



Full wwPDB X-ray Structure Validation Report ⓘ

Jul 29, 2024 – 02:54 PM EDT

PDB ID : 9BCM
Title : Crystal structure of the glycosyltransferase UGT95A1
Authors : Pereira, J.H.; Sirirungruang, S.; Shih, P.M.; Adams, P.D.
Deposited on : 2024-04-09
Resolution : 2.36 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.37.1
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.37.1

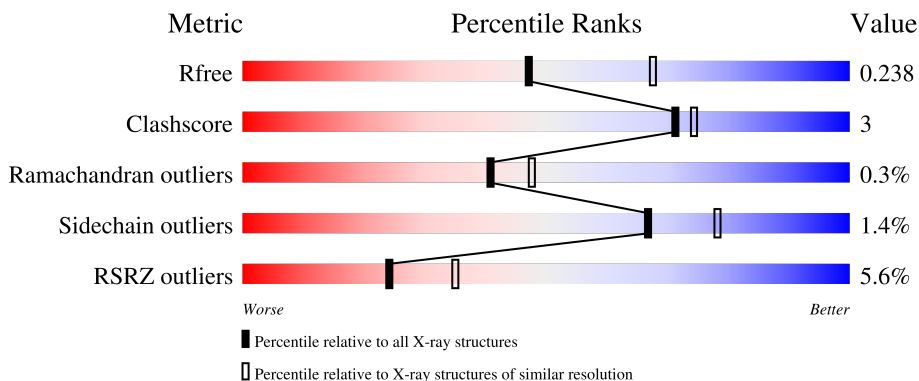
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.36 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	548	

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 6399 atoms, of which 3128 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Glycosyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	409	6302	2043	3128	525	592	14	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP B2CZL6
A	2	PRO	-	expression tag	UNP B2CZL6
A	3	LEU	-	expression tag	UNP B2CZL6

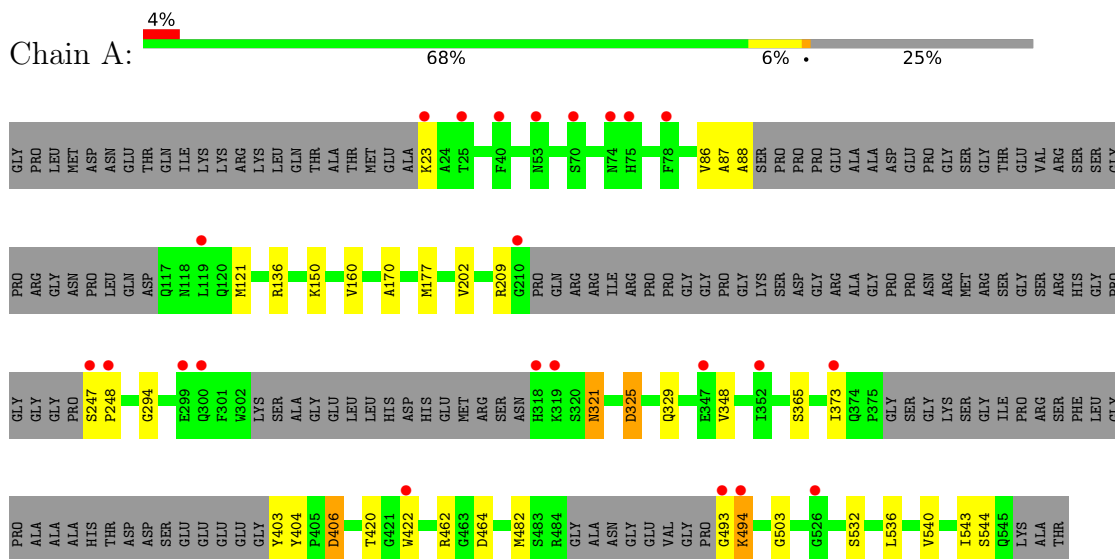
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	97	Total	O	0	0
			97	97		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Glycosyltransferase



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α , β , γ	62.52Å 92.22Å 92.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.11 – 2.36 46.11 – 2.36	Depositor EDS
% Data completeness (in resolution range)	100.0 (46.11-2.36) 100.0 (46.11-2.36)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.29 (at 2.37Å)	Xtrriage
Refinement program	PHENIX 1.21_5207	Depositor
R, R_{free}	0.205 , 0.236 0.206 , 0.238	Depositor DCC
R_{free} test set	1999 reflections (8.81%)	wwPDB-VP
Wilson B-factor (Å ²)	39.0	Xtrriage
Anisotropy	0.134	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.40 , 38.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.016 for -h,l,k	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6399	wwPDB-VP
Average B, all atoms (Å ²)	54.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z > 5$	RMSZ	$\# Z > 5$
1	A	0.37	0/3256	0.50	1/4413 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	209	ARG	NE-CZ-NH1	5.91	123.26	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	406	ASP	Peptide

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3174	3128	3126	22	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	97	0	0	3	0
All	All	3271	3128	3126	22	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 3.

All (22) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:464:ASP:OD2	2:A:601:HOH:O	2.10	0.68
1:A:150:LYS:HG3	1:A:160:VAL:HG11	1.80	0.63
1:A:177:MET:HG2	1:A:202:VAL:HG21	1.85	0.59
1:A:329:GLN:HG2	2:A:647:HOH:O	2.07	0.53
1:A:86:VAL:HG21	1:A:121:MET:HG2	1.91	0.53
1:A:543:ILE:HD12	1:A:544:SER:N	2.25	0.52
1:A:136:ARG:HH11	1:A:136:ARG:HG3	1.74	0.52
1:A:373:ILE:HD11	1:A:403:TYR:O	2.12	0.50
1:A:536:LEU:O	1:A:540:VAL:HG23	2.13	0.49
1:A:87:ALA:O	1:A:88:ALA:HB2	2.15	0.47
1:A:403:TYR:CG	1:A:404:TYR:N	2.84	0.46
1:A:294:GLY:O	1:A:532:SER:OG	2.29	0.44
1:A:170:ALA:CB	1:A:464:ASP:HB2	2.48	0.43
1:A:23:LYS:N	2:A:614:HOH:O	2.51	0.43
1:A:482:MET:CE	1:A:503:GLY:HA3	2.49	0.42
1:A:321:ASN:HB2	1:A:420:THR:H	1.84	0.42
1:A:543:ILE:HD12	1:A:543:ILE:C	2.39	0.42
1:A:348:VAL:HG21	1:A:462:ARG:HD2	2.03	0.41
1:A:325:ASP:OD1	1:A:325:ASP:N	2.52	0.41
1:A:247:SER:HB3	1:A:248:PRO:HD3	2.01	0.41
1:A:247:SER:N	1:A:248:PRO:CD	2.83	0.41
1:A:493:GLY:O	1:A:494:LYS:HB2	2.21	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	397/548 (72%)	387 (98%)	9 (2%)	1 (0%)	41	47

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	494	LYS

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	348/452 (77%)	343 (99%)	5 (1%)	67	78

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	321	ASN
1	A	325	ASP
1	A	365	SER
1	A	406	ASP
1	A	422	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	409/548 (74%)	0.46	23 (5%) 24 35	30, 46, 78, 115	0

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	493	GLY	5.0
1	A	247	SER	4.8
1	A	318	HIS	4.2
1	A	23	LYS	3.7
1	A	347	GLU	3.6
1	A	40	PHE	3.3
1	A	299	GLU	3.3
1	A	119	LEU	3.2
1	A	25	THR	3.0
1	A	422	TRP	3.0
1	A	300	GLN	3.0
1	A	373	ILE	2.7
1	A	75	HIS	2.7
1	A	526	GLY	2.6
1	A	74	ASN	2.5
1	A	210	GLY	2.5
1	A	78	PHE	2.4
1	A	70	SER	2.4
1	A	319	LYS	2.3
1	A	494	LYS	2.3
1	A	53	ASN	2.2
1	A	248	PRO	2.1
1	A	352	ILE	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.