



wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 10, 2023 – 12:15 PM JST

PDB ID : 8HNW
Title : Crystal structure of HpaCas9-sgRNA surveillance complex bound to double-stranded DNA
Authors : Sun, W.; Cheng, Z.; Wang, Y.
Deposited on : 2022-12-08
Resolution : 3.41 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) : 1.13
EDS : 2.34
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) : 2.34

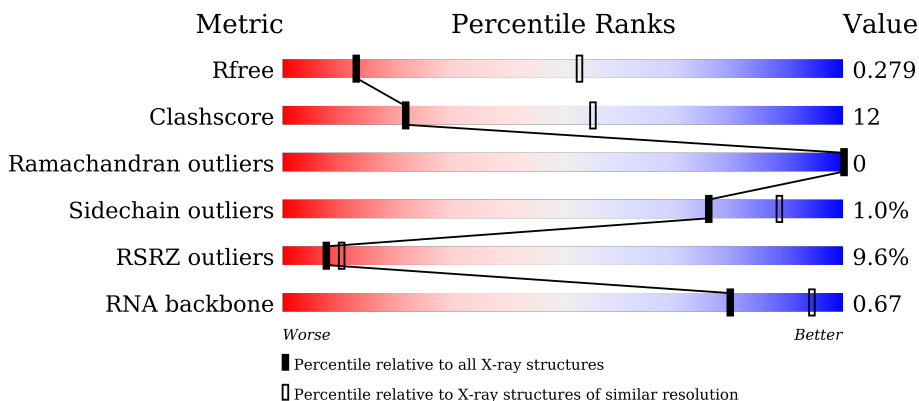
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.41 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	1486 (3.50-3.34)
Clashscore	141614	1572 (3.50-3.34)
Ramachandran outliers	138981	1534 (3.50-3.34)
Sidechain outliers	138945	1535 (3.50-3.34)
RSRZ outliers	127900	1395 (3.50-3.34)
RNA backbone	3102	1012 (3.88-2.96)

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	1055	 4% 57% 19% 23%
2	B	128	 31% 34% 47% 16%
3	C	35	 20% 43% 57%
4	D	11	 18% 100%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 9266 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 protein called CRISPR-associated endonuclease Cas9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	809	6071	3867	1084	1100	20	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP F0ET08
A	581	ALA	HIS	engineered mutation	UNP F0ET08

- Molecule 2 is a RNA chain called sgRNA.

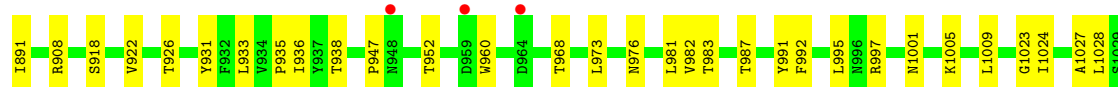
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	107	2251	1010	380	754	107	0	0	0

- Molecule 3 is a DNA chain called Target strand.

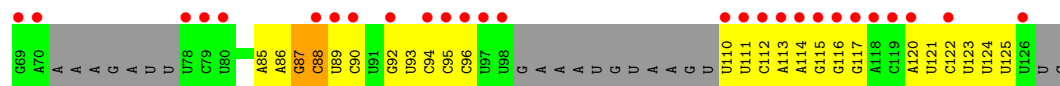
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	35	721	347	133	207	34	0	0	0

- Molecule 4 is a DNA chain called Non-target strand.

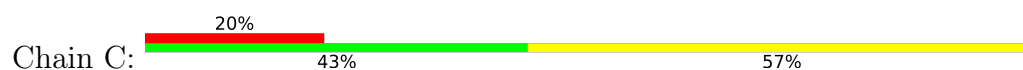
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	11	223	110	37	66	10	0	0	0



- Molecule 2: sgRNA



- Molecule 3: Target strand



- Molecule 4: Non-target strand



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	158.87Å 185.76Å 160.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.62 – 3.41 49.01 – 3.41	Depositor EDS
% Data completeness (in resolution range)	73.6 (44.62-3.41) 74.1 (49.01-3.41)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.78 (at 3.40Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.258 , 0.281 0.257 , 0.279	Depositor DCC
R_{free} test set	1185 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å ²)	39.6	Xtrriage
Anisotropy	0.557	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 52.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.74	EDS
Total number of atoms	9266	wwPDB-VP
Average B, all atoms (Å ²)	56.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.72% 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	A	0.25	0/6172	0.51	1/8365 (0.0%)
2	B	0.20	0/2508	0.81	0/3891
3	C	0.53	0/810	0.95	0/1250
4	D	0.52	0/249	1.04	0/383
All	All	0.28	0/9739	0.67	1/13889 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	791	LEU	CA-CB-CG	5.81	128.66	115.30

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	6071	0	5880	140	0
2	B	2251	0	1146	56	0
3	C	721	0	399	20	0
4	D	223	0	129	0	0
All	All	9266	0	7554	197	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 12.

The worst 5 of 197 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)
1:A:77:LYS:HG3	1:A:81:ARG:HH21	1.48	0.79
1:A:368:ASN:HB3	1:A:371:LEU:HB2	1.64	0.77
2:B:113:A:H2'	2:B:114:A:H8	1.52	0.73
1:A:104:ARG:NH2	1:A:201:GLU:OE1	2.24	0.71
1:A:60:ARG:NH1	2:B:18:A:OP2	2.24	0.71

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
1	A	788/1055 (75%)	731 (93%)	57 (7%)	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
1	A	593/918 (65%)	587 (99%)	6 (1%)	76 88

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	284	ARG
1	A	356	ASN
1	A	665	TYR
1	A	216	PHE
1	A	208	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	368	ASN
1	A	430	GLN
1	A	467	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	103/128 (80%)	5 (4%)	1 (0%)

All (5) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	43	G
2	B	44	A
2	B	65	A
2	B	88	C
2	B	117	G

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	87	G

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](#)

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	A	809/1055 (76%)	0.28	43 (5%) 26 28	5, 58, 99, 129	0
2	B	107/128 (83%)	1.79	40 (37%) 0 0	7, 56, 159, 190	0
3	C	35/35 (100%)	0.99	7 (20%) 1 1	13, 44, 80, 83	0
4	D	11/11 (100%)	1.07	2 (18%) 1 2	25, 29, 75, 76	0
All	All	962/1229 (78%)	0.48	92 (9%) 8 10	5, 57, 105, 190	0

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	115	G	6.2
2	B	97	U	6.0
2	B	43	G	5.8
2	B	117	G	5.8
1	A	22	ALA	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 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.