



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 31, 2023 – 09:06 PM JST

PDB ID : 8HCY
Title : Legionella glycosyltransferase
Authors : Chen, T.T.; Ouyang, S.Y.
Deposited on : 2022-11-03
Resolution : 2.64 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 : 2.36
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) : 2.36

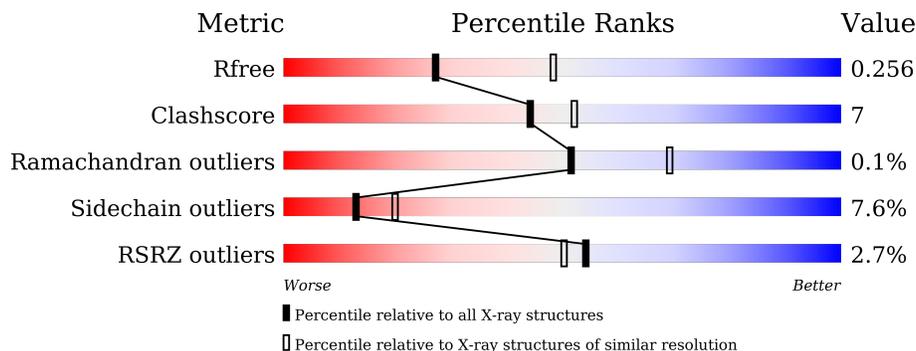
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.64 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	1426 (2.66-2.62)
Clashscore	141614	1472 (2.66-2.62)
Ramachandran outliers	138981	1446 (2.66-2.62)
Sidechain outliers	138945	1446 (2.66-2.62)
RSRZ outliers	127900	1408 (2.66-2.62)

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	865	
1	B	865	
1	C	865	
1	D	865	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 16216 atoms, of which 0 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 Dot/Icm secretion system substrate.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	C	499	3956	2500	676	767	13	0	0	0
1	A	499	3956	2500	676	767	13	0	0	0
1	B	499	3956	2500	676	767	13	0	0	0
1	D	499	3956	2500	676	767	13	0	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	91	Total 91	O 91	0	0
2	A	83	Total 83	O 83	0	0
2	B	99	Total 99	O 99	0	0
2	D	119	Total 119	O 119	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	92.36Å 112.61Å 125.80Å 90.00° 108.14° 90.00°	Depositor
Resolution (Å)	29.31 – 2.64 29.63 – 2.64	Depositor EDS
% Data completeness (in resolution range)	95.6 (29.31-2.64) 95.8 (29.63-2.64)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.59 (at 2.64Å)	Xtrriage
Refinement program	PHENIX 1.18_3845	Depositor
R, R_{free}	0.240 , 0.256 0.240 , 0.256	Depositor DCC
R_{free} test set	2001 reflections (2.90%)	wwPDB-VP
Wilson B-factor (Å ²)	28.3	Xtrriage
Anisotropy	0.537	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 45.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	16216	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 48.38 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.6758e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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	2/4027 (0.0%)	0.59	6/5441 (0.1%)
1	B	0.37	0/4027	0.63	4/5441 (0.1%)
1	C	0.37	0/4027	0.71	7/5441 (0.1%)
1	D	0.36	0/4027	0.58	4/5441 (0.1%)
All	All	0.37	2/16108 (0.0%)	0.63	21/21764 (0.1%)

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	C	0	1
1	D	0	1
All	All	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	160	GLU	CD-OE2	7.43	1.33	1.25
1	A	526	LYS	CE-NZ	5.32	1.62	1.49

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	120	ARG	NE-CZ-NH2	-16.25	112.18	120.30
1	B	189	LEU	CB-CG-CD2	-15.63	84.43	111.00
1	C	120	ARG	NE-CZ-NH1	13.59	127.09	120.30
1	C	569	ILE	CG1-CB-CG2	-10.05	89.29	111.40
1	C	200	GLU	CB-CA-C	9.91	130.22	110.40

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	258	GLN	Peptide
1	D	120	ARG	Sidechain

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	3956	0	3964	58	0
1	B	3956	0	3964	42	0
1	C	3956	0	3964	67	0
1	D	3956	0	3964	45	0
2	A	83	0	0	2	0
2	B	99	0	0	3	0
2	C	91	0	0	2	0
2	D	119	0	0	2	0
All	All	16216	0	15856	210	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 7.

The worst 5 of 210 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:C:166:TYR:HD2	1:C:167:ASN:N	1.39	1.20
1:C:166:TYR:HE2	1:C:167:ASN:CG	1.57	1.05
1:C:166:TYR:CD2	1:C:167:ASN:N	2.29	1.00
1:C:166:TYR:HE2	1:C:167:ASN:OD1	1.49	0.95
1:C:166:TYR:CE2	1:C:167:ASN:CG	2.41	0.93

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	497/865 (58%)	482 (97%)	15 (3%)	0	100	100
1	B	497/865 (58%)	483 (97%)	14 (3%)	0	100	100
1	C	497/865 (58%)	484 (97%)	12 (2%)	1 (0%)	47	64
1	D	497/865 (58%)	480 (97%)	17 (3%)	0	100	100
All	All	1988/3460 (58%)	1929 (97%)	58 (3%)	1 (0%)	51	69

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	488	ALA

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	439/763 (58%)	406 (92%)	33 (8%)	13	20
1	B	439/763 (58%)	412 (94%)	27 (6%)	18	28
1	C	439/763 (58%)	410 (93%)	29 (7%)	16	25
1	D	439/763 (58%)	395 (90%)	44 (10%)	7	10
All	All	1756/3052 (58%)	1623 (92%)	133 (8%)	13	20

5 of 133 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	449	GLU
1	D	469	GLN
1	D	532	LYS
1	A	497	LYS
1	A	465	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	287	ASN
1	D	258	GLN
1	B	281	GLN
1	D	257	ASN
1	A	539	GLN

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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	499/865 (57%)	0.21	21 (4%) 36 33	22, 34, 66, 103	0
1	B	499/865 (57%)	0.07	7 (1%) 75 73	21, 32, 53, 82	0
1	C	499/865 (57%)	0.15	11 (2%) 62 58	21, 33, 58, 116	0
1	D	499/865 (57%)	0.18	14 (2%) 53 49	23, 35, 56, 82	0
All	All	1996/3460 (57%)	0.15	53 (2%) 54 50	21, 34, 57, 116	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	528	PRO	6.0
1	A	160	GLU	5.4
1	A	527	THR	4.4
1	A	531	ILE	3.9
1	A	356	ASP	3.8

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.