



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 02:53 pm GMT

PDB ID : 3PKM  
Title : Crystal structure of Cas6 with its substrate RNA  
Authors : Wang, R.; Preamplume, G.; Li, H.  
Deposited on : 2010-11-11  
Resolution : 3.10 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  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

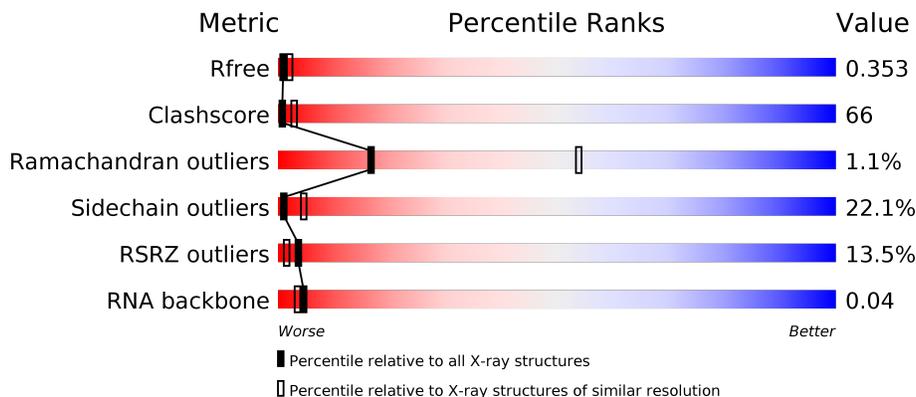
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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))
$R_{free}$	100719	1001 (3.12-3.08)
Clashscore	112137	1099 (3.12-3.08)
Ramachandran outliers	110173	1057 (3.12-3.08)
Sidechain outliers	110143	1057 (3.12-3.08)
RSRZ outliers	101464	1006 (3.12-3.08)
RNA backbone	2435	1112 (3.50-2.70)

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 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	269	
1	X	269	
2	G	10	
3	R	9	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4195 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 Cas6 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	229	1891	1250	309	325	7	0	0	0
1	X	233	1925	1270	318	330	7	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	HIS	-	EXPRESSION TAG	UNP Q8U1S4
A	3	HIS	-	EXPRESSION TAG	UNP Q8U1S4
A	4	HIS	-	EXPRESSION TAG	UNP Q8U1S4
A	5	HIS	-	EXPRESSION TAG	UNP Q8U1S4
A	6	GLY	-	EXPRESSION TAG	UNP Q8U1S4
A	7	SER	-	EXPRESSION TAG	UNP Q8U1S4
X	2	HIS	-	EXPRESSION TAG	UNP Q8U1S4
X	3	HIS	-	EXPRESSION TAG	UNP Q8U1S4
X	4	HIS	-	EXPRESSION TAG	UNP Q8U1S4
X	5	HIS	-	EXPRESSION TAG	UNP Q8U1S4
X	6	GLY	-	EXPRESSION TAG	UNP Q8U1S4
X	7	SER	-	EXPRESSION TAG	UNP Q8U1S4

- Molecule 2 is a RNA chain called 5'-R(\*AP\*UP\*UP\*AP\*CP\*AP\*AP\*UP\*AP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	G	9	189	86	34	60	9	0	0	0

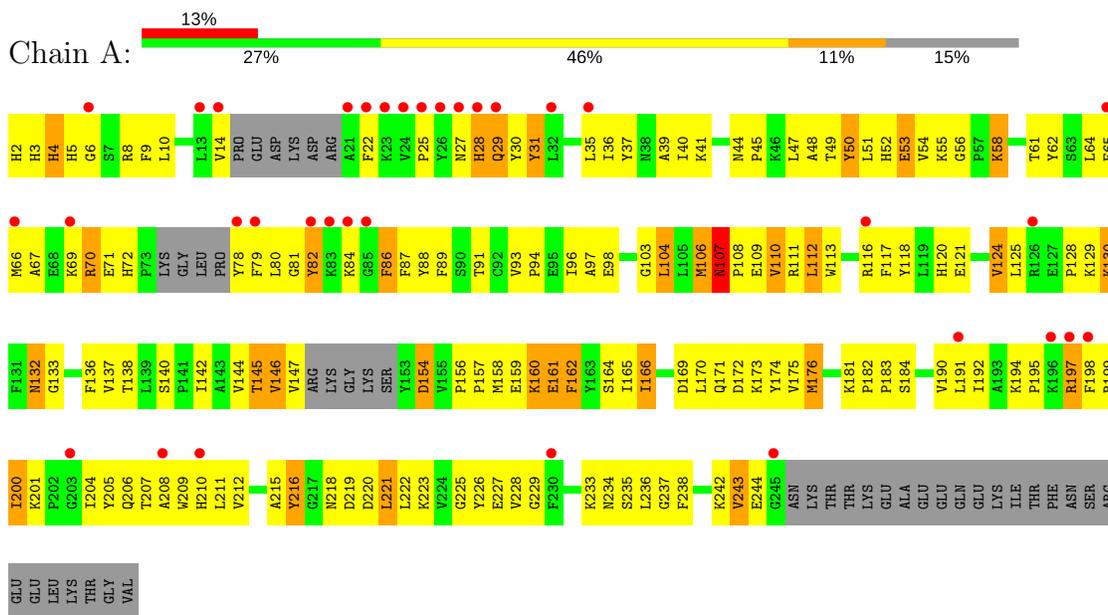
- Molecule 3 is a RNA chain called 5'-R(P\*UP\*UP\*AP\*CP\*AP\*AP\*UP\*AP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	R	9	190	86	34	61	9	0	0	0

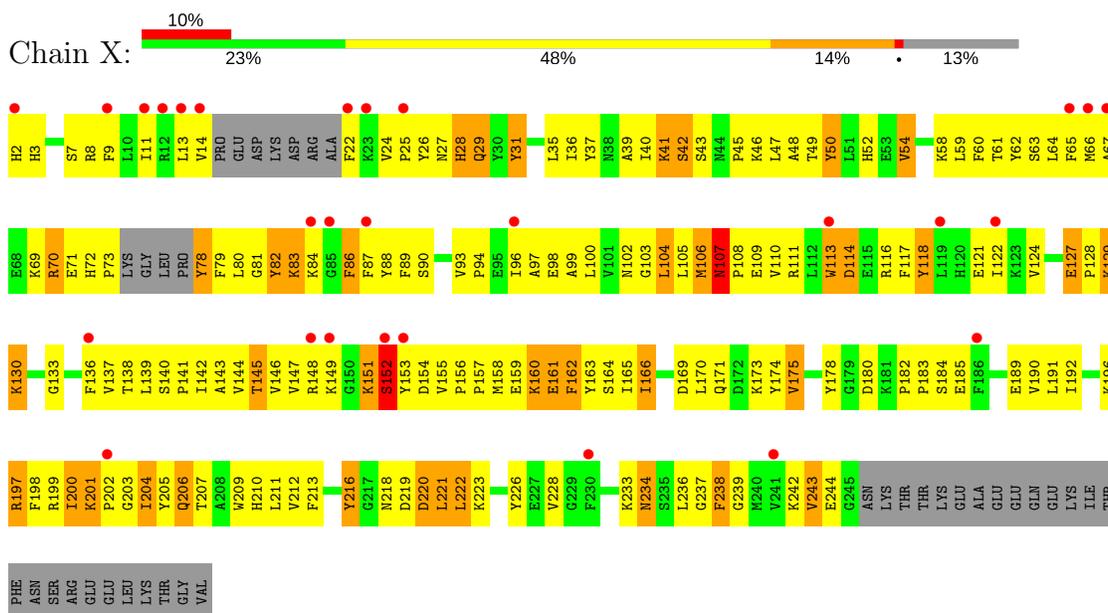
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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: Cas6 protein



- Molecule 1: Cas6 protein

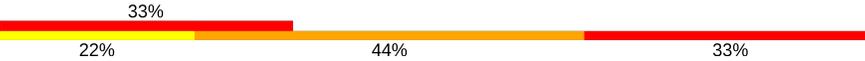


- Molecule 2: 5'-R(\*AP\*UP\*UP\*AP\*CP\*AP\*AP\*UP\*AP\*A)-3'

Chain G: 



- Molecule 3: 5'-R(P\*UP\*UP\*AP\*CP\*AP\*AP\*UP\*AP\*A)-3'

Chain R: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	96.56Å 96.56Å 165.49Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.35 – 3.10 46.35 – 3.10	Depositor EDS
% Data completeness (in resolution range)	69.4 (46.35-3.10) 69.7 (46.35-3.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.81 (at 3.12Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.6.1_357)	Depositor
R, $R_{free}$	0.276 , 0.318 0.317 , 0.353	Depositor DCC
$R_{free}$ test set	1482 reflections (14.15%)	DCC
Wilson B-factor (Å <sup>2</sup> )	80.9	Xtrriage
Anisotropy	0.747	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 115.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.051 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	4195	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	172.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 19.59% 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 [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	0.35	0/1948	0.68	2/2628 (0.1%)
1	X	0.41	0/1983	0.72	3/2673 (0.1%)
2	G	0.62	0/211	1.32	1/325 (0.3%)
3	R	0.70	0/212	1.40	4/327 (1.2%)
All	All	0.42	0/4354	0.80	10/5953 (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
1	X	0	1

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	X	107	ASN	C-N-CD	-5.86	107.70	120.60
1	A	107	ASN	C-N-CD	-5.82	107.80	120.60
3	R	5	C	C6-N1-C2	-5.70	118.02	120.30
3	R	10	A	C8-N9-C4	-5.70	103.52	105.80
1	X	152	SER	N-CA-C	5.56	126.02	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	X	42	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	1891	0	1882	213	0
1	X	1925	0	1925	310	0
2	G	189	0	97	20	0
3	R	190	0	97	43	0
All	All	4195	0	4001	543	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 66.

The worst 5 of 543 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:X:72:HIS:CE1	3:R:6:A:H1'	1.74	1.22
1:A:107:ASN:CB	1:A:108:PRO:HD3	1.62	1.21
1:X:107:ASN:CB	1:X:108:PRO:HD3	1.61	1.21
2:G:2:U:O2'	2:G:3:U:H5'	1.46	1.15
1:X:148:ARG:HD3	1:X:152:SER:OG	1.45	1.15

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	221/269 (82%)	198 (90%)	21 (10%)	2 (1%)	20 60
1	X	227/269 (84%)	200 (88%)	24 (11%)	3 (1%)	14 48

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	448/538 (83%)	398 (89%)	45 (10%)	5 (1%)	17 54

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	107	ASN
1	X	107	ASN
1	X	204	ILE
1	A	234	ASN
1	X	234	ASN

### 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
1	A	202/238 (85%)	161 (80%)	41 (20%)	1 6
1	X	206/238 (87%)	157 (76%)	49 (24%)	1 3
All	All	408/476 (86%)	318 (78%)	90 (22%)	1 4

5 of 90 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	243	VAL
1	X	70	ARG
1	X	210	HIS
1	X	3	HIS
1	X	41	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	132	ASN
1	A	206	GLN
1	X	52	HIS

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	52	HIS
1	X	28	HIS

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	G	8/10 (80%)	7 (87%)	0
3	R	8/9 (88%)	7 (87%)	0
All	All	16/19 (84%)	14 (87%)	0

5 of 14 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	G	3	U
2	G	4	A
2	G	5	C
2	G	6	A
2	G	7	A

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

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	229/269 (85%)	0.78	34 (14%) 3 1	83, 158, 220, 243	0
1	X	233/269 (86%)	0.67	28 (12%) 5 2	76, 160, 221, 257	0
2	G	9/10 (90%)	0.10	0 100 100	159, 211, 239, 251	0
3	R	9/9 (100%)	0.62	3 (33%) 0 0	165, 219, 249, 253	0
All	All	480/557 (86%)	0.71	65 (13%) 3 1	76, 161, 225, 257	0

The worst 5 of 65 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	14	VAL	7.2
1	A	29	GLN	6.8
1	A	21	ALA	6.3
1	X	23	LYS	6.0
1	A	69	LYS	5.8

### 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.