wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2016 – 10:37 PM GMT

PDB ID	:	5DS4
Title	:	Crystal structure the Escherichia coli Cas1-Cas2 complex bound to protospacer DNA
Authors	:	Nunez, J.K.; Harrington, L.B.; Kranzusch, P.J.; Engelman, A.N.; Doudna, J.A.
Deposited on	:	2015-09-16
Resolution	:	3.20 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

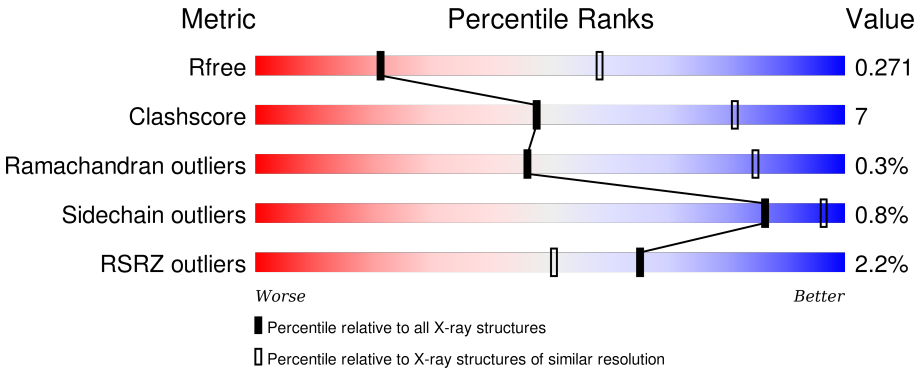
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

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 3.20 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}	91344	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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. 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% 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	306	2% 67% 17% 16%
1	B	306	% 71% 16% 12%
1	C	306	3% 68% 15% 17%
1	D	306	3% 67% 15% 17%
2	E	104	76% 13% 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	104	
3	G	28	
4	H	28	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10517 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 Cas1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	256	Total	C	N	O	S	0	0	0
			1953	1249	347	350	7			
1	B	268	Total	C	N	O	S	0	0	0
			2061	1319	367	368	7			
1	C	254	Total	C	N	O	S	0	0	0
			1941	1241	345	348	7			
1	D	253	Total	C	N	O	S	0	0	0
			1949	1253	344	345	7			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP Q46896
B	0	SER	-	expression tag	UNP Q46896
C	0	SER	-	expression tag	UNP Q46896
D	0	SER	-	expression tag	UNP Q46896

- Molecule 2 is a protein called CRISPR-associated endoribonuclease Cas2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	93	Total	C	N	O	S	0	0	0
			732	470	127	131	4			
2	F	94	Total	C	N	O	S	0	0	0
			739	475	128	132	4			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	0	MET	-	initiating methionine	UNP P45956
E	95	GLY	-	expression tag	UNP P45956
E	96	SER	-	expression tag	UNP P45956
E	97	SER	-	expression tag	UNP P45956

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	98	GLU	-	expression tag	UNP P45956
E	99	ASN	-	expression tag	UNP P45956
E	100	LEU	-	expression tag	UNP P45956
E	101	TYR	-	expression tag	UNP P45956
E	102	PHE	-	expression tag	UNP P45956
E	103	GLN	-	expression tag	UNP P45956
F	0	MET	-	initiating methionine	UNP P45956
F	95	GLY	-	expression tag	UNP P45956
F	96	SER	-	expression tag	UNP P45956
F	97	SER	-	expression tag	UNP P45956
F	98	GLU	-	expression tag	UNP P45956
F	99	ASN	-	expression tag	UNP P45956
F	100	LEU	-	expression tag	UNP P45956
F	101	TYR	-	expression tag	UNP P45956
F	102	PHE	-	expression tag	UNP P45956
F	103	GLN	-	expression tag	UNP P45956

- Molecule 3 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	28	Total	C	N	O	P	0	0	0
			578	275	118	158	27			

- Molecule 4 is a DNA chain called DNA (28-MER).

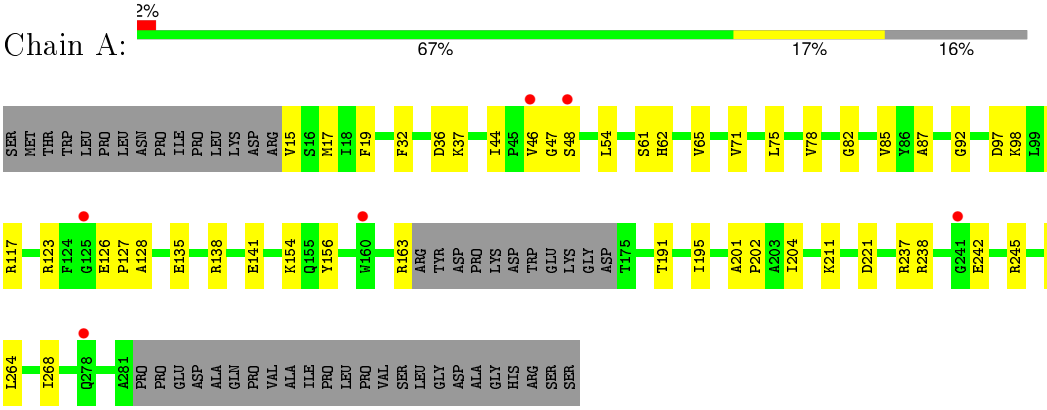
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	H	28	Total	C	N	O	P	0	0	0
			564	274	86	177	27			

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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.

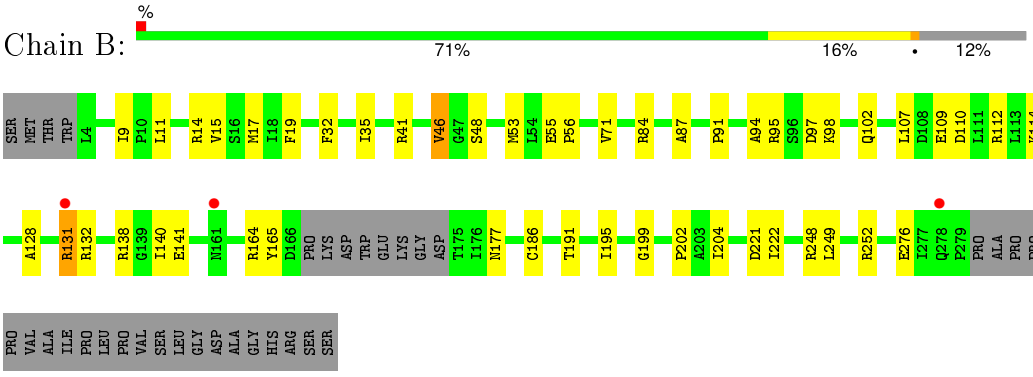
- Molecule 1: CRISPR-associated endonuclease Cas1

Chain A:



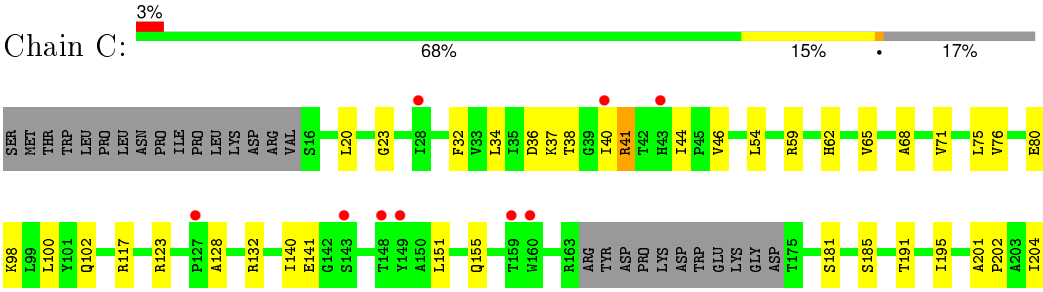
- Molecule 1: CRISPR-associated endonuclease Cas1

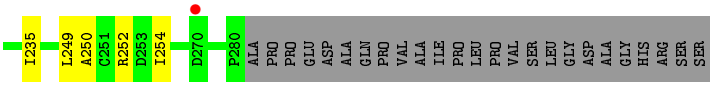
Chain B:



- Molecule 1: CRISPR-associated endonuclease Cas1

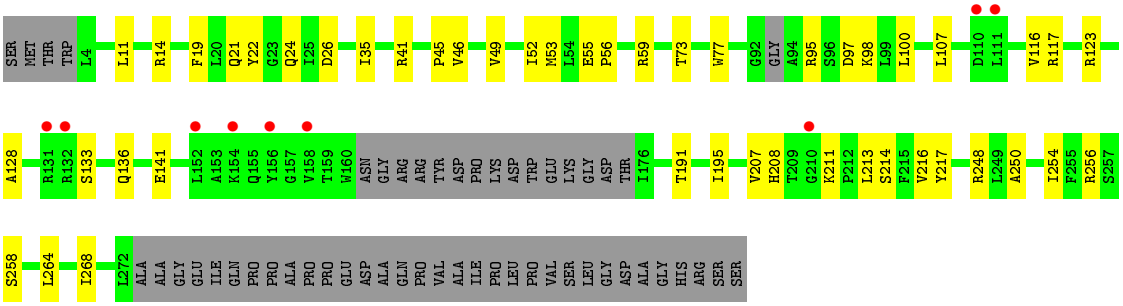
Chain C:





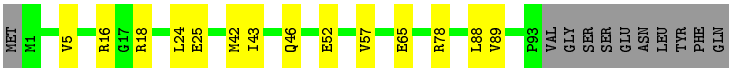
- Molecule 1: CRISPR-associated endonuclease Cas1

Chain D:



- Molecule 2: CRISPR-associated endoribonuclease Cas2

Chain E:



- Molecule 2: CRISPR-associated endoribonuclease Cas2

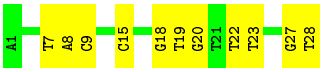
Chain F:

- Molecule 3: DNA (28-MER)

Chain G:

- Molecule 4: DNA (28-MER)

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	88.02Å 123.01Å 196.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.00 – 3.20 49.00 – 3.20	Depositor EDS
% Data completeness (in resolution range)	99.7 (49.00-3.20) 99.8 (49.00-3.20)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.73 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.242 , 0.270 0.251 , 0.271	Depositor DCC
R_{free} test set	2000 reflections (5.92%)	DCC
Wilson B-factor (Å^2)	64.6	Xtriage
Anisotropy	0.639	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 30.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Outliers	0 of 35808 reflections	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	10517	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.48% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.22	0/1988	0.44	0/2695
1	B	0.22	0/2099	0.41	0/2846
1	C	0.23	0/1976	0.44	0/2678
1	D	0.21	0/1984	0.42	0/2689
2	E	0.20	0/746	0.40	0/1014
2	F	0.20	0/753	0.38	0/1024
3	G	0.55	0/652	0.83	0/1005
4	H	0.65	1/627 (0.2%)	1.06	0/966
All	All	0.29	1/10825 (0.0%)	0.52	0/14917

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	H	28	DT	C1'-N1	7.50	1.59	1.49

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-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 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	1953	0	2019	32	0
1	B	2061	0	2134	34	0
1	C	1941	0	2005	34	0
1	D	1949	0	2030	29	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	732	0	747	12	0
2	F	739	0	756	16	0
3	G	578	0	314	15	0
4	H	564	0	324	8	0
All	All	10517	0	10329	151	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 7.

The worst 5 of 151 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:E:5:VAL:HG21	2:F:5:VAL:HG21	1.57	0.86
3:G:25:DG:H4'	3:G:26:DG:H5'	1.73	0.70
1:C:46:VAL:HG11	1:C:71:VAL:HG11	1.73	0.70
3:G:25:DG:N3	3:G:26:DG:N2	2.40	0.69
1:B:94:ALA:HA	1:B:199:GLY:HA2	1.74	0.69

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	252/306 (82%)	244 (97%)	7 (3%)	1 (0%)	39 80
1	B	264/306 (86%)	251 (95%)	13 (5%)	0	100 100
1	C	250/306 (82%)	240 (96%)	9 (4%)	1 (0%)	39 80
1	D	247/306 (81%)	239 (97%)	8 (3%)	0	100 100
2	E	91/104 (88%)	88 (97%)	3 (3%)	0	100 100
2	F	92/104 (88%)	88 (96%)	3 (3%)	1 (1%)	17 62

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1196/1432 (84%)	1150 (96%)	43 (4%)	3 (0%)	46 85

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	40	ILE
1	A	238	ARG
2	F	53	GLU

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	202/246 (82%)	201 (100%)	1 (0%)	92 97
1	B	215/246 (87%)	211 (98%)	4 (2%)	65 89
1	C	201/246 (82%)	199 (99%)	2 (1%)	82 95
1	D	205/246 (83%)	205 (100%)	0	100 100
2	E	78/88 (89%)	78 (100%)	0	100 100
2	F	79/88 (90%)	78 (99%)	1 (1%)	76 92
All	All	980/1160 (84%)	972 (99%)	8 (1%)	86 96

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	131	ARG
2	F	91	PHE
1	C	41	ARG
1	B	46	VAL
1	B	165	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	177	ASN

5.3.3 RNA

There are no RNA molecules 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

There are no ligands in this entry.

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, 95^{th} 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	256/306 (83%)	0.26	6 (2%) 64 49	41, 67, 85, 103	0
1	B	268/306 (87%)	0.12	3 (1%) 82 72	36, 59, 85, 102	0
1	C	254/306 (83%)	0.42	10 (3%) 43 28	45, 72, 96, 110	0
1	D	253/306 (82%)	0.38	9 (3%) 46 31	43, 73, 100, 112	0
2	E	93/104 (89%)	-0.10	0 100 100	36, 50, 67, 76	0
2	F	94/104 (90%)	-0.03	0 100 100	35, 51, 65, 71	0
3	G	28/28 (100%)	0.17	0 100 100	51, 75, 92, 97	0
4	H	28/28 (100%)	0.03	0 100 100	46, 78, 94, 105	0
All	All	1274/1488 (85%)	0.23	28 (2%) 65 50	35, 65, 93, 112	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	111	LEU	3.8
1	D	152	LEU	3.5
1	A	46	VAL	3.4
1	B	278	GLN	2.9
1	C	28	ILE	2.8

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.