



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2026 – 02:11 PM UTC

PDB ID : 9KJ6 / pdb_00009kj6
Title : Crystal Structure of SpCas9 ternary complex, amino acids (1242-1263) replaced with Gly-Ser linker
Authors : Kim, S.; Bae, J.; Choi, H.J.
Deposited on : 2024-11-12
Resolution : 3.59 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

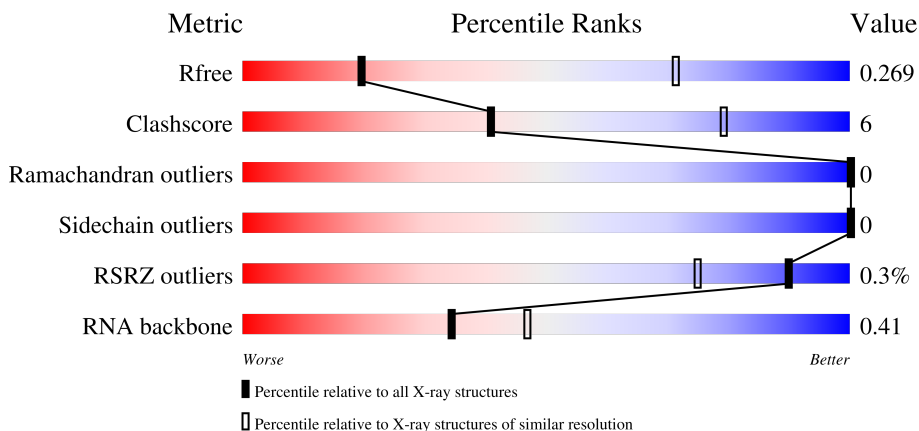
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.59 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}	180053	1747 (3.70-3.50)
Clashscore	190562	1827 (3.70-3.50)
Ramachandran outliers	187476	1773 (3.70-3.50)
Sidechain outliers	187428	1772 (3.70-3.50)
RSRZ outliers	180081	1745 (3.70-3.50)
RNA backbone	3983	1014 (4.14-3.06)

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	1357	
2	B	102	
3	C	43	
4	D	27	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13121 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/Csn1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1316	Total	C	N	O	S	0	0	0
			9700	6123	1692	1865	20			

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP Q99ZW2
A	-2	ALA	-	expression tag	UNP Q99ZW2
A	-1	ALA	-	expression tag	UNP Q99ZW2
A	0	SER	-	expression tag	UNP Q99ZW2
A	10	ALA	ASP	engineered mutation	UNP Q99ZW2
A	840	ALA	HIS	engineered mutation	UNP Q99ZW2
A	1242	GLY	-	linker	UNP Q99ZW2
A	1243	SER	-	linker	UNP Q99ZW2
A	1244	GLY	-	linker	UNP Q99ZW2
A	1245	SER	-	linker	UNP Q99ZW2
A	1246	GLY	-	linker	UNP Q99ZW2
A	1247	SER	-	linker	UNP Q99ZW2
A	1248	GLY	-	linker	UNP Q99ZW2

- Molecule 2 is a RNA chain called single-guide RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	97	Total	C	N	O	P	0	0	0
			2077	931	385	665	96			

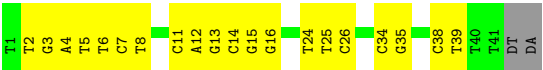
- Molecule 3 is a DNA chain called Target Strand of double-stranded DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	41	Total	C	N	O	P	0	0	0
			828	399	138	251	40			

- Molecule 4 is a DNA chain called Non-target Strand of double-stranded DNA.

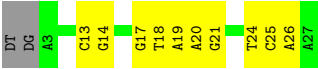
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	25	Total	C	N	O	P	0	0	0
			516	245	100	146	25			

Chain C: 49% 47% 5%



● Molecule 4: Non-target Strand of double-stranded DNA

Chain D: 56% 37% 7%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	88.55Å 146.85Å 222.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.90 – 3.59 29.90 – 3.59	Depositor EDS
% Data completeness (in resolution range)	99.1 (29.90-3.59) 99.0 (29.90-3.59)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.86 (at 3.56Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.228 , 0.271 0.226 , 0.269	Depositor DCC
R_{free} test set	1717 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	131.1	Xtriage
Anisotropy	0.230	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 92.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	13121	wwPDB-VP
Average B, all atoms (Å ²)	137.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.74% 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	0.13	0/9874	0.25	1/13446 (0.0%)
2	B	0.14	0/2329	0.30	0/3630
3	C	0.23	0/924	0.44	0/1423
4	D	0.18	0/580	0.36	0/893
All	All	0.14	0/13707	0.28	1/19392 (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	133	PRO	CB-CA-C	-5.13	104.86	112.55

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	9700	0	8697	108	0
2	B	2077	0	1041	30	0
3	C	828	0	468	17	0
4	D	516	0	281	7	0
All	All	13121	0	10487	153	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 6.

The worst 5 of 153 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:133:PRO:HG2	1:A:137:HIS:CE1	2.11	0.86
1:A:1210:ARG:HG2	1:A:1265:VAL:HA	1.61	0.81
1:A:869:ASN:HD21	1:A:908:LEU:H	1.33	0.73
1:A:820:ARG:NH2	1:A:825:ASP:O	2.23	0.71
1:A:840:ALA:O	1:A:864:ARG:NH1	2.27	0.68

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	1300/1357 (96%)	1240 (95%)	60 (5%)	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	905/1208 (75%)	905 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10

such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1044	ASN
1	A	1221	GLN
1	A	1302	ASN
1	A	754	HIS
1	A	758	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	96/102 (94%)	20 (20%)	1 (1%)

5 of 20 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	4	A
2	B	10	G
2	B	12	U
2	B	20	A
2	B	28	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	2	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	A	1316/1357 (96%)	-0.16	4 (0%) 90 75	73, 130, 176, 210	0
2	B	97/102 (95%)	-0.39	0 100 100	93, 131, 227, 234	0
3	C	41/43 (95%)	0.05	0 100 100	89, 147, 247, 251	0
4	D	25/27 (92%)	0.10	0 100 100	110, 211, 267, 270	0
All	All	1479/1529 (96%)	-0.16	4 (0%) 90 75	73, 130, 187, 270	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	42	SER	3.1
1	A	899	ASN	2.6
1	A	117	PRO	2.5
1	A	366	GLY	2.0

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 oligosaccharides 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.