



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 05:02 pm BST

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 X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 ⓘ](#)) 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

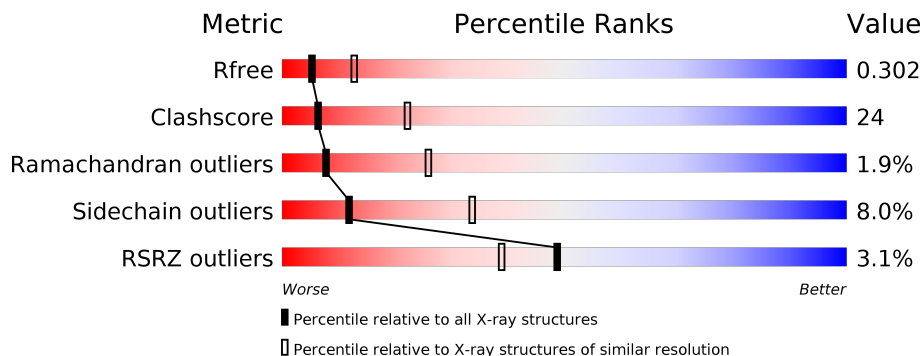
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

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))
$R_{free}$	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (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  $\geq 3$ , 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  $\leq 5\%$ . 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	376	
1	B	376	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 5644 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 ankyrin repeat family protein.

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

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

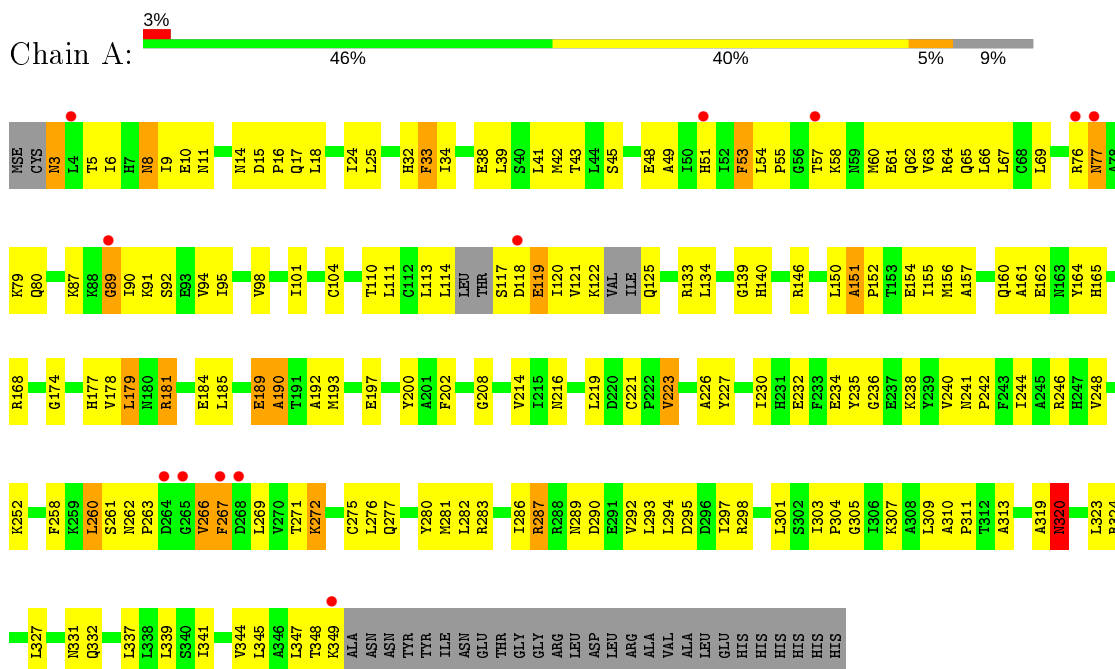
- Molecule 2 is water.

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

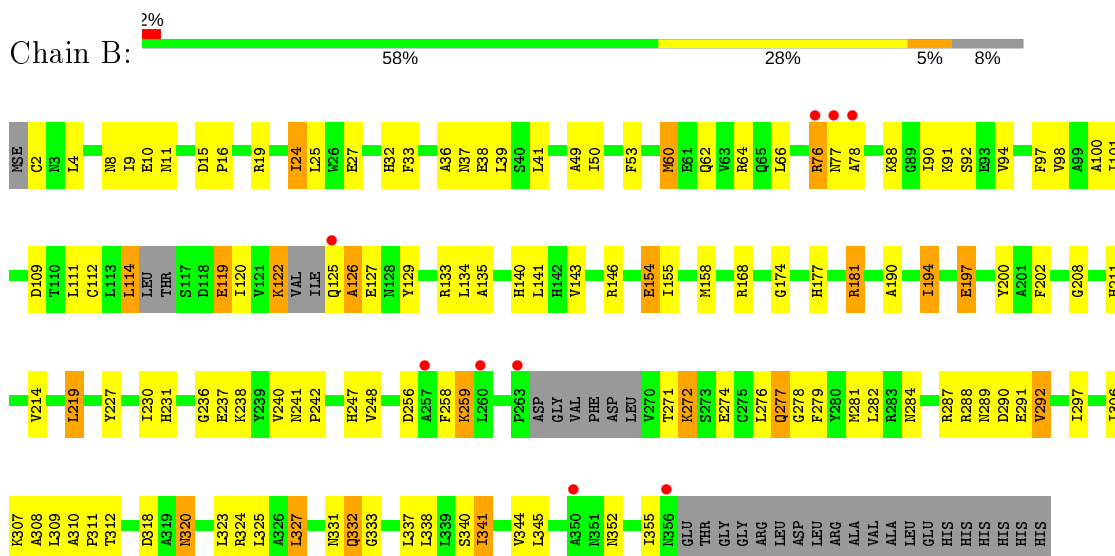
### 3 Residue-property plots [i](#)

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: ankyrin repeat family protein



- Molecule 1: ankyrin repeat family protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	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
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.65 (at 2.80Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.221 , 0.286 0.240 , 0.302	Depositor DCC
$R_{free}$ test set	2025 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.8	Xtrriage
Anisotropy	0.071	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 38.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	5644	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtrriage'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.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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	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 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	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

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 24.

The worst 5 of 265 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: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 [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	337/376 (90%)	281 (83%)	47 (14%)	9 (3%)	5	17
1	B	337/376 (90%)	308 (91%)	25 (7%)	4 (1%)	13	39
All	All	674/752 (90%)	589 (87%)	72 (11%)	13 (2%)	8	26

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 [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	292/311 (94%)	268 (92%)	24 (8%)	11	33
1	B	294/311 (94%)	271 (92%)	23 (8%)	12	35
All	All	586/622 (94%)	539 (92%)	47 (8%)	12	34

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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	335/376 (89%)	0.05	12 (3%) 42 32	11, 45, 61, 61	0
1	B	337/376 (89%)	-0.31	9 (2%) 54 44	11, 24, 57, 61	0
All	All	672/752 (89%)	-0.13	21 (3%) 49 39	11, 33, 60, 61	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	125	GLN	5.4
1	A	77	ASN	3.7
1	B	263	PRO	3.5
1	B	77	ASN	3.3
1	A	267	PHE	2.8

### 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.