



wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 02:35 AM GMT

PDB ID : 2AJA
Title : X-Ray structure of an ankyrin repeat family protein Q5ZSV0 from Legionella pneumophila. Northeast Structural Genomics Consortium target LgR21.
Authors : Kuzin, A.P.; Chen, Y.; Acton, T.; Xiao, R.; Conover, K.; Ma, C.; Kellie, R.; Montelione, G.T.; Tong, L.; Hunt, J.F.; Northeast Structural Genomics Consortium (NESG)
Deposited on : 2005-08-01
Resolution : 2.80 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

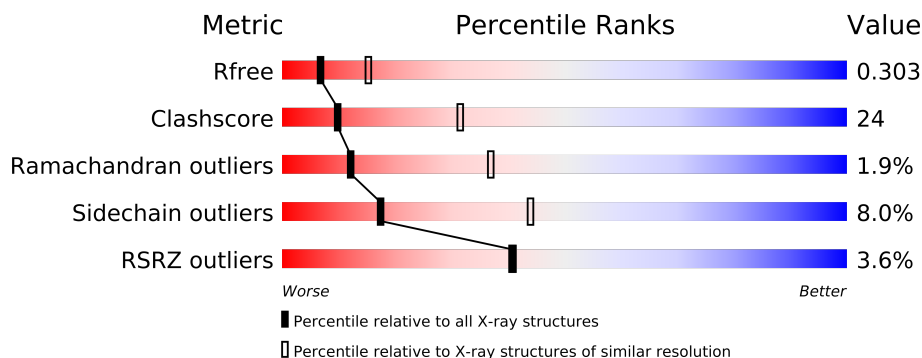
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.80 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}	66092	1799 (2.80-2.80)
Clashscore	79885	2295 (2.80-2.80)
Ramachandran outliers	78287	2252 (2.80-2.80)
Sidechain outliers	78261	2254 (2.80-2.80)
RSRZ outliers	66119	1802 (2.80-2.80)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	376	
1	B	376	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5644 atoms, of which 0 are hydrogen and 0 are deuterium.

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 ankyrin repeat family protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	343	Total	C	N	O	S	Se	0	0	0
			2742	1757	477	491	9	8			
1	B	345	Total	C	N	O	S	Se	0	0	0
			2763	1769	482	494	10	8			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	42	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	60	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	156	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	158	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	193	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	224	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	254	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	281	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
A	369	LEU	-	CLONING ARTIFACT	UNP Q5ZSV0
A	370	GLU	-	CLONING ARTIFACT	UNP Q5ZSV0
A	371	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
A	372	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
A	373	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
A	374	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
A	375	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
A	376	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
B	42	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
B	60	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
B	156	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
B	158	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
B	193	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
B	224	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
B	254	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	281	MSE	MET	MODIFIED RESIDUE	UNP Q5ZSV0
B	369	LEU	-	CLONING ARTIFACT	UNP Q5ZSV0
B	370	GLU	-	CLONING ARTIFACT	UNP Q5ZSV0
B	371	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
B	372	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
B	373	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
B	374	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
B	375	HIS	-	EXPRESSION TAG	UNP Q5ZSV0
B	376	HIS	-	EXPRESSION TAG	UNP Q5ZSV0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	47	Total O 47 47	0	0
2	B	92	Total O 92 92	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	79.61Å 90.46Å 118.72Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.92 – 2.80 29.88 – 2.81	Depositor EDS
% Data completeness (in resolution range)	92.2 (19.92-2.80) 98.5 (29.88-2.81)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.65 (at 2.80Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.221 , 0.286 0.240 , 0.303	Depositor DCC
R_{free} test set	1051 reflections (4.97%)	DCC
Wilson B-factor (Å ²)	38.8	Xtriage
Anisotropy	0.071	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 14.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 40240 reflections	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	5644	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.65% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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	A	0.38	0/2793	0.61	0/3775
1	B	0.43	0/2814	0.61	1/3803 (0.0%)
All	All	0.40	0/5607	0.61	1/7578 (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	B	126	ALA	N-CA-C	-5.70	95.61	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2742	0	2737	166	0
1	B	2763	0	2753	106	0
2	A	47	0	0	3	0
2	B	92	0	0	8	0
All	All	5644	0	5490	265	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 24.

The worst 5 of 265 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:272:LYS:HD3	1:A:272:LYS:H	1.16	1.08
1:B:146:ARG:NH1	2:B:466:HOH:O	1.90	1.02
1:A:287:ARG:HG2	1:A:287:ARG:HH11	1.29	0.97
1:A:208:GLY:HA2	1:B:24:ILE:HG12	1.46	0.97
1:A:62:GLN:NE2	1:A:91:LYS:H	1.68	0.91

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	A	337/376 (90%)	281 (83%)	47 (14%)	9 (3%)	8	25
1	B	337/376 (90%)	308 (91%)	25 (7%)	4 (1%)	19	54
All	All	674/752 (90%)	589 (87%)	72 (11%)	13 (2%)	12	37

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	154	GLU
1	B	77	ASN
1	A	320	ASN
1	B	236	GLY
1	A	53	PHE

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	292/311 (94%)	268 (92%)	24 (8%)	17	43
1	B	294/311 (94%)	271 (92%)	23 (8%)	18	45
All	All	586/622 (94%)	539 (92%)	47 (8%)	17	44

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	324	ARG
1	B	60	MSE
1	B	332	GLN
1	A	347	LEU
1	B	76	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	255	HIS
1	B	8	ASN
1	B	284	ASN
1	A	284	ASN
1	A	80	GLN

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates 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(Å ²)	Q<0.9
1	A	343/376 (91%)	0.07	12 (3%) 42 42	11, 45, 61, 61	0
1	B	345/376 (91%)	-0.28	13 (3%) 38 38	11, 25, 57, 61	0
All	All	688/752 (91%)	-0.11	25 (3%) 41 41	11, 34, 60, 61	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	125	GLN	6.1
1	A	77	ASN	4.3
1	A	267	PHE	4.0
1	B	260	LEU	3.8
1	B	77	ASN	3.7

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 carbohydrates 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.