



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 07:07 am BST

PDB ID : 5FQ5  
Title : Crystal structure of Cas9-sgRNA-DNA complex solved by native SAD phasing  
Authors : Olieric, V.; Weinert, T.; Finke, A.; Anders, C.; Jinek, M.; Wang, M.  
Deposited on : 2015-12-07  
Resolution : 2.14 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

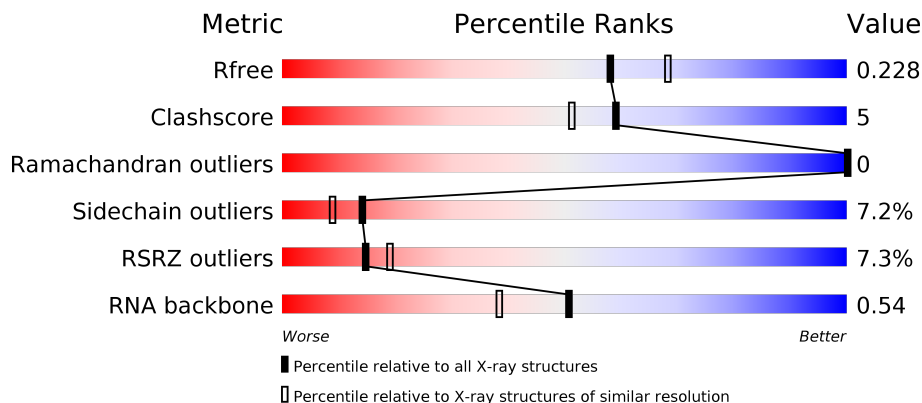
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

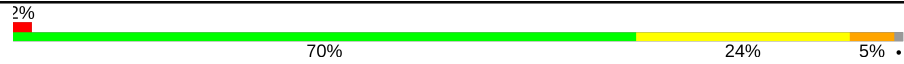



The reported resolution of this entry is 2.14 Å.

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	2523 (2.16-2.12)
Clashscore	141614	2653 (2.16-2.12)
Ramachandran outliers	138981	2618 (2.16-2.12)
Sidechain outliers	138945	2617 (2.16-2.12)
RSRZ outliers	127900	2485 (2.16-2.12)
RNA backbone	3102	1104 (2.60-1.68)

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	84	
2	B	1372	
3	C	11	
4	D	11	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	E	17	 88% 12%

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 26622 atoms, of which 12243 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	H	N	O	P			
1	A	83	2605	778	868	318	559	82	0	0	1

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	1318	21749	6877	10959	1874	2017	22	0	2	0

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
			Total	C	H	N	O	P			
3	C	9	279	86	102	34	49	8	0	0	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
4	D	10	323	100	117	35	62	9	0	0	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
5	E	17	543	169	197	59	102	16	0	0	0

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	2	Total	Mg	0	0
			2	2		

- Molecule 7 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	B	9	Total	K	0	0
			9	9		
7	A	3	Total	K	0	0
			3	3		
7	D	1	Total	K	0	0
			1	1		

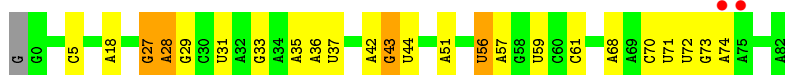
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	301	Total	O	0	0
			301	301		
8	B	744	Total	O	0	0
			744	744		
8	C	6	Total	O	0	0
			6	6		
8	D	10	Total	O	0	0
			10	10		
8	E	47	Total	O	0	0
			47	47		

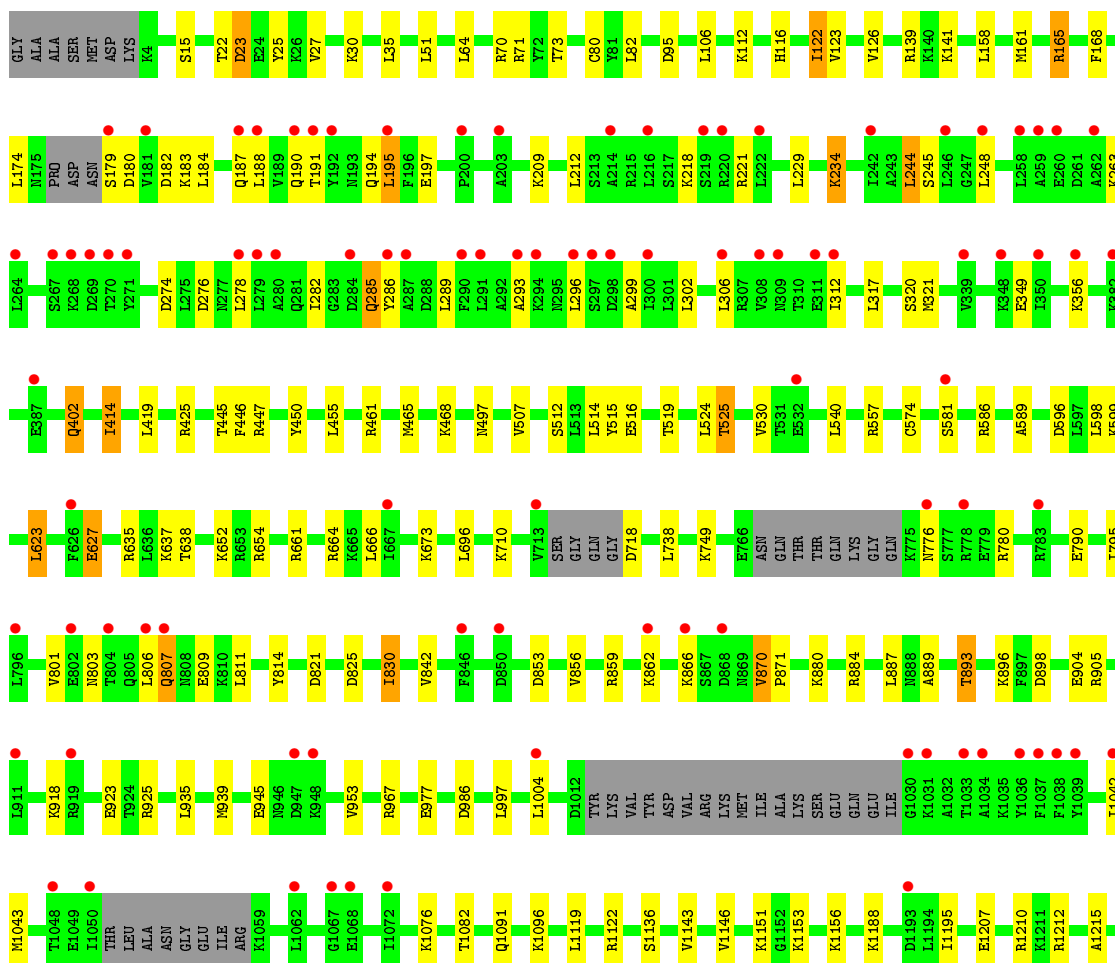
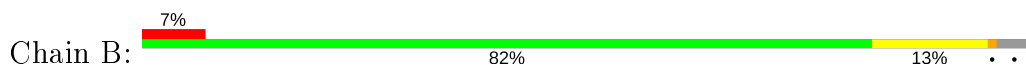
### 3 Residue-property plots [i](#)

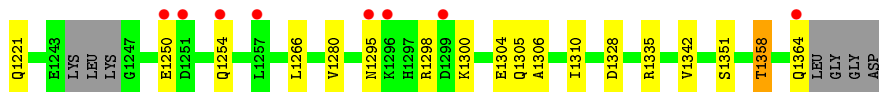
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are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

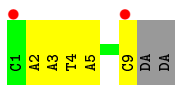


- Molecule 2: CRISPR-ASSOCIATED ENDONUCLEASE CAS9/CSN1





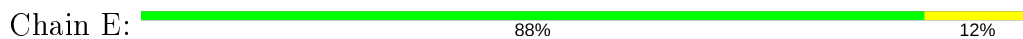
- Molecule 3: TARGET DNA STRAND PROXIMAL FRAGMENT



- Molecule 4: NON-TARGET DNA STRAND



- Molecule 5: TARGET DNA STRAND DISTAL FRAGMENT



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	177.74Å 67.57Å 188.19Å 90.00° 111.31° 90.00°	Depositor
Resolution (Å)	47.91 – 2.14 47.91 – 2.14	Depositor EDS
% Data completeness (in resolution range)	96.8 (47.91-2.14) 97.8 (47.91-2.14)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.00 (at 2.14Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.191 , 0.225 0.193 , 0.228	Depositor DCC
$R_{free}$ test set	1989 reflections (1.76%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.4	Xtrriage
Anisotropy	0.236	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 41.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	26622	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	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, 0.333 respectively for untwinned datasets, and 0.375, 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: K, 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	0.24	0/1947	0.75	0/3033
2	B	0.26	0/10984	0.48	0/14754
3	C	0.64	0/198	0.92	0/302
4	D	0.60	0/230	1.06	0/355
5	E	0.53	0/387	1.03	0/596
All	All	0.29	0/13746	0.58	0/19040

There are no bond length outliers.

There are no bond angle outliers.

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	1737	868	868	15	0
2	B	10790	10959	10957	101	0
3	C	177	102	102	4	0
4	D	206	117	117	5	0
5	E	346	197	197	1	0
6	A	2	0	0	0	0
7	A	3	0	0	0	0
7	B	9	0	0	1	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	D	1	0	0	0	0
8	A	301	0	0	6	0
8	B	744	0	0	46	2
8	C	6	0	0	0	0
8	D	10	0	0	0	0
8	E	47	0	0	0	0
All	All	14379	12243	12241	119	2

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 5.

The worst 5 of 119 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)
2:B:248:LEU:O	8:B:3181:HOH:O	1.75	1.01
2:B:229:LEU:O	8:B:3174:HOH:O	1.78	0.99
2:B:825:ASP:N	8:B:3473:HOH:O	2.05	0.89
2:B:321:MET:SD	8:B:3205:HOH:O	2.31	0.86
2:B:306:LEU:O	2:B:320:SER:OG	1.92	0.86

All (2) 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)
8:B:3346:HOH:O	8:B:3398:HOH:O[4_546]	2.14	0.06
8:B:3169:HOH:O	8:B:3419:HOH:O[4_546]	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	1306/1372 (95%)	1267 (97%)	39 (3%)	0	100 100

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	1183/1226 (96%)	1098 (93%)	85 (7%)	14 9

5 of 85 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	540	LEU
2	B	666	LEU
2	B	1266	LEU
2	B	586	ARG
2	B	627	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	982	HIS

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	80/84 (95%)	12 (15%)	2 (2%)

5 of 12 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	28	A
1	A	29	G

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	31	U
1	A	35	A
1	A	37	U

All (2) RNA pucker outliers are listed below:

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

Of 15 ligands modelled in this entry, 15 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	83/84 (98%)	-0.31	2 (2%) 59 65	26, 41, 136, 180	0
2	B	1318/1372 (96%)	0.52	100 (7%) 13 17	24, 49, 95, 137	0
3	C	9/11 (81%)	0.94	2 (22%) 0 0	48, 69, 114, 129	0
4	D	10/11 (90%)	0.44	1 (10%) 7 9	36, 58, 104, 125	0
5	E	17/17 (100%)	-0.15	0 100 100	35, 39, 47, 56	0
All	All	1437/1495 (96%)	0.47	105 (7%) 15 19	24, 49, 98, 180	0

The worst 5 of 105 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	278	LEU	6.4
2	B	293	ALA	6.1
2	B	259	ALA	4.8
2	B	290	PHE	4.8
2	B	286	TYR	4.6

### 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
7	K	B	2372	1/1	0.77	0.17	92,92,92,92	0
6	MG	A	1083	1/1	0.88	0.14	62,62,62,62	0
7	K	D	1013	1/1	0.89	0.43	121,121,121,121	0
7	K	B	2368	1/1	0.93	0.13	53,53,53,53	0
7	K	B	2366	1/1	0.94	0.16	38,38,38,38	1
7	K	B	2370	1/1	0.94	0.10	83,83,83,83	0
7	K	A	1085	1/1	0.94	0.11	49,49,49,49	1
7	K	B	2369	1/1	0.97	0.11	58,58,58,58	1
7	K	B	2371	1/1	0.97	0.08	62,62,62,62	0
7	K	B	2373	1/1	0.98	0.15	69,69,69,69	1
7	K	A	2084	1/1	0.98	0.14	40,40,40,40	1
7	K	B	2365	1/1	0.98	0.18	32,32,32,32	1
6	MG	A	1084	1/1	0.98	0.16	29,29,29,29	0
7	K	B	2367	1/1	0.99	0.16	34,34,34,34	1
7	K	A	2083	1/1	0.99	0.17	30,30,30,30	1

## 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.