



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 23, 2024 – 04:08 PM JST

PDB ID : 8Y0D
Title : Crystal structure of SauCas9 in complex with sgRNA and 20nt ssDNA target
Authors : Chen, Y.; Chen, J.; Liu, L.
Deposited on : 2024-01-22
Resolution : 3.92 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.21
EDS : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.004 (Gargrove)
Density-Fitness : 1.0.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.40

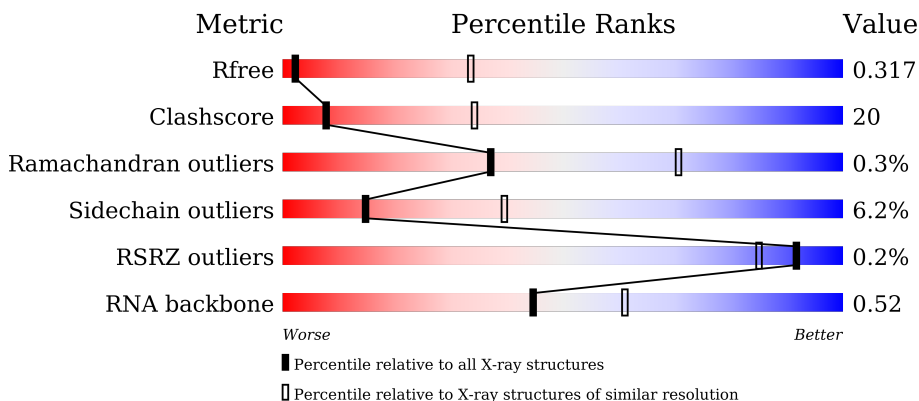
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.92 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}	164625	1175 (4.14-3.70)
Clashscore	180529	1045 (4.12-3.72)
Ramachandran outliers	177936	1006 (4.12-3.72)
Sidechain outliers	177891	1185 (4.14-3.70)
RSRZ outliers	164620	1175 (4.14-3.70)
RNA backbone	3690	1136 (4.84-3.00)

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 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	B	74	 19% 53% 23% 5%
2	C	20	 30% 50% 15% 5%
3	A	1053	 56% 37% 5%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 10023 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 a RNA chain called RNA (73-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	73	1530	684	265	508	73	0	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(P*GP*TP*AP*AP*AP*GP*TP*TP*AP*AP*AP*TP*AP*GP*CP*AP*GP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	19	395	188	79	109	19	0	0	0

- Molecule 3 is a protein called CRISPR-associated endonuclease Cas9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	1009	8093	5154	1376	1550	13	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	580	ALA	ASN	conflict	UNP J7RUA5
A	946	ALA	CYS	conflict	UNP J7RUA5

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	O	0	0
			1	1		
4	A	4	Total	O	0	0
			4	4		

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	68.96Å 145.57Å 97.53Å 90.00° 93.99° 90.00°	Depositor
Resolution (Å)	48.65 – 3.92 48.65 – 3.92	Depositor EDS
% Data completeness (in resolution range)	77.0 (48.65-3.92) 74.5 (48.65-3.92)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.86 (at 3.88Å)	Xtrriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.286 , 0.318 0.286 , 0.317	Depositor DCC
R_{free} test set	11962 reflections (9.97%)	wwPDB-VP
Wilson B-factor (Å ²)	-1.3	Xtrriage
Anisotropy	6.044	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.10 , 0.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.38$, $\langle L^2 \rangle = 0.21$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.86	EDS
Total number of atoms	10023	wwPDB-VP
Average B, all atoms (Å ²)	40.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.57% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	B	0.63	9/1709 (0.5%)	0.97	7/2657 (0.3%)
2	C	0.98	3/445 (0.7%)	0.90	0/685
3	A	0.30	0/8231	0.50	0/11115
All	All	0.42	12/10385 (0.1%)	0.64	7/14457 (0.0%)

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1	G	OP3-P	-10.65	1.48	1.61
1	B	12	U	C1'-N1	6.68	1.58	1.48
1	B	5	U	C1'-N1	6.39	1.58	1.48
1	B	18	U	C1'-N1	6.37	1.58	1.48
1	B	8	U	C1'-N1	6.15	1.57	1.48

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	45	C	C2-N1-C1'	8.17	127.79	118.80
1	B	15	C	O5'-P-OP1	-6.64	99.72	105.70
1	B	45	C	C6-N1-C1'	-5.74	113.91	120.80
1	B	63	A	O4'-C1'-N9	5.66	112.72	108.20
1	B	62	A	P-O3'-C3'	5.23	125.97	119.70

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	B	1530	0	769	72	0
2	C	395	0	214	26	0
3	A	8093	0	7933	310	0
4	A	4	0	0	0	0
4	B	1	0	0	0	0
All	All	10023	0	8916	369	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 20.

The worst 5 of 369 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)
3:A:907:LEU:HD13	3:A:908:SER:H	1.34	0.92
3:A:284:ILE:HA	3:A:288:PHE:HB2	1.51	0.90
1:B:45:C:OP2	3:A:792:ARG:NH1	2.07	0.88
3:A:911:PRO:HA	3:A:929:LYS:HA	1.57	0.87
3:A:3:ARG:HG3	3:A:22:ASP:OD2	1.77	0.84

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
3	A	998/1053 (95%)	938 (94%)	57 (6%)	3 (0%)	37 70

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	292	LYS
3	A	908	SER
3	A	964	ALA

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
3	A	864/970 (89%)	810 (94%)	54 (6%)	15 39

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	462	ASN
3	A	661	ARG
3	A	967	TYR
3	A	480	ARG
3	A	569	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	1037	ASN
3	A	846	GLN
3	A	645	ASN
3	A	557	HIS
3	A	835	GLN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	B	71/74 (95%)	13 (18%)	2 (2%)

5 of 13 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	B	17	U
1	B	36	A
1	B	39	C
1	B	43	A
1	B	45	C

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	B	35	G
1	B	62	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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, 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	B	73/74 (98%)	-1.42	0 100 100	19, 39, 90, 115	0
2	C	19/20 (95%)	-1.06	0 100 100	21, 34, 50, 52	0
3	A	1009/1053 (95%)	-1.18	2 (0%) 92 85	16, 39, 60, 149	0
All	All	1101/1147 (95%)	-1.19	2 (0%) 92 85	16, 39, 62, 149	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	937	GLU	3.1
3	A	691	LYS	2.1

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.