



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 15, 2023 – 10:43 PM JST

PDB ID : 6K4U  
Title : Crystal structure of xCas9 in complex with sgRNA and DNA (TGA PAM)  
Authors : Chen, W.; Zhang, H.; Zhang, Y.; Wang, Y.; Gan, J.; Ji, Q.  
Deposited on : 2019-05-27  
Resolution : 3.20 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

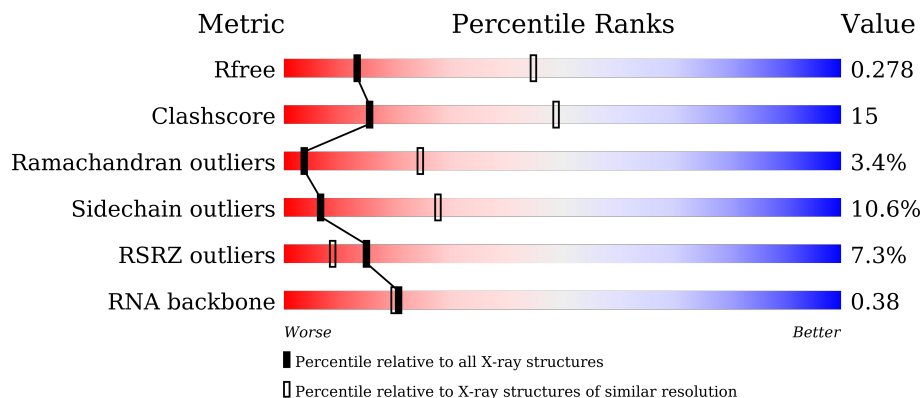
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)
RNA backbone	3102	1010 (3.50-2.90)

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 of 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	B	1368	
2	A	83	
3	C	28	
4	D	12	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 12814 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
			Total	C	N	O	S			
1	B	1312	10326	6583	1785	1941	17	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	262	THR	ALA	engineered mutation	UNP Q99ZW2
B	324	LEU	ARG	engineered mutation	UNP Q99ZW2
B	409	ILE	SER	engineered mutation	UNP Q99ZW2
B	480	LYS	GLU	engineered mutation	UNP Q99ZW2
B	543	ASP	GLU	engineered mutation	UNP Q99ZW2
B	694	ILE	MET	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
B	1219	VAL	GLU	engineered mutation	UNP Q99ZW2

- Molecule 2 is a RNA chain called sgRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	A	80	1710	768	313	549	80	0	0	0

- Molecule 3 is a DNA chain called target DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	27	552	268	92	165	27	0	0	0

- Molecule 4 is a DNA chain called non-target DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	11	226	110	43	63	10	0	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	178.00Å 70.31Å 190.08Å 90.00° 110.34° 90.00°	Depositor
Resolution (Å)	49.38 – 3.20 49.34 – 3.20	Depositor EDS
% Data completeness (in resolution range)	81.8 (49.38-3.20) 81.9 (49.34-3.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.94 (at 3.19Å)	Xtrriage
Refinement program	REFMAC 5.8.0238	Depositor
R, $R_{free}$	0.220 , 0.282 0.221 , 0.278	Depositor DCC
$R_{free}$ test set	1566 reflections (5.19%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	60.7	Xtrriage
Anisotropy	0.139	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 43.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	12814	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	84.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.40% 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	B	0.74	0/10504	0.90	0/14169
2	A	1.02	17/1917 (0.9%)	0.95	0/2984
3	C	0.70	1/617 (0.2%)	0.88	0/950
4	D	1.19	2/254 (0.8%)	0.87	0/391
All	All	0.80	20/13292 (0.2%)	0.91	0/18494

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	6	DG	O3'-P	-10.46	1.48	1.61
2	A	18	A	O3'-P	-9.44	1.49	1.61
2	A	49	A	O3'-P	-9.16	1.50	1.61
2	A	12	U	O3'-P	-8.95	1.50	1.61
4	D	5	DT	O3'-P	-8.53	1.50	1.61

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	B	10326	0	10134	301	0
2	A	1710	0	858	65	0
3	C	552	0	311	12	0
4	D	226	0	127	6	0

*Continued on next page...*



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	12814	0	11430	359	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 15.

The worst 5 of 359 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)
2:A:18:A:H5'	2:A:18:A:H8	1.11	1.15
2:A:37:U:C2	2:A:38:A:C8	2.48	1.00
2:A:18:A:H5'	2:A:18:A:C8	1.98	0.99
1:B:615:ILE:HD12	1:B:615:ILE:H	1.27	0.98
2:A:42:A:O2'	2:A:43:G:OP1	1.87	0.93

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	B	1296/1368 (95%)	1077 (83%)	175 (14%)	44 (3%)	<b>3</b> <b>24</b>

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	303	SER
1	B	584	GLU
1	B	653	ARG
1	B	670	ILE
1	B	685	SER

### 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	B	1074/1226 (88%)	960 (89%)	114 (11%)	<b>6</b> <b>27</b>

5 of 114 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	682	PHE
1	B	1335	ARG
1	B	853	ASP
1	B	1334	LYS
1	B	1252	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	1066	ASN
1	B	726	ASN
1	B	436	ASN
1	B	413	GLN
1	B	504	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	A	79/83 (95%)	17 (21%)	4 (5%)

5 of 17 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	A	9	A
2	A	10	U
2	A	18	A
2	A	19	A
2	A	27	G

All (4) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	A	8	A
2	A	9	A
2	A	18	A
2	A	42	A

#### 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 monosaccharides 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	B	1312/1368 (95%)	0.24	102 (7%) 13 7	24, 83, 157, 224	0
2	A	80/83 (96%)	-0.22	0 100 100	23, 50, 125, 143	0
3	C	27/28 (96%)	0.04	0 100 100	46, 67, 130, 155	0
4	D	11/12 (91%)	0.72	3 (27%) 0 0	49, 76, 162, 168	0
All	All	1430/1491 (95%)	0.21	105 (7%) 15 9	23, 82, 155, 224	0

The worst 5 of 105 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	225	LEU	9.2
1	B	252	PHE	7.5
1	B	1030	GLY	7.2
1	B	229	LEU	7.0
1	B	237	LEU	6.3

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.