.27**	**1.30**	**1.34**
d($HO4_{GlcA}$-O_P)	2.00	1.78	1.70	1.76	1.80
d($HO4_{GlcA}$-$O4_{GlcA}$)	0.98	1.00	1.01	1.00	1.00
d(WAT604-$O2B_{UDP}$)	3.76	1.79	1.74	1.74	1.73
d($HN2'_{GlcNAc}$-$O1B_{UDP}$)	3.50	3.36	3.51	3.72	3.69
d($HN2'_{GlcNAc}$-$O3_{GlcA}$)	2.24	1.86	1.82	1.85	1.88
d($O1B_{GlcNAc}$-$HO3_{GlcA}$)	1.78	1.52	1.50	1.50	1.49
d($NH1_{R293}$-$O6_{GlcNAc}$)	3.20	3.03	2.99	2.78	2.91
d($NH1_{R293}$-$C1_{GlcNAc}$)	**3.84**	**3.56**	**3.56**	**4.01**	**3.98**
d($NH1_{R293}$-$O5_{GlcNAc}$)	**3.31**	**2.95**	**2.95**	**3.01**	**2.93**
d($HH12_{R293}$- $OD2_{D246}$)	1.78	2.03	1.97	3.71	3.66
d($HH22_{R293}$- $OD1_{D246}$)	1.62	1.65	1.63	1.80	1.85
d(NE_{R293}-$O6B_{GlcA}$)	4.42	4.28	4.00	3.94	3.92
d($NH2_{R293}$-$O6B_{GlcA}$)	4.49	4.34	3.99	3.77	3.62
d($HH21_{R293}$-OH_{Y193})	2.00	1.94	1.93	1.90	1.91
d(HH_{Y193}-$O6A_{GlcA}$)	3.28	3.25	2.97	3.01	2.95
d(HH_{Y193}-$O6B_{GlcA}$)	1.73	1.72	1.71	1.68	1.67
d(HE_{R181}-$O6A_{GlcA}$)	1.70	1.71	1.71	1.73	1.72
d($HH21_{R181}$-$O5_{GlcA}$)	1.75	1.75	1.71	1.69	1.69
$q(C1+H1+O5)_{GlcNAc}$	0.00	0.39	0.39	0.19	0.03
ΔE	0.00	21.88	20.58	31.91	29.29
ν (cm^{-1})[1]		76.87	89.73	37.12*i*	

3. Results for the mutant forms of EXTL2.

Figure S8. Comparison of the front-side (blue) and the Asp246 attack (orange) energy profiles (left-hand vertical axis, in kcal/mol) for the Arg293Ala mutant of EXTL2. Distances of interest are also shown (right-hand axis, in Å): $C1_{GlcNAc}$–O_P (dashed colored lines), $C1_{GlcNAc}$–$O4_{GlcA}$ (dotted lines), O_P–$HO4_{GlcA}$ (solid colored lines) and $C1_{GlcNAc}$-$OD2_{D246}$ (solid gray lines for front-side and solid black lines for Asp246 attack).



Within the static approach, the formation of the CGE in the Arg293Ala mutant is quite easy, with a potential energy barrier of roughly 16 kcal/mol in the mutant, noticeably smaller than the corresponding value in the WT enzyme (26.8 kcal/mol). The acceptor substrate stays more than 3 Å away from $C1_{GlcNAc}$ during this first step of the mechanism. However, lacking Arg293, the second step of the mechanism to form the final products will present the same problems commented for the GlcA attack to the IP. Thus, again, the reaction would be much slower in the Arg293Ala mutant.

Figure S9. Root mean squared deviations (RMSDs) of the C_α along the 10 ns molecular dynamics simulation for the three mutants: Arg293Ala (red), Asp246Ala (blue) and Asp246Glu (black).

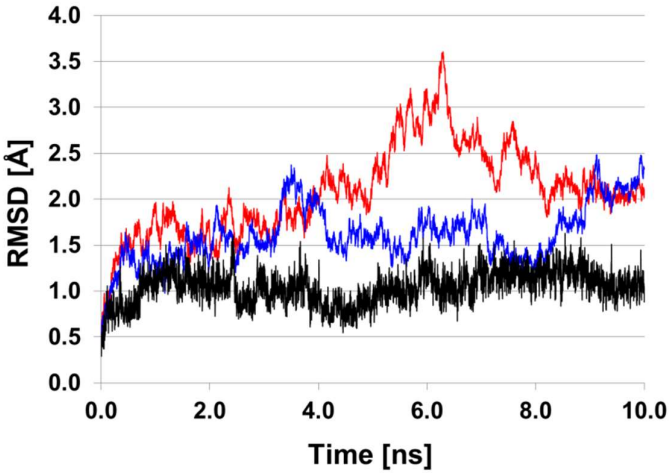


Figure S10. Distance evolution (in Å) along the molecular dynamics simulations for the Asp246Ala mutant: *d*(OH_{Tyr193}-$HH21_{Arg293}$) in gray, *d*(HH_{Tyr193}-$O6A_{GlcA}$) in yellow and *d*(HH_{Tyr193}-$O6B_{GlcA}$) in blue.

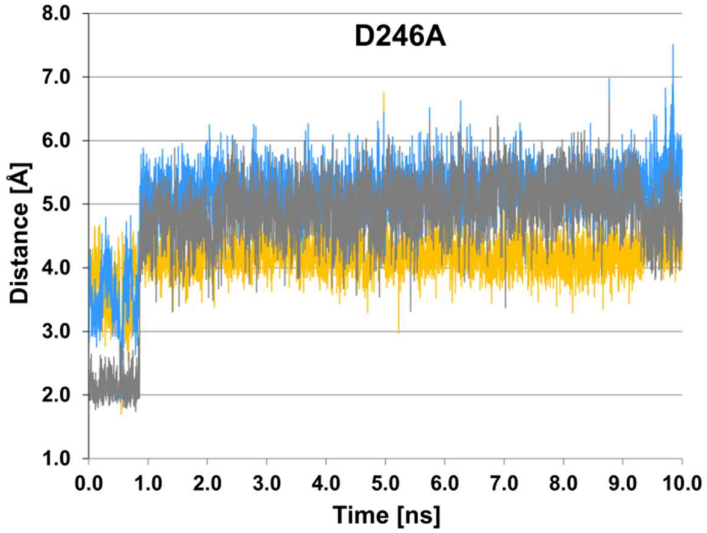


Figure S11. QM/MM energy profile (left-hand axis, in kcal/mol) for the front-side attack mechanism in wild type (green) and Arg293Ala/Asp246Ala double mutant (blue) of EXTL2. Distances (right-hand axis, in Å) of interest are also shown: $C1_{GlcNAc}$–O_P (dashed colored lines), $C1_{GlcNAc}$–$O4_{GlcA}$ (dotted lines), O_P–$HO4_{GlcA}$ (solid colored lines), $C1_{GlcNAc}$-$OD2_{D246}$ (solid black for WT) and $HO4_{GlcA}$–$O6B_{GlcA}$ (gray line for Arg293Ala).

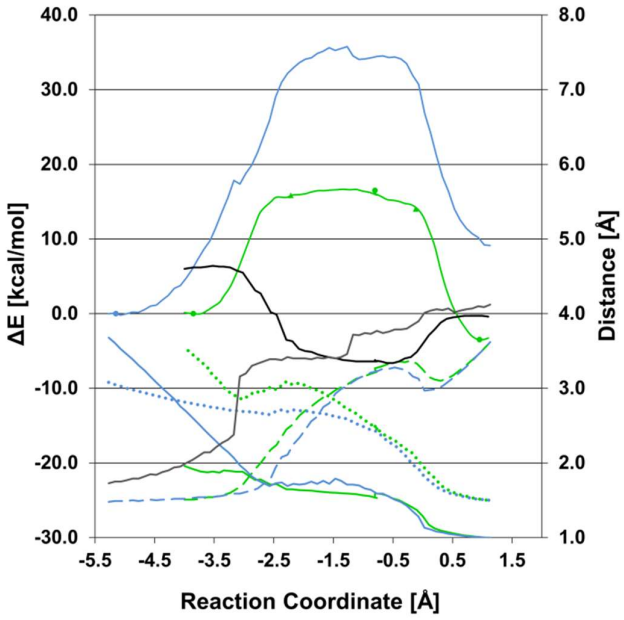
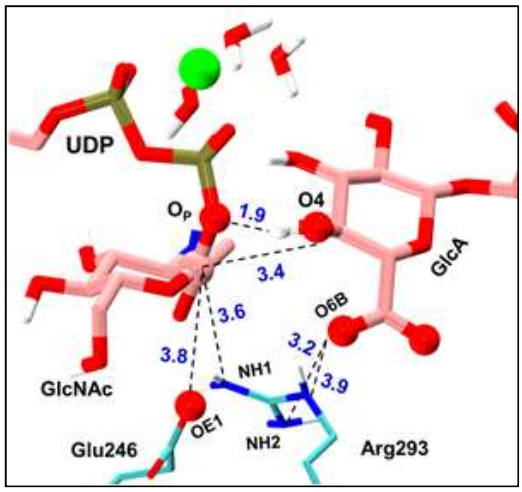


Figure S12. Molecular representation of the active site of the Asp246Glu mutant in the reactants structure built following the static approach (see main text). The substrates and some protein residues are represented in sticks, colored by element, with pink carbon atoms for the substrates and cyan ones for the enzyme. The Mg^{2+} is depicted as a green sphere.



4. Comparison of EXTL2 with enzymes from CAZy family GT6.

Figure S13. Molecular view for active site comparison for EXTL2 (blue, modelled in this work), hGTA (light green, PDB ID 3sxe), bacterial α3GalNAcT (lila, PDB ID 4cj8, chain E), hGTB (dark green, PDB ID 3sxc) and bovine α3GalT (tan, modelled in ref. 29 of the main text). The figure was prepared with the UCSF Chimera package, v.1.11. (ref. 82 of the main manuscript)