WORLDWIDE PDB

PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report ⓘ

Sep 13, 2022 – 10:58 pm BST

PDB ID	:	7PL5
Title	:	Crystal structure of choline-binding module (R1-R9) of LytB from Streptococcus pneumoniae
Authors	:	Molina, R.; Martinez Caballero, S.; Hermoso, J.A.
Deposited on	:	2021-08-28
Resolution	:	1.99 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:

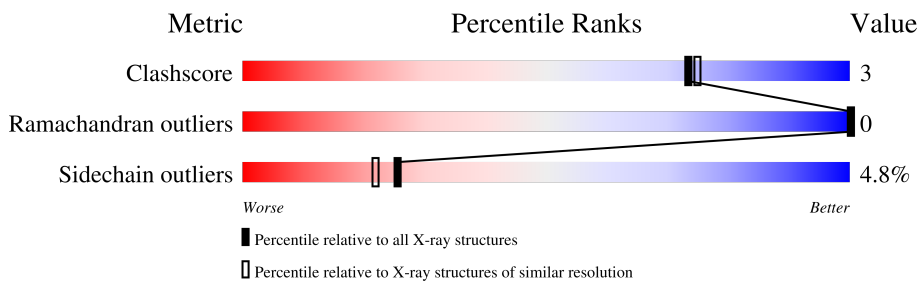|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.30

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.99 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	Percentile Ranks	Value
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		4.8%

Worse   *Better*

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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 <=5%

Note EDS failed to run properly.

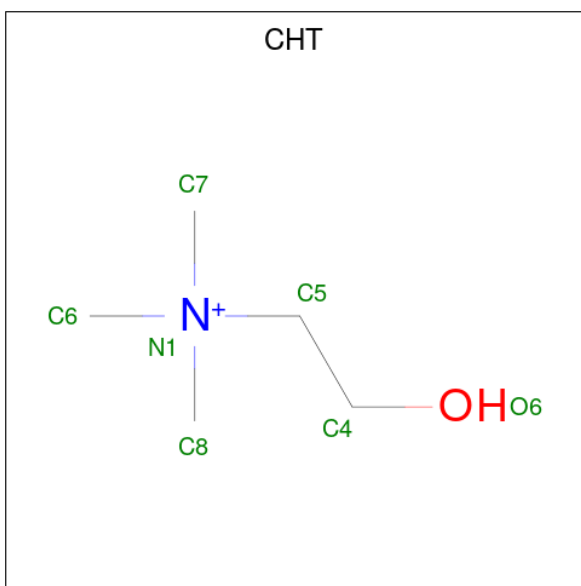| Mol	Chain	Length	Quality of chain
1	AAA	185	87% 8% · ·
1	BBB	185	91% 7% ··

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3352 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 Putative endo-beta-N-acetylglucosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	178	Total	C	N	O	S	0	0	0
			1494	974	236	282	2			
1	BBB	183	Total	C	N	O	S	0	2	0
			1548	1006	245	295	2			

- Molecule 2 is CHOLINE ION (three-letter code: CHT) (formula: $C_5H_{14}NO$).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	AAA	1	Total	C	N	O	0	0
			7	5	1	1		
2	AAA	1	Total	C	N	O	0	0
			7	5	1	1		
2	AAA	1	Total	C	N	O	0	0
			7	5	1	1		
2	AAA	1	Total	C	N	O	0	0
			7	5	1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	AAA	1	Total 7	C 5	N 1	O 1	0	0
2	AAA	1	Total 7	C 5	N 1	O 1	0	0
2	BBB	1	Total 7	C 5	N 1	O 1	0	0
2	BBB	1	Total 7	C 5	N 1	O 1	0	0
2	BBB	1	Total 7	C 5	N 1	O 1	0	0
2	BBB	1	Total 7	C 5	N 1	O 1	0	0
2	BBB	1	Total 7	C 5	N 1	O 1	0	0
2	BBB	1	Total 7	C 5	N 1	O 1	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	2	Total 2	Zn 2	0	0
3	BBB	2	Total 2	Zn 2	0	0

- Molecule 4 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).
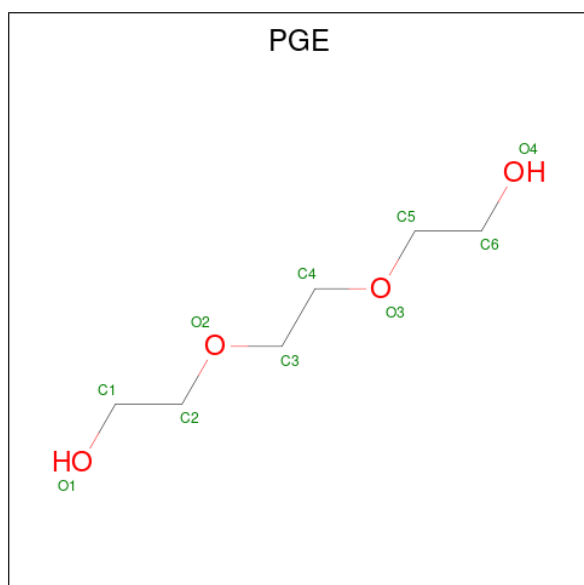
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	BBB	1	Total 10	C 6	O 4	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	AAA	113	Total 113	O 113	0	0
5	BBB	99	Total 99	O 99	0	0

3 Residue-property plots ⓘ

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. 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. 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.

Note EDS failed to run properly.

• Molecule 1: Putative endo-beta-N-acetylglucosaminidase

Chain AAA:

• Molecule 1: Putative endo-beta-N-acetylglucosaminidase

Chain BBB:

4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	45.30Å 96.14Å 49.10Å 90.00° 100.97° 90.00°	Depositor
Resolution (Å)	44.50 – 1.99	Depositor
% Data completeness (in resolution range)	99.2 (44.50-1.99)	Depositor
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.05 (at 1.98Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.172 , 0.233	Depositor
Wilson B-factor (Å^2)	19.2	Xtriage
Anisotropy	0.244	Xtriage
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3352	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.44% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: CHT, ZN, PGE

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	AAA	0.89	0/1546	0.99	0/2093
1	BBB	0.87	0/1601	0.96	0/2168
All	All	0.88	0/3147	0.98	0/4261

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

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	AAA	1494	0	1361	10	0
1	BBB	1548	0	1403	9	0
2	AAA	42	0	84	1	0
2	BBB	42	0	84	3	0
3	AAA	2	0	0	0	0
3	BBB	2	0	0	0	0
4	BBB	10	0	14	1	0
5	AAA	113	0	0	1	0
5	BBB	99	0	0	3	0
All	All	3352	0	2946	20	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 (20) 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:BBB:182:ASN:HB2	5:BBB:311:HOH:O	1.84	0.77
1:AAA:138:LYS:HG2	5:AAA:379:HOH:O	2.01	0.61
1:BBB:20:ASN:HD21	1:BBB:36:ALA:HB2	1.68	0.58
1:BBB:178:LYS:NZ	5:BBB:302:HOH:O	2.37	0.57
1:AAA:169:LYS:O	1:AAA:172:GLN:NE2	2.45	0.50
1:AAA:41:ALA:HB3	1:AAA:55:LEU:HD13	1.94	0.49
1:AAA:163:GLN:HA	1:AAA:177:ILE:O	2.13	0.49
1:BBB:62:ALA:HB3	1:BBB:76:LEU:HD13	1.96	0.47
1:AAA:45:TRP:CE2	2:AAA:205:CHT:H61	2.49	0.46
1:AAA:144:LEU:HD23	1:AAA:149:ILE:CD1	2.45	0.46
1:BBB:62:ALA:CB	1:BBB:76:LEU:HD13	2.46	0.46
1:AAA:148:TRP:O	1:AAA:149:ILE:HD13	2.18	0.44
2:BBB:202:CHT:H63	2:BBB:202:CHT:HC41	1.83	0.43
1:AAA:144:LEU:HD23	1:AAA:149:ILE:HD12	1.99	0.43
1:AAA:76:LEU:HD22	1:AAA:76:LEU:N	2.34	0.42
1:AAA:144:LEU:HA	1:AAA:144:LEU:HD12	1.78	0.42
1:BBB:58:GLY:O	2:BBB:202:CHT:H83	2.20	0.42
1:BBB:24:PHE:H	4:BBB:203:PGE:H22	1.85	0.41
1:BBB:147:GLN:NE2	5:BBB:314:HOH:O	2.54	0.40
1:BBB:22:TRP:CG	2:BBB:201:CHT:H61	2.56	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	AAA	176/185 (95%)	174 (99%)	2 (1%)	0	100	100
1	BBB	183/185 (99%)	178 (97%)	5 (3%)	0	100	100

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	359/370 (97%)	352 (98%)	7 (2%)	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	AAA	145/150 (97%)	138 (95%)	7 (5%)	25 22
1	BBB	151/150 (101%)	144 (95%)	7 (5%)	27 23
All	All	296/300 (99%)	282 (95%)	14 (5%)	25 22

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	50	ASP
1	AAA	55	LEU
1	AAA	63	LYS
1	AAA	102	GLU
1	AAA	144	LEU
1	AAA	162	GLN
1	AAA	178	LYS
1	BBB	8	LYS
1	BBB	50	ASP
1	BBB	76	LEU
1	BBB	101	ILE
1	BBB	170	GLN
1	BBB	173	SER
1	BBB	178	LYS

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

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 17 ligands modelled in this entry, 4 are monoatomic - leaving 13 for Mogul analysis.

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	$\#|Z| > 2$
2	CHT	BBB	206	-	6,6,6	0.38	0	8,8,8	0.37	0
2	CHT	AAA	203	-	6,6,6	0.29	0	8,8,8	0.56	0
2	CHT	BBB	205	-	6,6,6	0.31	0	8,8,8	0.28	0
2	CHT	BBB	204	-	6,6,6	0.32	0	8,8,8	0.64	0
2	CHT	AAA	204	-	6,6,6	0.51	0	8,8,8	0.45	0
2	CHT	BBB	207	-	6,6,6	0.42	0	8,8,8	0.57	0
4	PGE	BBB	203	-	9,9,9	0.34	0	8,8,8	0.21	0
2	CHT	BBB	202	-	6,6,6	0.49	0	8,8,8	0.25	0
2	CHT	AAA	202	-	6,6,6	0.74	0	8,8,8	0.58	0
2	CHT	AAA	205	-	6,6,6	0.32	0	8,8,8	0.20	0
2	CHT	AAA	206	-	6,6,6	0.60	0	8,8,8	0.67	0
2	CHT	BBB	201	-	6,6,6	0.09	0	8,8,8	0.35	0
2	CHT	AAA	201	-	6,6,6	0.69	0	8,8,8	0.64	0

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
2	CHT	BBB	206	-	-	0/4/4/4	-
2	CHT	AAA	203	-	-	0/4/4/4	-
2	CHT	BBB	205	-	-	1/4/4/4	-
2	CHT	BBB	204	-	-	0/4/4/4	-
2	CHT	AAA	204	-	-	0/4/4/4	-
2	CHT	BBB	207	-	-	1/4/4/4	-
4	PGE	BBB	203	-	-	4/7/7/7	-
2	CHT	BBB	202	-	-	0/4/4/4	-
2	CHT	AAA	202	-	-	3/4/4/4	-
2	CHT	AAA	205	-	-	0/4/4/4	-
2	CHT	AAA	206	-	-	0/4/4/4	-
2	CHT	BBB	201	-	-	3/4/4/4	-
2	CHT	AAA	201	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (12) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	BBB	201	CHT	C4-C5-N1-C7
2	AAA	202	CHT	C4-C5-N1-C7
2	BBB	201	CHT	C4-C5-N1-C6
2	BBB	201	CHT	C4-C5-N1-C8
4	BBB	203	PGE	O2-C3-C4-O3
2	AAA	202	CHT	C4-C5-N1-C6
2	AAA	202	CHT	C4-C5-N1-C8
2	BBB	205	CHT	O6-C4-C5-N1
2	BBB	207	CHT	O6-C4-C5-N1
4	BBB	203	PGE	C4-C3-O2-C2
4	BBB	203	PGE	C3-C4-O3-C5
4	BBB	203	PGE	C1-C2-O2-C3

There are no ring outliers.

4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	BBB	203	PGE	1	0
2	BBB	202	CHT	2	0
2	AAA	205	CHT	1	0
2	BBB	201	CHT	1	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates ⓘ

EDS failed to run properly - this section is therefore empty.

6.4 Ligands ⓘ

EDS failed to run properly - this section is therefore empty.

6.5 Other polymers ⓘ

EDS failed to run properly - this section is therefore empty.