



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2026 – 07:35 AM UTC

PDB ID : 5JOY / pdb_00005joy
Title : Bacteroides ovatus Xyloglucan PUL GH43A in complex with AraLOG
Authors : Thompson, A.J.; Hemsworth, G.R.; Stepper, J.; Sobala, L.F.; Coyle, T.; Larsbrink, J.; Spadiut, O.; Stubbs, K.A.; Brumer, H.; Davies, G.J.
Deposited on : 2016-05-03
Resolution : 1.90 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

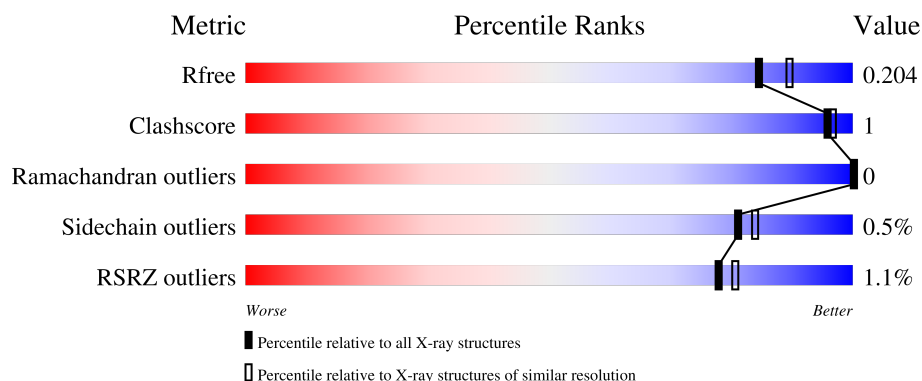
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

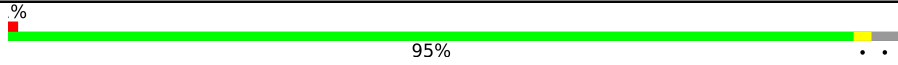
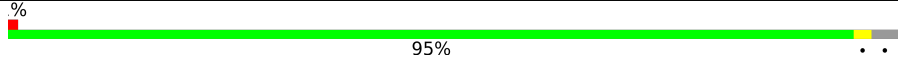
The reported resolution of this entry is 1.90 Å.

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}	180053	7789 (1.90-1.90)
Clashscore	190562	8410 (1.90-1.90)
Ramachandran outliers	187476	8333 (1.90-1.90)
Sidechain outliers	187428	8333 (1.90-1.90)
RSRZ outliers	180081	7790 (1.90-1.90)

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	516	
1	B	516	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 9085 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 Non-reducing end alpha-L-arabinofuranosidase BoGH43A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	502	Total	C	N	O	S	4	23	0
			4038	2584	671	767	16			
1	B	502	Total	C	N	O	S	0	19	0
			4008	2572	669	750	17			

There are 20 discrepancies between the modelled and reference sequences:

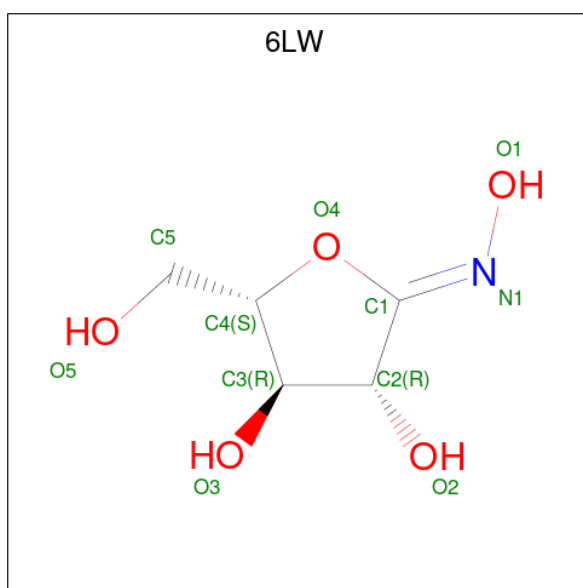
Chain	Residue	Modelled	Actual	Comment	Reference
A	11	MET	-	initiating methionine	UNP A7LXT8
A	12	GLY	-	expression tag	UNP A7LXT8
A	13	SER	-	expression tag	UNP A7LXT8
A	14	SER	-	expression tag	UNP A7LXT8
A	15	HIS	-	expression tag	UNP A7LXT8
A	16	HIS	-	expression tag	UNP A7LXT8
A	17	HIS	-	expression tag	UNP A7LXT8
A	18	HIS	-	expression tag	UNP A7LXT8
A	19	HIS	-	expression tag	UNP A7LXT8
A	20	HIS	-	expression tag	UNP A7LXT8
B	11	MET	-	initiating methionine	UNP A7LXT8
B	12	GLY	-	expression tag	UNP A7LXT8
B	13	SER	-	expression tag	UNP A7LXT8
B	14	SER	-	expression tag	UNP A7LXT8
B	15	HIS	-	expression tag	UNP A7LXT8
B	16	HIS	-	expression tag	UNP A7LXT8
B	17	HIS	-	expression tag	UNP A7LXT8
B	18	HIS	-	expression tag	UNP A7LXT8
B	19	HIS	-	expression tag	UNP A7LXT8
B	20	HIS	-	expression tag	UNP A7LXT8

- Molecule 2 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (CCD ID: TRS) (formula: C₄H₁₂NO₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			8	4	1	3		
2	B	1	Total	C	N	O	0	0
			8	4	1	3		

- Molecule 3 is (Z)-L-Arabinonhydroximo-1,4-lactone (CCD ID: 6LW) (formula: $C_5H_9NO_5$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			11	5	1	5		
3	B	1	Total	C	N	O	0	0
			11	5	1	5		

- Molecule 4 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: $C_2H_6O_2$).



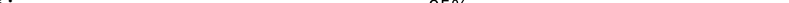
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			4	2	2		

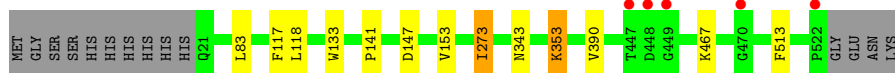
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	496	Total	O	0	0
			496	496		
5	B	500	Total	O	0	1
			501	501		

- Molecule 1: Non-reducing end alpha-L-arabinofuranosidase BoGH43A



- Chain B: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2 ₁	Depositor
Cell constants a, b, c, α , β , γ	75.59Å 91.50Å 154.16Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.49 – 1.90 46.49 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.6 (46.49-1.90) 99.6 (46.49-1.90)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.70 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.8.0124	Depositor
R, R_{free}	0.179 , 0.218 (Not available) , 0.204	Depositor DCC
R_{free} test set	4273 reflections (4.26%)	wwPDB-VP
Wilson B-factor (Å ²)	23.9	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 35.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	9085	wwPDB-VP
Average B, all atoms (Å ²)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.81% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: EDO, TRS, 6LW

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	0.63	0/4210	0.76	1/5735 (0.0%)
1	B	0.65	0/4183	0.76	0/5697
All	All	0.64	0/8393	0.76	1/11432 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	129	SER	N-CA-C	5.52	117.45	110.33

There are no chirality outliers.

There are no planarity outliers.

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	4038	0	3849	9	0
1	B	4008	0	3844	9	0
2	A	8	0	12	0	0
2	B	8	0	12	0	0
3	A	11	0	0	1	0
3	B	11	0	0	0	0
4	A	4	0	6	0	0
5	A	496	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	501	0	0	2	0
All	All	9085	0	7723	18	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 1.

All (18) 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:21[B]:GLN:NE2	1:A:21[B]:GLN:O	2.11	0.83
1:B:83:LEU:H	1:B:343:ASN:HD21	1.37	0.72
1:A:424:LYS:CB	5:A:1142:HOH:O	2.48	0.62
1:A:21[B]:GLN:HE21	1:A:21[B]:GLN:C	2.13	0.56
1:B:390[B]:VAL:HG12	1:B:513:PHE:CE2	2.42	0.55
1:A:117:PHE:HB3	1:A:133:TRP:CE3	2.48	0.48
1:B:117:PHE:HB3	1:B:133:TRP:CE3	2.52	0.44
1:B:390[B]:VAL:HG12	1:B:513:PHE:CD2	2.53	0.43
1:B:141:PRO:HA	1:B:153:VAL:O	2.19	0.43
1:A:273[B]:ILE:HA	1:A:273[B]:ILE:HD12	1.79	0.42
1:B:118:LEU:C	1:B:118:LEU:HD12	2.45	0.41
1:A:501[B]:THR:HG21	5:A:993:HOH:O	2.21	0.41
1:B:353[A]:LYS:NZ	5:B:713:HOH:O	2.53	0.41
1:B:467:LYS:CB	5:B:1085:HOH:O	2.69	0.41
1:A:21[B]:GLN:NE2	1:A:21[B]:GLN:C	2.76	0.40
1:A:210:GLU:OE2	3:A:602:6LW:O1	2.39	0.40
1:B:273[A]:ILE:HD12	1:B:273[A]:ILE:HA	1.81	0.40
1:A:270:ALA:HB3	1:A:286:TYR:CE1	2.56	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	523/516 (101%)	507 (97%)	16 (3%)	0	100	100
1	B	519/516 (101%)	503 (97%)	16 (3%)	0	100	100
All	All	1042/1032 (101%)	1010 (97%)	32 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	436/442 (99%)	434 (100%)	2 (0%)	81	84
1	B	431/442 (98%)	426 (99%)	5 (1%)	63	63
All	All	867/884 (98%)	860 (99%)	7 (1%)	81	75

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	273[A]	ILE
1	A	273[B]	ILE
1	B	147	ASP
1	B	273[A]	ILE
1	B	273[B]	ILE
1	B	353[A]	LYS
1	B	353[B]	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	116	ASN
1	A	136	GLN
1	A	172	GLN
1	A	228	GLN
1	A	398	ASN
1	A	473	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	136	GLN
1	B	172	GLN
1	B	301	ASN
1	B	343	ASN
1	B	379	GLN
1	B	398	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

5 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	6LW	B	602	-	9,11,11	2.27	2 (22%)	11,15,15	1.43	2 (18%)
2	TRS	A	601	-	7,7,7	0.27	0	9,9,9	0.51	0
4	EDO	A	603	-	3,3,3	0.43	0	2,2,2	0.01	0
2	TRS	B	601	-	7,7,7	0.55	0	9,9,9	0.75	0
3	6LW	A	602	-	9,11,11	2.28	1 (11%)	11,15,15	1.56	3 (27%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	6LW	B	602	-	-	0/4/20/20	0/1/1/1
2	TRS	A	601	-	-	3/9/9/9	-
4	EDO	A	603	-	-	0/1/1/1	-
2	TRS	B	601	-	-	2/9/9/9	-
3	6LW	A	602	-	-	0/4/20/20	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	602	6LW	O4-C1	6.57	1.45	1.35
3	B	602	6LW	O4-C1	6.30	1.45	1.35
3	B	602	6LW	O4-C4	-2.29	1.43	1.46

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	602	6LW	C4-C3-C2	2.99	104.75	101.11
3	B	602	6LW	C3-C2-C1	-2.78	99.03	101.78
3	B	602	6LW	C4-C3-C2	2.75	104.45	101.11
3	A	602	6LW	C3-C2-C1	-2.51	99.29	101.78
3	A	602	6LW	O4-C4-C3	2.23	106.33	103.98

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	601	TRS	N-C-C3-O3
2	B	601	TRS	N-C-C2-O2
2	A	601	TRS	C1-C-C3-O3
2	B	601	TRS	C3-C-C2-O2
2	A	601	TRS	C2-C-C3-O3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	602	6LW	1	0

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	502/516 (97%)	-0.13	6 (1%) 76 79	11, 22, 37, 60	23 (4%)
1	B	502/516 (97%)	-0.17	5 (0%) 79 82	9, 21, 39, 68	19 (3%)
All	All	1004/1032 (97%)	-0.15	11 (1%) 78 80	9, 22, 38, 68	42 (4%)

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	522	PRO	5.2
1	A	505[A]	LYS	3.8
1	B	447	THR	3.6
1	B	470	GLY	2.9
1	A	21[A]	GLN	2.6
1	B	448	ASP	2.6
1	A	20	HIS	2.5
1	B	449	GLY	2.4
1	A	506[A]	ASP	2.2
1	A	448	ASP	2.1
1	A	504[A]	SER	2.0

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 oligosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
4	EDO	A	603	4/4	0.86	0.12	42,42,43,45	0
3	6LW	B	602	11/11	0.89	0.09	31,33,33,34	0
3	6LW	A	602	11/11	0.90	0.09	31,32,33,33	0
2	TRS	B	601	8/8	0.93	0.08	17,19,20,21	0
2	TRS	A	601	8/8	0.93	0.08	20,21,21,22	0

6.5 Other polymers [i](#)

There are no such residues in this entry.