



wwPDB X-ray Structure Validation Summary Report

Mar 31, 2014 – 03:52 PM BST

PDB ID : 4CMP
Title : Crystal structure of *S. pyogenes* Cas9
Authors : Jinek, M.; Jiang, F.; Taylor, D.W.; Sternberg, S.H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A.T.; Charpentier, E.; Nogales, E.; Doudna, J.A.
Deposited on : 2014-01-16
Resolution : 2.62 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

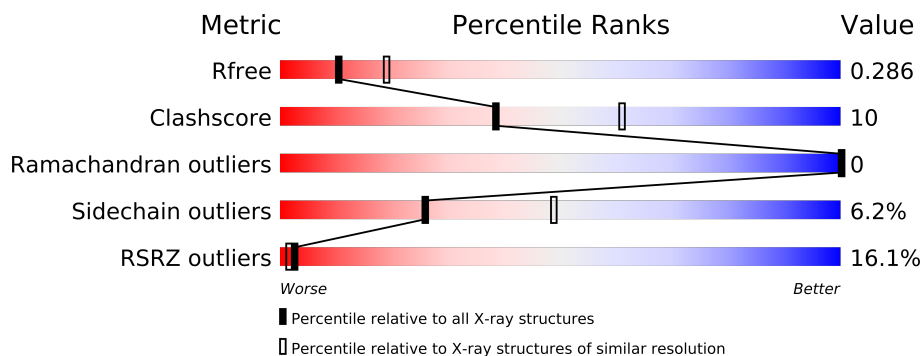
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable23004
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable23004

1 Overall quality at a glance

The reported resolution of this entry is 2.62 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}	66092	1940 (2.64-2.60)
Clashscore	79885	2404 (2.64-2.60)
Ramachandran outliers	78287	2360 (2.64-2.60)
Sidechain outliers	78261	2360 (2.64-2.60)
RSRZ outliers	66119	1939 (2.64-2.60)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	1372	
1	B	1372	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	MG	B	2367	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 38285 atoms, of which 19194 are hydrogens and 0 are deuterium.

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	1144	Total	C	H	N	O	S	0	0	0
			18772	5949	9456	1610	1738	19			
1	B	1166	Total	C	H	N	O	S	0	0	0
			19284	6101	9738	1648	1778	19			

There are 8 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
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

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



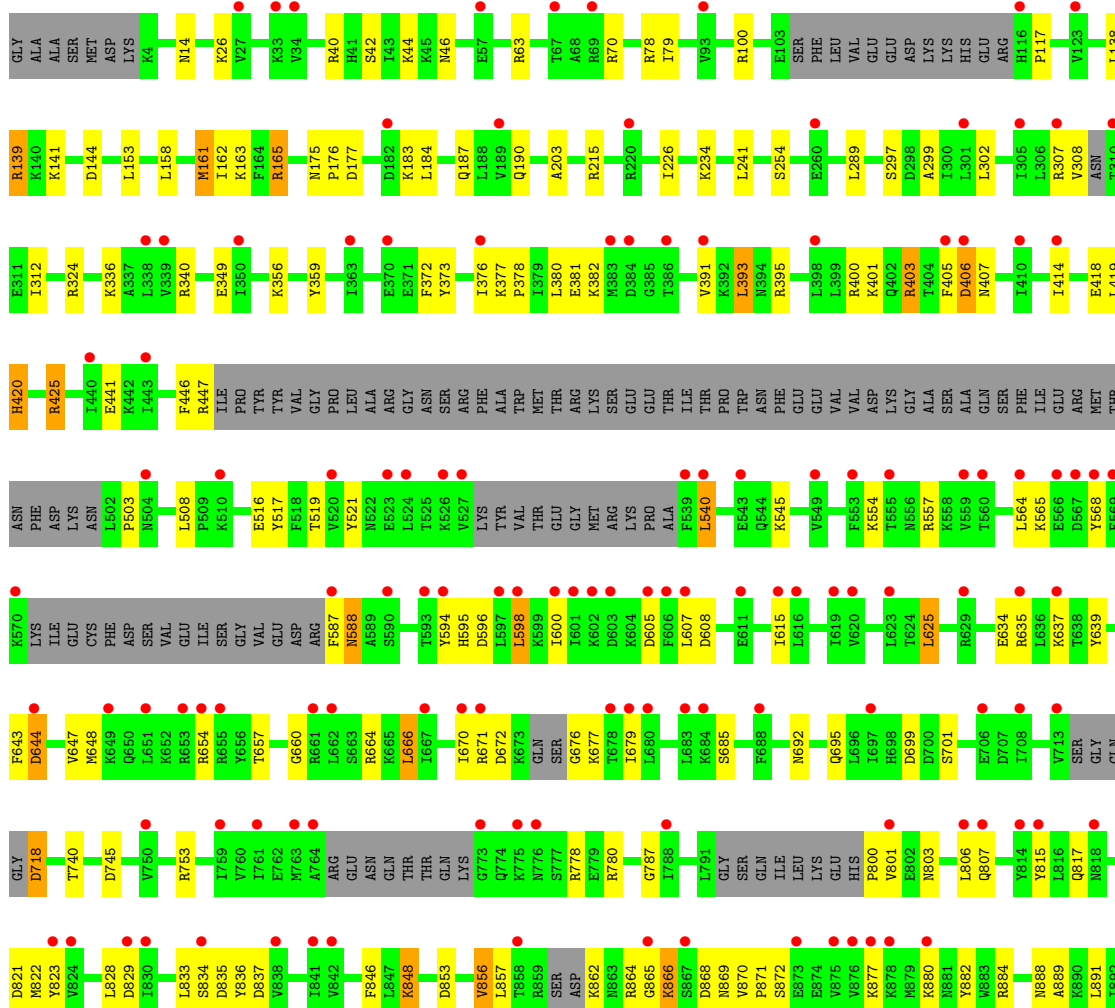
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		

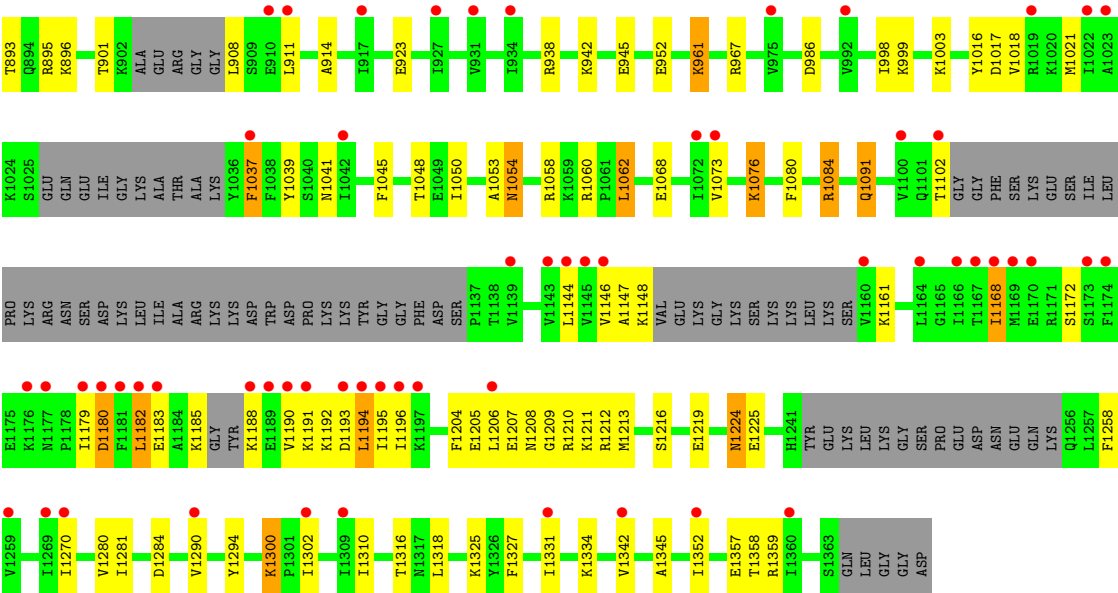
- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Mg	0	0
			1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	92	Total	O	0	0
			92	92		
4	B	111	Total	O	0	0
			111	111		





4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	159.78Å 209.62Å 91.26Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.48 – 2.62 47.48 – 2.62	Depositor EDS
% Data completeness (in resolution range)	99.6 (47.48-2.62) 99.6 (47.48-2.62)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.65 (at 2.61Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.252 , 0.286 0.252 , 0.286	Depositor DCC
R_{free} test set	2421 reflections (2.62%)	DCC
Wilson B-factor (Å ²)	64.8	Xtriage
Anisotropy	0.232	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 31.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 92492 reflections	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	38285	wwPDB-VP
Average B, all atoms (Å ²)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 19.91 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 9.7554e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, SO4

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	0.28	0/9463	0.51	1/12717 (0.0%)
1	B	0.30	0/9698	0.50	1/13022 (0.0%)
All	All	0.29	0/19161	0.50	2/25739 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1084	ARG	NE-CZ-NH1	-5.51	117.55	120.30
1	B	1193	ASP	CB-CG-OD2	5.17	122.95	118.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Close contacts

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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	9316	9456	0	168	0
1	B	9546	9738	0	206	0
2	A	10	0	0	0	0
2	B	15	0	0	5	0
3	B	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	92	0	0	7	0
4	B	111	0	0	13	0
All	All	19091	19194	0	374	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 10.

The worst 5 of 374 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:1179:ILE:O	1:B:1183:GLU:HG3	1.17	1.29
1:B:1191:LYS:HE2	1:B:1194:LEU:CD2	1.78	1.14
1:B:1179:ILE:HD12	1:B:1192:LYS:CD	1.80	1.12
1:B:1179:ILE:CD1	1:B:1192:LYS:HD2	1.81	1.10
1:B:1179:ILE:O	1:B:1183:GLU:CG	2.07	1.02

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

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	1108/1372 (81%)	1064 (96%)	44 (4%)	0	100	100
1	B	1132/1372 (82%)	1095 (97%)	37 (3%)	0	100	100
All	All	2240/2744 (82%)	2159 (96%)	81 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	1019/1228 (83%)	957 (94%)	62 (6%)	26	50
1	B	1050/1228 (86%)	984 (94%)	66 (6%)	25	48
All	All	2069/2456 (84%)	1941 (94%)	128 (6%)	26	49

5 of 128 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1274	SER
1	B	393	LEU
1	B	1180	ASP
1	A	1304	GLU
1	B	141	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 49 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1311	HIS
1	B	167	HIS
1	B	1177	ASN
1	B	46	ASN
1	B	187	GLN

5.3.3 RNA ⓘ

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	SO4	A	2364	-	4,4,4	0.32	0	6,6,6	0.10	0
2	SO4	A	2365	-	4,4,4	0.21	0	6,6,6	0.07	0
2	SO4	B	2364	-	4,4,4	0.30	0	6,6,6	0.15	0
2	SO4	B	2365	-	4,4,4	0.22	0	6,6,6	0.10	0
2	SO4	B	2366	-	4,4,4	0.13	0	6,6,6	0.36	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	SO4	A	2364	-	-	0/0/0/0	0/0/0/0
2	SO4	A	2365	-	-	0/0/0/0	0/0/0/0
2	SO4	B	2364	-	-	0/0/0/0	0/0/0/0
2	SO4	B	2365	-	-	0/0/0/0	0/0/0/0
2	SO4	B	2366	-	-	0/0/0/0	0/0/0/0

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.

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	1144/1372 (83%)	0.89	185 (16%) 2 1	17, 59, 125, 164	0
1	B	1166/1372 (84%)	0.93	188 (16%) 2 1	16, 54, 119, 162	0
All	All	2310/2744 (84%)	0.91	373 (16%) 2 1	16, 57, 122, 164	0

The worst 5 of 373 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	568	TYR	11.6
1	B	527	VAL	10.2
1	A	1195	ILE	7.9
1	B	1190	VAL	7.7
1	A	527	VAL	7.7

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(\AA^2)	Q<0.9
3	MG	B	2367	1/1	0.43	8.71	54,54,54,54	0
2	SO4	A	2365	5/5	0.24	1.58	88,99,106,106	0
2	SO4	A	2364	5/5	0.21	0.30	32,39,54,59	0
2	SO4	B	2365	5/5	0.20	0.07	80,91,102,103	0
2	SO4	B	2366	5/5	0.17	-0.18	30,30,30,30	0
2	SO4	B	2364	5/5	0.19	-0.34	21,25,39,57	0

6.5 Other polymers ⓘ

There are no such residues in this entry.