WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 18, 2020 – 09:22 pm BST

PDB ID	:	2HA9
Title	:	Crystal structure of protein SP0239 from Streptococcus pneumoniae
Authors	:	Chang, C.; Hatzos, C.; Abdullah, J.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2006-06-12
Resolution	:	2.70 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

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 2.70 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.
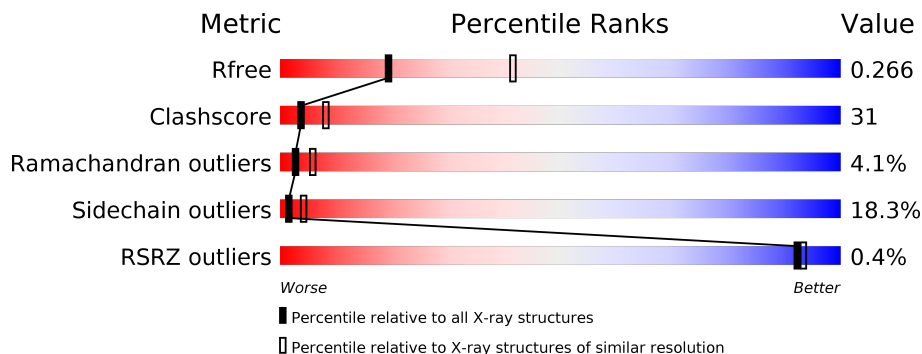
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	446	49% 31% 10% · 7%
1	B	446	47% 33% 10% · 7%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6011 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 UPF0210 protein SP0239.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	414	Total 2978	C 1867	N 504	O 586	S 4	Se 17	0	0	0
1	B	415	Total 2979	C 1865	N 508	O 587	S 3	Se 16	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	ALA	-	CLONING ARTIFACT	UNP Q97ST4
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	12	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	25	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	158	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	164	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	177	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	195	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	255	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	288	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	315	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	336	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	353	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	363	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	378	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	390	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	406	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
A	419	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	0	ALA	-	CLONING ARTIFACT	UNP Q97ST4
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	12	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	25	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	158	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	164	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	177	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	195	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	255	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	288	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	315	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	336	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	353	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	363	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	378	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	390	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	406	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4
B	419	MSE	MET	MODIFIED RESIDUE	UNP Q97ST4

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	27	Total O 27 27	0	0
2	B	27	Total O 27 27	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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: UPF0210 protein SP0239

Chain A:

- Molecule 1: UPF0210 protein SP0239

Chain B:
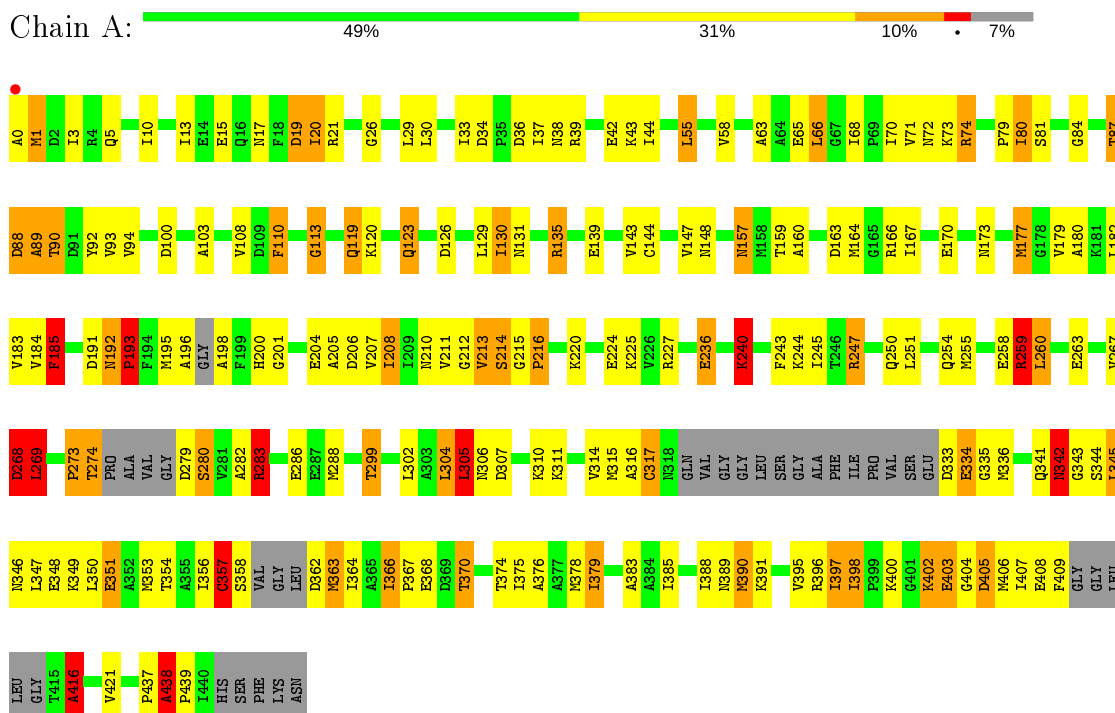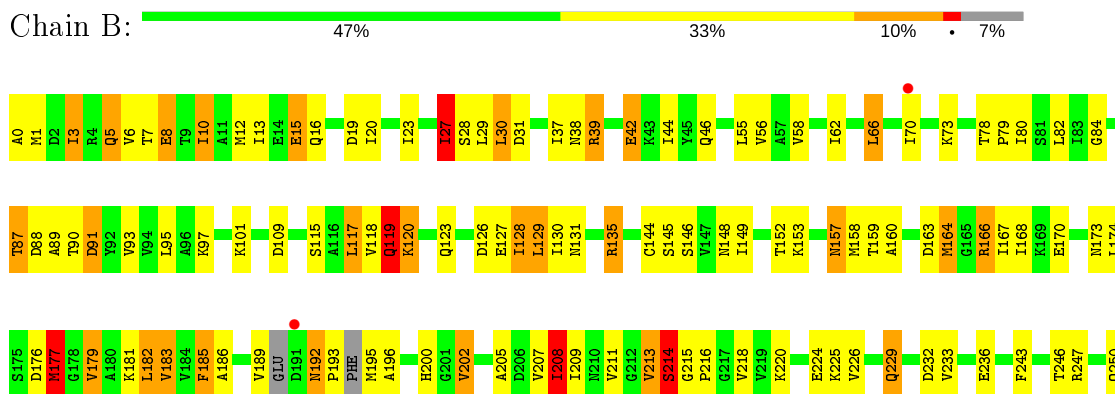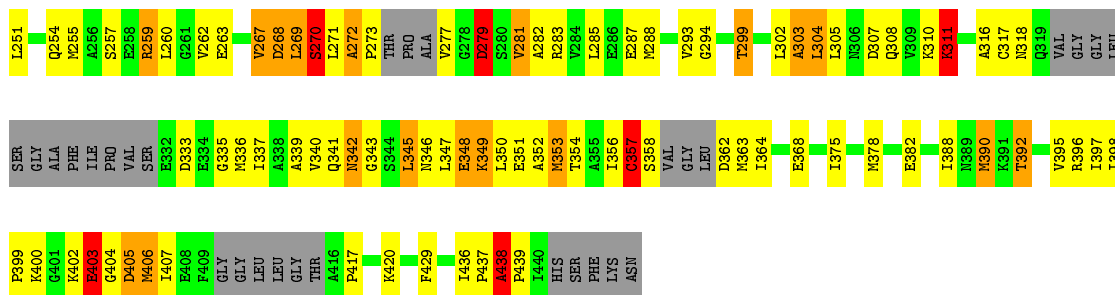
4   Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	119.91Å 139.96Å 148.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.70 43.58 – 2.70	Depositor EDS
% Data completeness (in resolution range)	95.4 (50.00-2.70) 95.4 (43.58-2.70)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.15 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.194 , 0.271 0.191 , 0.266	Depositor DCC
R_{free} test set	1672 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	59.2	Xtriage
Anisotropy	0.306	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 48.5	EDS
L-test for twinning[2]	$<|L|> = 0.48$, $<L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6011	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.26% 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 (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	1.51	22/2982 (0.7%)	1.39	28/4004 (0.7%)
1	B	1.59	31/2982 (1.0%)	1.43	31/4002 (0.8%)
All	All	1.55	53/5964 (0.9%)	1.41	59/8006 (0.7%)

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	A	0	5
1	B	0	3
All	All	0	8

The worst 5 of 53 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	185	PHE	CB-CG	-11.41	1.31	1.51
1	B	123	GLN	CG-CD	8.77	1.71	1.51
1	B	185	PHE	CB-CG	-8.31	1.37	1.51
1	B	144	CYS	CB-SG	-8.30	1.68	1.82
1	A	334	GLU	CG-CD	7.95	1.63	1.51

The worst 5 of 59 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	185	PHE	CB-CA-C	-10.06	90.28	110.40
1	A	208	ILE	CG1-CB-CG2	-9.71	90.05	111.40
1	A	185	PHE	CB-CA-C	-9.63	91.14	110.40
1	A	163	ASP	CB-CG-OD1	9.37	126.73	118.30
1	B	126	ASP	CB-CG-OD2	-9.17	110.05	118.30

There are no chirality outliers.

5 of 8 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	192	ASN	Peptide
1	A	268	ASP	Peptide
1	A	269	LEU	Peptide
1	A	416	ALA	Peptide
1	A	437	PRO	Peptide

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	2978	0	3041	205	0
1	B	2979	0	3035	193	0
2	A	27	0	0	2	0
2	B	27	0	0	5	0
All	All	6011	0	6076	375	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 31.

The worst 5 of 375 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:208:ILE:CD1	1:A:208:ILE:CG1	1.74	1.60
1:B:364:ILE:CG1	1:B:364:ILE:CD1	1.84	1.50
1:B:177:MSE:CG	1:B:177:MSE:SE	2.19	1.39
1:B:251:LEU:HG	1:B:255:MSE:CE	1.68	1.22
1:A:438:ALA:HB2	1:B:311:LYS:HZ3	1.10	1.13

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	402/446 (90%)	356 (89%)	31 (8%)	15 (4%)	3 7
1	B	401/446 (90%)	357 (89%)	26 (6%)	18 (4%)	2 5
All	All	803/892 (90%)	713 (89%)	57 (7%)	33 (4%)	3 6

5 of 33 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	193	PRO
1	A	214	SER
1	A	273	PRO
1	A	403	GLU
1	A	405	ASP

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	310/329 (94%)	253 (82%)	57 (18%)	1 4
1	B	308/329 (94%)	252 (82%)	56 (18%)	1 4
All	All	618/658 (94%)	505 (82%)	113 (18%)	1 4

5 of 113 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	390	MSE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	15	GLU
1	B	337	ILE
1	A	391	LYS
1	B	1	MSE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	342	ASN
1	A	389	ASN
1	B	250	GLN
1	A	341	GLN
1	B	308	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 carbohydrates 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(Å²)	Q<0.9
1	A	397/446 (89%)	-0.51	1 (0%)	94	95	23, 40, 74, 93	0
1	B	398/446 (89%)	-0.49	2 (0%)	91	92	22, 37, 71, 86	0
All	All	795/892 (89%)	-0.50	3 (0%)	92	93	22, 39, 72, 93	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	70	ILE	3.1
1	A	0	ALA	2.7
1	B	191	ASP	2.4

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 carbohydrates 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.