Full wwPDB X-ray Structure Validation Report ⓘ

Jun 27, 2023 – 04:35 pm BST

PDB ID	:	8P6K
Title	:	Structure of the hypervariable region of Streptococcus pyogenes M3 protein
Authors	:	Wojnowska, M.; Schwarz-Linek, U.
Deposited on	:	2023-05-26
Resolution	:	1.92 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:

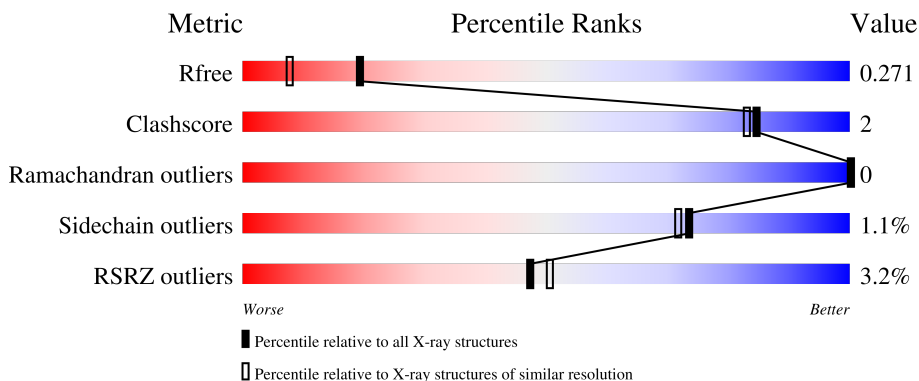|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.33
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

1 Overall quality at a glance ⓘ

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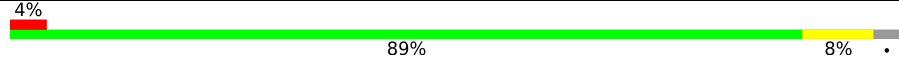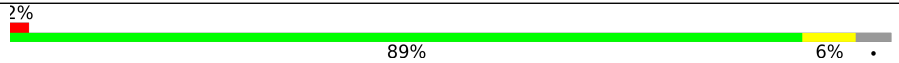*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.92 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
Rfree		0.271
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		1.1%
RSRZ outliers		3.2%

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))
R_{free}	130704	7937 (1.94-1.90)
Clashscore	141614	8644 (1.94-1.90)
Ramachandran outliers	138981	8530 (1.94-1.90)
Sidechain outliers	138945	8530 (1.94-1.90)
RSRZ outliers	127900	7793 (1.94-1.90)

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 <=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	BBB	113	4% — 89% — 8% ·
1	CCC	113	2% — 89% — 6% ·

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 3650 atoms, of which 1782 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 Antiphagocytic M protein, type 3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	BBB	110	Total	C	H	N	O	S	47	0	0
			1798	563	898	160	176	1			
1	CCC	108	Total	C	H	N	O	S	46	0	0
			1771	557	884	153	176	1			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	1	GLY	-	expression tag	UNP A0A0H2UWN1
BBB	2	ALA	-	expression tag	UNP A0A0H2UWN1
BBB	3	MET	-	expression tag	UNP A0A0H2UWN1
BBB	113	CYS	LEU	engineered mutation	UNP A0A0H2UWN1
CCC	1	GLY	-	expression tag	UNP A0A0H2UWN1
CCC	2	ALA	-	expression tag	UNP A0A0H2UWN1
CCC	3	MET	-	expression tag	UNP A0A0H2UWN1
CCC	113	CYS	LEU	engineered mutation	UNP A0A0H2UWN1

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	BBB	37	Total	O	0	0
			37	37		
2	CCC	44	Total	O	0	0
			44	44		

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 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: Antiphagocytic M protein, type 3

Chain BBB:

• Molecule 1: Antiphagocytic M protein, type 3

Chain CCC:
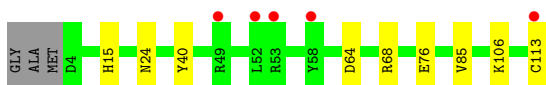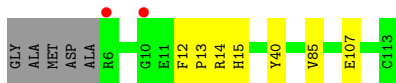
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	150.47Å 24.11Å 78.01Å 90.00° 105.29° 90.00°	Depositor
Resolution (Å)	75.25 – 1.92 75.25 – 1.92	Depositor EDS
% Data completeness (in resolution range)	99.7 (75.25-1.92) 99.7 (75.25-1.92)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.71 (at 1.92Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.230 , 0.268 0.239 , 0.271	Depositor DCC
R_{free} test set	1054 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	42.1	Xtriage
Anisotropy	0.051	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 42.3	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3650	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.84% 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 ⓘ

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	BBB	0.70	0/911	0.75	0/1220
1	CCC	0.71	0/898	0.75	0/1202
All	All	0.71	0/1809	0.75	0/2422

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	BBB	900	898	887	7	0
1	CCC	887	884	875	6	0
2	BBB	37	0	0	1	0
2	CCC	44	0	0	0	0
All	All	1868	1782	1762	8	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 2.

All (8) 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:64:ASP:OD1	1:CCC:14:ARG:NH2	2.24	0.70

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:BBB:24:ASN:ND2	2:BBB:202:HOH:O	2.26	0.62
1:BBB:64:ASP:O	1:BBB:68:ARG:HG3	2.05	0.57
1:BBB:85:VAL:HG12	1:CCC:85:VAL:HG12	1.87	0.57
1:BBB:15:HIS:CE1	1:CCC:40:TYR:HB2	2.48	0.48
1:BBB:40:TYR:HB2	1:CCC:15:HIS:CE1	2.53	0.44
1:CCC:12:PHE:HB3	1:CCC:13:PRO:HD3	2.00	0.43
1:BBB:106:LYS:NZ	1:CCC:107:GLU:OE2	2.50	0.41

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	BBB	108/113 (96%)	105 (97%)	3 (3%)	0	100	100
1	CCC	106/113 (94%)	104 (98%)	2 (2%)	0	100	100
All	All	214/226 (95%)	209 (98%)	5 (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	BBB	95/100 (95%)	93 (98%)	2 (2%)	53	46

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	CCC	95/100 (95%)	95 (100%)	0	100	100
All	All	190/200 (95%)	188 (99%)	2 (1%)	73	72

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	76	GLU
1	BBB	113	CYS

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 ⓘ

There are no ligands in this entry.

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 ⓘ

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(Å²)	Q<0.9
1	BBB	110/113 (97%)	0.38	5 (4%)	33	36	29, 48, 76, 95	0
1	CCC	108/113 (95%)	0.28	2 (1%)	66	69	30, 44, 66, 83	0
All	All	218/226 (96%)	0.33	7 (3%)	47	50	29, 46, 75, 95	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	CCC	10	GLY	4.1
1	CCC	6	ARG	3.6
1	BBB	53	ARG	3.2
1	BBB	113	CYS	3.0
1	BBB	52	LEU	2.1
1	BBB	58	TYR	2.0
1	BBB	49	ARG	2.0

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.