



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 7, 2021 – 01:55 pm BST

PDB ID : 7OXA
Title : Target-bound SpCas9 complex with AAVS1 chimeric RNA-DNA guide
Authors : Donohoue, P.; Pacesa, M.; Lau, E.; Vidal, B.; Irby, M.J.; Nyer, D.B.; Rotstein, T.; Banh, L.; Toh, M.T.; Gibson, J.; Kohrs, B.; Baek, K.; Owen, A.L.G.; Slorach, E.M.; van Overbeek, M.; Fuller, C.K.; May, A.P.; Jinek, M.; Cameron, P.
Deposited on : 2021-06-22
Resolution : 2.15 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

<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 : 2.23.1
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) : 2.23.1

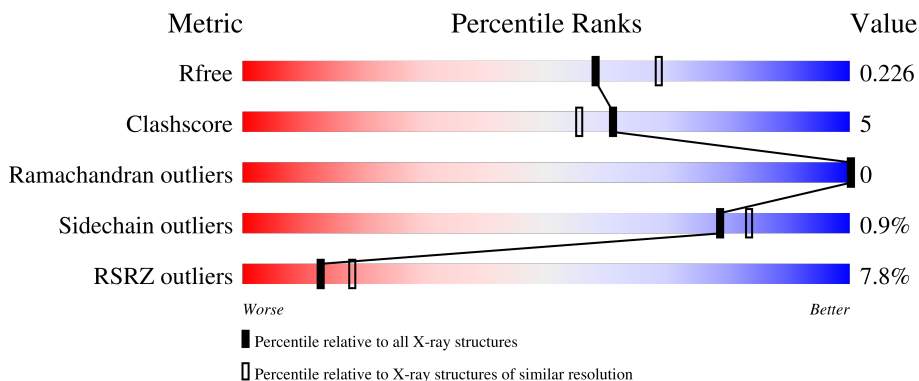
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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 $\leq 5\%$. 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	81	<div style="display: flex; align-items: center;"> <div style="width: 2%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 72%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 28%; height: 10px; background-color: yellow; margin-right: 5px;"></div> </div>
2	B	1372	<div style="display: flex; align-items: center;"> <div style="width: 8%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 86%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 10%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 2%; height: 10px; background-color: grey; margin-right: 5px;"></div> </div>
3	C	28	<div style="display: flex; align-items: center;"> <div style="width: 57%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 43%; height: 10px; background-color: yellow; margin-right: 5px;"></div> </div>
4	D	12	<div style="display: flex; align-items: center;"> <div style="width: 8%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 67%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 25%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 8%; height: 10px; background-color: grey; margin-right: 5px;"></div> </div>

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 13928 atoms, of which 0 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 DNA/RNA hybrid called chimeric RNA-DNA guide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	81	1732	781	326	545	80	0	0	0

- Molecule 2 is a protein called CRISPR-associated endonuclease Cas9/Csn1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1320	10808	6892	1875	2019	22	0	0	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 AAVS1 target DNA strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	28	559	268	98	166	27	0	0	0

- Molecule 4 is a DNA chain called AAVS1 non-target DNA strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	11	209	100	35	64	10	0	0	1

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

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

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	5	Total K 5 5	0	0
6	B	8	Total K 8 8	0	0
6	C	1	Total K 1 1	0	0

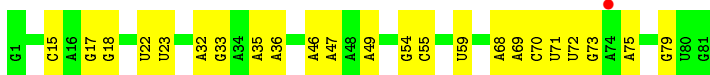
- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	164	Total O 164 164	0	0
7	B	394	Total O 394 394	0	0
7	C	34	Total O 34 34	0	0
7	D	12	Total O 12 12	0	0

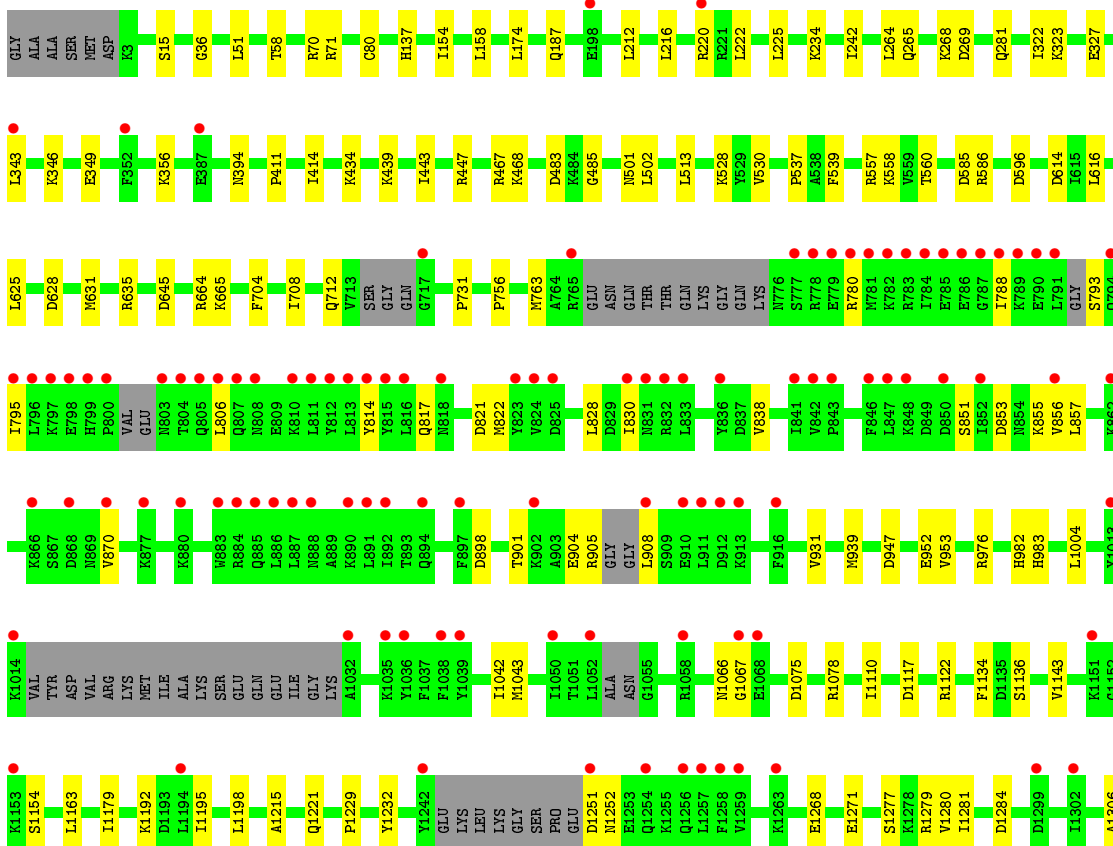
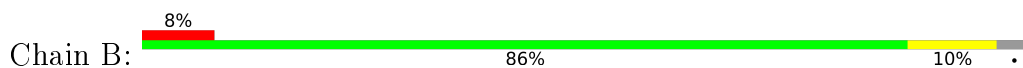
3 Residue-property plots [i](#)

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: chimeric RNA-DNA guide



- Molecule 2: CRISPR-associated endonuclease Cas9/Csn1

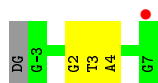




- Molecule 3: AAVS1 target DNA strand



- Molecule 4: AAVS1 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	177.93Å 67.57Å 189.23Å 90.00° 112.52° 90.00°	Depositor
Resolution (Å)	47.66 – 2.15 47.66 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.66-2.15) 99.9 (47.66-2.15)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.15 (at 2.16Å)	Xtrriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.198 , 0.226 0.198 , 0.226	Depositor DCC
R_{free} test set	5666 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	46.8	Xtrriage
Anisotropy	0.300	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 43.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	13928	wwPDB-VP
Average B, all atoms (Å ²)	70.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.12% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² 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	0.40	0/1944	0.84	0/3027
2	B	0.26	0/10995	0.43	0/14767
3	C	0.69	0/624	0.97	0/958
4	D	0.63	0/233	1.05	0/360
All	All	0.33	0/13796	0.57	0/19112

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	0	876	22	0
2	B	10808	0	10981	88	0
3	C	559	0	316	9	0
4	D	209	0	116	3	0
5	A	2	0	0	0	0
6	A	5	0	0	0	0
6	B	8	0	0	0	0
6	C	1	0	0	0	0
7	A	164	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	B	394	0	0	7	0
7	C	34	0	0	0	0
7	D	12	0	0	0	0
All	All	13928	0	12289	115	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 5.

The worst 5 of 115 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:343:LEU:HD13	2:B:346:LYS:HB2	1.33	1.06
2:B:343:LEU:O	2:B:343:LEU:HD12	1.56	1.02
2:B:343:LEU:CD1	2:B:346:LYS:HB2	1.93	0.98
2:B:80:CYS:SG	7:B:1884:HOH:O	2.49	0.70
2:B:1335:ARG:NH2	4:D:2:DG:O6	2.24	0.70

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	1302/1372 (95%)	1262 (97%)	40 (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	1186/1226 (97%)	1175 (99%)	11 (1%)	78 83

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	947	ASP
2	B	952	GLU
2	B	1284	ASP
2	B	1154	SER
2	B	635	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	240	ASN
2	B	255	ASN
2	B	394	ASN
2	B	511	HIS

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 16 ligands modelled in this entry, 16 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 [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	81/81 (100%)	-0.48	1 (1%) 79 83	37, 56, 164, 234	0
2	B	1320/1372 (96%)	0.41	111 (8%) 11 15	34, 60, 133, 200	0
3	C	28/28 (100%)	-0.20	0 100 100	44, 60, 129, 143	0
4	D	11/12 (91%)	0.34	1 (9%) 9 13	53, 75, 133, 150	0
All	All	1440/1493 (96%)	0.35	113 (7%) 13 18	34, 60, 138, 234	0

The worst 5 of 113 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	806	LEU	10.7
2	B	788	ILE	10.2
2	B	911	LEU	8.2
2	B	795	ILE	8.0
2	B	784	ILE	7.9

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 monosaccharides 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, 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	B-factors(\AA^2)	Q<0.9
6	K	B	1407	1/1	0.75	0.06	112,112,112,112	0
6	K	B	1402	1/1	0.77	0.24	102,102,102,102	0
6	K	A	107	1/1	0.88	0.10	117,117,117,117	0
6	K	B	1404	1/1	0.89	0.09	68,68,68,68	0
6	K	A	106	1/1	0.90	0.24	106,106,106,106	0
5	MG	A	101	1/1	0.92	0.03	74,74,74,74	0
6	K	B	1401	1/1	0.93	0.14	85,85,85,85	0
6	K	B	1408	1/1	0.94	0.07	84,84,84,84	0
6	K	B	1406	1/1	0.95	0.08	65,65,65,65	0
6	K	A	103	1/1	0.98	0.24	65,65,65,65	0
6	K	B	1405	1/1	0.98	0.06	68,68,68,68	0
6	K	A	105	1/1	0.98	0.10	46,46,46,46	0
5	MG	A	102	1/1	0.98	0.07	39,39,39,39	0
6	K	B	1403	1/1	0.98	0.09	65,65,65,65	0
6	K	A	104	1/1	0.99	0.08	67,67,67,67	0
6	K	C	101	1/1	0.99	0.14	72,72,72,72	0

6.5 Other polymers [i](#)

There are no such residues in this entry.