



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 04:28 am GMT

PDB ID : 4ZT0  
Title : Crystal structure of catalytically-active Streptococcus pyogenes Cas9 in complex with single-guide RNA at 2.9 Angstrom resolution  
Authors : Jiang, F.; Doudna, J.A.  
Deposited on : 2015-05-14  
Resolution : 2.90 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

<http://wwpdb.org/validation/2016/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  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

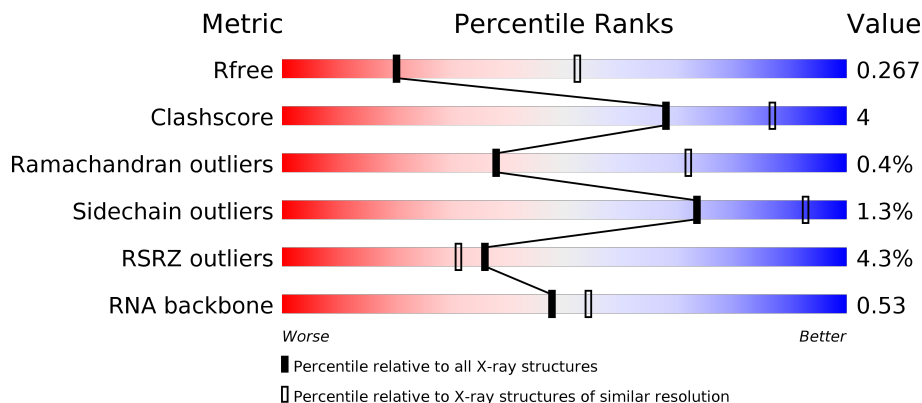
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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))
$R_{free}$	100719	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)
RNA backbone	2435	1004 (3.20-2.60)

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. 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	1369	<div> <div>3%</div> <div>87%</div> <div>12%</div> </div>
1	C	1369	<div> <div>5%</div> <div>83%</div> <div>11%</div> <div>5%</div> </div>
2	B	85	<div> <div>%</div> <div>49%</div> <div>31%</div> <div>5%</div> <div>15%</div> </div>
2	D	85	<div> <div>45%</div> <div>36%</div> <div>•</div> <div>15%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 23790 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 CRISPR-associated endonuclease Cas9.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1364	Total	C	N	O	S	Se	0	0	0
			10654	6778	1829	2025	2	20			
1	C	1296	Total	C	N	O	S	Se	0	0	0
			10046	6393	1728	1904	2	19			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP A0A0C6FZC2
A	1	MSE	-	expression tag	UNP A0A0C6FZC2
C	0	SER	-	expression tag	UNP A0A0C6FZC2
C	1	MSE	-	expression tag	UNP A0A0C6FZC2

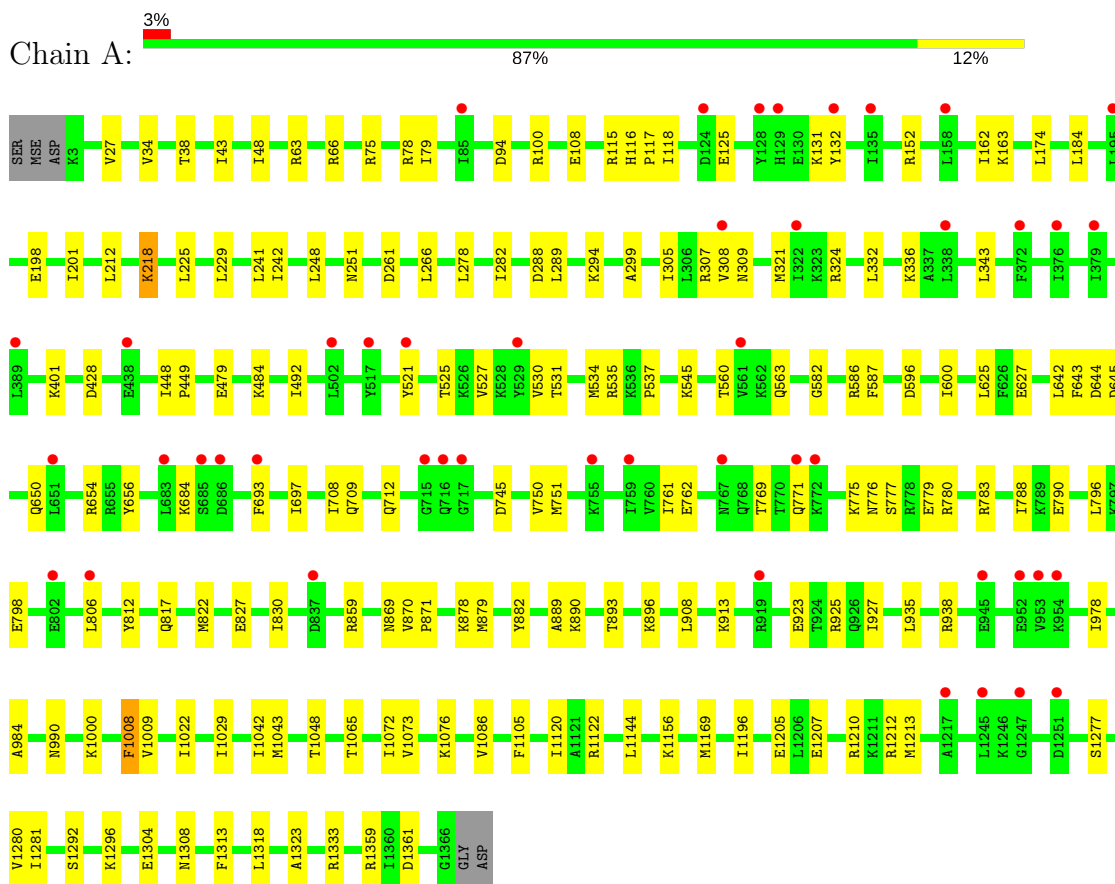
- Molecule 2 is a RNA chain called single-guide RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	72	Total	C	N	O	P	0	0	0
			1545	692	285	496	72			
2	D	72	Total	C	N	O	P	0	0	0
			1545	692	285	496	72			

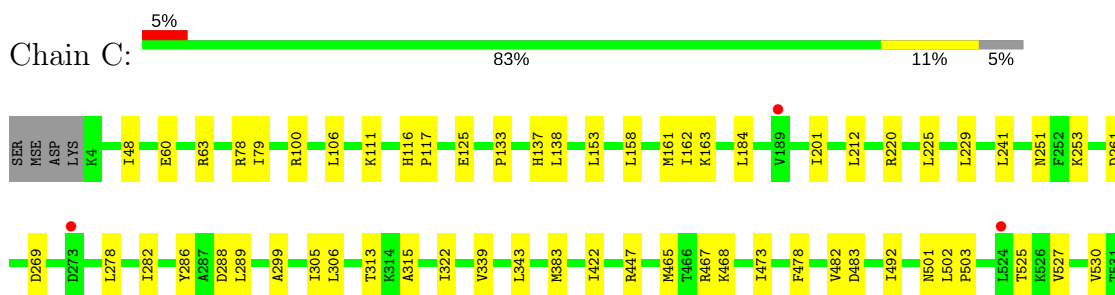
### 3 Residue-property plots [i](#)

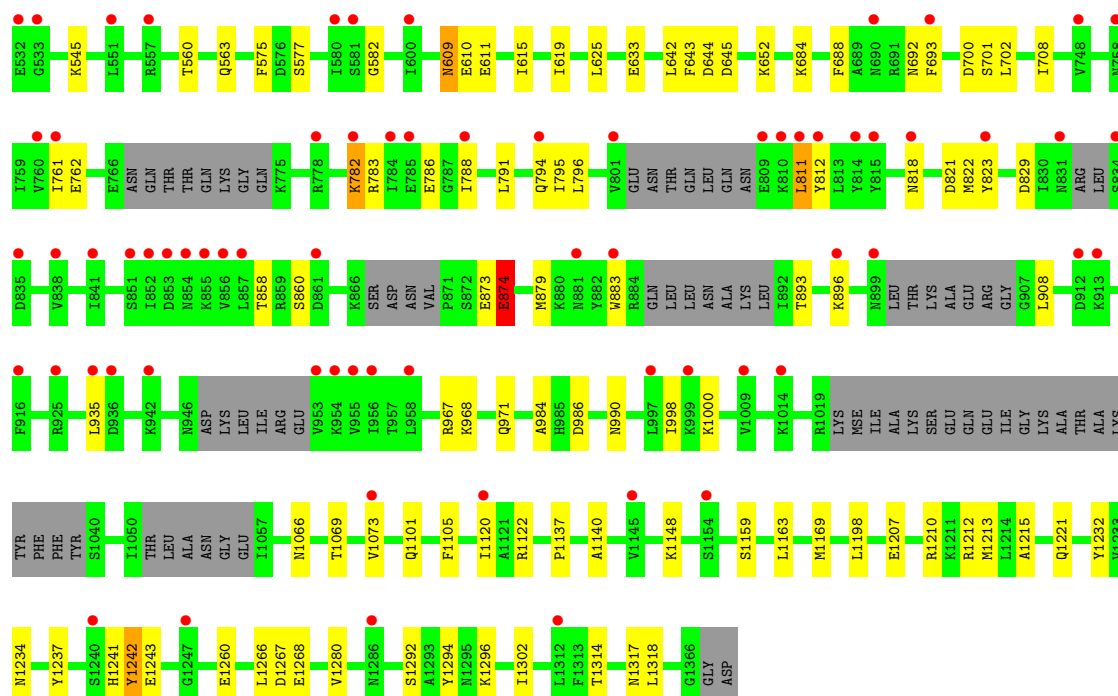
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: CRISPR-associated endonuclease Cas9

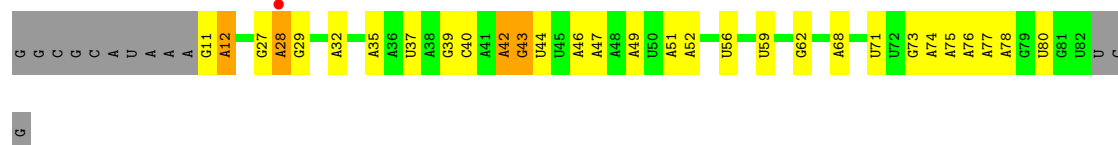


#### • Molecule 1: CRISPR-associated endonuclease Cas9

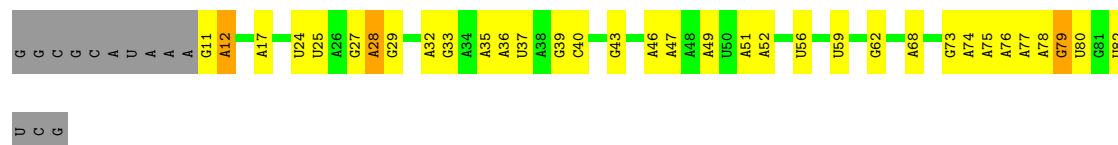




• Molecule 2: single-guide RNA



• Molecule 2: single-guide RNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	103.24Å 143.00Å 294.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.15 – 2.90 102.66 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.3 (49.15-2.90) 99.5 (102.66-2.90)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.88 (at 2.91Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.238 , 0.264 0.244 , 0.267	Depositor DCC
$R_{free}$ test set	4809 reflections (4.96%)	DCC
Wilson B-factor (Å <sup>2</sup> )	69.7	Xtriage
Anisotropy	0.112	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 46.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	23790	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 28.42 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.8491e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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	0.25	0/10825	0.37	0/14614
1	C	0.26	0/10200	0.38	0/13759
2	B	0.17	0/1732	0.70	0/2698
2	D	0.17	0/1732	0.70	0/2698
All	All	0.24	0/24489	0.44	0/33769

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 [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	10654	0	10330	92	0
1	C	10046	0	9673	84	0
2	B	1545	0	774	15	0
2	D	1545	0	774	13	0
All	All	23790	0	21551	199	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 4.

The worst 5 of 199 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:C:1120:ILE:HD11	1:C:1137:PRO:HG3	1.66	0.78
2:B:32:A:H61	2:B:37:U:H3	1.31	0.77
1:C:1122:ARG:NH2	2:D:49:A:N3	2.35	0.74
1:A:218:LYS:HG3	1:A:248:LEU:HD21	1.70	0.73
1:A:762:GLU:OE1	1:A:990:ASN:ND2	2.21	0.73

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	1362/1369 (100%)	1301 (96%)	59 (4%)	2 (0%)	55	85
1	C	1276/1369 (93%)	1216 (95%)	51 (4%)	9 (1%)	25	60
All	All	2638/2738 (96%)	2517 (95%)	110 (4%)	11 (0%)	38	72

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	308	VAL
1	C	874	GLU
1	C	688	PHE
1	C	1242	TYR
1	C	692	ASN

### 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	1106/1206 (92%)	1094 (99%)	12 (1%)	78	94
1	C	1031/1206 (86%)	1015 (98%)	16 (2%)	68	90
All	All	2137/2412 (89%)	2109 (99%)	28 (1%)	73	93

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	288	ASP
1	C	642	LEU
1	C	998	ILE
1	C	530	VAL
1	C	609	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	71/85 (83%)	14 (19%)	0
2	D	71/85 (83%)	16 (22%)	0
All	All	142/170 (83%)	30 (21%)	0

5 of 30 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	12	A
2	B	28	A
2	B	29	G
2	B	35	A
2	B	39	G

There are no RNA pucker outliers to report.

## 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 [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	1343/1369 (98%)	0.22	46 (3%) 46 39	37, 97, 152, 216	0
1	C	1276/1369 (93%)	0.36	72 (5%) 25 20	39, 93, 181, 232	0
2	B	72/85 (84%)	0.06	1 (1%) 75 74	46, 78, 221, 260	0
2	D	72/85 (84%)	0.09	0 100 100	47, 65, 156, 194	0
All	All	2763/2908 (95%)	0.28	119 (4%) 36 31	37, 94, 170, 260	0

The worst 5 of 119 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	716	GLN	8.6
1	C	811	LEU	7.0
1	C	856	VAL	6.9
1	A	767	ASN	6.5
1	C	835	ASP	6.4

### 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.