wwPDB X-ray Structure Validation Summary Report (i)

Jun 3, 2024 – 02:29 PM JST

PDB ID	:	8KAM
Title	:	Crystal structure of SpyCas9 in complex with sgRNA and 16nt target DNA
Authors	:	Chen, Y.; Chen, J.; Liu, L.
Deposited on	:	2023-08-03
Resolution	:	3.91 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 (i) 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 (i)) were used in the production of this report:

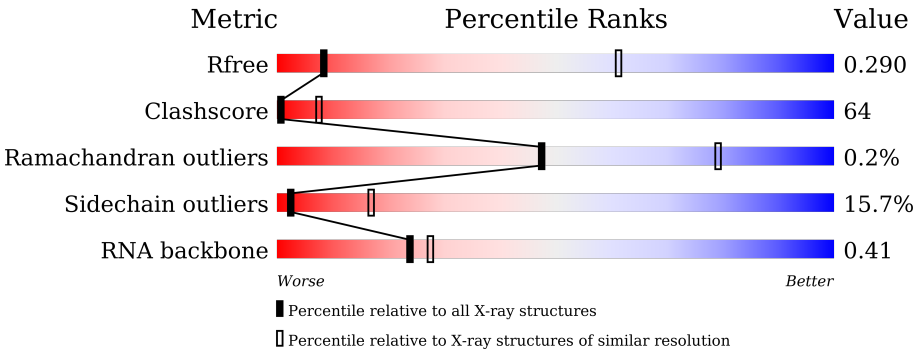
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 3.91 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	1019 (4.18-3.66)
Clashscore	141614	1016 (4.16-3.68)
Ramachandran outliers	138981	1039 (4.18-3.66)
Sidechain outliers	138945	1032 (4.18-3.66)
RNA backbone	3102	1041 (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 <=5%

Mol	Chain	Length	Quality of chain
1	A	98	33% 43% 20% .
2	B	1368	39% 46% 12% . .
3	C	24	38% 63%
4	D	11	64% 36%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13534 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 (98-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	94	2009	899	362	654	94	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	1326	10816	6890	1877	2027	22	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	ASP	engineered mutation	UNP Q99ZW2
B	840	ALA	HIS	engineered mutation	UNP Q99ZW2

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	24	483	233	88	139	23	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(*TP*TP*TP*AP*GP*GP*TP*AP*TP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	11	225	110	37	68	10	0	0	0

- Molecule 5 is water.

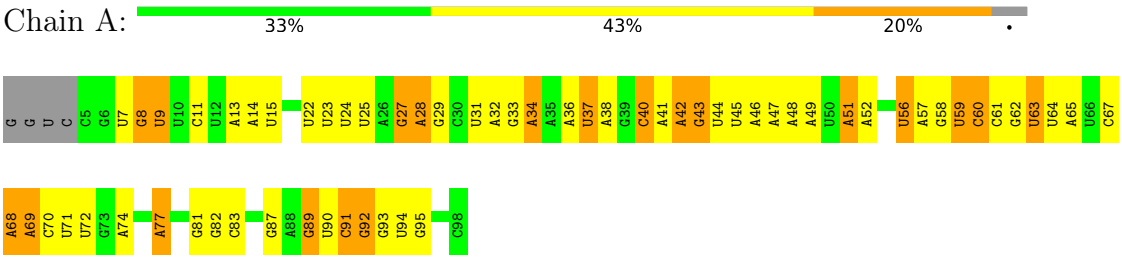
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total O 1 1	0	0

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. 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. 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.

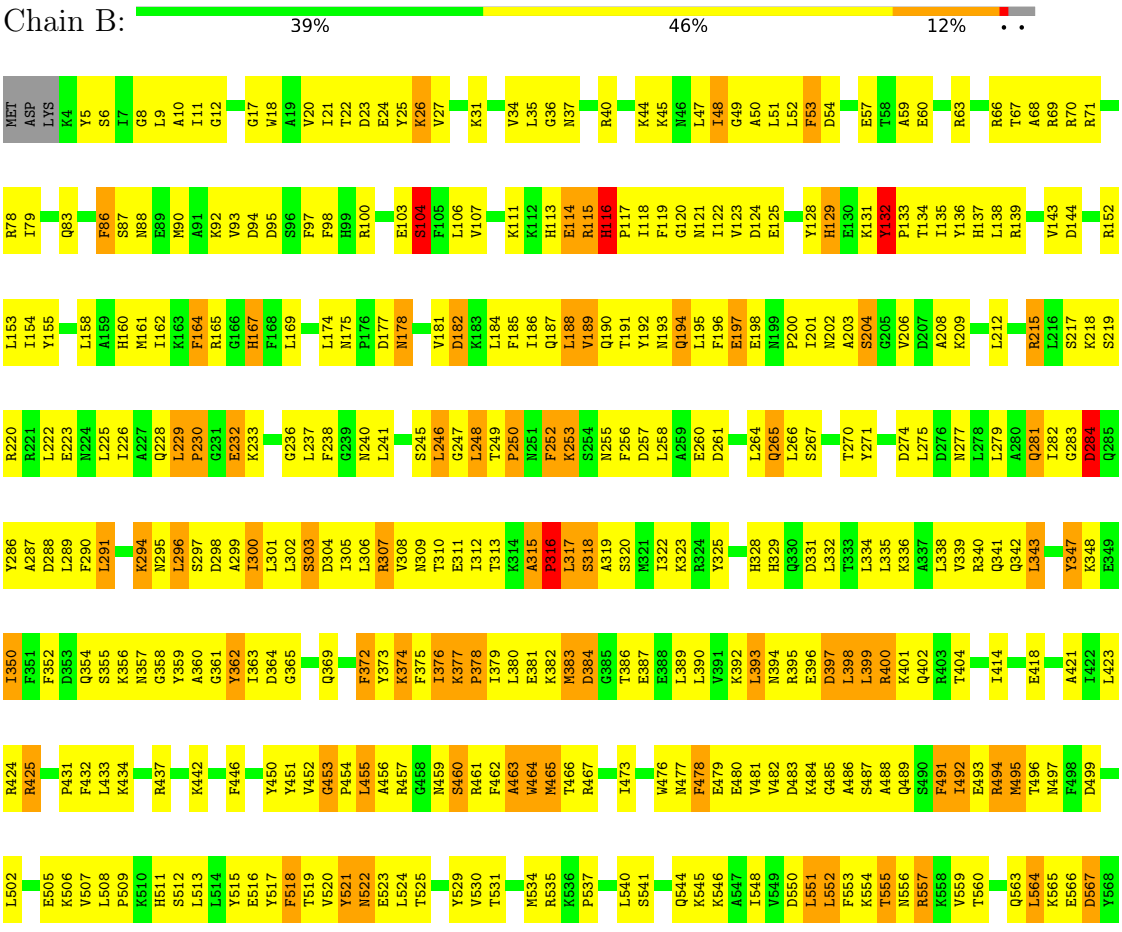
• Molecule 1: RNA (98-MER)

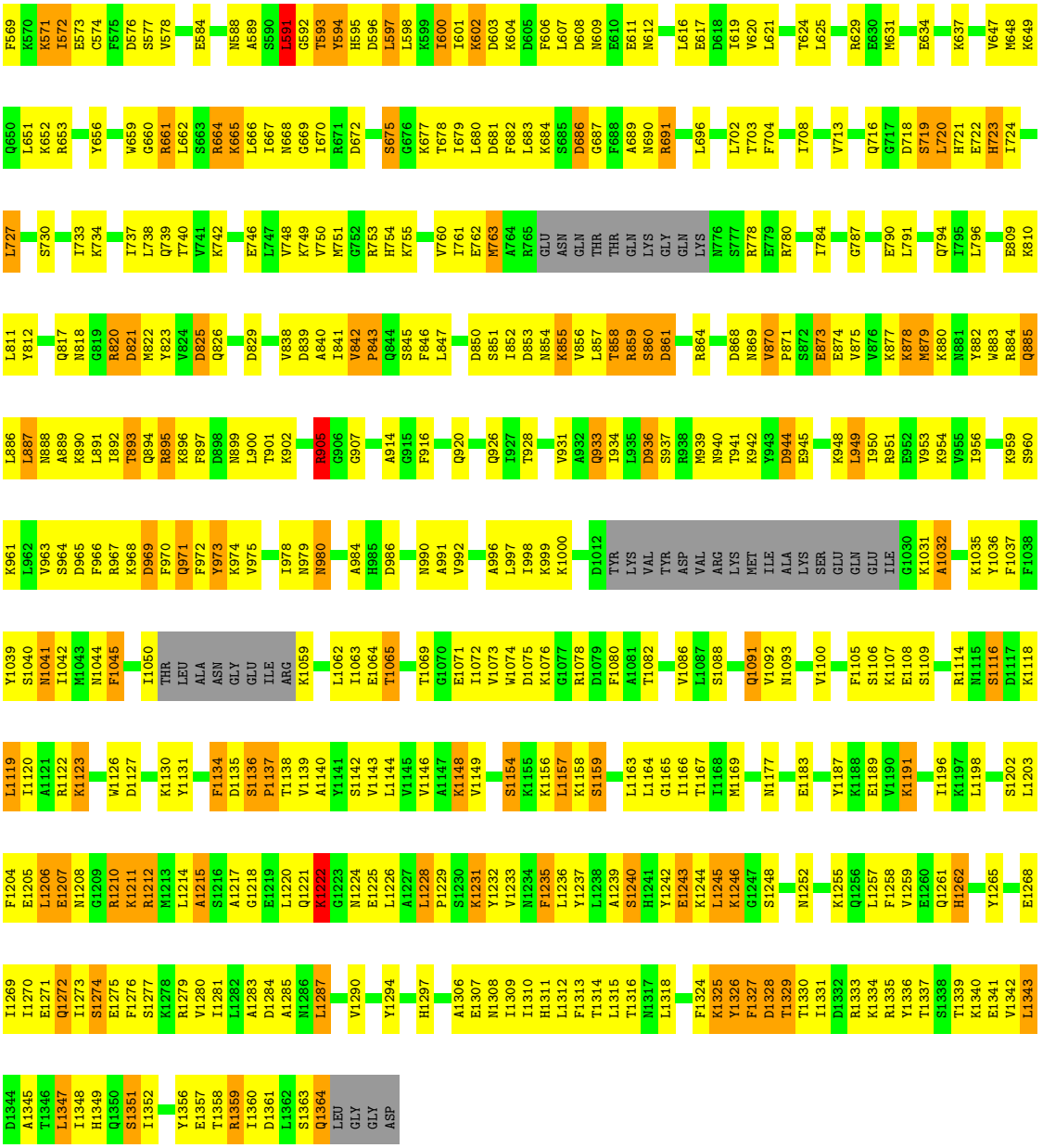
Chain A:



• Molecule 2: CRISPR-associated endonuclease Cas9/Csn1

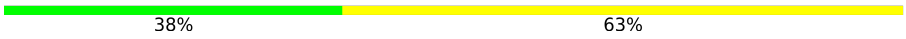
Chain B:





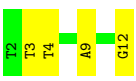
• Molecule 3: DNA (5'-D(*CP*AP*AP*TP*AP*CP*CP*TP*CP*TP*TP*CP*AP*AP*TP*TP*AP*GP*AP*AP*CP*AP*CP*G)-3')

Chain C:



• Molecule 4: DNA (5'-D(*TP*TP*TP*AP*GP*GP*TP*AP*TP*TP*G)-3')

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	199.03Å 70.39Å 186.96Å 90.00° 109.62° 90.00°	Depositor
Resolution (Å)	38.78 – 3.91 48.71 – 3.91	Depositor EDS
% Data completeness (in resolution range)	36.0 (38.78-3.91) 47.6 (48.71-3.91)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.86 (at 3.88Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.260 , 0.290 0.260 , 0.290	Depositor DCC
R_{free} test set	1138 reflections (10.01%)	wwPDB-VP
Wilson B-factor (Å^2)	-2.7	Xtriage
Anisotropy	8.165	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.16 , -10.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.71	EDS
Total number of atoms	13534	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.79% 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

5.1 Standard geometry

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	0/2249	0.81	0/3503
2	B	0.39	6/11008 (0.1%)	0.71	48/14794 (0.3%)
3	C	0.64	0/541	1.01	0/831
4	D	0.55	0/251	1.05	0/387
All	All	0.39	6/14049 (0.0%)	0.75	48/19515 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	B	0	2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1137	PRO	N-CD	5.32	1.55	1.47
2	B	378	PRO	N-CD	5.26	1.55	1.47
2	B	230	PRO	N-CD	5.15	1.55	1.47
2	B	133	PRO	N-CD	5.07	1.54	1.47
2	B	843	PRO	N-CD	5.07	1.54	1.47

The worst 5 of 48 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	675	SER	CB-CA-C	-19.99	72.11	110.10
2	B	1222	LYS	CB-CA-C	-14.01	82.39	110.40
2	B	284	ASP	CB-CA-C	-10.87	88.67	110.40
2	B	164	PHE	N-CA-C	-10.23	83.39	111.00
2	B	1222	LYS	N-CA-C	9.77	137.38	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	1222	LYS	Peptide
2	B	675	SER	Peptide

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	2009	0	1009	88	0
2	B	10816	0	10957	1624	3
3	C	483	0	272	48	0
4	D	225	0	129	6	0
5	B	1	0	0	0	0
All	All	13534	0	12367	1663	3

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 64.

The worst 5 of 1663 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:1204:PHE:CE1	2:B:1347:LEU:HD12	1.32	1.59
2:B:557:ARG:HG2	2:B:595:HIS:CD2	1.33	1.57
2:B:520:VAL:HG21	2:B:591:LEU:CD2	1.32	1.57
2:B:446:PHE:CZ	2:B:478:PHE:CD1	1.92	1.56
2:B:178:ASN:CB	2:B:299:ALA:HA	1.26	1.56

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:232:GLU:O	2:B:573:GLU:OE2[4_445]	1.99	0.21
2:B:226:ILE:CD1	2:B:574:CYS:SG[4_445]	2.08	0.12
2:B:611:GLU:O	2:B:859:ARG:NH2[1_565]	2.12	0.08

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	1318/1368 (96%)	1254 (95%)	62 (5%)	2 (0%)	47 79

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	250	PRO
2	B	316	PRO

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	1183/1225 (97%)	997 (84%)	186 (16%)	2 17

5 of 186 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	829	ASP
2	B	1037	PHE
2	B	860	SER
2	B	936	ASP
2	B	1123	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	980	ASN
2	B	990	ASN
2	B	1262	HIS
2	B	1091	GLN
2	B	459	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	93/98 (94%)	27 (29%)	4 (4%)

5 of 27 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	9	U
1	A	11	C
1	A	27	G
1	A	28	A
1	A	29	G

All (4) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	8	G
1	A	27	G
1	A	42	A
1	A	68	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 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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.