



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 30, 2017 – 07:11 PM EST

PDB ID : 5WTK
Title : Crystal structure of RNP complex
Authors : Liu, L.; Wang, Y.
Deposited on : 2016-12-13
Resolution : 2.65 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 ⓘ 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-20028442
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-20028442

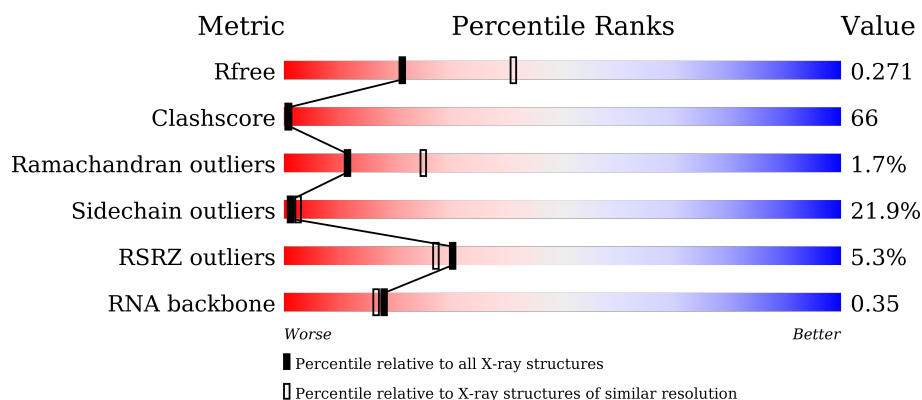
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.65 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	3152 (2.70-2.62)
Clashscore	102246	3524 (2.70-2.62)
Ramachandran outliers	100387	3469 (2.70-2.62)
Sidechain outliers	100360	3469 (2.70-2.62)
RSRZ outliers	91569	3161 (2.70-2.62)
RNA backbone	2183	1001 (3.08-2.24)

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 $\leq 5\%$. 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	1397	
2	B	58	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 10671 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 endoribonuclease C2c2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1215	Total	C	N	O	S	0	0	0
			9757	6260	1633	1846	18			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1390	LEU	-	expression tag	UNP P0DOC6
A	1391	GLU	-	expression tag	UNP P0DOC6
A	1392	HIS	-	expression tag	UNP P0DOC6
A	1393	HIS	-	expression tag	UNP P0DOC6
A	1394	HIS	-	expression tag	UNP P0DOC6
A	1395	HIS	-	expression tag	UNP P0DOC6
A	1396	HIS	-	expression tag	UNP P0DOC6
A	1397	HIS	-	expression tag	UNP P0DOC6

- Molecule 2 is a RNA chain called RNA (58-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	40	Total	C	N	O	P	0	0	0
			849	383	161	266	39			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	60	Total	O	0	0
			60	60		
3	B	5	Total	O	0	0
			5	5		

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 ($\text{RSRZ} > 2$). 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.

- Chain A:

The sequence logo displays the conservation of amino acids across 270 positions. The color-coded regions are: 1-40 (green), 41-120 (yellow), 121-200 (orange), and 201-270 (grey). The horizontal bar at the top shows the overall conservation score, with segments of 5%, 31%, 44%, 12%, and 13%. Amino acids are represented by single-letter codes in colored boxes, with their relative frequency indicated by the height of the box. Red dots above the sequence highlight specific positions of interest.



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	86.59Å 137.05Å 153.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.41 – 2.65 40.41 – 2.65	Depositor EDS
% Data completeness (in resolution range)	98.5 (40.41-2.65) 98.5 (40.41-2.65)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.44 (at 2.65Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.259 , 0.272 0.257 , 0.271	Depositor DCC
R_{free} test set	2620 reflections (4.94%)	DCC
Wilson B-factor (Å ²)	56.2	Xtriage
Anisotropy	0.112	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 44.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	10671	wwPDB-VP
Average B, all atoms (Å ²)	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% 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

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	0.59	0/9906	0.74	14/13336 (0.1%)
2	B	0.90	4/949 (0.4%)	0.83	1/1471 (0.1%)
All	All	0.63	4/10855 (0.0%)	0.75	15/14807 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	16	A	O3'-P	-8.13	1.51	1.61
2	B	15	A	O3'-P	-6.30	1.53	1.61
2	B	14	G	O3'-P	-6.15	1.53	1.61
2	B	13	C	O3'-P	-5.82	1.54	1.61

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	24	A	O5'-P-OP1	-7.24	99.19	105.70
1	A	219	ARG	NE-CZ-NH2	-6.76	116.92	120.30
1	A	4	LEU	CA-CB-CG	6.00	129.10	115.30
1	A	1358	VAL	CB-CA-C	-5.88	100.23	111.40
1	A	1289	ASN	C-N-CD	5.85	140.68	128.40

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	9757	0	9502	1294	0
2	B	849	0	438	78	0
3	A	60	0	0	66	0
3	B	5	0	0	5	0
All	All	10671	0	9940	1347	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 1347 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:25:VAL:HG11	1:A:40:ILE:CD1	1.36	1.52
1:A:898:LEU:HD13	1:A:1059:TYR:CD2	1.56	1.37
1:A:604:ILE:HA	1:A:605:SER:CB	1.55	1.34
1:A:633:GLU:CB	1:A:825:THR:HG21	1.58	1.31
1:A:906:LYS:NZ	1:A:1022:LYS:HA	1.49	1.28

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	1207/1397 (86%)	1100 (91%)	86 (7%)	21 (2%)	11 25

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	211	GLU
1	A	268	ASP
1	A	269	LYS
1	A	694	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	695	THR

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	1024/1332 (77%)	800 (78%)	224 (22%)	1 2

5 of 224 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	630	SER
1	A	736	LEU
1	A	1275	GLU
1	A	646	ASP
1	A	672	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	665	ASN
1	A	734	GLN
1	A	1279	ASN
1	A	726	ASN
1	A	286	ASN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	38/58 (65%)	18 (47%)	6 (15%)

5 of 18 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	3	A
2	B	8	A
2	B	9	A
2	B	14	G
2	B	15	A

5 of 6 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	22	C
2	B	35	U
2	B	29	G
2	B	21	A
2	B	34	A

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, 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	1215/1397 (86%)	0.16	65 (5%) 30 27	28, 69, 112, 156	0
2	B	40/58 (68%)	-0.38	1 (2%) 61 59	32, 55, 111, 123	0
All	All	1255/1455 (86%)	0.14	66 (5%) 30 27	28, 68, 112, 156	0

The worst 5 of 66 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	804	GLU	8.0
1	A	177	ILE	7.6
1	A	1037	SER	7.5
1	A	181	ASP	6.1
1	A	540	ASN	5.2

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.