



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 03:37 pm BST

PDB ID : 6KC7  
Title : Crystal structure of Nme1Cas9 in complex with sgRNA and target DNA (ATATGATT PAM) in seed-base paring state  
Authors : Sun, W.; Yang, J.; Cheng, Z.; Liu, C.; Wang, K.; Huang, X.; Wang, Y.  
Deposited on : 2019-06-27  
Resolution : 3.30 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

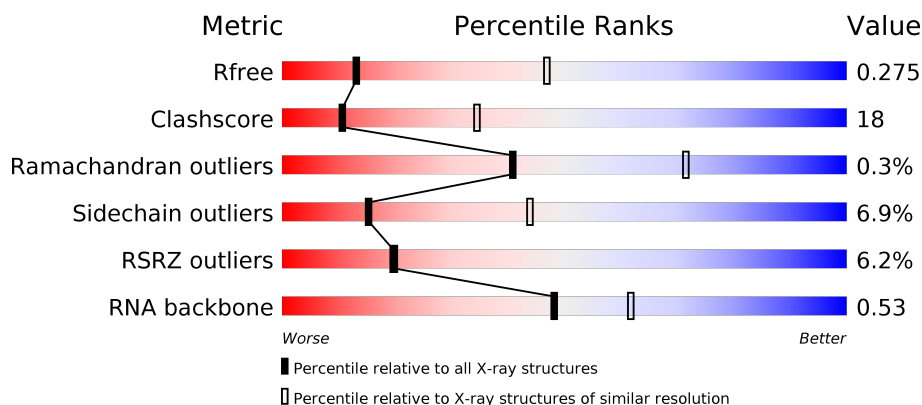
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)
RNA backbone	3102	1117 (3.70-2.90)

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 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	1083	<div> <div>6%</div> <div> <div>55%</div> <div>31%</div> <div>•</div> <div>11%</div> </div> </div>
2	B	135	<div> <div>2%</div> <div> <div>34%</div> <div>40%</div> <div>7%</div> <div>•</div> <div>18%</div> </div> </div>
3	C	19	<div> <div>53%</div> <div>47%</div> </div>
4	D	11	<div> <div>82%</div> <div>18%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10502 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	963	Total	C	N	O	S	0	0	0
			7540	4775	1393	1352	20			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP C9X1G5
A	588	ALA	HIS	engineered mutation	UNP C9X1G5

- Molecule 2 is a RNA chain called sgRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	111	Total	C	N	O	P	0	0	0
			2350	1052	409	778	111			

- Molecule 3 is a DNA chain called DNA (5'-D(\*TP\*AP\*AP\*AP\*AP\*TP\*CP\*AP\*TP\*AP\*TP\*GP\*TP\*AP\*AP\*AP\*GP\*TP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	19	Total	C	N	O	P	0	0	0
			389	189	72	110	18			

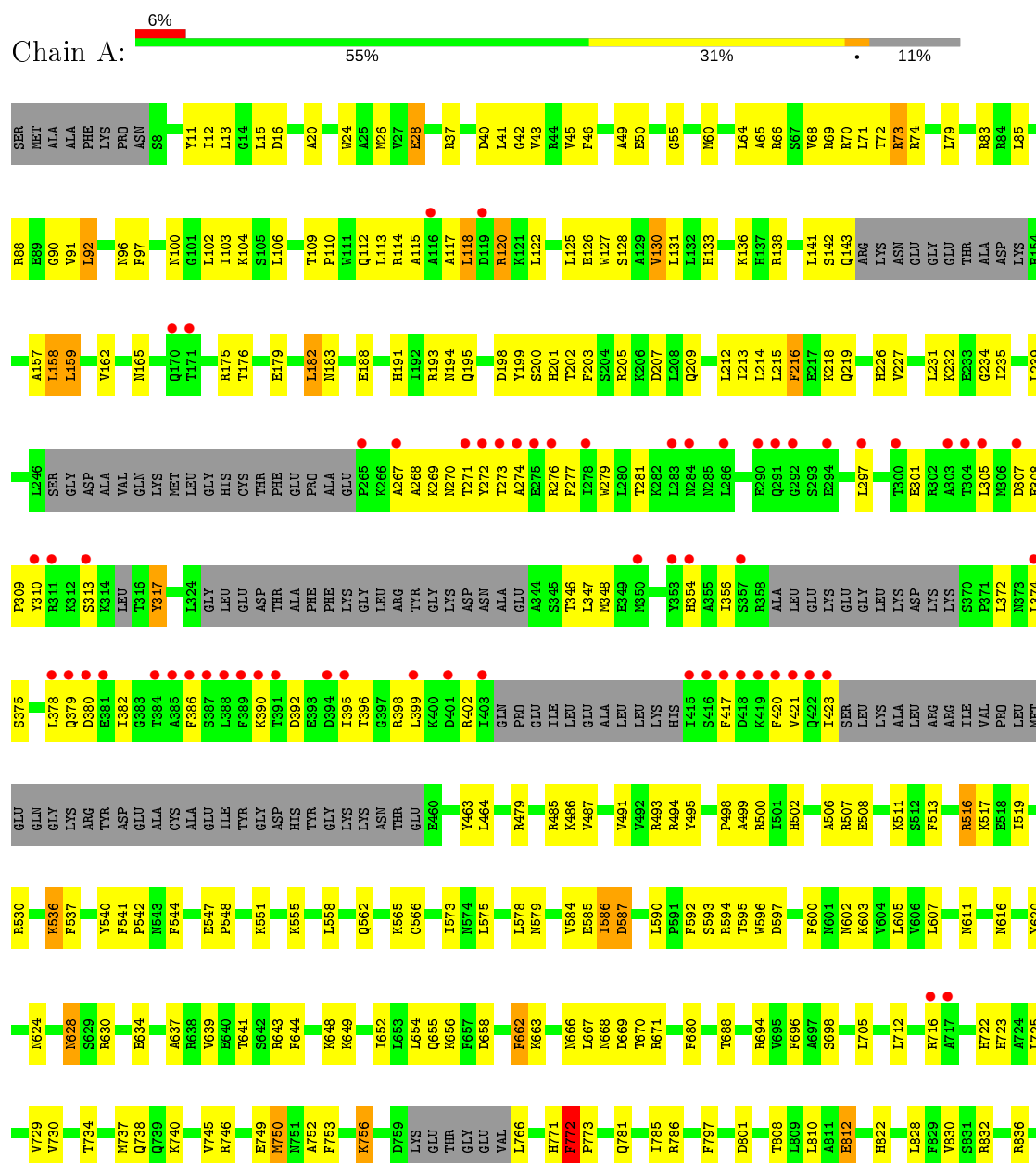
- Molecule 4 is a DNA chain called DNA (5'-D(\*AP\*TP\*AP\*TP\*GP\*AP\*TP\*TP\*TP\*TP\*A)-3').

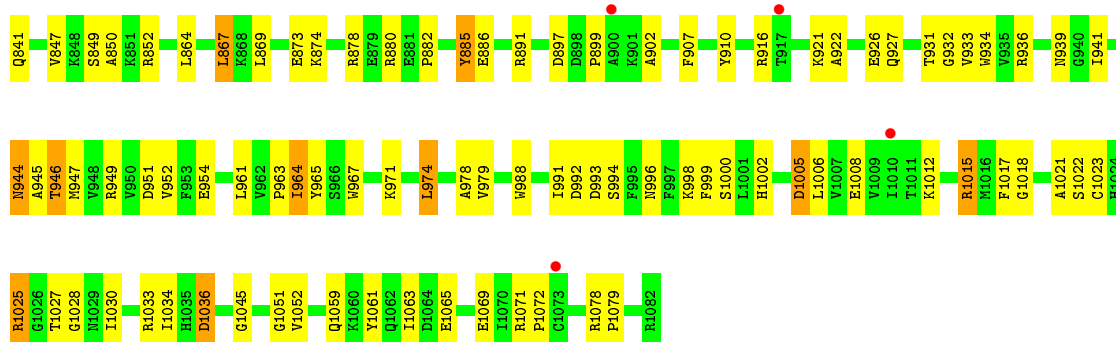
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	11	Total	C	N	O	P	0	0	0
			223	110	37	66	10			

### 3 Residue-property plots

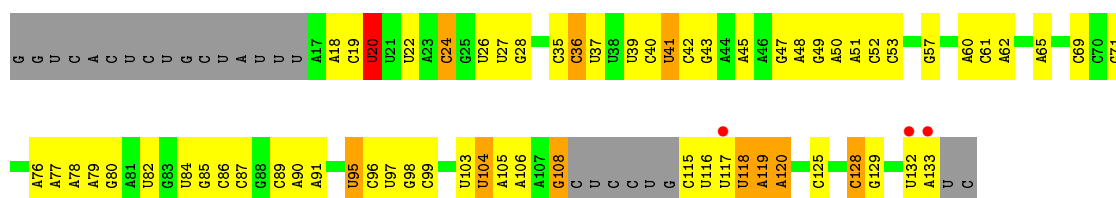
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are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





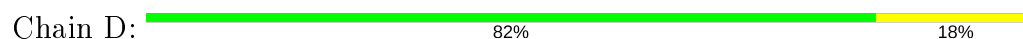
• Molecule 2: sgRNA



• Molecule 3: DNA (5'-D(\*TP\*AP\*AP\*AP\*AP\*TP\*CP\*AP\*TP\*AP\*TP\*GP\*TP\*AP\*AP\*AP\*GP\*TP\*T)-3')



• Molecule 4: DNA (5'-D(\*AP\*TP\*AP\*TP\*GP\*AP\*TP\*TP\*TP\*TP\*A)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	235.58Å 72.51Å 115.15Å 90.00° 104.67° 90.00°	Depositor
Resolution (Å)	46.20 – 3.30 46.20 – 3.30	Depositor EDS
% Data completeness (in resolution range)	86.0 (46.20-3.30) 86.0 (46.20-3.30)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.35 (at 3.32Å)	Xtriage
Refinement program	PHENIX (1.14_3247: ???)	Depositor
R, $R_{free}$	0.253 , 0.275 0.253 , 0.275	Depositor DCC
$R_{free}$ test set	1251 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	51.1	Xtriage
Anisotropy	0.068	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 27.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.84	EDS
Total number of atoms	10502	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.44% 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	0.59	4/7684 (0.1%)	0.68	8/10368 (0.1%)
2	B	0.31	0/2623	0.97	3/4079 (0.1%)
3	C	0.65	0/437	1.03	0/673
4	D	0.63	0/249	1.08	0/383
All	All	0.54	4/10993 (0.0%)	0.80	11/15503 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	934	TRP	CD2-CE2	-6.44	1.33	1.41
1	A	934	TRP	CE3-CZ3	-5.19	1.29	1.38
1	A	933	VAL	C-O	-5.19	1.13	1.23
1	A	934	TRP	CZ3-CH2	-5.06	1.31	1.40

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	92	LEU	CB-CG-CD1	7.79	124.25	111.00
2	B	20	U	C5-C6-N1	6.26	125.83	122.70
2	B	41	U	P-O3'-C3'	5.64	126.47	119.70
1	A	1036	ASP	CB-CG-OD2	5.45	123.20	118.30
1	A	716	ARG	N-CA-C	5.32	125.37	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	7540	0	7382	325	0
2	B	2350	0	1193	52	0
3	C	389	0	218	8	0
4	D	223	0	129	2	0
All	All	10502	0	8922	357	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 18.

The worst 5 of 357 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:267:ALA:HB3	1:A:421:VAL:CG2	1.62	1.29
1:A:267:ALA:CB	1:A:421:VAL:HG21	1.64	1.26
1:A:267:ALA:CB	1:A:421:VAL:CG2	2.15	1.24
1:A:269:LYS:HG2	1:A:421:VAL:CG1	1.84	1.05
1:A:175:ARG:HB2	1:A:179:GLU:OE2	1.58	1.02

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	945/1083 (87%)	864 (91%)	78 (8%)	3 (0%)	41 71

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	270	ASN
1	A	897	ASP
1	A	772	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	753/932 (81%)	701 (93%)	52 (7%)	15	44

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	586	ILE
1	A	628	ASN
1	A	1012	LYS
1	A	587	ASP
1	A	594	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	195	GLN
1	A	822	HIS
1	A	944	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	109/135 (80%)	31 (28%)	1 (0%)

5 of 31 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	20	U
2	B	24	C
2	B	36	C
2	B	37	U
2	B	42	C

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	41	U

#### 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	963/1083 (88%)	0.18	66 (6%) 16 16	12, 42, 71, 110	117 (12%)
2	B	111/135 (82%)	0.16	3 (2%) 54 52	13, 49, 138, 158	0
3	C	19/19 (100%)	-0.22	0 100 100	15, 36, 47, 50	0
4	D	11/11 (100%)	-0.36	0 100 100	14, 22, 61, 69	0
All	All	1104/1248 (88%)	0.16	69 (6%) 20 20	12, 42, 74, 158	117 (10%)

The worst 5 of 69 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	378	LEU	8.1
1	A	291	GLN	7.8
1	A	292	GLY	7.8
1	A	395	ILE	6.9
1	A	388	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 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.