WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 3, 2024 – 02:17 am GMT

PDB ID	:	5FW2
Title	:	Crystal structure of SpCas9 variant EQR bound to sgRNA and TGAG PAM target DNA
Authors	:	Anders, C.; Bargsten, K.; Jinek, M.
Deposited on	:	2016-02-11
Resolution	:	2.68 Å(reported)

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

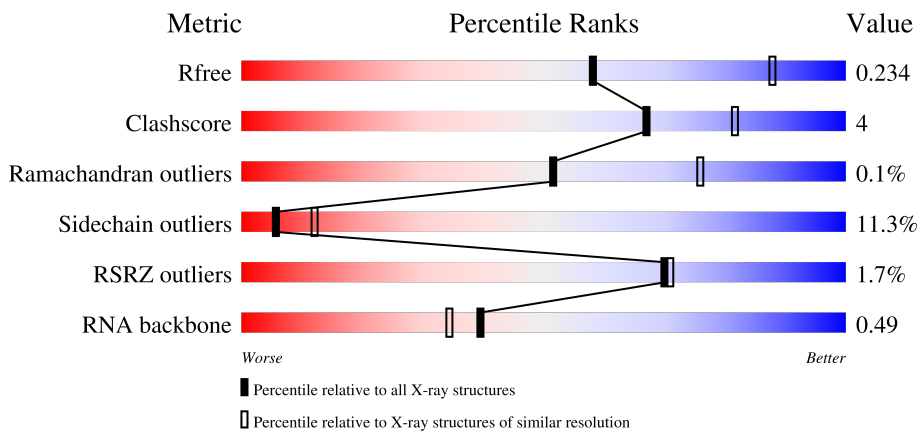|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.68 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		0.234
Clashscore		4
Ramachandran outliers		0.1%
Sidechain outliers		11.3%
RSRZ outliers		1.7%
RNA backbone		0.49

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$_{free}$	130704	3863 (2.70-2.66)
Clashscore	141614	4210 (2.70-2.66)
Ramachandran outliers	138981	4141 (2.70-2.66)
Sidechain outliers	138945	4141 (2.70-2.66)
RSRZ outliers	127900	3780 (2.70-2.66)
RNA backbone	3102	1007 (2.98-2.38)

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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	63% 25% 10% ·
2	B	1372	2% 79% 15% · ·
3	C	28	89% 11%
4	D	12	8% 50% 42% 8%

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 25767 atoms, of which 12261 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
			2601	778	869	318	555	81			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	1316	Total	C	H	N	O	S	0	0	0
			21707	6862	10943	1869	2011	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	GLU	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 354	C 110	H 126	N 46	O 62	P 10	0	0	0

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

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

- Molecule 6 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	2	Total 2	K 2	0	0
6	B	8	Total 8	K 8	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	87	Total 87	O 87	0	0
7	B	107	Total 107	O 107	0	0
7	C	10	Total 10	O 10	0	0

3   Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.
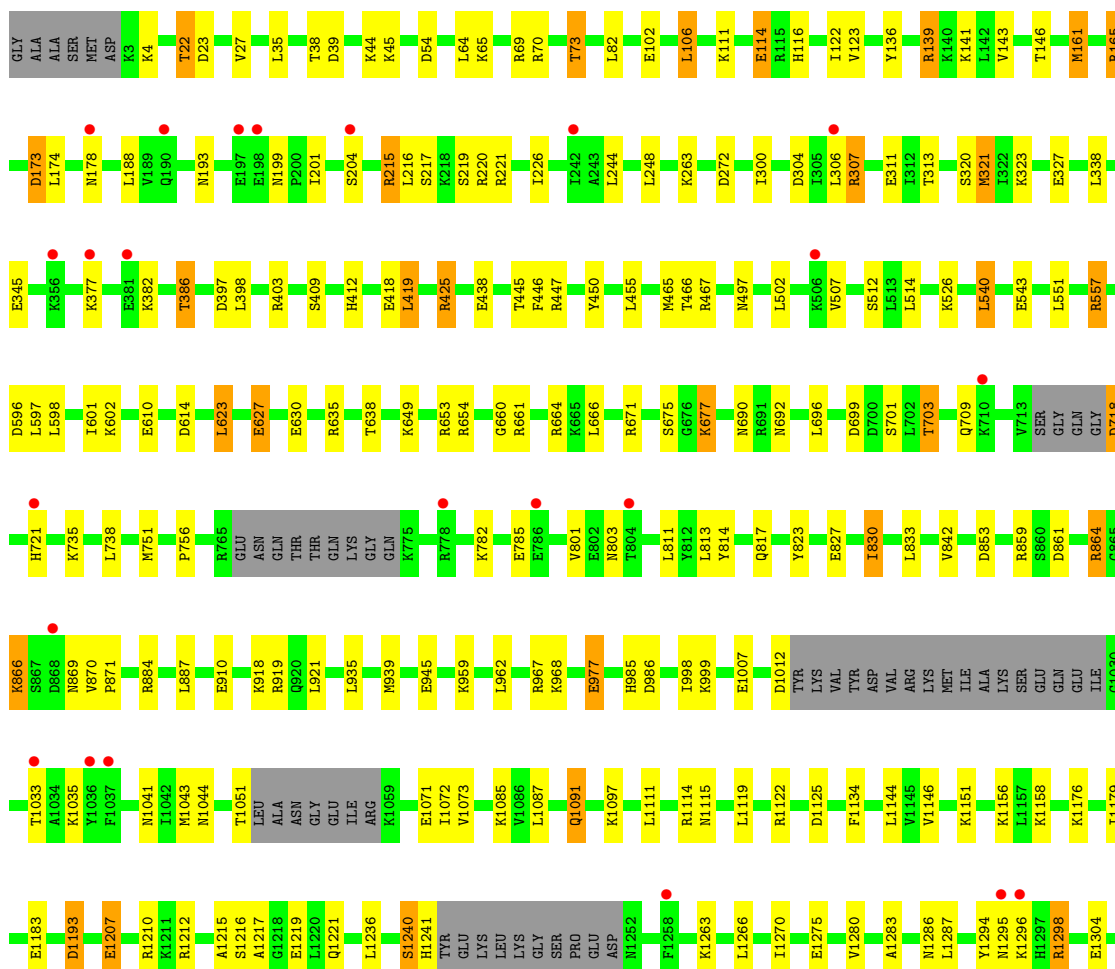
- Molecule 1: SGRNA

Chain A:
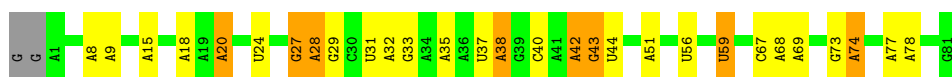
- Molecule 2: CRISPR-ASSOCIATED ENDONUCLEASE CAS9/CSN1

Chain B:

N1308 I1309 I1310 T1314 K1325 R1333 K1334 S1338 L1347 E1357 T1358 R1359 L1365 GLY GLY ASP

• Molecule 3: TARGET DNA STRAND

Chain C:

89%   11%

C1 A2 A3 A19 A27 T28
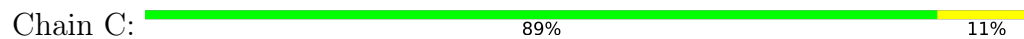
• Molecule 4: NON-TARGET DNA STRAND

Chain D:

8%
50%   42%   8%

DA A2 G6 A7 G8 A9 T10 T11 G12

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	177.74Å 68.24Å 188.22Å 90.00° 110.95° 90.00°	Depositor
Resolution (Å)	48.25 – 2.68 48.25 – 2.68	Depositor EDS
% Data completeness (in resolution range)	99.4 (48.25-2.68) 99.5 (48.25-2.68)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.53 (at 2.69Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.205 , 0.231 0.207 , 0.234	Depositor DCC
R_{free} test set	2983 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	44.7	Xtriage
Anisotropy	0.615	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 34.3	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	25767	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% 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, 0.333 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: MG, K

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	0.31	0/1942	0.79	0/3023
2	B	0.27	0/10952	0.46	0/14713
3	C	0.58	0/632	1.06	0/973
4	D	0.70	0/257	0.93	0/396
All	All	0.31	0/13783	0.58	0/19105

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	1732	869	869	16	0
2	B	10764	10943	10940	83	0
3	C	566	323	323	5	0
4	D	228	126	126	6	0
5	A	2	0	0	0	0
6	A	2	0	0	0	0
6	B	8	0	0	0	0
7	A	87	0	0	3	0
7	B	107	0	0	8	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	C	10	0	0	2	0
All	All	13506	12261	12258	98	0

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 4.

The worst 5 of 98 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:557:ARG:NH2	2:B:596:ASP:OD1	2.05	0.90
2:B:1286:ASN:N	7:B:2097:HOH:O	2.05	0.90
2:B:1334:LYS:NZ	3:C:3:DA:OP2	2.04	0.89
2:B:106:LEU:O	2:B:111:LYS:NZ	2.08	0.86
2:B:1283:ALA:O	7:B:2097:HOH:O	1.93	0.86

There are no symmetry-related clashes.

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	1304/1372 (95%)	1275 (98%)	28 (2%)	1 (0%)	51 76

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	398	LEU

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	1181/1226 (96%)	1047 (89%)	134 (11%)	6 12

5 of 134 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	1156	LYS
2	B	1207	GLU
2	B	1325	LYS
2	B	455	LEU
2	B	445	THR

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

Mol	Chain	Res	Type
2	B	674	GLN
2	B	1101	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	80/83 (96%)	18 (22%)	5 (6%)

5 of 18 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	9	A
1	A	20	A
1	A	24	U
1	A	28	A
1	A	29	G

All (5) RNA pucker outliers are listed below:

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

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	28	A
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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 12 ligands modelled in this entry, 12 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 ⓘ

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(Å^2)	Q<0.9
1	A	81/83 (97%)	-0.28	0	100	100	20, 38, 93, 100	0
2	B	1316/1372 (95%)	0.18	24 (1%)	68	69	16, 48, 85, 115	0
3	C	28/28 (100%)	-0.31	0	100	100	26, 43, 73, 85	0
4	D	11/12 (91%)	0.44	1 (9%)	9	7	34, 55, 95, 105	0
All	All	1436/1495 (96%)	0.15	25 (1%)	70	71	16, 48, 85, 115	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	178	ASN	3.6
2	B	1037	PHE	3.5
2	B	1365	LEU	3.4
2	B	868	ASP	3.2
2	B	306	LEU	3.2

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 monosaccharides 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. 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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
6	K	B	2370	1/1	0.30	0.36	50,50,50,50	0
6	K	A	1183	1/1	0.88	0.15	50,50,50,50	0
6	K	B	2372	1/1	0.88	0.12	50,50,50,50	0
6	K	B	2371	1/1	0.89	0.25	50,50,50,50	0
6	K	A	1182	1/1	0.91	0.24	79,79,79,79	0
5	MG	A	1084	1/1	0.93	0.22	45,45,45,45	0
5	MG	A	1083	1/1	0.93	0.26	63,63,63,63	0
6	K	B	2367	1/1	0.96	0.11	57,57,57,57	0
6	K	B	2373	1/1	0.96	0.15	50,50,50,50	0
6	K	B	2366	1/1	0.97	0.11	39,39,39,39	0
6	K	B	2369	1/1	0.98	0.08	43,43,43,43	0
6	K	B	2368	1/1	0.99	0.12	32,32,32,32	0

6.5 Other polymers ⓘ

There are no such residues in this entry.