Comparing Hydrolysis and Transglycosylation Reactions Catalyzed by *Thermus thermophilus* β-glycosidase. A Combined MD and QM/MM Study

Sonia Romero-Téllez¹², José M. Lluch¹², Àngels González-Lafont¹²,* Laura Masgrau¹²,*

¹Departament de Química, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain

²Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain

**Supplementary information**

**Supplementary Scheme S1.** Atom type scheme of the pNP-Fuc substrate of the glycosylation step. Atom types from fucose moiety belong to GLYCAM06j force field and pNP atom types from gaff force field.

![Atom type scheme of the pNP-Fuc substrate](image)

**Supplementary Table S1.** Frcmod file with missing parameters of the system.

**BOND**

| Bond       | Value 1 | Value 2 |
|------------|---------|---------|
| os-Cg      | 285.0   | 1.460   |

**ANGLE**

| Angle       | Value 1 | Value 2 | Value 3 |
|-------------|---------|---------|---------|
| os-Cg-H2    | 60.00   | 110.00  |
| os-Cg-Os    | 100.00  | 112.00  |
| ca-os-Cg    | 66.103  | 117.960 |

**DIHE**

| DIHE        | Value 1 | Value 2 | Value 3 | Value 4 |
|-------------|---------|---------|---------|---------|
| Oh-Cg-Cg-os | 1.010   | 0.000   | -3      |
| Oh-Cg-Cg-os | 0.000   | 0.000   | -2      |
| Oh-Cg-Cg-os | 0.020   | 180.000 | 1       |
| ca-os-Cg-H2 | 3.150   | 0.000   | 3.000   |
| ca-os-Cg-Os | 3.150   | 0.000   | 3.000   |
| ca-os-Cg-Cg | 3.150   | 0.000   | 3.000   |
| os-Cg-Cg-H1 | 0.05    | 0.00    | 3.      |
Supplementary Table S2. Analysis over the MD simulation of the most populated hydrogen-bonds between pNP-Fuc acting as a donor and protein residues at the active site. H-bond occupancies are defined as the fraction of frames the bond is present. Only H-bond occupancies above 10% are included. The average donor to acceptor heavy atoms distances of the bonds when present are also given.

| Acceptor     | Donor | H-bond occupancy % | Average Distance (Å) |
|--------------|-------|--------------------|----------------------|
| OE1GLN18     | O3FUC | 79.33              | 2.71                 |
| OE1GLU392    | O4FUC | 62.53              | 2.79                 |
| OE1GLU338    | O2FUC | 46.60              | 2.84                 |
| O2GLU338     | O2FUC | 16.29              | 2.84                 |
| OE1GLN18     | O4FUC | 0.19               | 2.87                 |

Supplementary Table S3. Analysis over the MD simulation of the most populated hydrogen-bonds between substrate atoms and protein residues at the active site. H-bond occupancies are defined as the fraction of frames the bond is present. Only H-bond occupancies above 10% are included. The average distance of the bonds when present between donor and acceptor heavy atoms are also given.

| Substrate atom | Protein atom | H-bond occupancy % | Average Distance (Å) |
|----------------|--------------|--------------------|----------------------|
| O4D<sub>pNP428</sub> | OE2GLU164    | 90.61              | 2.76                 |
| O2FUC          | ND2ASN163    | 46.81              | 2.89                 |
| O3FUC          | OE2HIE119    | 34.44              | 2.86                 |
| O3FUC          | OE1TRP393    | 16.6               | 2.91                 |
| H3OFUC         | OE1TRP393    | 14.13              | 2.90                 |
| H2OFUC         | ND2ASN163    | 3.96               | 2.93                 |
| H4OFUC         | NE2GLN18     | 1.6                | 2.91                 |
| O2FUC          | OE2GLU164    | 1.17               | 2.76                 |
**Supplementary Table S4.** Analysis over the MD simulation of the most populated hydrogen-bonds between protein residues at the active site. H-bond occupancies are defined as the fraction of frames the bond is present. Only H-bond occupancies above 10% are included. The average distance of the bonds when present between donor and acceptor heavy atoms are also given.

| Acceptor Donor       | H-bond occupancy % | Average Distance (Å) |
|----------------------|--------------------|----------------------|
| OE1GLU164 ND2ASN282  | 92.53              | 2.81                 |
| OE2GLU338 OHTYR284   | 88.99              | 2.76                 |

**Supplementary Table S5.** Analysis over the MD simulation of the most populated hydrogen-bonds between pNP-Fuc substrate and Solvent molecules. H-bond occupancies are defined as the fraction of frames the bond is present. Only H-bond occupancies above 10% are included. The average distance of the bonds when present between donor and acceptor heavy atoms are also given.

| Acceptor Donor | H-bond occupancy % | Average Distance (Å) |
|----------------|--------------------|----------------------|
| O01pNP428 Solvent | 99.64              | 2.80                 |
| O5FUC Solvent   | 24.00              | 2.85                 |
| O4FUC Solvent   | 20.44              | 2.84                 |
| O02pNP428 Solvent | 4.95               | 2.88                 |
| Solvent O4FUC   | 4.35               | 2.81                 |
| O4DpNP428 Solvent | 0.93               | 2.89                 |
| N01pNP428 Solvent | 0.07               | 2.97                 |
Supplementary Table S6. Distances (in Å) between selected atoms involved in the glycosylation step (G), with the QM(large)/MM partition and at the PBE0/TZVP level for the QM description. Subscripts I and II stand for snapshot I and II, respectively.

|                  | Reactant | TS  | Product |
|------------------|----------|-----|---------|
|                  | G_I      | G_II| G_I     | G_II   |
| d(C1FUC-OE2GLU338) | 3.33     | 3.41| 2.61     | 2.67   | 1.50   | 1.50   |
| d(C1FUC-O4D_pNP)  | 1.41     | 1.41| 2.33     | 2.33   | 3.17   | 3.63   |
| d(C1FUC-O5FUC)    | 1.38     | 1.38| 1.24     | 1.24   | 1.35   | 1.36   |
| d(HGLU164-O4D_pNP)| 1.99     | 2.02| 1.82     | 1.88   | 0.98   | 0.97   |
| d(HGLU164-OE2GLU164)| 0.96    | 0.95| 0.97     | 0.96   | 1.64   | 1.67   |
| d(H1TYR284-O5FUC) | 4.08     | 4.05| 2.80     | 2.82   | 2.22   | 2.21   |
| d(H1TYR284-OE2GLU338)| 1.81   | 1.83| 1.97     | 1.87   | 2.29   | 2.31   |
| d(HARG75-OE1GLU338)| 2.02    | 1.96| 2.00     | 2.08   | 2.16   | 2.24   |
| d(H2OFUC-OE1GLU338)| 2.09    | 2.37| 1.82     | 1.85   | 1.78   | 1.78   |
| d(H4OFUC-OGLU392) | 1.96     | 1.89| 2.22     | 2.23   | 2.30   | 2.42   |
| d(HASN163-O2FUC)  | 2.04     | 2.21| 2.00     | 2.12   | 1.97   | 2.06   |

Supplementary Table S7. Distances (in Å) between selected atoms involved in the glycosylation step (G), with the QM(small)/MM partition and at the PBE0/TZVP levels for the QM description. Subscripts I and II stand for snapshot I and II, respectively.

|                  | Reactant | TS  | Product |
|------------------|----------|-----|---------|
|                  | G_I      | G_II| G_I     | G_II   |
| d(C1FUC-OE2GLU338) | 3.28     | 3.29| 2.56     | 2.48   | 1.47   | 1.48   |
| d(C1FUC-O4D_pNP)  | 1.40     | 1.40| 2.31     | 1.76   | 3.52   | 3.66   |
| d(C1FUC-O5FUC)    | 1.39     | 1.40| 1.24     | 1.32   | 1.37   | 1.37   |
| d(HGLU164-O4D_pNP)| 2.07     | 2.49| 1.89     | 1.45   | 0.98   | 0.98   |
| d(HGLU164-OE2GLU164)| 0.95    | 0.96| 0.96     | 1.01   | 1.63   | 1.64   |
| d(H1TYR284-O5FUC) | 4.09     | 4.02| 2.79     | 2.90   | 2.23   | 2.21   |
| d(H1TYR284-OE2GLU338)| 1.79   | 1.74| 1.93     | 1.91   | 2.48   | 2.39   |
| d(HARG75-OE1GLU338)| 1.97    | 2.00| 1.92     | 1.94   | 1.99   | 2.01   |
| d(H2OFUC-OE1GLU338)| 1.89    | 1.96| 1.73     | 1.74   | 1.76   | 1.77   |
| d(H4OFUC-OGLU392) | 1.80     | 1.80| 2.17     | 2.31   | 2.32   | 2.83   |
| d(HASN163-O2FUC)  | 1.98     | 2.54| 1.95     | 2.01   | 1.95   | 1.95   |
**Supplementary Table S8.** Distances (in Å) between selected atoms involved in each reaction step for the reactant, transition state (TS) and product of the hydrolysis (H) and transglycosylation (T_I and T_II, corresponding to the two different frames studied) steps. The results correspond to QM(small)/MM and at the PBE0/TZVP level. The Acc subscript refers to the acceptor water and glucose moieties in hydrolysis and transglycosylation processes, respectively.

|                | Reactant       | TS             | Product        |
|----------------|----------------|----------------|----------------|
|                | H  | T_I | T_II | H  | T_I | T_II | H  | T_I | T_II |
| $d(C1_{FUC}-OE2_{GLU338})$ | 1.47 | 1.48 | 1.47 | 2.99 | 3.34 | 3.35 | 3.18 | 3.36 | 3.36 |
| $d(C1_{FUC}-O_{Acc})$       | 3.67 | 3.38 | 3.38 | 2.25 | 2.15 | 2.17 | 1.39 | 1.41 | 1.42 |
| $d(C1_{FUC}-O5_{FUC})$      | 1.37 | 1.37 | 1.37 | 1.25 | 1.25 | 1.25 | 1.40 | 1.39 | 1.39 |
| $d(H_{Acc}-OE2_{GLU164})$   | 1.78 | 1.76 | 1.71 | 1.45 | 1.44 | 1.40 | 0.96 | 0.96 | 0.96 |
| $d(H_{TYR284}-O5_{FUC})$    | 2.11 | 2.24 | 2.21 | 2.93 | 3.13 | 3.17 | 3.83 | 4.04 | 4.09 |
| $d(H_{TYR284}-OE2_{GLU338})$| 2.50 | 2.49 | 2.45 | 1.89 | 1.85 | 1.85 | 1.77 | 1.76 | 1.76 |
| $d(H_{ARG75}-OE1_{GLU338})$ | 1.97 | 2.00 | 1.98 | 1.99 | 2.10 | 1.98 | 1.99 | 2.06 | 1.92 |
| $d(H_{2O_{FUC}}-OE1_{GLU338})$ | 1.76 | 1.77 | 1.74 | 1.77 | 1.77 | 1.68 | 1.90 | 1.98 | 1.86 |
| $d(H_{4O_{FUC}}-O_{GLU392})$ | 2.41 | 3.03 | 4.25 | 2.21 | 3.18 | 4.24 | 1.84 | 1.88 | 3.51 |
| $d(H_{ASN163}-O2_{FUC})$    | 1.97 | 1.91 | 1.92 | 1.90 | 1.87 | 1.87 | 1.90 | 1.94 | 1.88 |
| $d(H_{3O_{GLC}}-O4_{FUC})$  | -   | 2.99 | 2.96 | -   | 2.52 | 2.56 | -   | 3.23 | 3.12 |
| $d(H_{3O_{GLC}}-O5_{FUC})$  | -   | 3.96 | 4.12 | -   | 3.41 | 3.62 | -   | 1.89 | 2.00 |
| $d(H_{3O_{GLC}}-O4_{GLC})$  | -   | 2.30 | 2.30 | -   | 2.37 | 2.37 | -   | 2.52 | 2.52 |
| $d(O3_{GLC}-H2_{WAT433})$   | -   | 2.02 | 2.17 | -   | 2.52 | 2.19 | -   | 3.23 | 2.14 |
| $d(O_{WAT431}-H2_{WAT432})$ | 1.76 | -   | -   | 1.82 | -   | -   | 2.66 | -   | -   |
| $d(H2_{WAT431}-O_{WAT433})$ | 2.06 | -   | -   | 2.26 | -   | -   | 2.73 | -   | -   |
| $d(H2_{WAT432}-O5_{FUC})$   | 3.64 | -   | -   | 3.36 | -   | -   | 2.02 | -   | -   |
| $d(H1_{WAT432}-O4_{FUC})$   | 2.38 | -   | -   | 2.33 | -   | -   | 2.57 | -   | -   |
**Supplementary Table S9.** Selected NPA atom charges (in a.u) for the large QM region at QM(PBE0/TZVP) level. The Acc subscript refers to the acceptor water and glucose moieties in hydrolysis and transglycosylation processes, respectively.

|       | Reactant |       |       | TS     |       |       | Product |
|-------|----------|-------|-------|--------|-------|-------|---------|
|       |          | H     | T_1   | T_II  | H     | T_1   | T_II   |
| OE1   | -0.89    | -0.90 | -0.89 | -0.88 | -0.88 | -0.87 | -0.74   |
| GLU164|          |       |       |        |       |       |         |
| OE2   | -0.84    | -0.85 | -0.85 | -0.85 | -0.86 | -0.85 | -0.74   |
| GLU164|          |       |       |        |       |       |         |
| H     | 0.51     | 0.50  | 0.50  | 0.53   | 0.53  | 0.53  | 0.54    |
| TYR284|          |       |       |        |       |       |         |
| OE1   | -0.77    | -0.77 | -0.78 | -0.90 | -0.90 | -0.90 | -0.90   |
| GLU338|          |       |       |        |       |       |         |
| OE2   | -0.63    | -0.63 | -0.62 | -0.88 | -0.87 | -0.87 | -0.88   |
| GLU338|          |       |       |        |       |       |         |
| C1    | 0.45     | 0.46  | 0.45  | 0.72   | 0.70  | 0.70  | 0.42    |
| FUC   |          |       |       |        |       |       |         |
| H     | 0.20     | 0.20  | 0.21  | 0.22   | 0.22  | 0.23  | 0.15    |
| OE1   | -0.57    | -0.57 | -0.57 | -0.51  | -0.52 | -0.51 | -0.61   |
| GLU164|          |       |       |        |       |       |         |
| OE2   | -0.62    | -0.63 | -0.62 | -0.91  | -0.90 | -0.90 | -0.90   |
| GLU338|          |       |       |        |       |       |         |
| C1    | 0.50     | 0.50  | 0.51  | 0.53   | 0.53  | 0.54  | 0.52    |
| FUC   |          |       |       |        |       |       |         |
| H     | -1.04    | -0.79 | -0.80 | -1.05  | -0.81 | -0.82 | -0.77   |
| ACC   |          |       |       |        |       |       |         |
| H     | 0.53     | 0.53  | 0.53  | 0.54   | 0.55  | 0.54  | 0.54    |

**Supplementary Table S10.** Selected NPA atom charges (in a.u) for the small QM region at QM(PBE0/TZVP) level. The Acc subscript refers to the acceptor water and glucose moieties in hydrolysis and transglycosylation processes, respectively.

|       | Reactant |       |       | TS     |       |       | Product |
|-------|----------|-------|-------|--------|-------|-------|---------|
|       |          | H     | T_1   | T_II  | H     | T_1   | T_II   |
| OE1   | -0.91    | -0.92 | -0.91 | -0.89 | -0.88 | -0.88 | -0.75   |
| GLU164|          |       |       |        |       |       |         |
| OE2   | -0.85    | -0.86 | -0.86 | -0.85 | -0.86 | -0.86 | -0.74   |
| GLU164|          |       |       |        |       |       |         |
| H     | 0.51     | 0.50  | 0.50  | 0.54   | 0.54  | 0.53  | 0.54    |
| TYR284|          |       |       |        |       |       |         |
| OE1   | -0.74    | -0.75 | -0.76 | -0.89 | -0.89 | -0.90 | -0.89   |
| GLU338|          |       |       |        |       |       |         |
| OE2   | -0.62    | -0.63 | -0.62 | -0.91 | -0.90 | -0.90 | -0.90   |
| GLU338|          |       |       |        |       |       |         |
| C1    | 0.43     | 0.44  | 0.43  | 0.73   | 0.71  | 0.71  | 0.41    |
| FUC   |          |       |       |        |       |       |         |
| H     | 0.21     | 0.21  | 0.22  | 0.24   | 0.23  | 0.24  | 0.16    |
| FUC   |          |       |       |        |       |       |         |
| O     | -0.57    | -0.57 | -0.56 | -0.51  | -0.51 | -0.50 | -0.61   |
| ACC   |          |       |       |        |       |       |         |
| H     | 0.51     | 0.51  | 0.51  | 0.54   | 0.54  | 0.54  | 0.53    |
| ACC   | -1.04    | -0.80 | -0.80 | -1.06  | -0.82 | -0.83 | -0.76   |
| H     | 0.53     | 0.53  | 0.53  | 0.54   | 0.54  | 0.54  | 0.54    |
**Supplementary Figure S1.** RMSD analysis of (A) the protein backbone over 100 ns, excluding the first 4 residues and the last one, and (B) the heavy atoms of the pNP-Fuc substrate.

![Supplementary Figure S1](image)

**Supplementary Figure S2.** Evolution of C\textsubscript{1FUC} - O4D\textsubscript{pNP} (in blue), C\textsubscript{1FUC} - OE2\textsubscript{GLU338} (in orange), H\textsubscript{GLU164} - O4D\textsubscript{pNP} (in purple) and H\textsubscript{GLU164} - OE2\textsubscript{GLU164} (in green) distances along the glycosylation reaction coordinate of G\textsubscript{1} with the QM(small)/MM partition at the (A) PBE0/SVP and (B) PBE0/TZVP levels for the QM description. Distances are in Å.

![Supplementary Figure S2](image)
Supplementary Figure S3. Potential energy profile (in kcal/mol) for the glycosylation step, starting from an optimized, distorted conformation of the fucose ring of ρNP-Fuc substrate. Calculations were performed with QM(PBE0/TZVP)/MM using the large QM region.

Supplementary Figure S4. Structural comparison of transglycosylation substrate (A) before and (B) after performing the refinement protocol involving QM(SCC-DFTB)/MM molecular dynamics described in the Results section for transglycosylation.
Supplementary Figure S5. Free energy profile (in kcal/mol) for the hydrolysis step (blue) and transglycosylation step (red). Calculations were performed by umbrella sampling calculations at the QM(SCC-DFTB)/MM level using the catalytic residues, substrates and one water (in Hydrolysis) as QM region.

Supplementary Figure S6. Histogram from the umbrella sampling simulations of the (A) hydrolysis and (B) transglycosylation steps.