



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:02 PM GMT

PDB ID : 4UN3  
Title : Crystal structure of Cas9 bound to PAM-containing DNA target  
Authors : Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M.  
Deposited on : 2014-05-25  
Resolution : 2.59 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

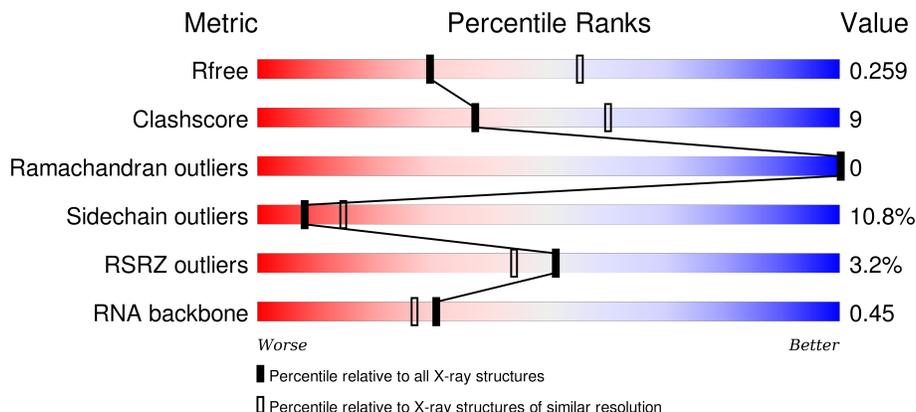
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.59 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))
$R_{free}$	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)
RNA backbone	2183	1022 (3.00-2.20)

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	83	 45% 31% 22% 3%
2	B	1372	 68% 24% 5% 3%
3	C	28	 82% 14% 8%
4	D	12	 75% 17% 8%

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 13629 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 SGRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	81	1732	778	318	555	81	0	0	0

- Molecule 2 is a protein called CRISPR-ASSOCIATED ENDONUCLEASE CAS9/CSN1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1306	10690	6816	1854	1998	22	0	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	EXPRESSION TAG	UNP Q99ZW2
B	-2	ALA	-	EXPRESSION TAG	UNP Q99ZW2
B	-1	ALA	-	EXPRESSION TAG	UNP Q99ZW2
B	0	SER	-	EXPRESSION TAG	UNP Q99ZW2
B	840	ALA	HIS	ENGINEERED MUTATION	UNP Q99ZW2

- Molecule 3 is a DNA chain called TARGET DNA STRAND.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	28	567	276	96	168	27	0	0	0

- Molecule 4 is a DNA chain called NON-TARGET DNA STRAND.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	11	227	110	43	64	10	0	0	0

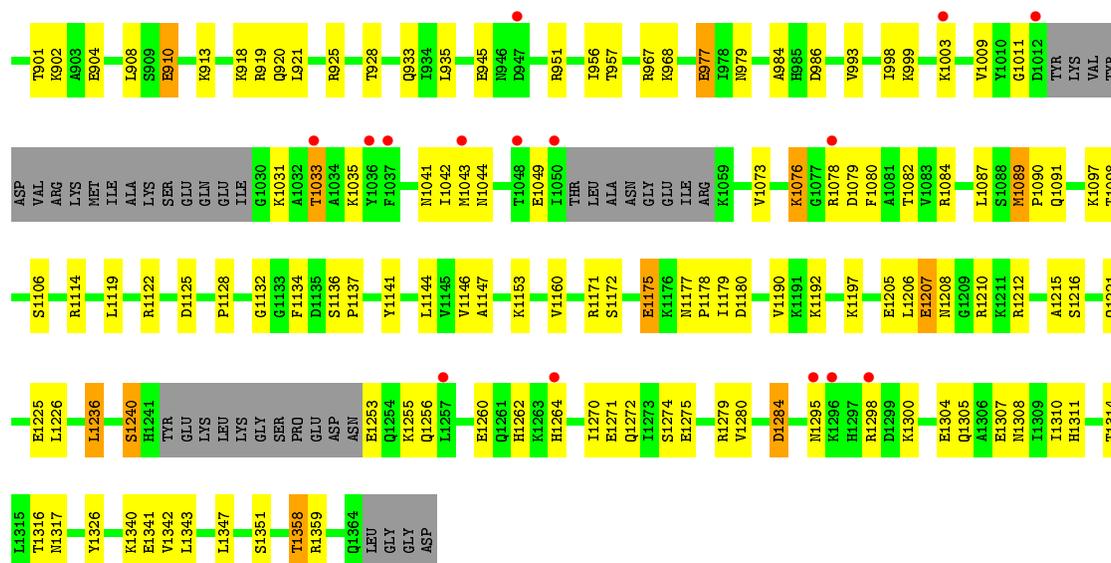
- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	4	Total Mg 4 4	0	0
5	A	4	Total Mg 4 4	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	146	Total O 146 146	0	0
6	B	246	Total O 246 246	0	0
6	C	11	Total O 11 11	0	0
6	D	2	Total O 2 2	0	0





- Molecule 3: TARGET DNA STRAND

Chain C: 82% 14%



- Molecule 4: NON-TARGET DNA STRAND

Chain D: 8% 75% 17% 8%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	177.72Å 68.14Å 188.23Å 90.00° 111.17° 90.00°	Depositor
Resolution (Å)	48.15 – 2.59 48.15 – 2.59	Depositor EDS
% Data completeness (in resolution range)	99.0 (48.15-2.59) 99.3 (48.15-2.59)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.03 (at 2.58Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.2_1309)	Depositor
R, $R_{free}$	0.217 , 0.252 0.232 , 0.259	Depositor DCC
$R_{free}$ test set	3263 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	46.0	Xtrriage
Anisotropy	0.620	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 47.1	EDS
Estimated twinning fraction	No twinning to report.	Xtrriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Outliers	0 of 65243 reflections	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	13629	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.18% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: MG

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	0.28	0/1942	0.81	0/3023
2	B	0.24	0/10877	0.43	0/14612
3	C	0.48	0/634	1.27	3/976 (0.3%)
4	D	0.48	0/255	1.10	0/393
All	All	0.27	0/13708	0.60	3/19004 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	2	DA	O4'-C1'-N9	6.68	112.67	108.00
3	C	11	DT	O4'-C4'-C3'	-5.99	102.11	104.50
3	C	22	DG	O4'-C1'-N9	5.22	111.66	108.00

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	1732	0	869	40	0
2	B	10690	0	10858	202	0
3	C	567	0	322	5	0
4	D	227	0	127	3	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	4	0	0	0	0
5	B	4	0	0	0	0
6	A	146	0	0	10	0
6	B	246	0	0	27	1
6	C	11	0	0	0	0
6	D	2	0	0	0	0
All	All	13629	0	12176	234	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 234 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:9:A:OP1	6:A:2016:HOH:O	1.83	0.97
2:B:587:PHE:O	6:B:2124:HOH:O	1.86	0.92
2:B:1084:ARG:O	6:B:2182:HOH:O	1.92	0.85
2:B:1225:GLU:OE2	6:B:2230:HOH:O	1.94	0.85
6:A:2121:HOH:O	2:B:69:ARG:NH2	2.11	0.82

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)
6:B:2016:HOH:O	6:B:2092:HOH:O[3_445]	2.18	0.02

## 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
2	B	1292/1372 (94%)	1249 (97%)	43 (3%)	0	<b>100</b> <b>100</b>

There are no Ramachandran outliers to report.

### 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
2	B	1174/1227 (96%)	1047 (89%)	127 (11%)	<b>8</b> <b>15</b>

5 of 127 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	623	LEU
2	B	735	LYS
2	B	1240	SER
2	B	630	GLU
2	B	666	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	402	GLN
2	B	1256	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	80/83 (96%)	21 (26%)	4 (5%)

5 of 21 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	9	A
1	A	20	A
1	A	28	A
1	A	29	G
1	A	31	U

All (4) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	8	A
1	A	27	G
1	A	42	A
1	A	68	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](#)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

#### 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	81/83 (97%)	-0.09	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	16, 38, 96, 112	0
2	B	1306/1372 (95%)	0.23	45 (3%) <span style="border: 1px solid gray; padding: 2px;">49</span> <span style="border: 1px solid gray; padding: 2px;">41</span>	12, 43, 76, 100	0
3	C	28/28 (100%)	-0.25	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	24, 38, 66, 85	0
4	D	11/12 (91%)	0.41	1 (9%) <span style="border: 1px solid red; padding: 2px;">11</span> <span style="border: 1px solid red; padding: 2px;">7</span>	37, 47, 86, 97	0
All	All	1426/1495 (95%)	0.20	46 (3%) <span style="border: 1px solid gray; padding: 2px;">51</span> <span style="border: 1px solid gray; padding: 2px;">44</span>	12, 43, 77, 112	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	178	ASN	3.8
2	B	1296	LYS	3.5
2	B	200	PRO	3.4
2	B	803	ASN	3.1
2	B	804	THR	3.1

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	MG	A	1082	1/1	0.86	0.17	1.00	27,27,27,27	0
5	MG	B	2368	1/1	0.58	0.25	0.60	36,36,36,36	0
5	MG	B	2367	1/1	0.71	0.17	-0.09	15,15,15,15	0
5	MG	B	2366	1/1	0.87	0.11	-0.99	25,25,25,25	0
5	MG	A	1083	1/1	0.98	0.11	-2.87	10,10,10,10	0
5	MG	B	2365	1/1	0.58	0.28	-	53,53,53,53	0
5	MG	A	1084	1/1	0.91	0.12	-	27,27,27,27	0
5	MG	A	1085	1/1	0.79	0.51	-	53,53,53,53	0

## 6.5 Other polymers

There are no such residues in this entry.