



# wwPDB X-ray Structure Validation Summary Report i

Sep 26, 2014 – 12:25 PM EDT

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 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/ValidationPDFNotes.html>

---

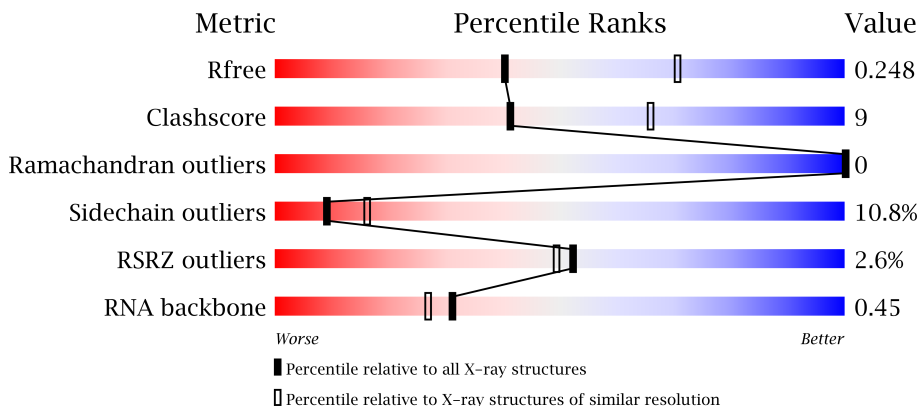
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : trunk23956  
Percentile statistics : 23426  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk23956

# 1 Overall quality at a glance

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))
$R_{free}$	77520	1997 (2.60-2.60)
Clashscore	88313	2341 (2.60-2.60)
Ramachandran outliers	86584	2297 (2.60-2.60)
C $\alpha$ geometry	86677	2302 (2.60-2.60)
Sidechain outliers	86556	2297 (2.60-2.60)
RSRZ outliers	77580	1998 (2.60-2.60)
RNA backbone	2044	1091 (3.10-2.10)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	83	
2	B	1372	
3	C	28	
4	D	12	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
5	MG	A	1082	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Geometry	Electron density
5	MG	A	1085	-	X
5	MG	B	2365	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 13629 atoms, of which 0 are hydrogen and 0 are deuterium.

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
1	A	81	Total	C	N	O	P	0	0	0
			1732	778	318	555	81			

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

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

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
3	C	28	Total	C	N	O	P	0	0	0
			567	276	96	168	27			

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

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

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

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

- Molecule 6 is water.

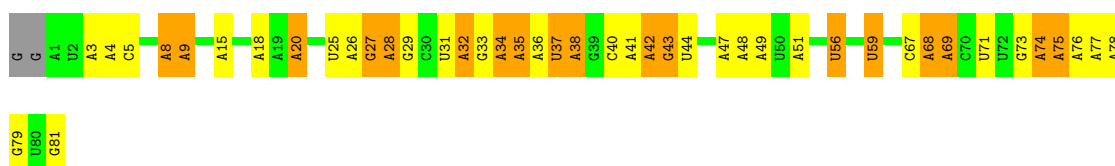
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	146	Total 146	O 146	0	0
6	B	246	Total 246	O 246	0	0
6	C	11	Total 11	O 11	0	0
6	D	2	Total 2	O 2	0	0

### 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 errors 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.

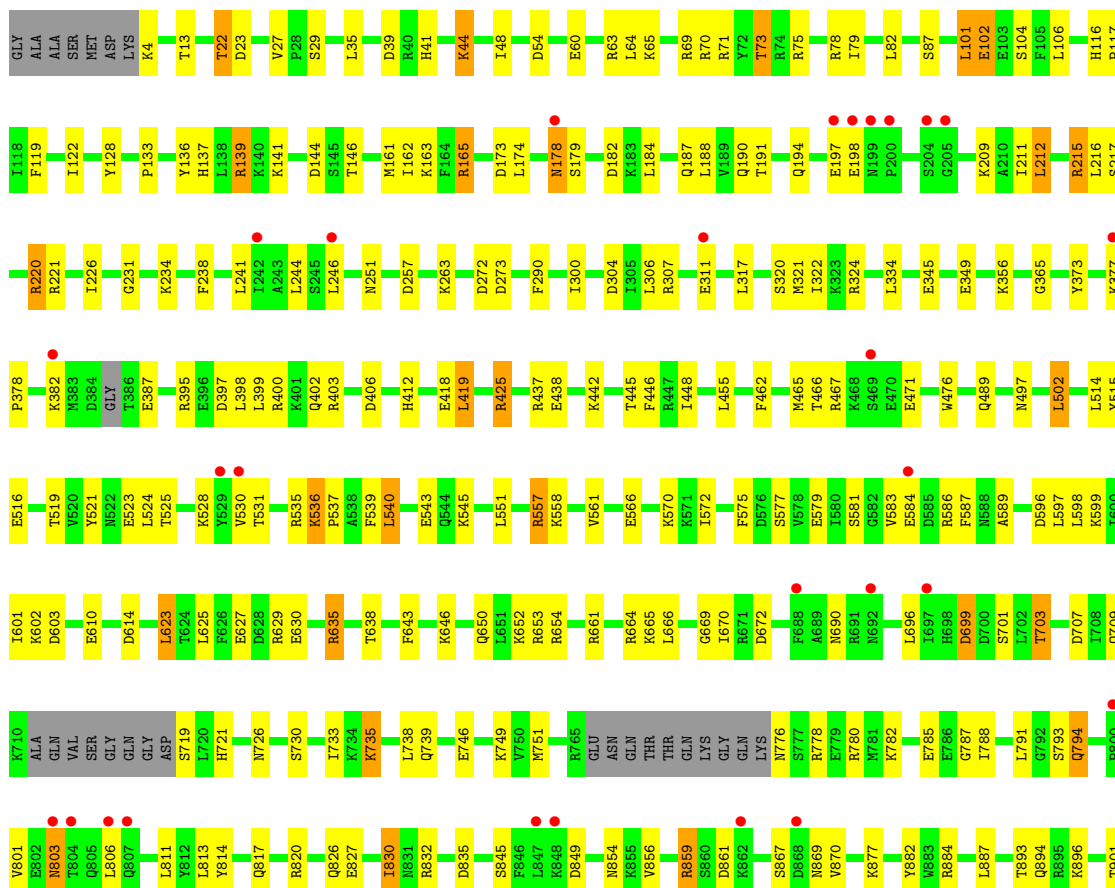
#### • Molecule 1: SGRNA

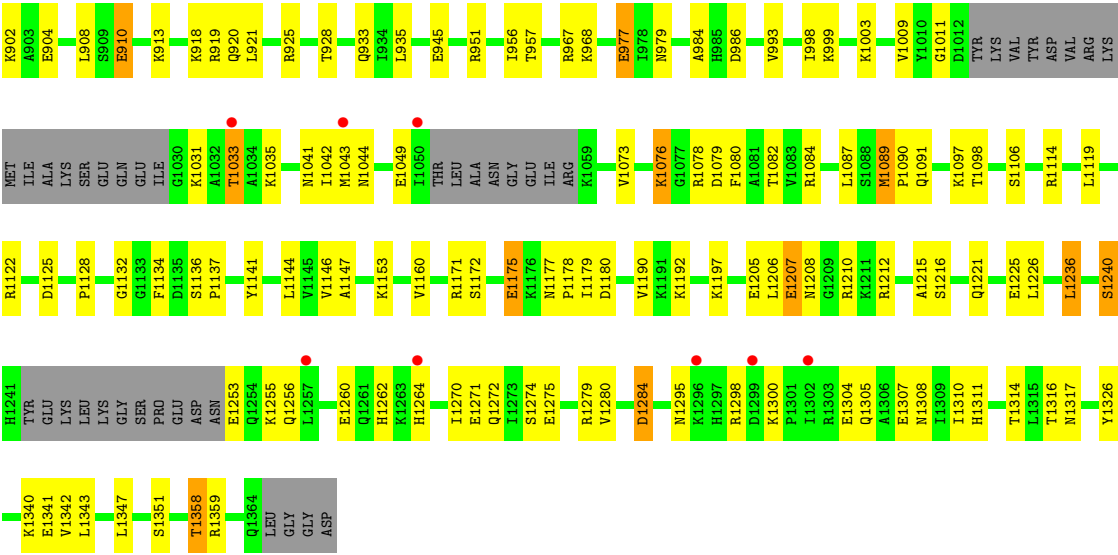
Chain A:



#### • Molecule 2: CRISPR-ASSOCIATED ENDONUCLEASE CAS9/CSN1

Chain B:





● Molecule 3: TARGET DNA STRAND



● Molecule 4: NON-TARGET DNA STRAND



## 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) 99.4 (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)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.2_1309)	Depositor
R, $R_{free}$	0.217 , 0.252 0.215 , 0.248	Depositor DCC
$R_{free}$ test set	3260 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	46.0	Xtriage
Anisotropy	0.620	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 18.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 65243 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	13629	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage'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

### 5.1 Standard geometry

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 Close contacts

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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

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

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	567	0	322	5	0
4	D	227	0	127	3	0
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

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 9.

The worst 5 of 234 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(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	Distance(Å)	Clash(Å)
6:B:2016:HOH:O	6:B:2092:HOH:O[3_445]	2.18	0.02

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	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
2	B	1174/1227 (96%)	1047 (89%)	127 (11%)	9 16

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 ⓘ

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 ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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%)	0.05	0 100 100	16, 38, 96, 112	0
2	B	1306/1372 (95%)	0.18	36 (2%) 49 47	12, 43, 76, 100	0
3	C	28/28 (100%)	-0.31	0 100 100	24, 38, 66, 85	0
4	D	11/12 (91%)	0.38	1 (9%) 9 6	37, 47, 86, 97	0
All	All	1426/1495 (95%)	0.16	37 (2%) 52 49	12, 43, 77, 112	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	803	ASN	3.1
2	B	804	THR	3.0
2	B	868	ASP	3.0
2	B	204	SER	3.0
2	B	242	ILE	2.9

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	MG	A	1085	1/1	0.50	19.43	53,53,53,53	0
5	MG	B	2365	1/1	0.26	13.20	53,53,53,53	0
5	MG	A	1082	1/1	0.18	2.29	27,27,27,27	0
5	MG	B	2368	1/1	0.24	0.53	36,36,36,36	0
5	MG	B	2367	1/1	0.17	0.21	15,15,15,15	0
5	MG	B	2366	1/1	0.11	-1.12	25,25,25,25	0
5	MG	A	1083	1/1	0.11	-1.83	10,10,10,10	0
5	MG	A	1084	1/1	0.12	-3.25	27,27,27,27	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.