



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 8, 2016 – 05:57 PM EDT

PDB ID : 5KK5  
Title : AsCpf1(E993A)-crRNA-DNA ternary complex  
Authors : Gao, P.; Yang, H.; Rajashankar, K.R.; Huang, Z.; Patel, D.J.  
Deposited on : 2016-06-21  
Resolution : 3.29 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
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-20027939

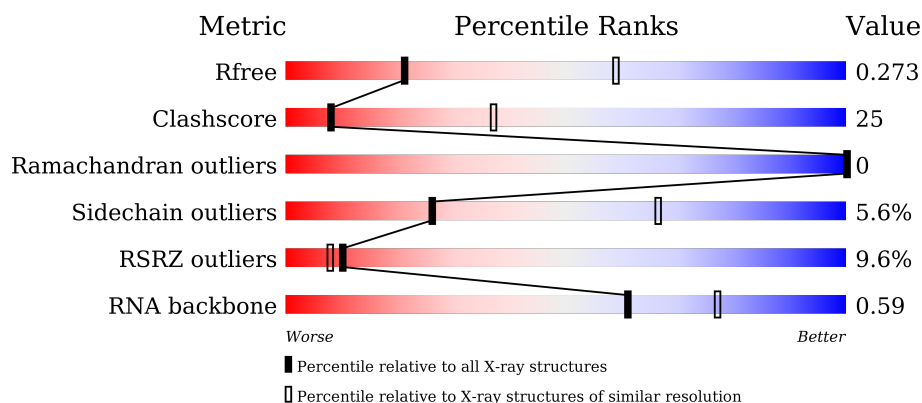
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.29 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	1756 (3.36-3.20)
Clashscore	102246	1941 (3.36-3.20)
Ramachandran outliers	100387	1905 (3.36-3.20)
Sidechain outliers	100360	1903 (3.36-3.20)
RSRZ outliers	91569	1764 (3.36-3.20)
RNA backbone	2183	1022 (3.80-2.76)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	1308	<div> <div>9%</div> <div>52%</div> <div>36%</div> <div>8%</div> </div>
2	B	45	<div> <div>20%</div> <div>60%</div> <div>9%</div> <div>11%</div> </div>
3	C	33	<div> <div>30%</div> <div>55%</div> <div>15%</div> </div>
4	D	8	<div> <div>50%</div> <div>50%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11201 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 Cpf1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1201	Total	C	N	O	S	0	0	0
			9646	6189	1621	1814	22			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP U2UMQ6
A	993	ALA	GLU	engineered mutation	UNP U2UMQ6

- Molecule 2 is a RNA chain called RNA (40-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	40	Total	C	N	O	P	0	0	0
			828	371	143	275	39			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	28	Total	C	N	O	P	0	0	0
			571	274	101	168	28			

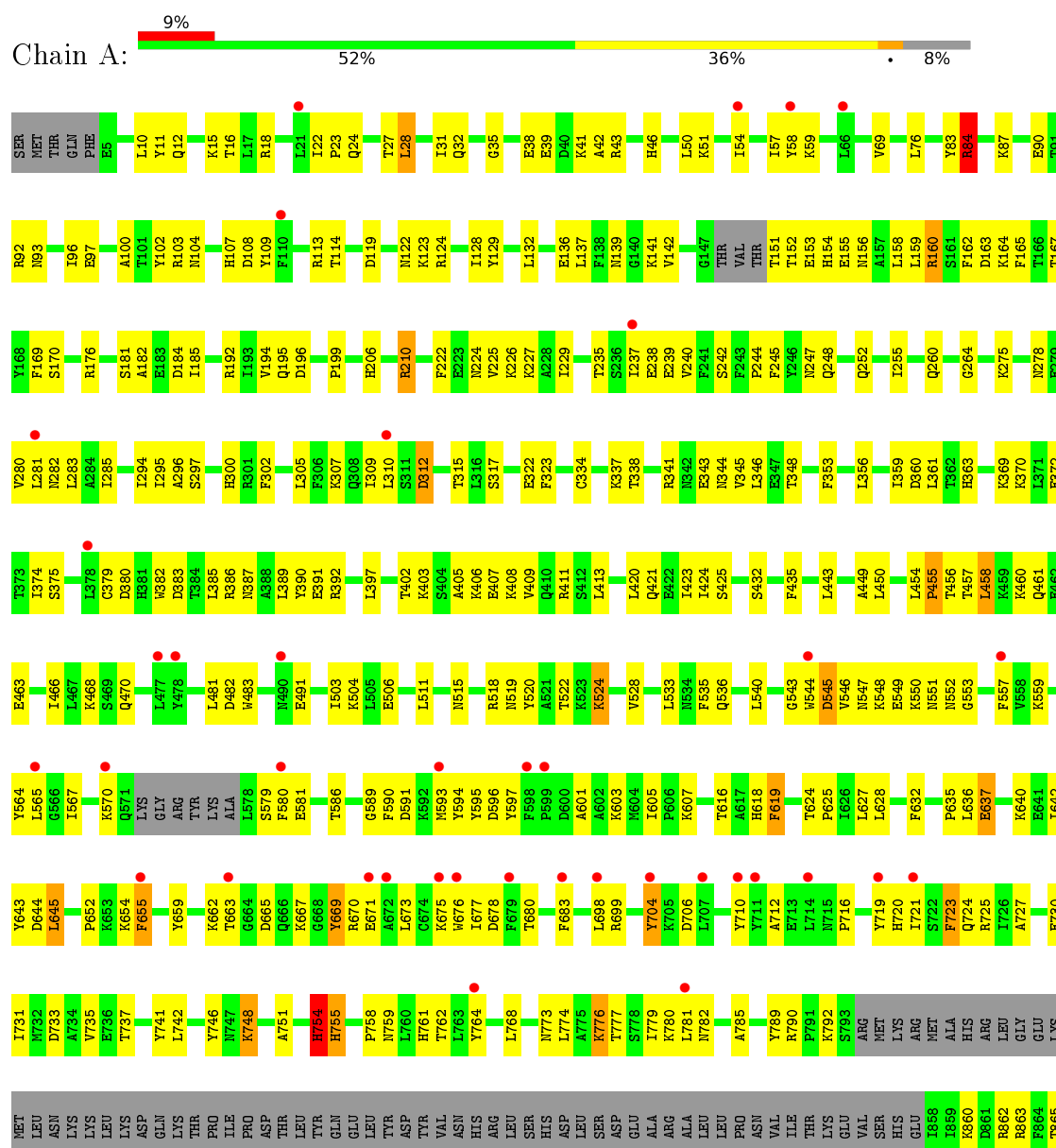
- Molecule 4 is a DNA chain called DNA (8-mer).

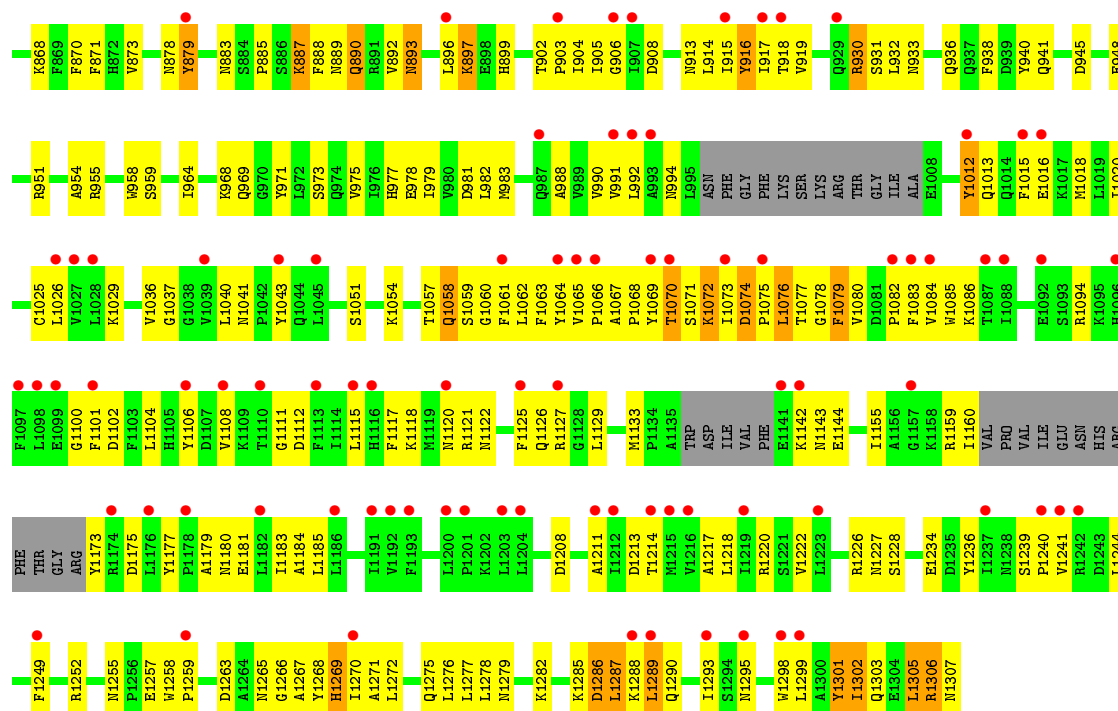
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	8	Total	C	N	O	P	0	0	0
			156	77	24	48	7			

### 3 Residue-property plots

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.

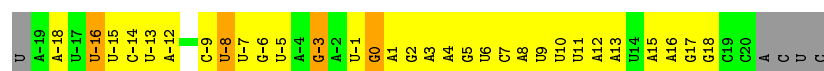
- Molecule 1: CRISPR-associated endonuclease Cpf1





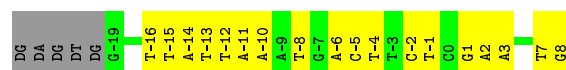
• Molecule 2: RNA (40-MER)

Chain B: 20% 60% 9% 11%



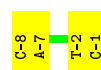
• Molecule 3: DNA (28-MER)

Chain C: 30% 55% 15%



• Molecule 4: DNA (8-mer)

Chain D: 50% 50%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	195.74Å 195.74Å 125.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	105.47 – 3.29 105.47 – 3.29	Depositor EDS
% Data completeness (in resolution range)	99.0 (105.47-3.29) 90.2 (105.47-3.29)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.36 (at 3.26Å)	Xtriage
Refinement program	PHENIX (1.10 _2155: ???)	Depositor
R, $R_{free}$	0.216 , 0.255 0.242 , 0.273	Depositor DCC
$R_{free}$ test set	1812 reflections (5.36%)	DCC
Wilson B-factor (Å <sup>2</sup> )	119.0	Xtriage
Anisotropy	0.160	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 143.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	11201	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	162.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

### 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	0.33	0/9861	0.60	5/13350 (0.0%)
2	B	0.31	0/925	0.94	4/1439 (0.3%)
3	C	0.60	0/639	0.92	0/983
4	D	0.73	0/173	1.17	0/265
All	All	0.36	0/11598	0.67	9/16037 (0.1%)

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
1	A	0	7

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	-8	U	C2-N1-C1'	9.05	128.56	117.70
1	A	28	LEU	CA-CB-CG	7.34	132.18	115.30
1	A	1067	ALA	C-N-CD	5.93	140.86	128.40
2	B	-8	U	C6-N1-C1'	-5.88	112.96	121.20
2	B	-8	U	N3-C2-O2	-5.78	118.16	122.20

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	455	PRO	Peptide
1	A	545	ASP	Peptide
1	A	754	HIS	Peptide
1	A	755	HIS	Peptide

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	84	ARG	Peptide

## 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	9646	0	9423	495	0
2	B	828	0	416	38	0
3	C	571	0	318	19	0
4	D	156	0	90	3	0
All	All	11201	0	10247	528	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 25.

The worst 5 of 528 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:893:ASN:HB3	1:A:896:LEU:HG	1.10	1.09
1:A:1073:ILE:HD11	1:A:1078:GLY:HA2	1.38	1.06
1:A:1070:THR:HG21	1:A:1270:ILE:HG13	1.40	1.03
1:A:1306:ARG:HG3	1:A:1306:ARG:HH11	1.23	1.03
1:A:1079:PHE:CD2	1:A:1241:VAL:HG22	1.96	0.98

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	1187/1308 (91%)	1129 (95%)	58 (5%)	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	1035/1170 (88%)	977 (94%)	58 (6%)	26	66

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	776	LYS
1	A	916	TYR
1	A	1299	LEU
1	A	860	LYS
1	A	887	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	759	ASN
1	A	890	GLN
1	A	977	HIS
1	A	754	HIS
1	A	933	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	38/45 (84%)	6 (15%)	0

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	-18	A
2	B	-16	U
2	B	-7	U
2	B	-6	G
2	B	-3	G

There are no RNA pucker outliers to report.

#### 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 carbohydrates 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	1201/1308 (91%)	0.50	123 (10%) 9 7	35, 155, 289, 390	0
2	B	40/45 (88%)	-0.05	0 100 100	97, 132, 256, 301	0
3	C	28/33 (84%)	-0.33	0 100 100	95, 115, 187, 250	0
4	D	8/8 (100%)	-0.25	0 100 100	139, 149, 214, 261	0
All	All	1277/1394 (91%)	0.46	123 (9%) 10 8	35, 153, 288, 390	0

The worst 5 of 123 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1088	ILE	12.1
1	A	1212	ILE	10.6
1	A	1087	THR	8.4
1	A	1216	VAL	7.2
1	A	1125	PHE	5.9

### 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 carbohydrates 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.