Full wwPDB X-ray Structure Validation Report ⓘ

Aug 29, 2020 – 06:17 PM BST

PDB ID	:	6SE5
Title	:	Y830A mutant from Mycoplasma genitalium P110 adhesin
Authors	:	Fita, I.; Aparicio, D.
Deposited on	:	2019-07-29
Resolution	:	2.21 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 following versions of software and data (see references ①) were used in the production of this report:

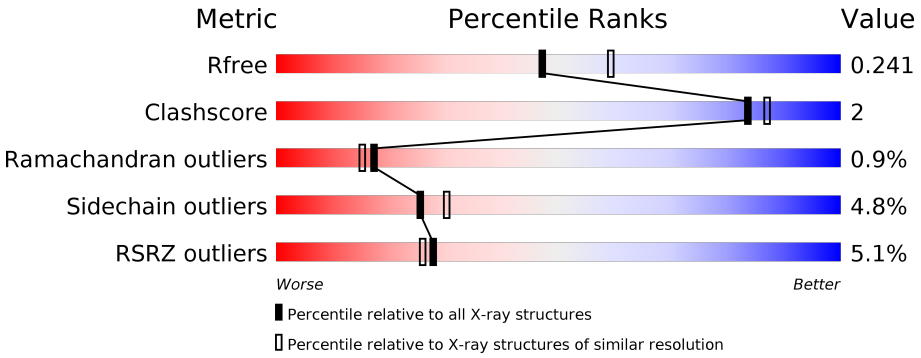
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.13
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13

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 2.21 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	5912 (2.24-2.20)
Clashscore	141614	6646 (2.24-2.20)
Ramachandran outliers	138981	6543 (2.24-2.20)
Sidechain outliers	138945	6544 (2.24-2.20)
RSRZ outliers	127900	5797 (2.24-2.20)

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 on 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	A	923	5% 85% 10%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7104 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 Mgp-operon protein 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	885	Total	C	N	O	S	0	1	0
			6812	4273	1137	1396	6			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	830	ALA	TYR	engineered mutation	UNP P22747
A	939	HIS	-	expression tag	UNP P22747
A	940	HIS	-	expression tag	UNP P22747
A	941	HIS	-	expression tag	UNP P22747
A	942	HIS	-	expression tag	UNP P22747
A	943	HIS	-	expression tag	UNP P22747
A	944	HIS	-	expression tag	UNP P22747

- Molecule 2 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	K	0	0
			1	1		

- Molecule 3 is water.

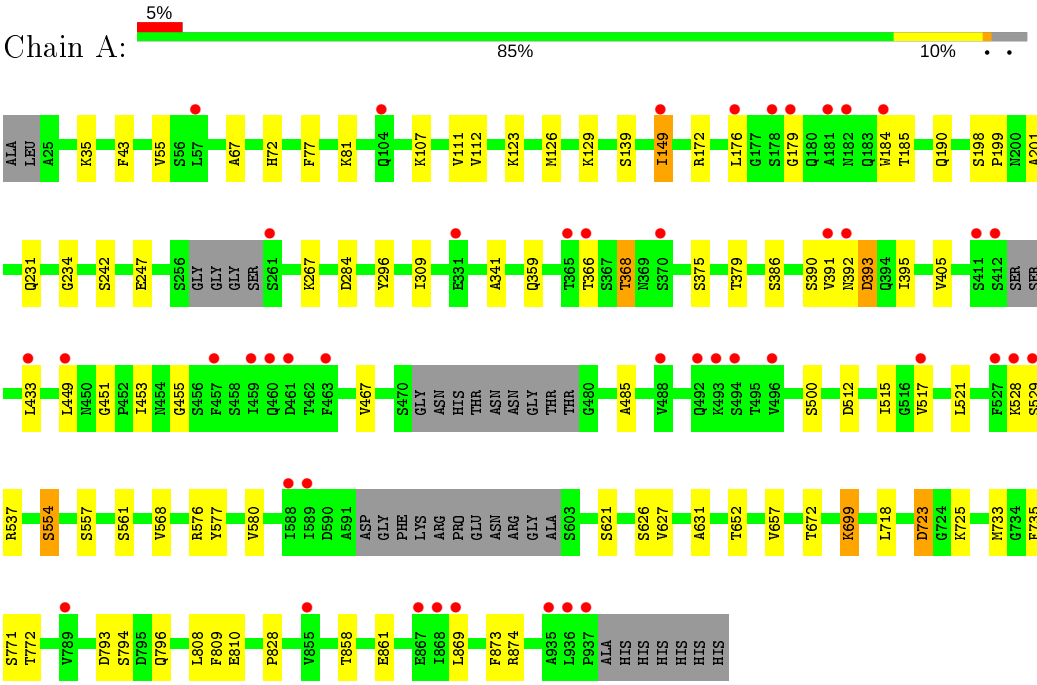
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	291	Total	O	0	0
			291	291		

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: Mgp-operon protein 3

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	109.32Å 153.12Å 169.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.30 – 2.21 91.93 – 2.21	Depositor EDS
% Data completeness (in resolution range)	98.8 (32.30-2.21) 98.8 (91.93-2.21)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.39 (at 2.20Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.192 , 0.227 0.205 , 0.241	Depositor DCC
R_{free} test set	3571 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	51.9	Xtriage
Anisotropy	0.625	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 70.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	7104	wwPDB-VP
Average B, all atoms (Å^2)	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.05% 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: K

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	0.50	0/6954	0.72	1/9459 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	766	GLN	C-N-CA	5.85	136.32	121.70

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	A	6812	0	6598	32	0
2	A	1	0	0	0	0
3	A	291	0	0	0	0
All	All	7104	0	6598	32	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 (32) 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:554:SER:HB2	1:A:568:VAL:HG13	1.59	0.84
1:A:126:MET:HE1	1:A:129:LYS:HD2	1.72	0.71
1:A:391:VAL:CG2	1:A:631:ALA:HB1	2.32	0.59
1:A:231:GLN:HG2	1:A:267:LYS:HB3	1.85	0.57
1:A:184:TRP:CD2	1:A:199:PRO:HB3	2.43	0.54
1:A:111:VAL:HG11	1:A:309:ILE:HG21	1.92	0.51
1:A:390:SER:HB3	1:A:395:ILE:HD13	1.94	0.49
1:A:528:LYS:HD3	1:A:529:SER:H	1.76	0.49
1:A:392[A]:ASN:CG	1:A:393:ASP:H	2.16	0.49
1:A:72:HIS:HB3	1:A:77:PHE:CE1	2.47	0.49
1:A:580:VAL:HG21	1:A:621:SER:O	2.15	0.46
1:A:723:ASP:HB3	1:A:725:LYS:H	1.82	0.45
1:A:517:VAL:O	1:A:521:LEU:HG	2.17	0.45
1:A:35:LYS:HG2	1:A:55:VAL:HG21	1.99	0.45
1:A:626:SER:HB3	1:A:652:THR:HG22	1.99	0.44
1:A:211:ASN:O	1:A:234:GLY:HA2	2.18	0.44
1:A:433:LEU:HD23	1:A:534:LEU:HD21	2.00	0.43
1:A:341:ALA:O	1:A:379:THR:HG22	2.18	0.43
1:A:858:THR:HG23	1:A:861:GLU:H	1.83	0.43
1:A:767:SER:O	1:A:809:PHE:HB2	2.19	0.43
1:A:172:ARG:HG3	1:A:176:LEU:HD12	2.00	0.43
1:A:451:GLY:HA3	1:A:485:ALA:O	2.19	0.43
1:A:81:LYS:HE3	1:A:296:TYR:CE2	2.54	0.42
1:A:198:SER:HB3	1:A:201:ALA:HB2	2.02	0.41
1:A:699:LYS:HE2	1:A:718:LEU:HD11	2.02	0.41
1:A:43:PHE:HB3	1:A:828:PRO:HD3	2.03	0.41
1:A:533:ARG:HD3	1:A:577:TYR:HE2	1.85	0.41
1:A:67:ALA:HB2	1:A:81:LYS:HD3	2.03	0.41
1:A:735:PHE:CE2	1:A:771:SER:HB3	2.56	0.41
1:A:359:GLN:HG2	1:A:405:VAL:HG12	2.02	0.40
1:A:149:ILE:HG23	1:A:213:TYR:HB3	2.04	0.40
1:A:112:VAL:HG22	1:A:129:LYS:HG3	2.04	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	876/923 (95%)	837 (96%)	31 (4%)	8 (1%)	17 15

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	179	GLY
1	A	512	ASP
1	A	767	SER
1	A	284	ASP
1	A	368	THR
1	A	453	ILE
1	A	794	SER
1	A	455	GLY

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	774/801 (97%)	737 (95%)	37 (5%)	25 30

All (37) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	107	LYS
1	A	123	LYS
1	A	139	SER
1	A	149	ILE
1	A	185	THR
1	A	190	GLN
1	A	242	SER
1	A	247	GLU
1	A	366	THR
1	A	368	THR
1	A	375	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	386	SER
1	A	393	ASP
1	A	449	LEU
1	A	467	VAL
1	A	500	SER
1	A	515	ILE
1	A	534	LEU
1	A	537	ARG
1	A	554	SER
1	A	557	SER
1	A	561	SER
1	A	576	ARG
1	A	627	VAL
1	A	657	VAL
1	A	672	THR
1	A	699	LYS
1	A	723	ASP
1	A	733	MET
1	A	772	THR
1	A	793	ASP
1	A	796	GLN
1	A	808	LEU
1	A	810	GLU
1	A	869	LEU
1	A	873	PHE
1	A	874	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	285	GLN
1	A	674	ASN
1	A	891	GLN

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 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 95^{th} 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	885/923 (95%)	0.60	45 (5%)	28	26	53, 80, 130, 186	0

All (45) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	460	GLN	10.3
1	A	366	THR	8.8
1	A	261	SER	6.4
1	A	527	PHE	5.6
1	A	937	PRO	5.5
1	A	529	SER	5.1
1	A	459	ILE	5.1
1	A	492	GLN	4.9
1	A	417	SER	4.8
1	A	461	ASP	4.4
1	A	365	THR	4.1
1	A	181	ALA	3.7
1	A	433	LEU	3.6
1	A	528	LYS	3.5
1	A	589	ILE	3.1
1	A	104	GLN	3.0
1	A	182	ASN	3.0
1	A	463	PHE	3.0
1	A	494	SER	2.8
1	A	412	SER	2.7
1	A	179	GLY	2.7
1	A	457	PHE	2.7
1	A	855	VAL	2.6
1	A	517	VAL	2.6
1	A	936	LEU	2.5
1	A	331	GLU	2.5
1	A	493	LYS	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	370	SER	2.4
1	A	935	ALA	2.4
1	A	588	ILE	2.4
1	A	496	VAL	2.3
1	A	867	GLU	2.3
1	A	488	VAL	2.2
1	A	411	SER	2.2
1	A	178	SER	2.2
1	A	184	TRP	2.2
1	A	869	LEU	2.2
1	A	149	ILE	2.2
1	A	57	LEU	2.1
1	A	449	LEU	2.1
1	A	392[A]	ASN	2.1
1	A	176	LEU	2.1
1	A	868	ILE	2.1
1	A	391	VAL	2.1
1	A	789	VAL	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)

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, 95^{th} 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(Å^2)	Q<0.9
2	K	A	1001	1/1	0.97	0.15	58,58,58,58	0

6.5 Other polymers (i)

There are no such residues in this entry.