



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 14, 2016 – 11:01 PM EDT

PDB ID : 5FW1
Title : Crystal structure of SpyCas9 variant VQR bound to sgRNA and TGAG PAM target DNA
Authors : Anders, C.; Bargsten, K.; Jinek, M.
Deposited on : 2016-02-11
Resolution : 2.50 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
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 : unknown
Xtriage (Phenix) : 1.9-1692
EDS : rb-20027107
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) : rb-20027107

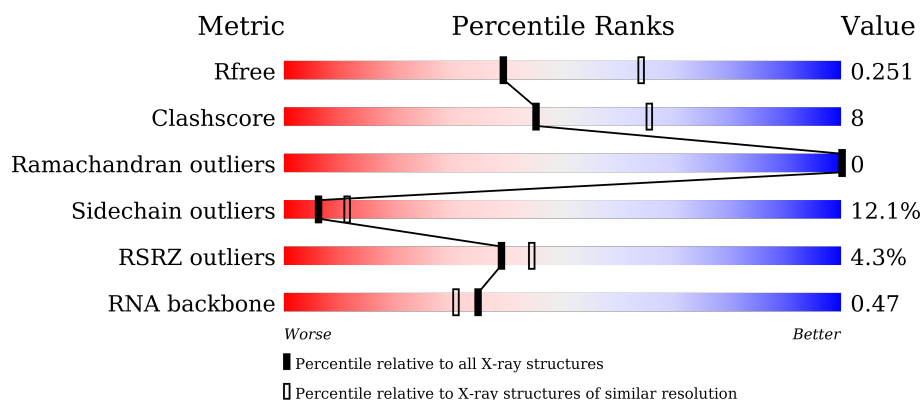
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)
RNA backbone	2183	1172 (3.00-2.00)

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. 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	<div> <div>6%</div> <div> <div>41%</div> <div>35%</div> <div>20%</div> <div>..</div> </div> </div>
2	B	1372	<div> <div>4%</div> <div> <div>72%</div> <div>21%</div> <div>5%</div> </div> </div>
3	C	28	<div> <div>68%</div> <div>32%</div> </div>
4	D	12	<div> <div>17%</div> <div> <div>50%</div> <div>42%</div> <div>8%</div> </div> </div>

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 25684 atoms, of which 12170 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
1	A	81	Total	C	H	N	O	P	0	0	0
			2600	778	868	318	555	81			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	1304	Total	C	H	N	O	S	0	0	0
			21525	6809	10853	1852	1989	22			

There are 9 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
B	1135	VAL	ASP	ENGINEERED MUTATION	UNP Q99ZW2
B	1335	GLN	ARG	ENGINEERED MUTATION	UNP Q99ZW2
B	1337	ARG	THR	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	H	N	O	P	0	0	0
			889	276	323	93	170	27			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	D	11	Total	C	H	N	O	P	0	0	0
			354	110	126	46	62	10			

- Molecule 5 is POTASSIUM ION (three-letter code: K) (formula: K).

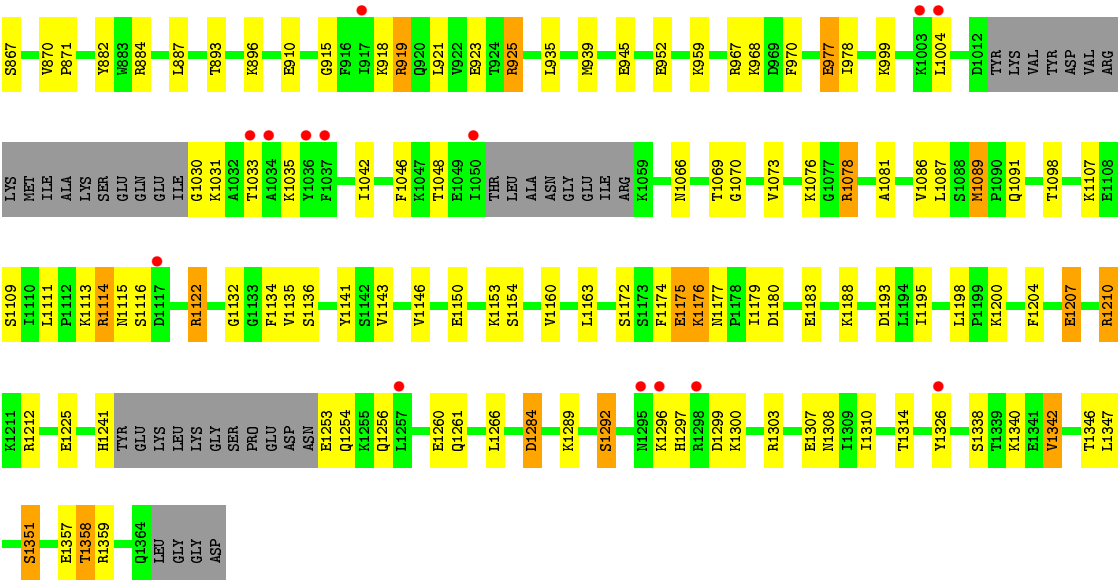
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	10	Total	K	0	0
			10	10		
5	A	2	Total	K	0	0
			2	2		
5	D	1	Total	K	0	0
			1	1		

- 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 water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	100	Total	O	0	0
			100	100		
7	B	187	Total	O	0	0
			187	187		
7	C	10	Total	O	0	0
			10	10		
7	D	4	Total	O	0	0
			4	4		



● 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, α , β , γ	176.64Å 66.98Å 187.23Å 90.00° 111.12° 90.00°	Depositor
Resolution (Å)	47.62 – 2.50 47.62 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (47.62-2.50) 99.8 (47.62-2.50)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.80 (at 2.51Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.218 , 0.254 0.216 , 0.251	Depositor DCC
R_{free} test set	3557 reflections (5.00%)	DCC
Wilson B-factor (Å ²)	48.2	Xtriage
Anisotropy	0.334	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 35.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	1 of 71152 reflections (0.001%)	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	25684	wwPDB-VP
Average B, all atoms (Å ²)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.10% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: 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	0.40	0/1942	0.98	2/3023 (0.1%)
2	B	0.28	0/10857	0.52	0/14581
3	C	0.64	0/632	1.09	0/973
4	D	0.66	0/257	1.08	1/396 (0.3%)
All	All	0.34	0/13688	0.67	3/18973 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	74	A	C6-N1-C2	5.84	122.10	118.60
1	A	32	A	N9-C4-C5	-5.71	103.52	105.80
4	D	7	DA	O4'-C1'-N9	5.51	111.86	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	868	869	44	0
2	B	10672	10853	10852	154	1
3	C	566	323	323	9	0
4	D	228	126	126	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	2	0	0	0	0
5	B	10	0	0	0	0
5	D	1	0	0	0	0
6	A	2	0	0	0	0
7	A	100	0	0	12	0
7	B	187	0	0	29	0
7	C	10	0	0	0	0
7	D	4	0	0	2	0
All	All	13514	12170	12170	198	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 8.

The worst 5 of 198 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:140:LYS:NZ	7:B:3036:HOH:O	1.90	1.01
2:B:60:GLU:OE1	7:B:3020:HOH:O	1.89	0.90
2:B:1212:ARG:NH1	7:B:3164:HOH:O	2.03	0.90
2:B:557:ARG:NH2	2:B:596:ASP:OD1	2.08	0.85
1:A:12:U:OP1	7:A:3014:HOH:O	1.93	0.85

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)
2:B:618:ASP:OD1	2:B:832:ARG:NH2[1_565]	2.07	0.13

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	1287/1372 (94%)	1264 (98%)	23 (2%)	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	1171/1226 (96%)	1029 (88%)	142 (12%)	6	11

5 of 142 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	623	LEU
2	B	751	MET
2	B	1266	LEU
2	B	630	GLU
2	B	665	LYS

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	497	ASN
2	B	1262	HIS

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	80/83 (96%)	23 (28%)	5 (6%)

5 of 23 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	9	A

Continued on next page...

Continued from previous page...

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

All (5) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	8	A
1	A	27	G
1	A	28	A
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 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 ⓘ

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, 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(Å ²)	Q<0.9
1	A	81/83 (97%)	0.26	5 (6%) 24 27	28, 43, 118, 162	0
2	B	1304/1372 (95%)	0.44	54 (4%) 41 46	24, 49, 81, 118	0
3	C	28/28 (100%)	-0.09	0 100 100	35, 43, 74, 83	0
4	D	11/12 (91%)	0.74	2 (18%) 2 2	37, 59, 97, 100	0
All	All	1424/1495 (95%)	0.42	61 (4%) 39 44	24, 49, 83, 162	0

The worst 5 of 61 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	269	ASP	14.7
2	B	179	SER	8.2
1	A	77	A	8.1
2	B	291	LEU	7.2
2	B	1037	PHE	5.4

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, 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	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
5	K	B	3365	1/1	0.93	0.19	0.20	45,45,45,45	0
6	MG	A	2084	1/1	0.91	0.17	-0.12	34,34,34,34	0
5	K	B	3372	1/1	0.90	0.12	-1.50	70,70,70,70	0
5	K	B	3368	1/1	0.94	0.10	-1.93	46,46,46,46	0
5	K	B	3366	1/1	0.94	0.09	-2.33	51,51,51,51	0
5	K	B	3367	1/1	0.97	0.17	-2.36	31,31,31,31	0
5	K	B	3370	1/1	0.67	0.10	-2.73	79,79,79,79	0
5	K	B	3369	1/1	0.88	0.08	-2.86	72,72,72,72	0
5	K	B	3371	1/1	0.93	0.08	-	77,77,77,77	0
5	K	A	2082	1/1	0.84	0.20	-	88,88,88,88	0
5	K	D	2013	1/1	0.86	0.38	-	92,92,92,92	0
5	K	A	2081	1/1	0.83	0.18	-	83,83,83,83	0
5	K	B	3373	1/1	0.73	0.41	-	92,92,92,92	0
6	MG	A	2083	1/1	0.92	0.34	-	45,45,45,45	0
5	K	B	3374	1/1	0.97	0.11	-	52,52,52,52	0

6.5 Other polymers [i](#)

There are no such residues in this entry.