wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 22, 2023 – 10:26 am BST

PDB ID	:	7PDV
Title	:	Crystal structure of RBM5 RRM1-zinc finger in complex with RNA
Authors	:	Soni, K.; Jagtap, P.K.A.; Sattler, M.
Deposited on	:	2021-08-08
Resolution	:	3.49 Å(reported)

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

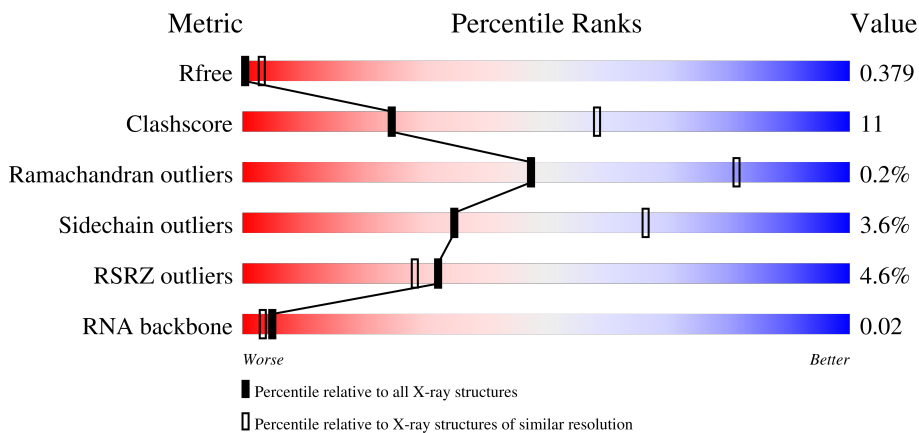|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.49 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.
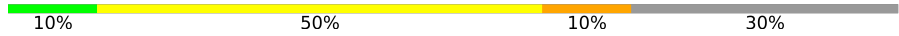
Metric	Percentile Ranks	Value
Rfree		0.379
Clashscore		11
Ramachandran outliers		0.2%
Sidechain outliers		3.6%
RSRZ outliers		4.6%
RNA backbone		0.02

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)
RNA backbone	3102	1002 (4.00-3.00)

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 <=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	119	2% 66% 24% • 10%
1	C	119	5% 67% 21% • 10%
1	E	119	5% 67% 20% • 10%
1	G	119	5% 64% 25% • 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	10	10% 50% 10% 30%

2 Entity composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 3426 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 RNA binding motif protein 5 isoform 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	107	Total 824	C 521	N 148	O 145	S 10	0	0	0
1	C	107	Total 824	C 521	N 148	O 145	S 10	0	0	0
1	E	107	Total 818	C 518	N 145	O 145	S 10	0	0	0
1	G	107	Total 818	C 518	N 145	O 145	S 10	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP A0A0S2Z549
A	2	GLY	-	expression tag	UNP A0A0S2Z549
A	16	THR	ILE	conflict	UNP A0A0S2Z549
A	100	GLY	CYS	conflict	UNP A0A0S2Z549
C	1	MET	-	initiating methionine	UNP A0A0S2Z549
C	2	GLY	-	expression tag	UNP A0A0S2Z549
C	16	THR	ILE	conflict	UNP A0A0S2Z549
C	100	GLY	CYS	conflict	UNP A0A0S2Z549
E	1	MET	-	initiating methionine	UNP A0A0S2Z549
E	2	GLY	-	expression tag	UNP A0A0S2Z549
E	16	THR	ILE	conflict	UNP A0A0S2Z549
E	100	GLY	CYS	conflict	UNP A0A0S2Z549
G	1	MET	-	initiating methionine	UNP A0A0S2Z549
G	2	GLY	-	expression tag	UNP A0A0S2Z549
G	16	THR	ILE	conflict	UNP A0A0S2Z549
G	100	GLY	CYS	conflict	UNP A0A0S2Z549

- Molecule 2 is a RNA chain called RNA (5'-R(P*UP*GP*GP*CP*UP*CP*UP*UP*CP*U)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	7	Total 138	C 61	N 20	O 50	P 7	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Zn 1	0	0
3	C	1	Total 1	Zn 1	0	0
3	E	1	Total 1	Zn 1	0	0
3	G	1	Total 1	Zn 1	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.
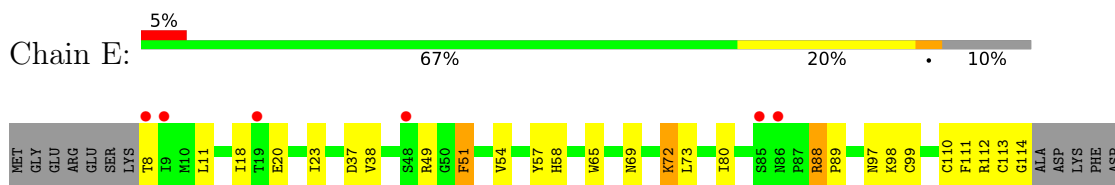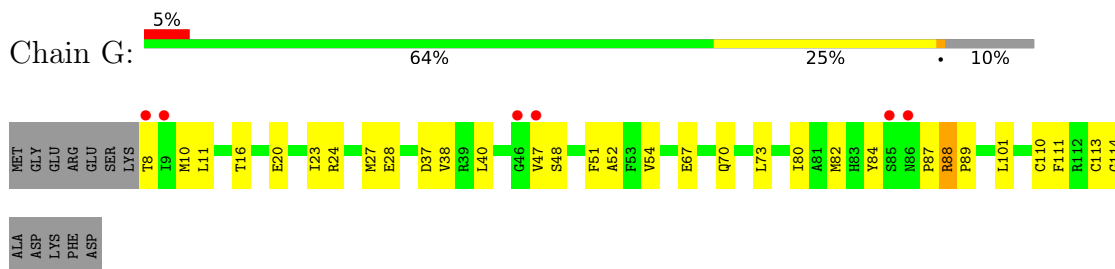
- Molecule 1: RNA binding motif protein 5 isoform 1
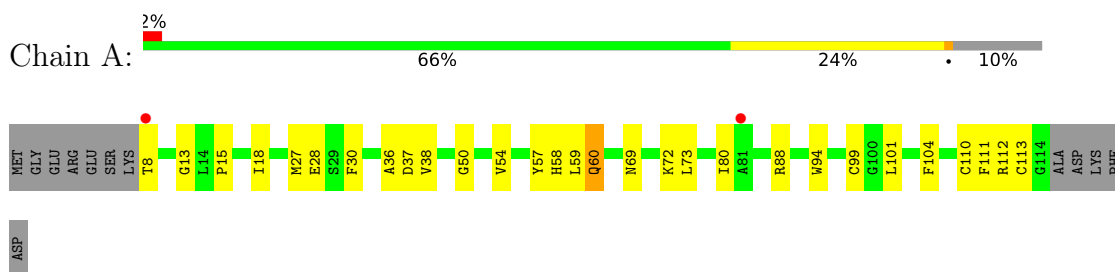
- Molecule 1: RNA binding motif protein 5 isoform 1
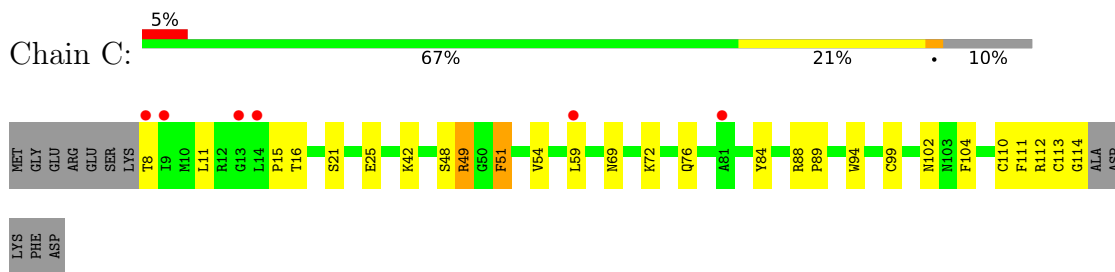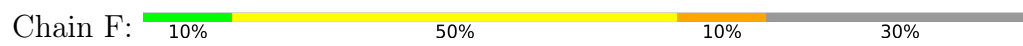

- Molecule 1: RNA binding motif protein 5 isoform 1

- Molecule 1: RNA binding motif protein 5 isoform 1

- Molecule 2: RNA (5'-R(P*UP*GP*GP*CP*UP*CP*UP*UP*CP*U)-3')

Chain F:　10%　　　　　　50%　　　　　　10%　　　　　30%

U2 G3 G4 C5 U6 C7 U8 U C U

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.23Å 65.89Å 94.91Å 90.00° 90.05° 90.00°	Depositor
Resolution (Å)	47.45 – 3.49 47.45 – 3.49	Depositor EDS
% Data completeness (in resolution range)	90.7 (47.45-3.49) 94.0 (47.45-3.49)	Depositor EDS
R_{merge}	0.29	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.55 (at 3.48Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.337 , 0.390 0.330 , 0.379	Depositor DCC
R_{free} test set	537 reflections (6.98%)	wwPDB-VP
Wilson B-factor (Å^2)	67.9	Xtriage
Anisotropy	0.509	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 107.4	EDS
L-test for twinning[2]	$< |L| > = 0.36, < L^2 > = 0.19$	Xtriage
Estimated twinning fraction	0.337 for h,-k,-l	Xtriage
Reported twinning fraction	0.500 for h,-k,-l	Depositor
Outliers	2 of 7688 reflections (0.026%)	Xtriage
F_o,F_c correlation	0.78	EDS
Total number of atoms	3426	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 53.02 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.4771e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.25	0/842	0.52	0/1135
1	C	0.25	0/842	0.51	0/1135
1	E	0.25	0/836	0.52	0/1128
1	G	0.27	0/836	0.53	0/1128
2	F	0.28	0/152	0.91	0/234
All	All	0.26	0/3508	0.54	0/4760

There are no bond length outliers.

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	824	0	789	17	0
1	C	824	0	789	18	0
1	E	818	0	778	21	0
1	G	818	0	779	24	0
2	F	138	0	71	3	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
3	E	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	1	0	0	0	0
All	All	3426	0	3206	72	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 11.

The worst 5 of 72 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:G:11:LEU:O	1:G:51:PHE:HA	1.60	0.99
1:E:11:LEU:O	1:E:51:PHE:HA	1.74	0.87
1:E:65:TRP:O	1:E:69:ASN:ND2	2.16	0.79
2:F:2:U:O2	2:F:3:G:N2	2.23	0.72
1:A:99:CYS:O	1:A:112:ARG:NH2	2.23	0.71

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	105/119 (88%)	94 (90%)	11 (10%)	0	100	100
1	C	105/119 (88%)	94 (90%)	11 (10%)	0	100	100
1	E	105/119 (88%)	94 (90%)	11 (10%)	0	100	100
1	G	105/119 (88%)	86 (82%)	18 (17%)	1 (1%)	15	54
All	All	420/476 (88%)	368 (88%)	51 (12%)	1 (0%)	47	81

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	28	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	84/105 (80%)	81 (96%)	3 (4%)	35	66
1	C	84/105 (80%)	80 (95%)	4 (5%)	25	60
1	E	83/105 (79%)	80 (96%)	3 (4%)	35	66
1	G	83/105 (79%)	81 (98%)	2 (2%)	49	76
All	All	334/420 (80%)	322 (96%)	12 (4%)	35	66

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	51	PHE
1	E	72	LYS
1	G	88	ARG
1	E	88	ARG
1	C	49	ARG

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

Mol	Chain	Res	Type
1	E	69	ASN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	F	6/10 (60%)	5 (83%)	0

All (5) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	F	3	G
2	F	4	G
2	F	5	C

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	F	7	C
2	F	8	U

There are no RNA pucker outliers to report.

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 (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(Å²)	Q<0.9
1	A	107/119 (89%)	0.15	2 (1%)	66	61	14, 31, 50, 61	0
1	C	107/119 (89%)	0.25	6 (5%)	24	22	17, 30, 43, 48	0
1	E	107/119 (89%)	0.11	6 (5%)	24	22	13, 28, 41, 44	0
1	G	107/119 (89%)	0.16	6 (5%)	24	22	16, 31, 47, 67	0
2	F	7/10 (70%)	0.18	0	100	100	19, 30, 35, 36	0
All	All	435/486 (89%)	0.17	20 (4%)	32	28	13, 30, 47, 67	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	8	THR	4.0
1	G	8	THR	4.0
1	E	8	THR	3.6
1	A	8	THR	3.5
1	C	59	LEU	3.3

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 monosaccharides in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum,

median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	ZN	A	201	1/1	0.97	0.12	28,28,28,28	0
3	ZN	G	201	1/1	0.97	0.10	21,21,21,21	0
3	ZN	E	201	1/1	0.98	0.06	16,16,16,16	0
3	ZN	C	201	1/1	0.99	0.06	32,32,32,32	0

6.5 Other polymers ⓘ

There are no such residues in this entry.