



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 12:09 PM UTC

PDB ID : 5VVJ / pdb_00005vvj
Title : Cas1-Cas2 bound to half-site intermediate
Authors : Wright, A.V.; Knott, G.J.; Doxzen, K.W.; Doudna, J.A.
Deposited on : 2017-05-19
Resolution : 3.89 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

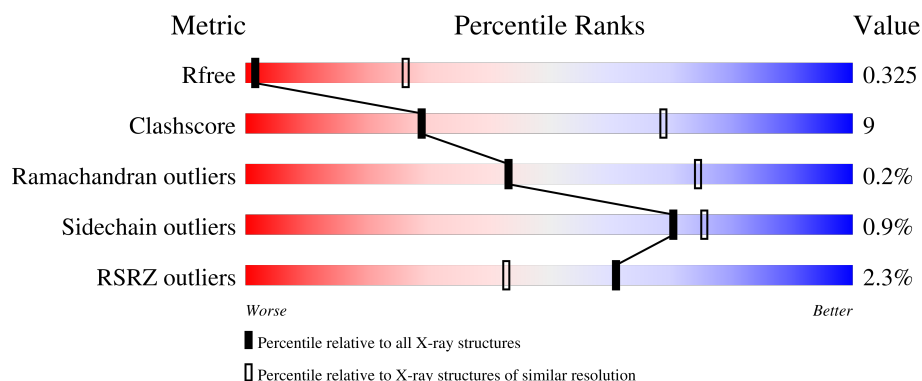
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.89 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	1270 (4.10-3.70)
Clashscore	190562	1034 (4.08-3.72)
Ramachandran outliers	187476	1251 (4.10-3.70)
Sidechain outliers	187428	1243 (4.10-3.70)
RSRZ outliers	180081	1269 (4.10-3.70)

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	305	68% 17% 14%
1	B	305	70% 17% 12%
1	C	305	3% 70% 13% 17%
1	D	305	3% 70% 17% 13%
2	E	95	4% 75% 23%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	95	4% 77% 23%
3	G	28	36% 64%
4	H	112	39% 38% 22%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11887 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	262	Total	C	N	O	S	0	0	0
			2018	1290	358	363	7			
1	B	269	Total	C	N	O	S	0	0	0
			2073	1328	369	369	7			
1	C	253	Total	C	N	O	S	0	0	0
			1934	1238	340	349	7			
1	D	265	Total	C	N	O	S	0	0	0
			2035	1303	363	362	7			

- 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
			731	470	127	131	3			
2	F	95	Total	C	N	O	S	0	0	0
			743	477	129	133	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	95	GLY	-	expression tag	UNP P45956
F	95	GLY	-	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			

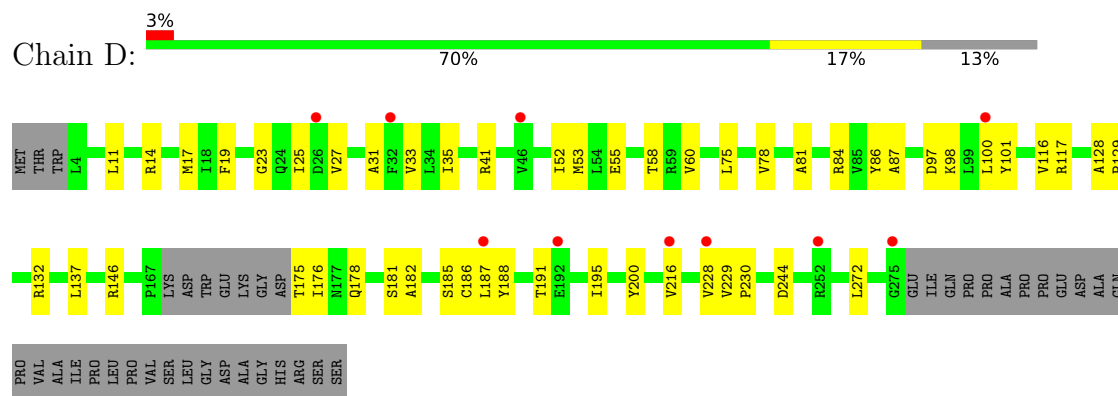
- Molecule 4 is a DNA chain called DNA (112-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	H	87	Total	C	N	O	P	0	0	0
			1775	845	313	531	86			

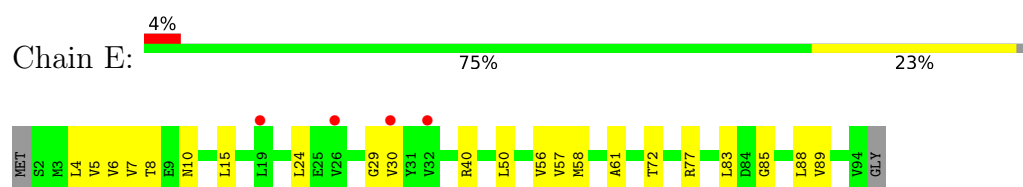
- Molecule 1: CRISPR-associated endonuclease Cas1



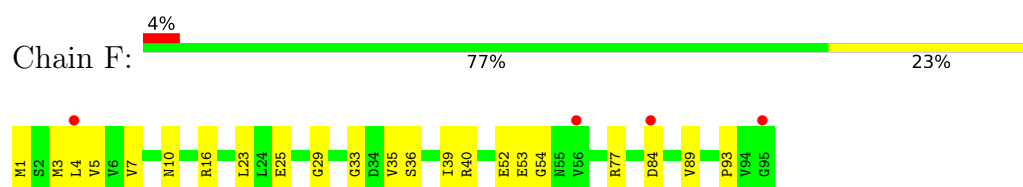
- Molecule 1: CRISPR-associated endonuclease Cas1



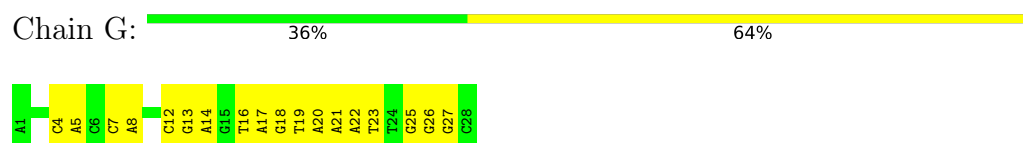
- Molecule 2: CRISPR-associated endoribonuclease Cas2



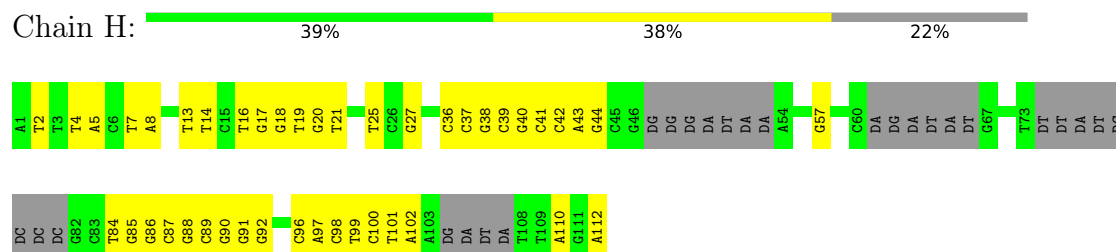
- Molecule 2: CRISPR-associated endoribonuclease Cas2



- Molecule 3: DNA (28-MER)



- Molecule 4: DNA (112-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	75.07Å 183.11Å 196.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	134.10 – 3.89 134.10 – 3.89	Depositor EDS
% Data completeness (in resolution range)	99.8 (134.10-3.89) 100.0 (134.10-3.89)	Depositor EDS
R_{merge}	0.67	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.99 (at 3.26Å)	Xtriage
Refinement program	PHENIX 1.11.1	Depositor
R, R_{free}	0.288 , 0.329 0.290 , 0.325	Depositor DCC
R_{free} test set	2090 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å ²)	88.6	Xtriage
Anisotropy	0.487	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 97.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.36$, $\langle L^2 \rangle = 0.19$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	11887	wwPDB-VP
Average B, all atoms (Å ²)	158.0	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.*

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	0.15	0/2057	0.33	0/2789
1	B	0.15	0/2113	0.37	1/2866 (0.0%)
1	C	0.14	0/1968	0.31	0/2666
1	D	0.14	0/2073	0.33	0/2811
2	E	0.14	0/745	0.31	0/1014
2	F	0.16	0/757	0.35	0/1029
3	G	0.21	0/652	0.37	0/1005
4	H	0.23	0/1982	0.43	0/3048
All	All	0.16	0/12347	0.36	1/17228 (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	B	71	VAL	N-CA-C	-5.50	107.12	112.29

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	2018	0	2069	35	0
1	B	2073	0	2143	37	1
1	C	1934	0	1989	28	1
1	D	2035	0	2109	33	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	731	0	744	15	0
2	F	743	0	759	16	0
3	G	578	0	314	20	0
4	H	1775	0	987	37	1
All	All	11887	0	11114	199	2

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 9.

The worst 5 of 199 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:C:59:ARG:NH1	4:H:2:DT:OP1	2.16	0.79
3:G:19:DT:H2'	3:G:20:DA:C8	2.20	0.76
1:C:59:ARG:HH12	4:H:2:DT:H5''	1.54	0.73
1:D:187:LEU:HD22	1:D:228:VAL:HG21	1.74	0.69
1:B:14:ARG:NH2	1:B:71:VAL:O	2.28	0.67

All (2) 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)
1:B:146:ARG:NH1	4:H:110:DA:OP1[1_655]	2.01	0.19
1:C:108:ASP:OD2	1:D:146:ARG:NH1[4_555]	2.04	0.16

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	258/305 (85%)	247 (96%)	11 (4%)	0	100 100
1	B	265/305 (87%)	251 (95%)	13 (5%)	1 (0%)	30 64

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	247/305 (81%)	243 (98%)	4 (2%)	0	100	100
1	D	261/305 (86%)	248 (95%)	13 (5%)	0	100	100
2	E	91/95 (96%)	87 (96%)	4 (4%)	0	100	100
2	F	93/95 (98%)	91 (98%)	1 (1%)	1 (1%)	11	43
All	All	1215/1410 (86%)	1167 (96%)	46 (4%)	2 (0%)	43	74

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	16	SER
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	209/245 (85%)	207 (99%)	2 (1%)	68	75
1	B	216/245 (88%)	213 (99%)	3 (1%)	59	71
1	C	200/245 (82%)	197 (98%)	3 (2%)	57	70
1	D	212/245 (86%)	211 (100%)	1 (0%)	81	82
2	E	78/79 (99%)	78 (100%)	0	100	100
2	F	79/79 (100%)	79 (100%)	0	100	100
All	All	994/1138 (87%)	985 (99%)	9 (1%)	70	75

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	190	VAL
1	D	176	ILE
1	B	51	CYS
1	B	176	ILE
1	C	51	CYS

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

Mol	Chain	Res	Type
1	A	177	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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	262/305 (85%)	0.07	3 (1%) 78 57	80, 134, 175, 194	0
1	B	269/305 (88%)	0.07	2 (0%) 84 67	75, 143, 197, 229	0
1	C	253/305 (82%)	0.21	8 (3%) 50 35	78, 163, 221, 248	0
1	D	265/305 (86%)	0.24	10 (3%) 44 32	81, 176, 258, 287	0
2	E	93/95 (97%)	0.27	4 (4%) 40 30	89, 118, 151, 168	0
2	F	95/95 (100%)	0.40	4 (4%) 40 30	87, 120, 150, 165	0
3	G	28/28 (100%)	-0.03	0 100 100	111, 137, 181, 183	0
4	H	87/112 (77%)	0.19	0 100 100	102, 227, 283, 392	0
All	All	1352/1550 (87%)	0.17	31 (2%) 61 43	75, 142, 245, 392	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	228	VAL	4.8
1	A	82	GLY	4.3
1	D	192	GLU	4.2
1	D	32	PHE	3.4
1	C	53	MET	3.0

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.