W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 26, 2014 – 12:54 PM EDT

PDB ID : 4UN5
Title : Crystal structure of Cas9 bound to PAM-containing DNA target containing mismatches at positions 1-3
Authors : Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M.
Deposited on : 2014-05-25
Resolution : 2.40 Å(reported)

1 Overall quality at a glance (i)

The reported resolution of this entry is 2.40 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	Percentile Ranks	Value
Rfree		0.244
Clashscore		6
Ramachandran outliers		0
Sidechain outliers		7.1%
RSRZ outliers		3.6%
RNA backbone		0.56

Worse ... *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
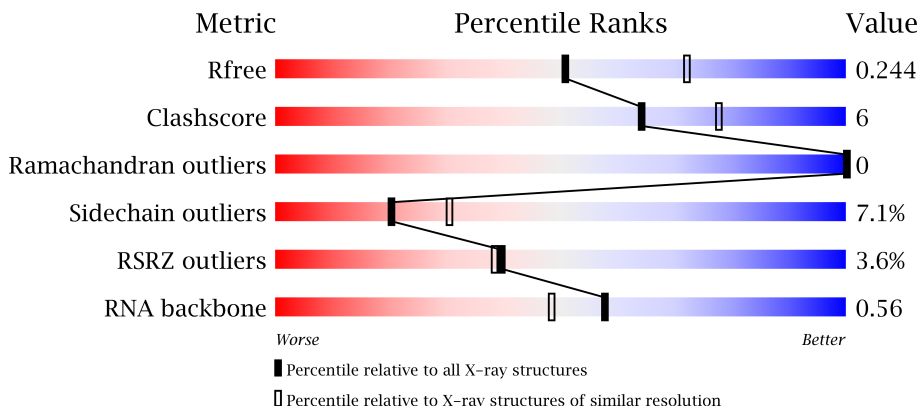
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	77520	2546 (2.40-2.40)
Clashscore	88313	3035 (2.40-2.40)
Ramachandran outliers	86584	2982 (2.40-2.40)
Cα geometry	86677	2981 (2.40-2.40)
Sidechain outliers	86556	2983 (2.40-2.40)
RSRZ outliers	77580	2550 (2.40-2.40)
RNA backbone	2044	1127 (3.00-1.80)

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 >=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	11	
4	D	11	
5	E	17	

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

Mol	Type	Chain	Res	Geometry	Electron density
6	MG	A	1082	-	X
6	MG	A	1084	-	X

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 13943 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	82	Total	C	N	O	P	0	0	1
			1733	778	318	556	81			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1310	Total	C	N	O	S	0	0	0
			10716	6832	1860	2002	22			

There are 6 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	10	ALA	ASP	ENGINEERED MUTATION	UNP Q99ZW2
B	840	ALA	HIS	ENGINEERED MUTATION	UNP Q99ZW2

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	9	Total	C	N	O	P	0	0	0
			162	77	31	46	8			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	10	Total	C	N	O	P	0	0	0
			206	100	35	62	9			

- Molecule 5 is a DNA chain called TARGET DNA STRAND DISTAL FRAGMENT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	17	Total 346	C 169	N 59	O 102	P 16	0	0	0

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	233	Total 233	O 233	0	0
7	B	491	Total 491	O 491	0	0
7	C	6	Total 6	O 6	0	0
7	D	15	Total 15	O 15	0	0
7	E	27	Total 27	O 27	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.
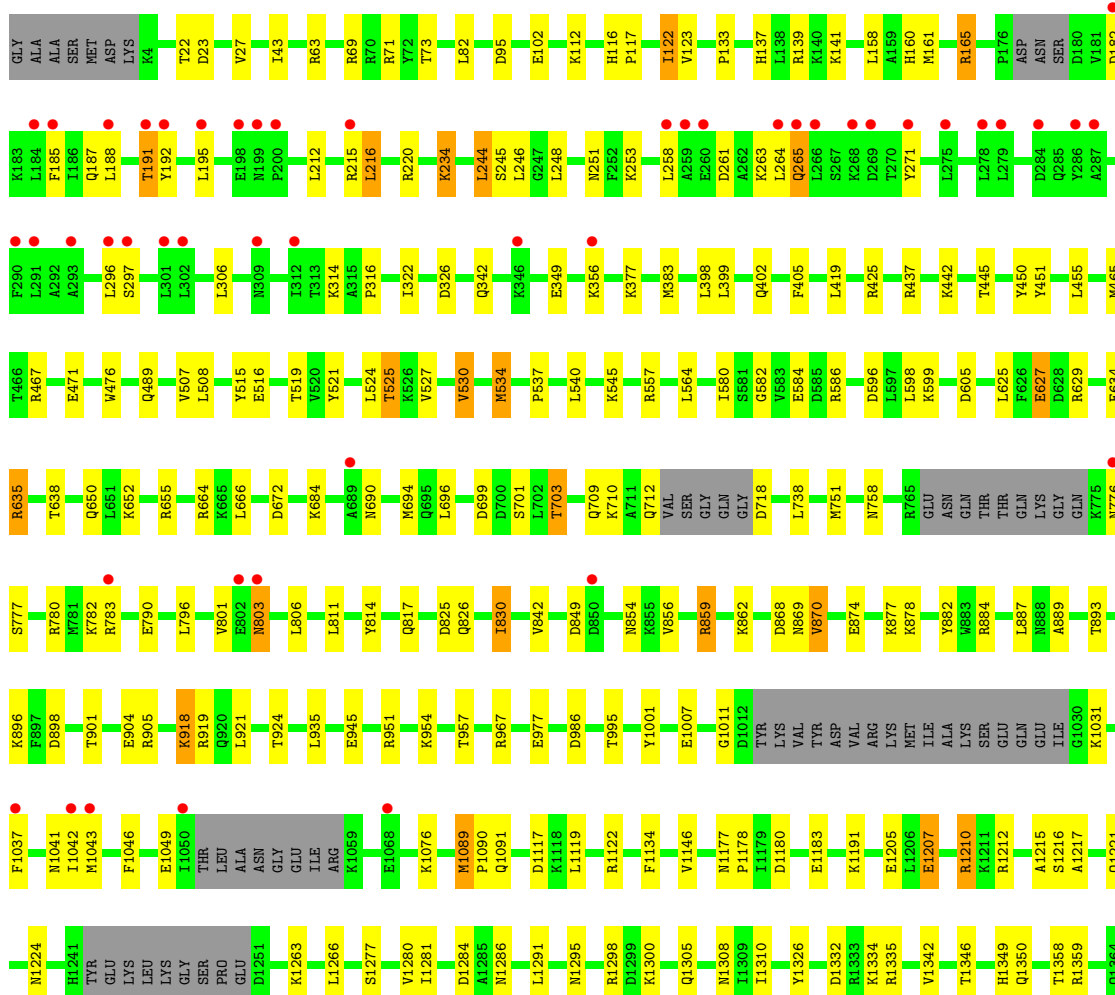
- Molecule 1: SGRNA

Chain A:

- Molecule 2: CRISPR-ASSOCIATED ENDONUCLEASE CAS9/CSN1

Chain B:

```
LEU
GLY
GLY
ASP
```

• Molecule 3: TARGET DNA STRAND PROXIMAL FRAGMENT

Chain C:

```
C1
A2
A3
T4
C9
DA
DA
```

• Molecule 4: NON-TARGET DNA STRAND

Chain D:

```
DT
T3
G4
G7
G12
```

• Molecule 5: TARGET DNA STRAND DISTAL FRAGMENT

Chain E:

```
T12
A17
A18
T28
```

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	180.18Å 68.07Å 190.26Å 90.00° 111.37° 90.00°	Depositor
Resolution (Å)	48.35 – 2.40 48.35 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.7 (48.35-2.40) 99.9 (48.35-2.40)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.17 (at 2.39Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.2_1309)	Depositor
R, R_{free}	0.215 , 0.246 0.213 , 0.244	Depositor DCC
R_{free} test set	4219 reflections (5.00%)	DCC
Wilson B-factor (Å^2)	37.2	Xtriage
Anisotropy	0.344	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 31.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning [2]	$<|L|> = 0.48$, $<L^2> = 0.31$	Xtriage
Outliers	0 of 84391 reflections	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	13943	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.16% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $<|L|>$, $<L^2>$ 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	0.19	0/1943	0.77	0/3026
2	B	0.24	0/10903	0.40	0/14646
3	C	0.48	0/181	1.17	2/277 (0.7%)
4	D	0.46	0/230	1.17	0/355
5	E	0.44	0/387	1.14	0/596
All	All	0.25	0/13644	0.56	2/18900 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	4	DT	C1'-O4'-C4'	-5.56	104.54	110.10
3	C	4	DT	O4'-C1'-C2'	-5.34	101.63	105.90

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	1733	0	869	19	0
2	B	10716	0	10887	123	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	162	0	90	4	0
4	D	206	0	117	3	0
5	E	346	0	197	1	0
6	A	4	0	0	0	0
6	B	4	0	0	0	0
7	A	233	0	0	10	0
7	B	491	0	0	33	0
7	C	6	0	0	1	0
7	D	15	0	0	0	0
7	E	27	0	0	0	0
All	All	13943	0	12160	144	0

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 6.

The worst 5 of 144 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:B:859:ARG:NH1	7:B:3316:HOH:O	2.03	0.89
2:B:1332:ASP:OD2	7:B:3481:HOH:O	1.93	0.87
2:B:437:ARG:NH2	7:B:3169:HOH:O	2.10	0.84
2:B:1212:ARG:NH1	7:B:3443:HOH:O	2.13	0.81
7:A:2187:HOH:O	2:B:69:ARG:NH2	2.15	0.79

There are no symmetry-related clashes.

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	1296/1372 (94%)	1249 (96%)	47 (4%)	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	1175/1226 (96%)	1092 (93%)	83 (7%)	20 30

5 of 83 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	634	GLU
2	B	712	GLN
2	B	1210	ARG
2	B	635	ARG
2	B	696	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	80/83 (96%)	20 (25%)	3 (3%)

5 of 20 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	28	A
1	A	29	G
1	A	32	A
1	A	34	A
1	A	35	A

All (3) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
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, 95th 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(Å²)	Q<0.9
1	A	82/83 (98%)	-0.17	3 (3%)	39	37	16, 32, 109, 144	0
2	B	1310/1372 (95%)	-0.02	48 (3%)	39	37	14, 37, 75, 132	0
3	C	9/11 (81%)	0.08	1 (11%)	5	5	34, 45, 85, 132	0
4	D	10/11 (90%)	-0.10	0	100	100	28, 44, 89, 100	0
5	E	17/17 (100%)	-0.79	0	100	100	23, 28, 35, 38	0
All	All	1428/1494 (95%)	-0.04	52 (3%)	40	38	14, 36, 77, 144	0

The worst 5 of 52 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	269	ASP	10.4
2	B	293	ALA	6.0
2	B	264	LEU	3.7
2	B	271	TYR	3.7
2	B	291	LEU	3.5

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, 95th 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(Å^2)	Q<0.9
6	MG	A	1084	1/1	0.27	5.58	50,50,50,50	0
6	MG	A	1082	1/1	0.15	2.32	29,29,29,29	0
6	MG	B	2368	1/1	0.21	0.80	31,31,31,31	0
6	MG	B	2365	1/1	0.17	0.42	22,22,22,22	0
6	MG	B	2366	1/1	0.13	-0.25	13,13,13,13	0
6	MG	B	2367	1/1	0.14	-0.47	43,43,43,43	0
6	MG	A	1085	1/1	0.12	-0.85	21,21,21,21	0
6	MG	A	1083	1/1	0.10	-3.38	6,6,6,6	0

6.5 Other polymers ⓘ

There are no such residues in this entry.