



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2020 – 02:02 pm BST

PDB ID : 6AEG  
Title : Crystal structure of xCas9 in complex with sgRNA and target DNA (GAT PAM)  
Authors : Guo, M.; Ren, K.; Zhu, Y.; Huang, Z.  
Deposited on : 2018-08-04  
Resolution : 2.70 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
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

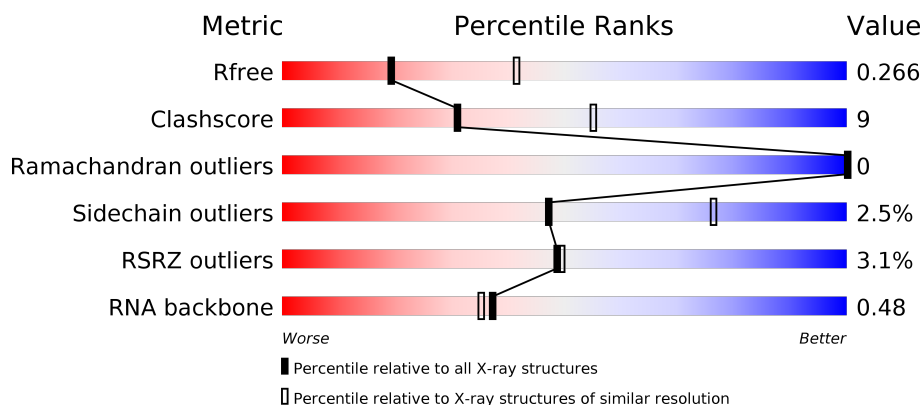
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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)
RNA backbone	3102	1159 (3.00-2.40)

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	100	<div> <div>9%</div> <div>54% 26% 13% 5%</div> </div>
2	C	28	<div> <div>4%</div> <div>43% 46% 11%</div> </div>
3	D	11	<div> <div>9%</div> <div>64% 36%</div> </div>
4	B	1368	<div> <div>3%</div> <div>76% 20% ..</div> </div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 13362 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 RNA chain called RNA (95-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	95	Total	C	N	O	P	0	0	0
			2032	911	372	654	95			

- Molecule 2 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	25	Total	C	N	O	P	0	0	0
			507	247	86	150	24			

- Molecule 3 is a DNA chain called DNA (5'-D(\*AP\*AP\*AP\*GP\*AP\*TP\*TP\*AP\*TP\*TP\*G)-3').

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

- Molecule 4 is a protein called DNA nuclease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	1322	Total	C	N	O	S	0	0	0
			10475	6690	1817	1947	21			

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

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2
B	1219	VAL	GLU	engineered mutation	UNP Q99ZW2

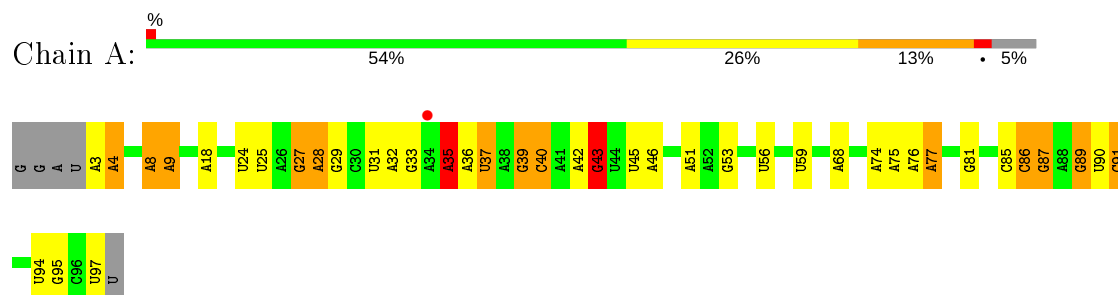
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	26	Total O 26 26	0	0
5	C	10	Total O 10 10	0	0
5	D	6	Total O 6 6	0	0
5	B	80	Total O 80 80	0	0

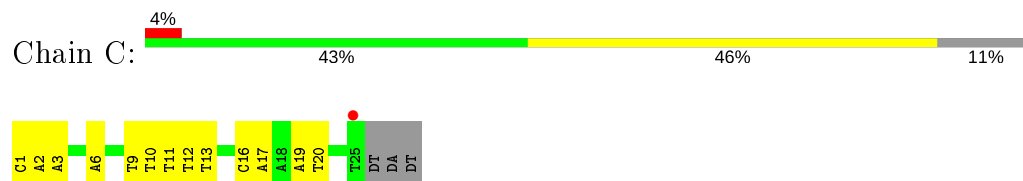
### 3 Residue-property plots [i](#)

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.

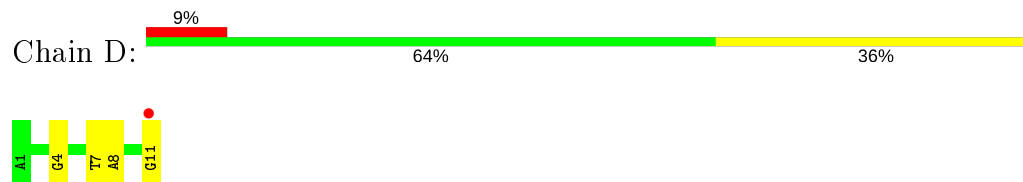
- Molecule 1: RNA (95-MER)



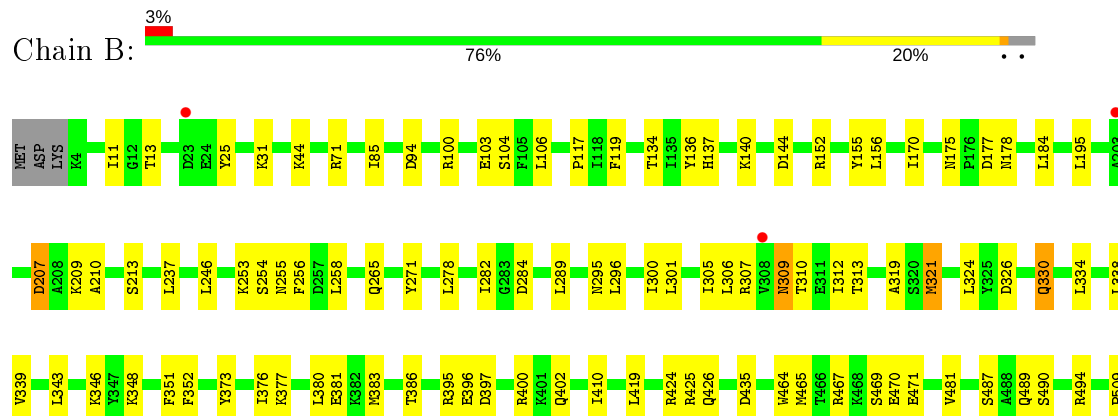
- Molecule 2: DNA (25-MER)

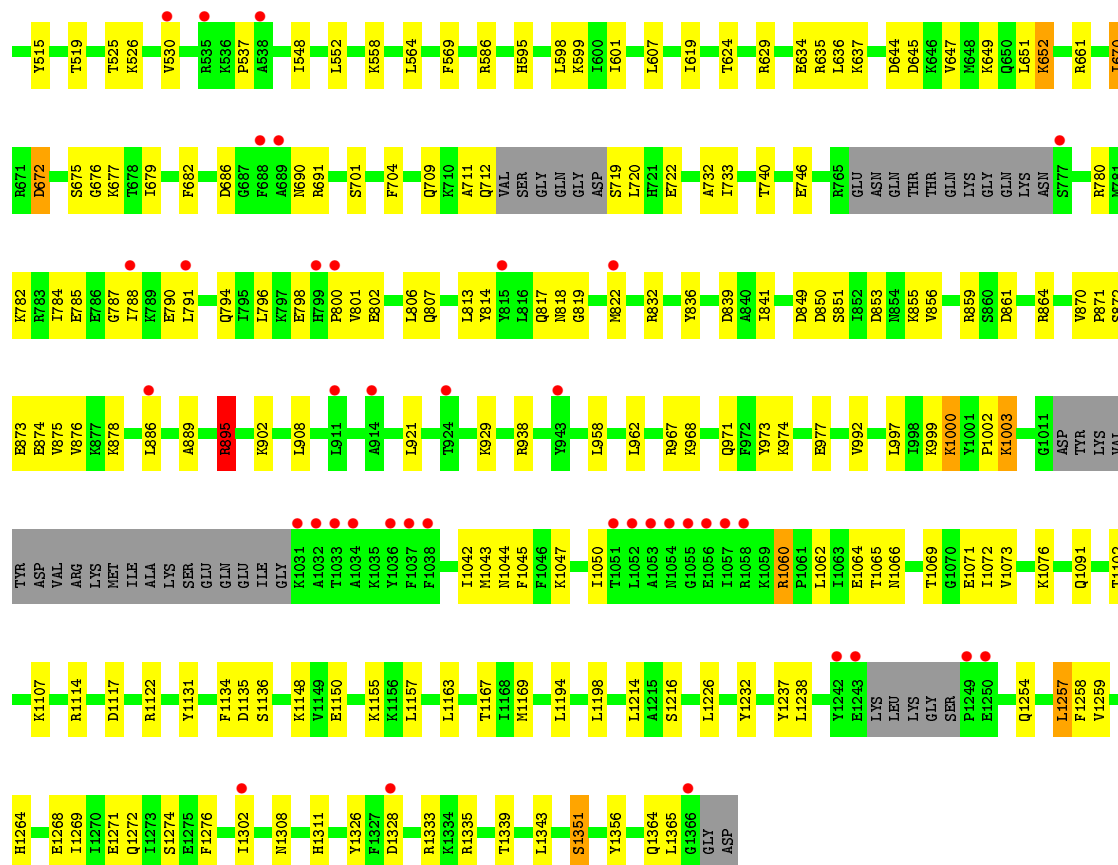


- Molecule 3: DNA (5'-D(\*AP\*AP\*AP\*GP\*AP\*TP\*TP\*AP\*TP\*TP\*G)-3')



- Molecule 4: DNA nuclease





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	190.48Å 70.46Å 183.53Å 90.00° 106.48° 90.00°	Depositor
Resolution (Å)	44.97 – 2.70 47.11 – 2.70	Depositor EDS
% Data completeness (in resolution range)	74.8 (44.97-2.70) 83.5 (47.11-2.70)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.10 (at 2.69Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, $R_{free}$	0.215 , 0.263 0.217 , 0.266	Depositor DCC
$R_{free}$ test set	2000 reflections (3.55%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.6	Xtriage
Anisotropy	0.008	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 40.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	13362	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.47% 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	0.40	0/2277	0.87	2/3546 (0.1%)
2	C	0.85	1/567 (0.2%)	1.13	2/873 (0.2%)
3	D	0.76	0/254	0.97	0/391
4	B	0.40	1/10664 (0.0%)	0.53	4/14393 (0.0%)
All	All	0.44	2/13762 (0.0%)	0.66	8/19203 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	6	DA	O3'-P	-6.33	1.53	1.61
4	B	672	ASP	CB-CG	-5.35	1.40	1.51

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	672	ASP	N-CA-C	-7.17	91.63	111.00
4	B	1257	LEU	CA-CB-CG	6.66	130.63	115.30
2	C	12	DT	O4'-C4'-C3'	-6.51	101.90	104.50
1	A	35	A	C8-N9-C4	-6.06	103.38	105.80
1	A	43	G	C4-C5-C6	5.83	122.30	118.80

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	2032	0	1021	26	1
2	C	507	0	288	9	0
3	D	226	0	127	4	0
4	B	10475	0	10378	200	1
5	A	26	0	0	3	0
5	B	80	0	0	5	0
5	C	10	0	0	1	0
5	D	6	0	0	1	0
All	All	13362	0	11814	227	1

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 9.

The worst 5 of 227 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)
4:B:1308:ASN:OD1	4:B:1326:TYR:O	1.54	1.19
4:B:530:VAL:HG22	4:B:537:PRO:HA	1.32	1.10
4:B:870:VAL:HG22	4:B:871:PRO:HD2	1.41	0.99
4:B:1069:THR:HG22	4:B:1071:GLU:H	1.40	0.86
4:B:958:LEU:HD22	4:B:962:LEU:HD13	1.55	0.86

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:94:U:OP1	4:B:424:ARG:NH1[3_445]	2.12	0.08

## 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	
4	B	1312/1368 (96%)	1226 (93%)	86 (7%)	0	100	100

There are no Ramachandran outliers to report.

### 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	
4	B	1103/1226 (90%)	1075 (98%)	28 (2%)	47	76

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

Mol	Chain	Res	Type
4	B	670	ILE
4	B	861	ASP
4	B	1333	ARG
4	B	832	ARG
4	B	853	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	650	GLN
4	B	1308	ASN
4	B	668	ASN
4	B	330	GLN
4	B	1297	HIS

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	94/100 (94%)	25 (26%)	4 (4%)

5 of 25 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	4	A
1	A	9	A
1	A	24	U
1	A	28	A
1	A	29	G

All (4) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	8	A
1	A	27	G
1	A	28	A
1	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 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	95/100 (95%)	-0.23	1 (1%) 80 82	12, 33, 82, 111	0
2	C	25/28 (89%)	0.17	1 (4%) 38 37	15, 26, 71, 91	0
3	D	11/11 (100%)	0.55	1 (9%) 9 7	22, 36, 78, 83	0
4	B	1322/1368 (96%)	0.14	42 (3%) 47 48	8, 35, 69, 108	0
All	All	1453/1507 (96%)	0.12	45 (3%) 49 49	8, 35, 71, 111	0

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	1054	ASN	8.2
4	B	1058	ARG	6.4
4	B	1052	LEU	6.1
4	B	1057	ILE	6.0
4	B	1032	ALA	5.5

### 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.