SUPPORTING INFORMATION

*O*-/*N*-/*S*-specificity in glycosyltransferases: from mechanistic understanding to engineering

1,2David Teze\*,¶, 3Joan Coines¶, 2Folmer Fredslund, 3Kshatresh D. Dubey,  2Gonzalo N. Bidart, 4,5,6Paul D. Adams, 4,7John E. Dueber, 1Birte Svensson,3, 8Carme Rovira\*, 1,2,5,6Ditte H. Welner\*

AUTHOR ADDRESS: 1Department of Biotechnology and Biomedicine, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark. 2The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark. 3Departament de Química Inorgànica i Orgànica and Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Barcelona, Spain. ﻿4Department of Bioengineering, University of California, Berkeley, California, USA. 5Joint BioEnergy Institute, Emeryville, California, USA. 6Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. 7Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. 8Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08018, Barcelona, Spain. ¶These authors contributed equally to the study.

**Corresponding Author**

\*[datez@dtu.dk](mailto:datez@dtu.dk), [diwel@biosustain.dtu.dk](mailto:diwel@biosustain.dtu.dk) and [c.rovira@ub.edu](mailto:c.rovira@ub.edu)

CONTENTS:

Materials and Methods

Fig. S1: pH profile of *Pt*UGT1 activity

Fig. S2: RMSD from MD simulations of DCA complexes

Fig. S3: RMSD from MD simulations of DCP complexes

Fig. S4: RMSD from MD simulations of DCT complexes

Fig. S5: Distances analysis from MD simulations of DCA complexes

Fig. S6: Distances analysis from MD simulations of DCP complexes

Fig. S7: Deprotonation of DCP by wild-type and His26Asp variant during QM/MM MD

Fig. S8: Nucleophilic attack angle of the enzymatic reactions

Fig. S9: Distances analysis from MD simulations of DCT complexes

Fig. S10: Structural overview of the active site cavity

Table S1: *k*cat values of *Pt*UGT1 wild-type and variants on DCA, DCP, and DCT

Table S2: Crystallographic statistics

**Materials and Methods**

Plasmids and mutagenesis

Wild-type *Pt*UGT1 was produced from the pTMH307 plasmid (accession no. MF688772), containing the N-terminally His6-tagged *PtUGT1* gene under the T7 promoter and the ampicillin resistance gene. *Pt*UGT1 point mutants were obtained from this plasmid with the Quick-Change site-directed mutagenesis kit (Agilent) following the manufacturer protocol.

Protein expression and purification

*Pt*UGT1 wild-type and variants were expressed in *E. coli* Rosetta(DE3) or BL21R (DE3) and purified on a nickel resin as previously described.1 The final buffer was 25 mM HEPES pH 7.0, 50 mM NaCl, and 1 mM DTT. The proteins were concentrated to 15 mg·mL⁻1 using a 30.000 MWCO Amicon Ultra-15 Centrifugal Filter Unit (EMD Millipore).

Kinetic assays

Acceptor consumption was monitored via reverse phase HPLC, using an Ultimate 3000 Series apparatus (Thermo Scientific) and a kinetex 2.6 µm C18 100 Å 100x4.6 mm analytical column (phenomenex). MilliQ water and acetonitrile containing 0.1% trifluoroacetic acid were used as mobile phases A and B, respectively. Vials containing 1 mL mixtures of 1 mM UDP-Glc, 50 mM HEPES buffer pH 7, 50 mM NaCl, 500 µM acceptor and 10–3500 nM enzyme were analyzed every 10 min for 180 min at 290 nm (DCP, DCA) or 250 nm (DCT) using a multi-step program (0% B until 1.45 min, ramp to 25% B at 1.5 min, ramp to 30% B at 3.5 min, ramp to 100% B at 6.25 min, stay at 100% B until 8 min, decrease to 0% B at 8.01 min, stay at 0% B until 9 min). Monitoring and data handling was operated using the Chromeleon software (Thermo Scientific). *k*cat values were calculated using the linear regression of acceptor concentration against time. Duplicates were performed at different days, from different protein aliquots. We reported no detectable activity if no product peak was observed after 24 h using 10 µM enzyme, which corresponds to activities <10–4 min–1.

Sequence analysis

20.000 sequences were retrieved by BLAST searches from the non-redundant protein database2 using *Pt*UGT1 as a query on February 16th, 2018. The sequences were clustered to limit pairwise sequence identity at 80 % by iterative cd-hit runs.3 Iterative multiple sequence alignments using ClustalΩ4 were performed to increase minimum pairwise sequence identity by 5 % increments until convergence is reached, until a 20 % threshold.

Crystallography

The *Pt*UGT1:UDP-Glc complex structure was obtained by overnight soaking of *Pt*UGT1 crystals - formed in previously identified conditions and identically set up1 - with 10 mM UDP-Glc in water. Briefly, 0.2 μL protein solution was mixed with 0.2 μL crystallization buffer (0.2 M MgCl2\*6H20, 0.1 M HEPES pH 7.5, 25% PEG 3350) in sitting drops, using the Phoenix crystallization robot (Art Robbins), and equilibrated against 50 μL crystallization buffer. Crystals grew to their final size in 4 days before 0.3 μL of a 10 mM UDP-Glc solution was added. The next day, crystals were mounted in nylon cryoloops (Hampton Research) and cryoprotected by briefly transferring to a drop consisting of 9 μL mother liquor and 1 μL glycerol. 270o of data were collected at 100 K, 0.9765 Å, at the Berkeley Center for Structural Biology beamline 5.0.3 of the Advanced Light Source in Berkeley, California, with a 1o oscillation and 10 s exposure time. The resolution at the edge of the detector was 2.5 Å and data could be processed to 2.4 Å with a completeness above 90%, using xia25 with XDS.6

For the *Pt*UGT:DCA complex structure, identical buffers and protein concentration were used, but the drops were set up by hand with 2 μL protein solution and 2 μL crystallization buffer. After the crystals had grown for several days, the sealing tape was removed and a small piece of DCA powder was placed on top of the drop, before resealing the drops. Crystals were harvested after 3–5 days of incubation. Cryoprotection was achieved by addition of a mixture of 100% PEG 400 and mother liquor in 1:3 ratio to the drops before harvesting the crystals. A total of 700 images with a 0.2o oscillation were collected to 2.75 Å resolution at 100 K, 0.9677 Å on MASSIF-3 at the ESRF, France. Data were processed with xia25 using DIALS.7

The structure of *Pt*UGT1 with indoxyl sulfate (PDB ID: 5NLM) was used as the starting model. For *Pt*UGT1:UDP-Glc, the unit cell and space group were almost identical to those of 5NLM and thus this model could be used directly with a first step of rigid body refinement in PHENIX.refine.8 For *Pt*UGT1:DCA, molecular replacement with PHASER9 was used to solve the phase problem. To complete the models, iterative rounds of model building with Coot10 and refinement and validation with PHENIX.refine8 were performed. For statistics on data processing and refinement see Table S2.

Molecular Dynamics (MD)

A *Pt*UGT1 monomer from the X-ray crystal structure of *Pt*UGT1 wild-type (PDB ID: 5NLM, two chains in the asymmetric unit) was used as the starting structure. MD simulations were performed with the Amber 18 GPU code,11 using the Amber ff14SB force field12 for the protein, TIP3P model13 for water and ions and GLYCAM0614 for the carbohydrate moiety. The protonation state of all residues was assigned considering the experimental pH of 7. The protonation of His residues was further assessed according their chemical environment. Specifically, His residues 26, 155, 300 and 382 were considered neutral with their proton located at Nδ; while His residues 15, 183, 204, 205, 233, 361 and 370 were protonated at Nε. Asp122 was considered as protonated in the simulations of the His26Asp and His26Glu variant, because its closely located to the acidic residue at position 26. Crystallographic water molecules were retained. Additional water molecules were added to solvate the protein. The partial atomic charges for the substrate DCA, DCT and DCP were calculated using RESP charge method for a QM optimized geometry using Gaussian 09 package at HF/6-31G\* level of theory.15 The crystal structure contains the holoenzyme with substrate DCA and the same site was also used for other two substrate DCT and DCP. Missing hydrogens were added by Tleap module of Amber 18. In this way we prepared three holoenzymes with different substrates i.e. DCA, DCT and DCP. For each WT complex, we prepared two different mutants for His26Asp and His26Glu by *in silico* mutating the histidine into the corresponding amino acid. For DCP we also prepared complex for His26Phe mutant. We used the deprotonated sulfur for DCT (p*K*a(DCA)=2.97).

Once the parameterizations and system setup were completed, the resulting structures were minimized using steepest descent and subsequently conjugate gradient algorithm with 5000 steps for each. The system was then gently annealed from 10 to 300 K under the isothermal (NVT) ensemble for 50 ps with a weak restraint of 5 kcal·mol−1·Å−2 applied on proteins and ligands, while keeping water and ions relaxed. This was followed by 1 ns of density equilibration in the isothermal-isobaric (NPT) ensemble at a target temperature of 300 K and a target pressure of 1.0 atm. We applied Langevin thermostat16 and the Monte-Carlo barostat17 with collision frequency of 2 ps and a pressure relaxation time of 1 ps with a weak restraint of 1 kcal·mol−1·Å−2. After density equilibrations all the restraints which were applied during heating and density dynamics were removed. The production dynamics was performed in two steps; an initial 3 ns production dynamics to get a well-settled pressure and temperature, and subsequently a final 100 ns for results analysis. The results from the initial 3 ns were discarded. The hydrogen bonds were constrained using the SHAKE algorithm,18 while the long-range electrostatic interactions were treated by Particle Mesh Ewald. Snapshot of the MD simulations in which the acceptor nucleophilic atom is close to the donor sugar was chosen as starting point for the forthcoming QM/MM simulations.

Quantum mechanics / Molecular mechanics (QM/MM) MD

After classical MD, QM/MM MD calculations were performed for the equilibrated structures of DCP in both wild-type and His26Asp variant; and DCT in the His26Asp mutant active site. The method developed by Laio et al19 was employed, within CPMD program, which couples classical MD with Car–Parrinello MD, based on DFT. For DCP catalysis, the QM region included the acceptor molecule (DCP), part of the donor (Glc and phosphate groups from UDP), the side chain of Asp26 and, only in the case of the wild-type enzyme, the side chain of Asp122. This sums up a total number of 67, 55 and 47 QM atoms for DCP:wt, DCP:His26Asp and DCT:His26Asp simulations, respectively, including capping hydrogens. Whereas in the DCT:His26Asp mutant complex just the C5’ atom was capped, in the DCP:His26Asp system the valence of Cα of Asp26 side chain was also saturated. In addition, the Cα of Asp122 side chain was saturated too for DCP-wt. The MM region included the protein (excluding QM atoms), part of the UDP moiety and the solvent, for a total number of 76920, 76927 and 76980 atoms for DCP-wt, DCP:His26Asp and DCT:His26Asp ternary complexes, respectively. The electronic structure was described employing the Density Functional Theory (DFT) level of theory. Norm-conserving Troullier-Martins pseudopotentials20 were used for all elements and Kohn-Sham orbitals were expanded in a plane-wave basis set with a kinetic energy cutoff of 70 Ry. Perdew, Burke and Ernzerhoff generalized gradient-corrected approximation (PBE)21 was used within the DFT calculations, which has given a good performance in previous works on glycosyl transferases.22,23 All systems were equilibrated according the following protocol. First, the structure was optimized with annealing of the nuclei until a maximal component of the nuclear gradient reached a value of 5·10-4 a.u. Afterwards, 3 ps of Car-Parrinello MD were performed at 300 K using the Nosé-Hoover thermostat.24,25 A fictitious electron mass of 800 a.u. and a time step of 0.12 fs were used. A snapshot of these simulations was chosen as starting point for the QM/MM metadynamics simulations.

QM/MM metadynamics

Metadynamics26 calculations were performed using the PLUMED plugin27 to compute the enzymatic reactions of both DCP and DCT. The collective variable (CV) employed to drive the reactions was the difference between the following two distances: (1) the distance between the nucleophilic atom (O or S, for DCP and DCT, respectively) and the anomeric carbon (C1) of Glc; (2) the distance between the C1 atom of Glc and the (initially bound) oxygen atom of the UDP phosphate. Heights and width of 1.0 kcal·mol−1 and 0.15 Åwere used, respectively, according to the oscillations observed in free dynamics. The deposition time was set to 350 MD steps. The simulations where stopped after one re-crossing event, following literature recommendations for chemical reactions.29 The reaction mechanism and the corresponding free energy profiles are shown in Figures 3-4 of the main text. The analysis of the nucleophilic attack angle is provides in Fig. S8

|  |  |  |
| --- | --- | --- |
| A | B | C |
| D  C:\Users\datez\AppData\Local\Microsoft\Windows\INetCache\Content.Word\20201103__ONS-kinetics_DCA.png | EC:\Users\datez\AppData\Local\Microsoft\Windows\INetCache\Content.Word\20201103__ONS-kinetics_DCP.png | FC:\Users\datez\AppData\Local\Microsoft\Windows\INetCache\Content.Word\20201103__ONS-kinetics_DCT.png |

Fig S1. Top, pH profiles of the activity of *Pt*UGT1 wild-type against DCA (A), DCP (B) and DCT (C). The activities are measured in duplicate in different buffers at different pH values (50 mM Citrate-phosphate; 50 mM HEPES; 50mM Tris-HCl; 50 mM Glycine NaOH), and normalized to the activity in the optimal condition. Error bars indicate standard deviations. Tris has an inhibitory effect on DCA and DCP, but not on DCT. Bottom, reaction monitoring of *Pt*UGT1 variants against DCA (D), DCP (E) and DCT (F).

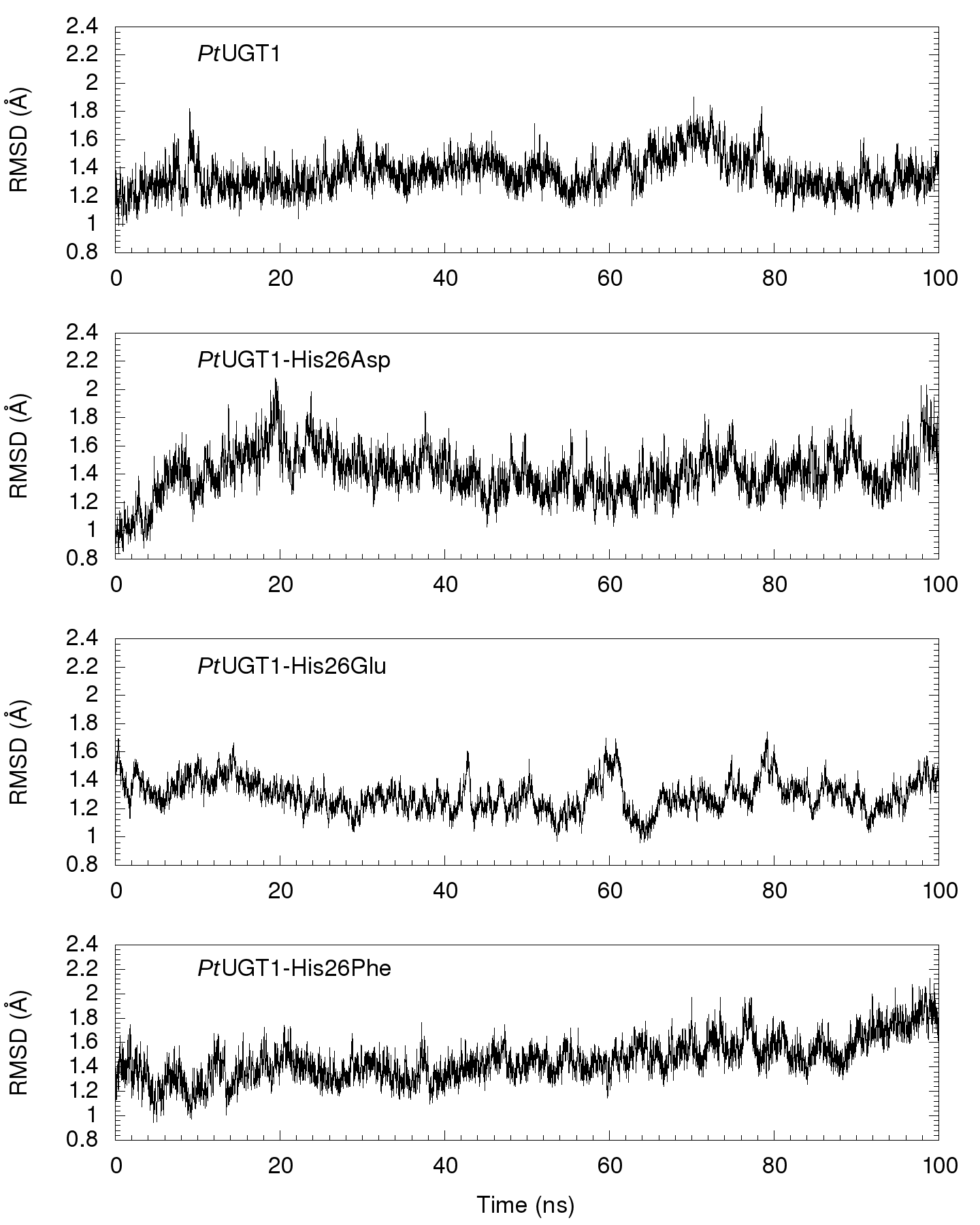


Fig S2. Root-mean-square deviation (RMSD) of the protein backbone during the production runs from simulations of DCA bound to the wild-type *Pt*UGT1 and its His26Asp, His26Glu and His26Phe variants.

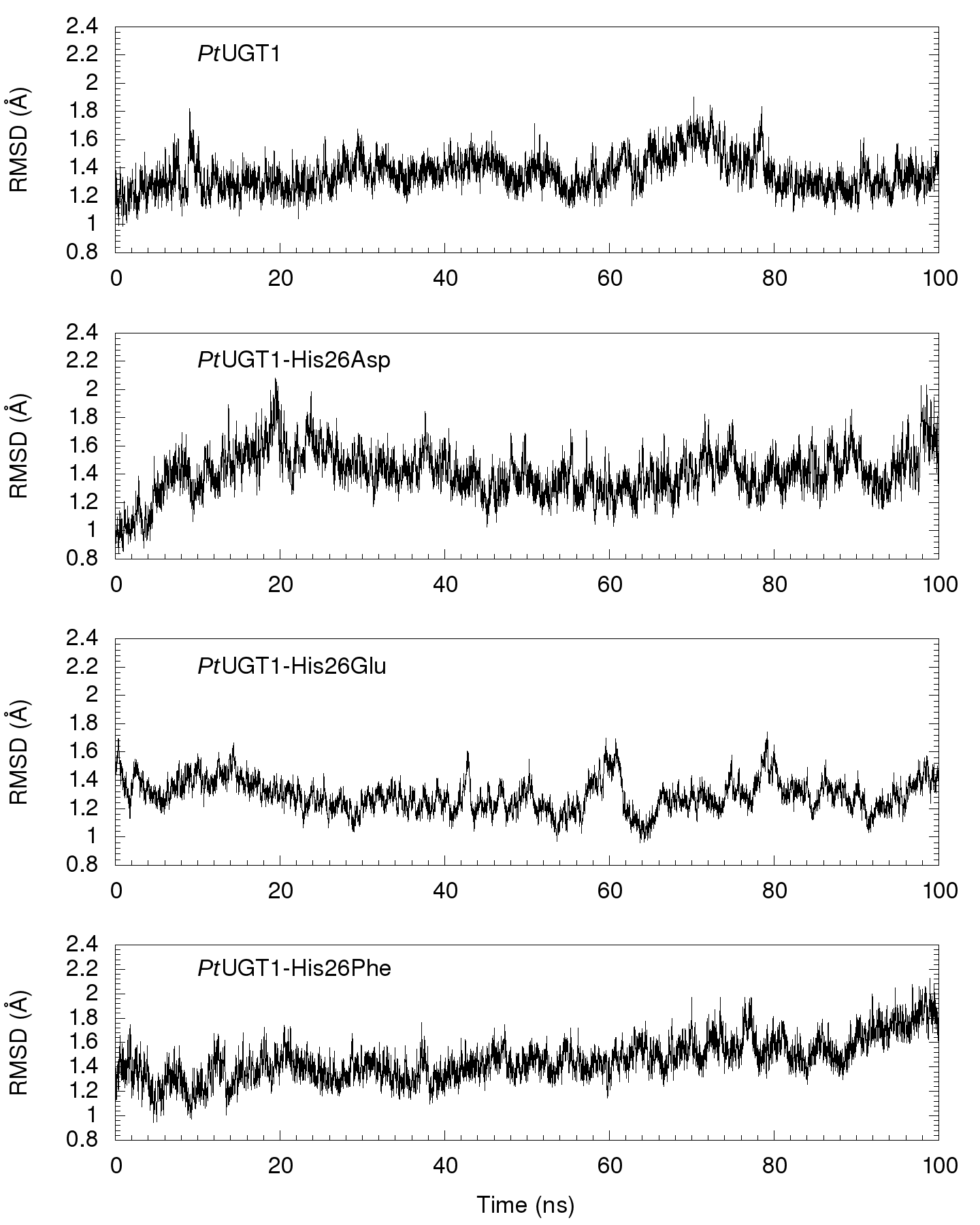


Fig S3. RMSD of the protein backbone during the production runs from simulations of DCP bound to the wild-type *Pt*UGT1 and its His26Asp and His26Glu variants.

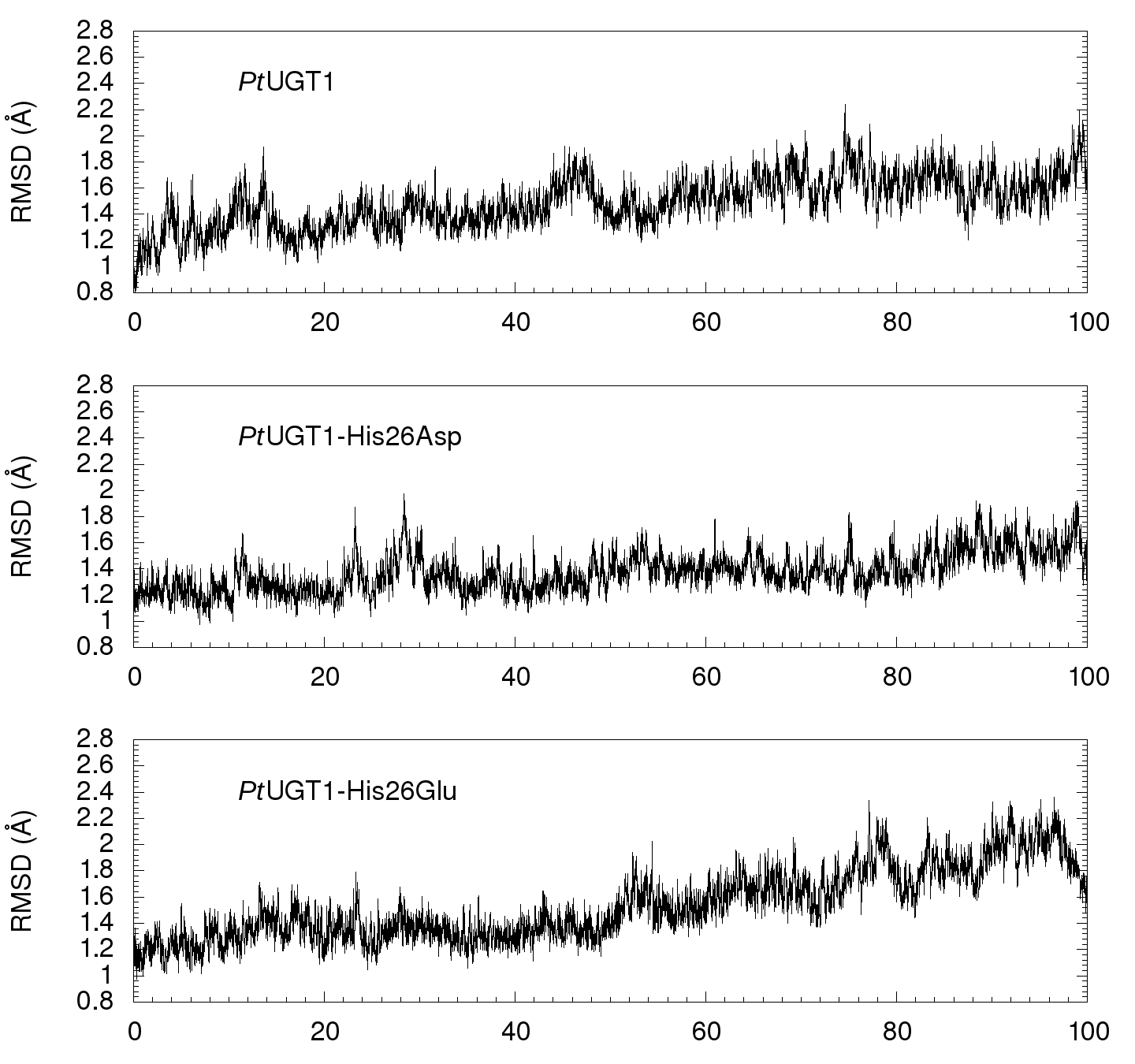


Fig S4. RMSD of the protein backbone during the production runs from simulations of DCT bound to the wild-type *Pt*UGT1 and its His26Asp and His26Glu variants.

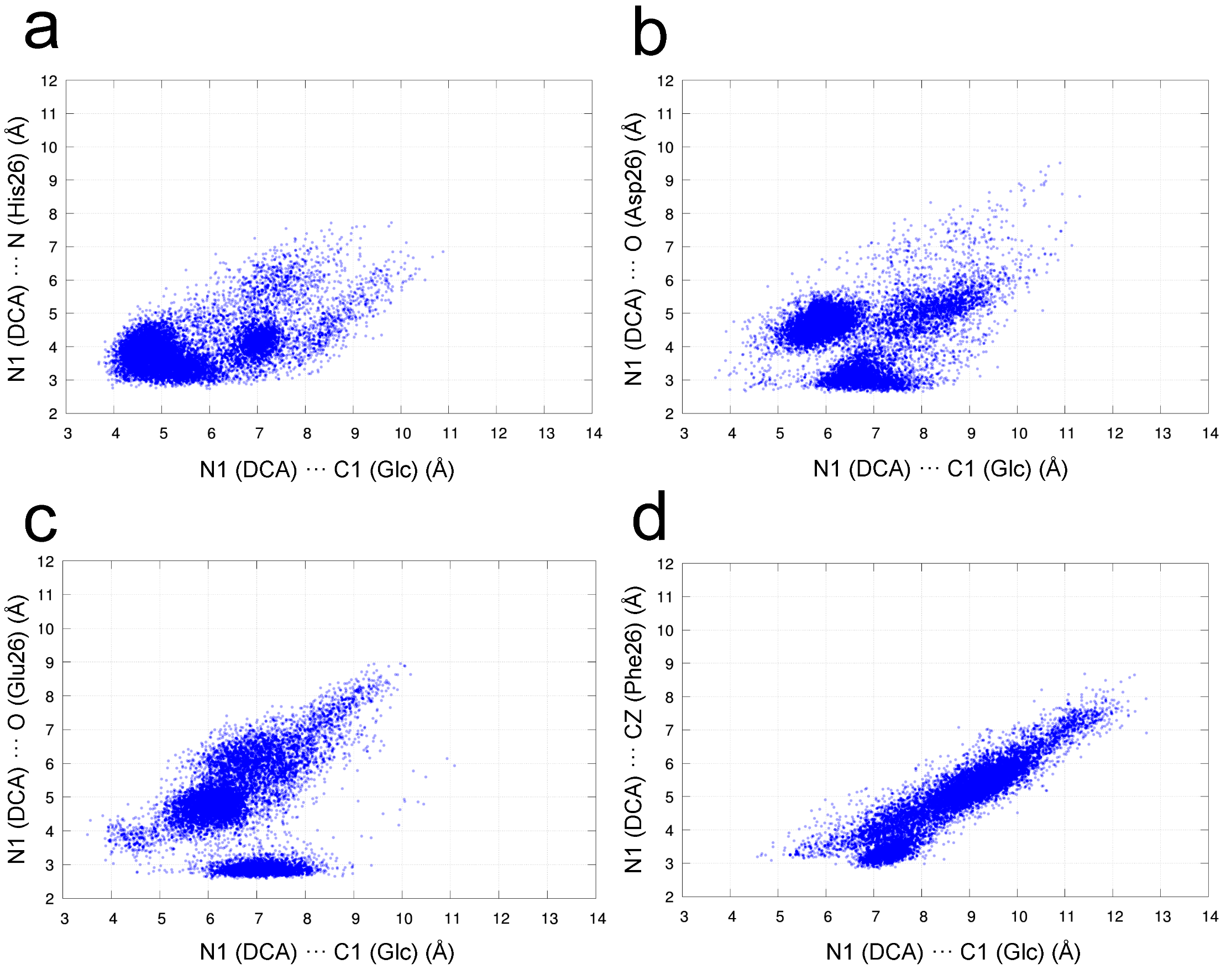


Fig S5. Scatter plots of the main active site distances that are relevant for *Pt*UGT1 catalysis: nucleophilic attack (x axis) and the interaction between DCA and the residue located at position 26 (y axis). a) wild-type; b) His26Asp; c) His26Glu; d) His26Phe.

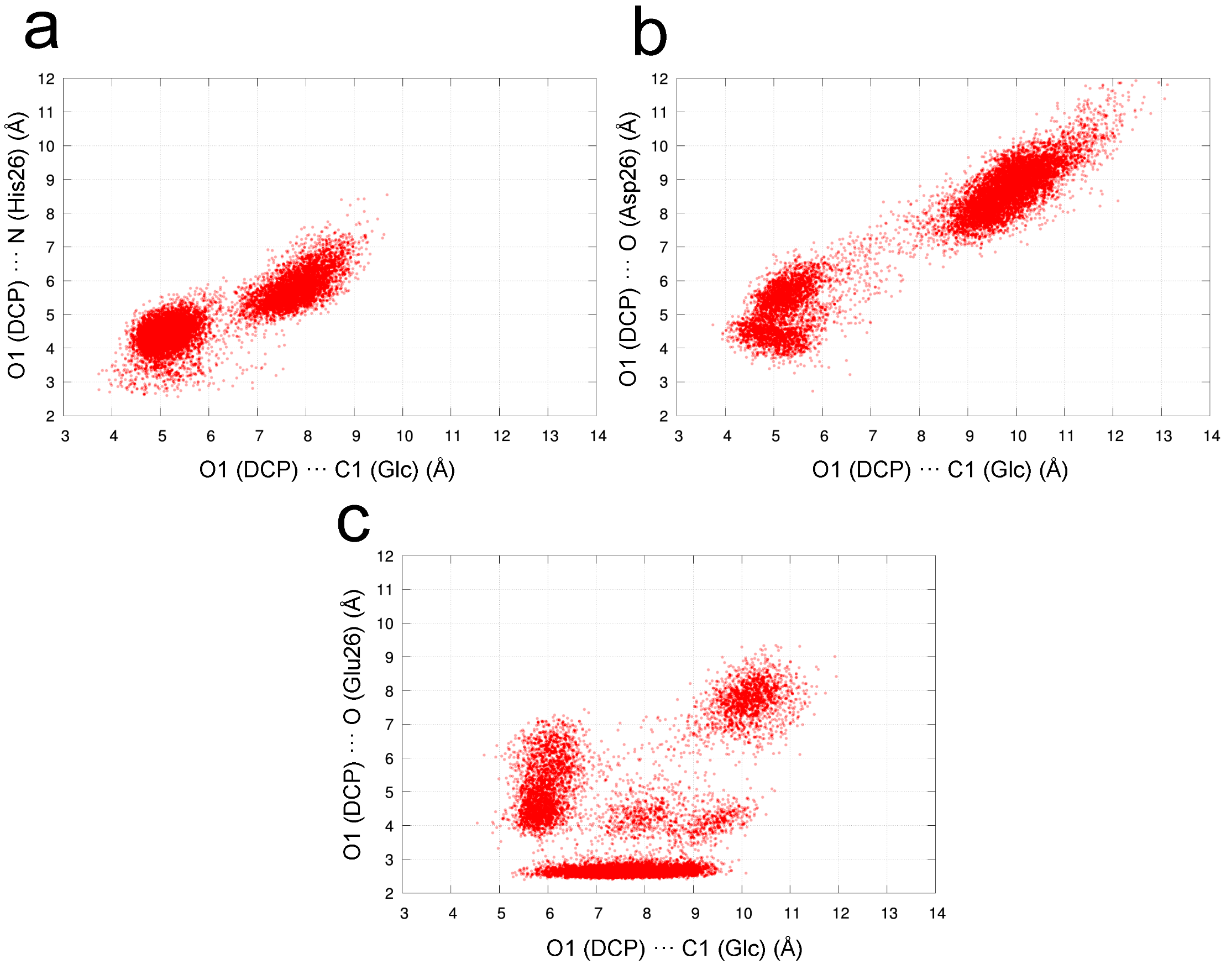


Fig S6. Scatter plots of the main active site distances that are relevant for *Pt*UGT1 catalysis: nucleophilic attack (x axis) and the interaction between DCP and the residue located at position 26 (y axis). a) wild-type; b) His26Asp; c) His26Glu.

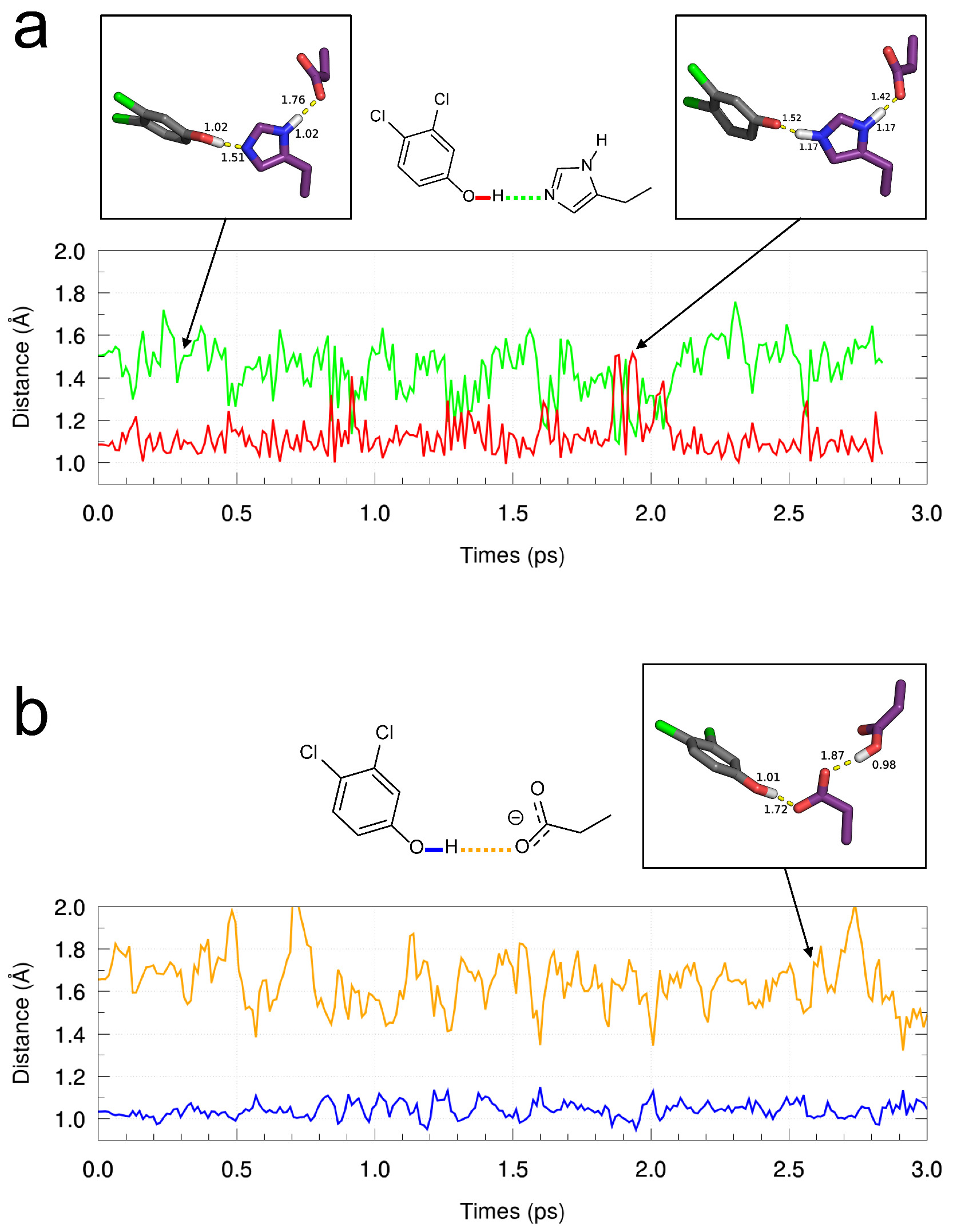


Fig S7. Bond distances involving the hydroxyl group of DCP and either a) the N atom of His26 or b) the O atom of Asp26 (wild-type *Pt*UGT1 and its His26Asp variant, respectively), during the QM/MM MD equilibration calculations.

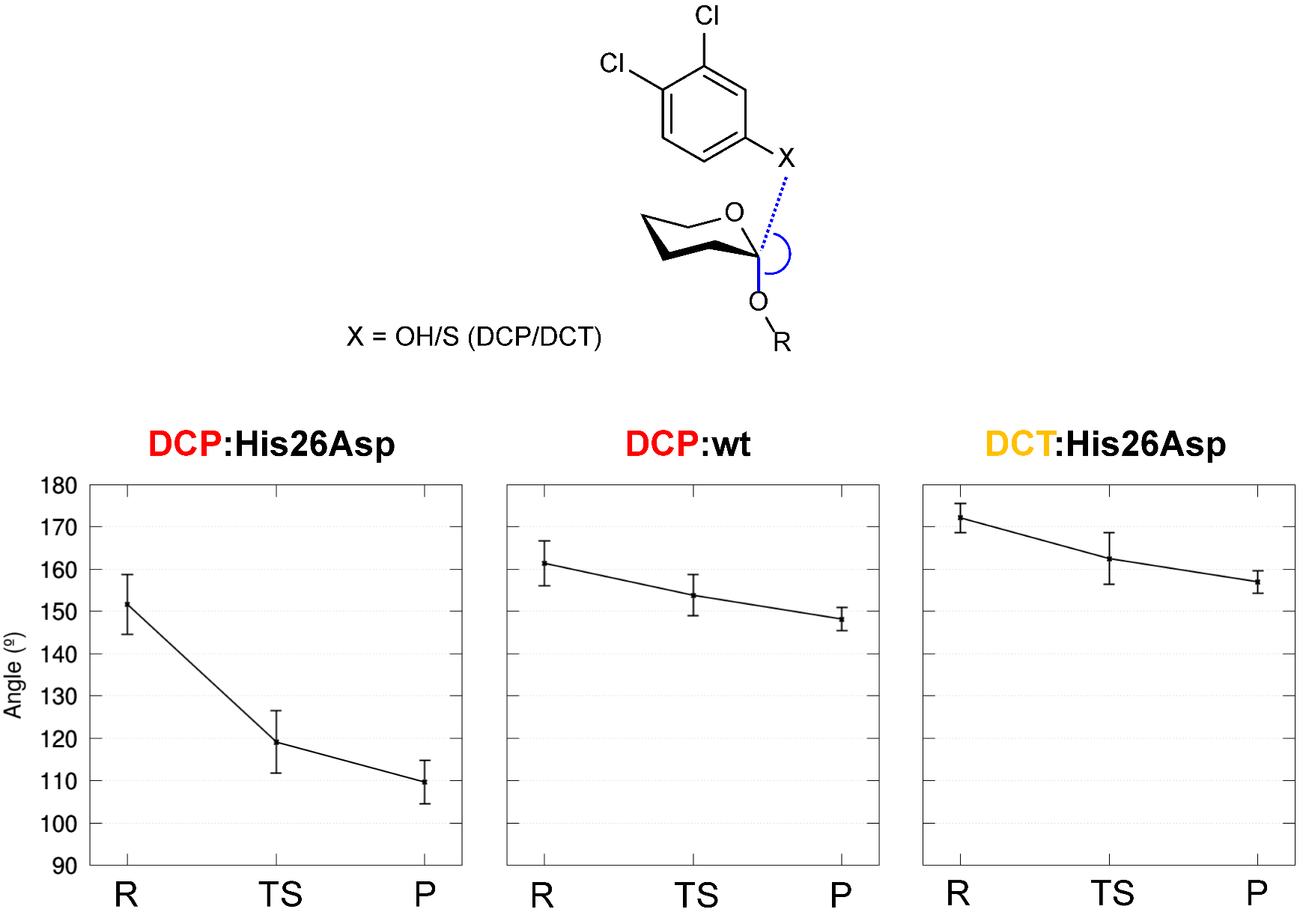


Fig S8. Angle formed by the nucleophilic atom (O/S depending on the substrate DCP/DCT, respectively), the anomeric carbon C1 from Glc and the oxygen from UDP (shown in blue) obtained along the enzymatic reactions obtained by QM/MM metadynamics. Error bars show the standard deviation.

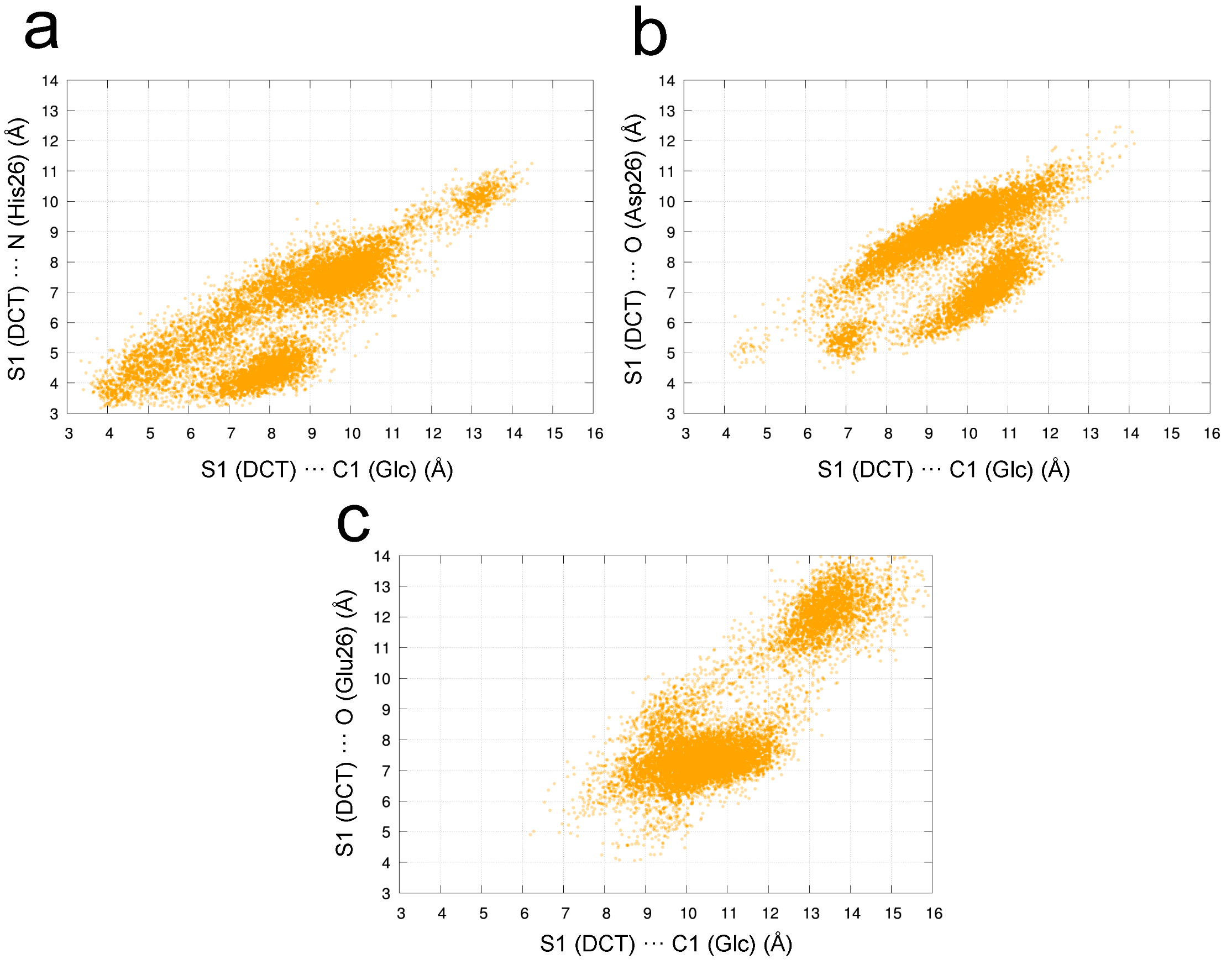


Fig S9. Scatter plots of the main active site distances that are relevant for *Pt*UGT1 catalysis: nucleophilic attack (x axis) and the interaction between DCT and the residue located at position 26 (y axis). a) wild-type; b) His26Asp; c) His26Glu.

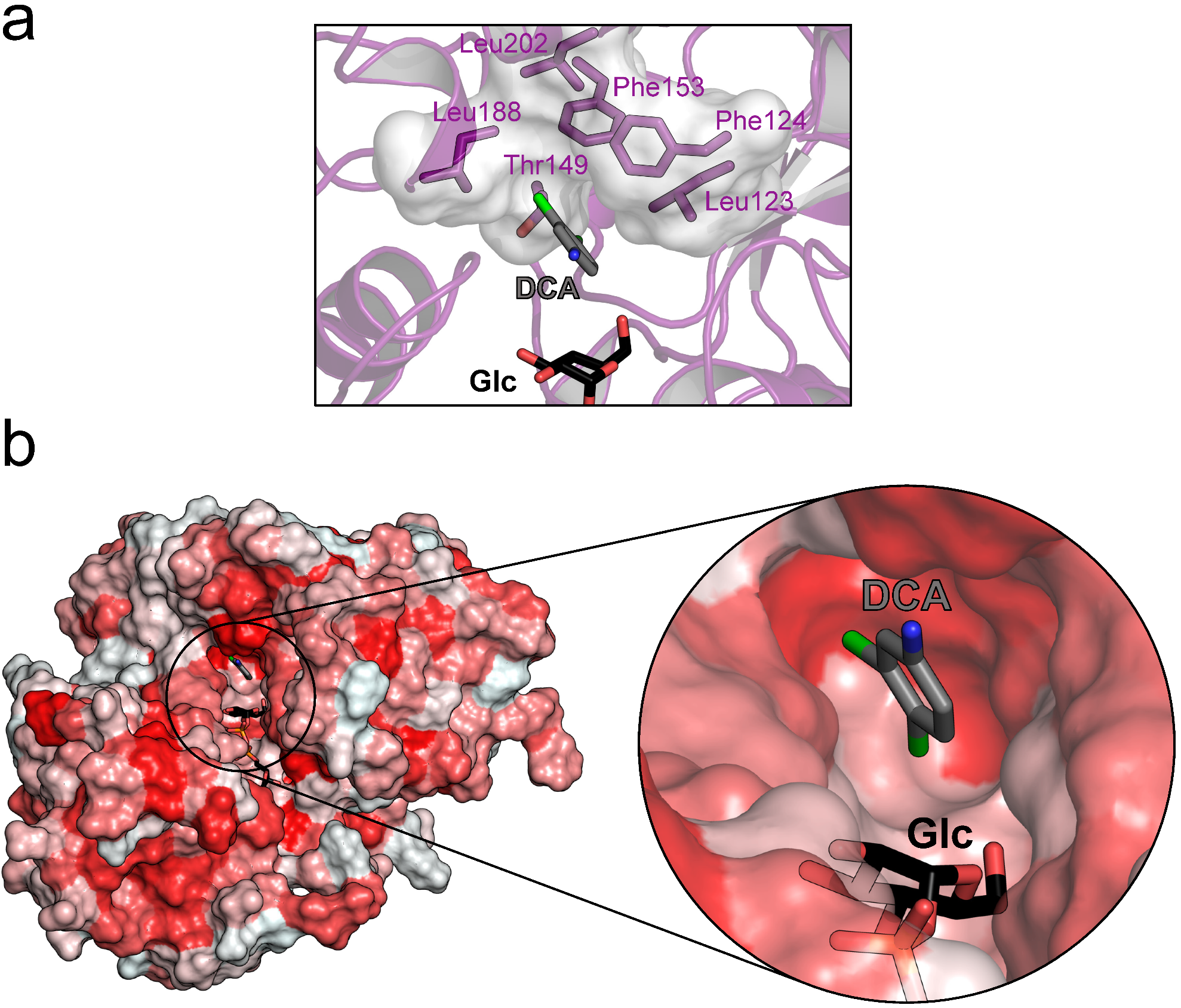


Fig S10. View of the *Pt*UGT1 active site cavity (region above donor glucose) in the ternary complex of *Pt*UGT1 with DCA and Glc-UDP (X-ray crystallography, this work). a) The residues forming an hydrophobic environment that allows conformational exploration of the acceptor aromatic ring (shown in both surface and sticks representation). b) Surface representation of *Pt*UGT1 with amino acids colored according the normalized consensus hydrophobicity scale28 (hydrophobic residues in red). Note that the orientation in b) is different from a).

Table S1. Catalytic constants of *Pt*UGt1 and its variants against DCA, DCP and DCT. Stdev: standard deviation between duplicates.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DCA | | DCP | | DCT | |
| *k*cat (min−1) | stdev | *k*cat (min−1) | stdev | *k*cat (min−1) | stdev |
| WT | 71.5 | 3.8 | 54.8 | 5.4 | 30.0 | 2.0 |
| His26Asp | 10.0 | 0.6 | < 10–4 | | 11.7 | 1.6 |
| His26Glu | 292.6 | 66.3 | < 10–4 | | 6.1 | 0.6 |
| His26Phe | 0.06 | 0.01 | < 10–4 | | 0.47 | 0.04 |
| His26Gln | 35.6 | 3.3 | < 10–4 | | 21.5 | 1.1 |
| His26Ala | 0.96 | 0.01 | < 10–4 | | < 10–4 | |
| Asp122Asn | 0.06 | 0.02 | < 10–4 | | 16.0 | 1.6 |

Table S2. Crystallography statistics.

|  |  |  |
| --- | --- | --- |
|  | ***Pt*UGT1:UDP-Glc** | ***Pt*UGT1:DCA** |
| PDB code | 6SU6 | 6SU7 |
| **Data collection** |  |  |
| Wavelength | 0.9765 | 0.9677 |
| Resolution range | 52.36 - 2.4 (2.49 - 2.4) | 48.9 - 2.75 (2.84 - 2.75 ) |
| Space group | *P*21212 | *P*212121 |
| Unit cell |  |  |
| a,b,c (Å) | 121.2 174.2 48.5 | 93.4 119.4 170.6 |
| Total reflections | 423418 (31479) | 265040 (26713) |
| Unique reflections | 40705 (3679) | 50246 (4826) |
| Multiplicity | 10.4 (8.6) | 5.3 (5.5) |
| Completeness (%) | 98.8 (91.0) | 99.3 (97.1) |
| Mean I/σ(I) | 17.13 (1.98) | 5.10 (1.27) |
| Wilson B-factor | 47.88 | 62.47 |
| Rmerge | 0.106 (1.15) | 0.152 (0.82) |
| Rmeas | 0.112 (1.23) | 0.169 (0.91) |
| Rpim | 0.034 (0.403) | 0.073 (0.382) |
| CC1/2 | 0.999 (0.657) | 0.995 (0.538) |
| CC\* | 1 (0.89) | 0.999 (0.837) |
| **Refinement** |  |  |
| Reflections used in refinement | 40705 (3672) | 50246 (4822) |
| Reflections used for Rfree | 2037 (183) | 2471 (227) |
| Rwork | 0.231 (0.301) | 0.256 (0.375) |
| Rfree | 0.263 (0.329) | 0.287 (0.373) |
| CC(work) | 0.938 (0.745) | 0.937 (0.630) |
| CC(free) | 0.933 (0.707) | 0.891 (0.632) |
| Number of non-hydrogen atoms | 7274 | 13820 |
| macromolecules | 7080 | 13811 |
| ligands | 72 | 9 |
| solvent | 122 | None |
| Protein residues | 922 | 1794 |
| RMS(bonds) | 0.006 | 0.002 |
| RMS(angles) | 0.60 | 0.53 |
| Ramachandran favored (%) | 94.20 | 94.53 |
| Ramachandran allowed (%) | 5.80 | 5.41 |
| Ramachandran outliers (%) | 0.00 | 0.06 |
| Rotamer outliers (%) | 5.42 | 1.78 |
| Clashscore | 3.49 | 3.68 |
| Average B-factor | 78.69 | 91.29 |
| macromolecules | 79.15 | 91.31 |
| ligands | 76.28 | 54.14 |
| solvent | 53.79 | None |
| Number of TLS groups | 6 | 14 |

Table S3. Mutagenesis primers.

|  |  |  |
| --- | --- | --- |
| Mutation | Forward primer | Reverse |
| H26A | GCATCGTCTCACTCATCCCCCTCGCCGAG | ATGCCGTCTCATGAGGGCGCCCATGCCGGCG |
| H26E | GCATCGTCTCACTCATCCCCCTCGCCGAG | ATGCCGTCTCATGAGTTCGCCCATGCCGGCG |
| H26Q | GCATCGTCTCACTCATCCCCCTCGCCGAG | ATGCCGTCTCATGAGTTGGCCCATGCCGGC |
| H26F | ACGAACGGUGGTGGAGCGGTGGTTGGT | ACCGTTCGUCATAATCGTGCCCTCCGC |
| H26D | ACGTCCGGUGGTGGAGCGGTGGTTGGT | ACCGGACGUCATAATCGTGCCCTCCGC |
| D122N | ATGAGGTUGCGGAGCGAGGGGAGGGA | AACCTCAUTGCCTCCTACTCCGCCTC |

References:

(1) Hsu, T. M.; Welner, D. H.; Russ, Z. N.; Cervantes, B.; Prathuri, R. L.; Adams, P. D.; Dueber, J. E. Employing a Biochemical Protecting Group for a Sustainable Indigo Dyeing Strategy. *Nat. Chem. Biol.* **2018**, *14*, 256–261.

(2) Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2004**, *32* (Database issue), D115-9.

(3) Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. CD-HIT Suite: A Web Server for Clustering and Comparing Biological Sequences. *Bioinformatics* **2010**, *26*, 680–682.

(4) Sievers, F.; Higgins, D. G. Clustal Omega. *Curr. Protoc. Bioinforma.* **2014**, *2014* (December), 3.13.1-3.13.16.

(5) Winter, G. Xia2: An Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Crystallogr.* **2010**, *43*, 186–190.

(6) Kabsch, W.; T., B. A.; K., D.; A., K. P.; K., D.; S., M.; G., R. R. B.; P., E.; S., F.; K., W.; et al. *XDS*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132.

(7) Winter, G.; Waterman, D. G.; Parkhurst, J. M.; Brewster, A. S.; Gildea, R. J.; Gerstel, M.; Fuentes-Montero, L.; Vollmar, M.; Michels-Clark, T.; Young, I. D.; et al. DIALS: Implementation and Evaluation of a New Integration Package. *Acta Crystallogr. Sect. D Struct. Biol.* **2018**, *74*, 85-97.

(8) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68*, 352–367.

(9) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

(10) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501.

(11) DA. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGran, D. M. Y. and P. A. K. AMBER 18. **2018**, University of California, San Francisco.

(12) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(13) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.

(14) Kirschner, K. N.; Yongye, A.; Tschampel M., S.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655.

(15) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had, J. C. and D. J. F. Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford CT. **2013**.

(16) Lzaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin Stabilization of Molecular Dynamics. *J. Chem. Phys.* **2001**, *114*, 2090–2098.

(17) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(18) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(19) Laio, A.; VandeVondele, J.; Rothlisberger, U. A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car-Parrinello Molecular Dynamics Simulations. *J. Chem. Phys.* **2002**, *116*, 6941–6947.

(20) Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. II. Operators for Fast Iterative Diagonalization. *Phys. Rev. B* **1991**, *43*, 8861–8869.

(21) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(22) Ardevol, A.; Iglesias-Fernandez, J.; Rojas-Cervellera, V.; Rovira, C. The Reaction Mechanism of Retaining Glycosyltransferases. *Biochem. Soc. Trans.* **2016**, *44*, 51–60.

(23) Bilyard, M. K.; Bailey, H. J.; Raich, L.; Gafitescu, M. A.; Machida, T.; Iglésias-Fernández, J.; Lee, S. S.; Spicer, C. D.; Rovira, C.; Yue, W. W.; et al. Palladium-Mediated Enzyme Activation Suggests Multiphase Initiation of Glycogenesis. *Nature* **2018**, *563*, 235–240.

(24) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(25) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(26) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(27) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(28) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *J. Mol. Biol.* **1984**, *179*, 125–142.

(29) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L.A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions *J. Phys. Chem. B* **2005**, 109, 6676-6687.