wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 12:51 PM EDT

PDB ID	:	7UJ1
Title	:	Crystal structure of PSF-RNA complex
Authors	:	Sachpatzidis, A.; Wang, J.; Konigsberg, W.H.
Deposited on	:	2022-03-30
Resolution	:	3.50 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 (i) 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 (i)) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
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

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 3.50 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		0.243
Clashscore		43
Ramachandran outliers		4.9%
Sidechain outliers		29.1%
RSRZ outliers		4.7%
RNA backbone		0.08

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
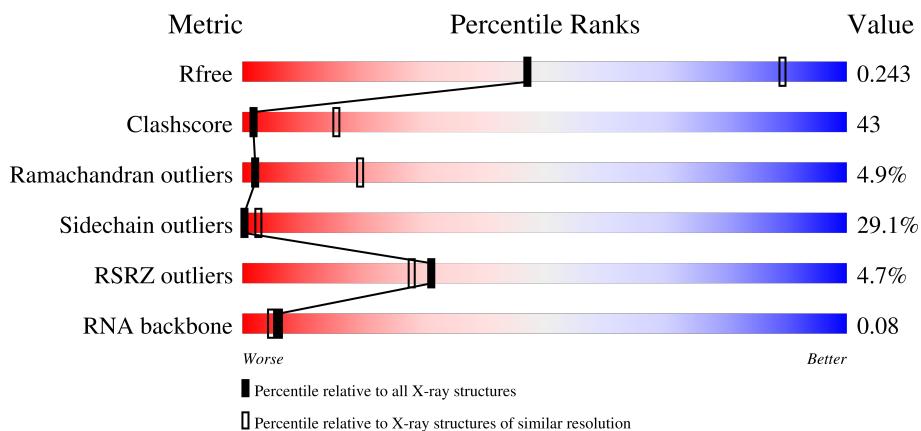
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)
RNA backbone	3102	1002 (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	412	3% 19% 33% 12% • 35%
1	B	412	2% 21% 32% 11% • 34%
2	D	30	3% 17% 17% 67%
2	E	30	7% 10% 17% • 70%
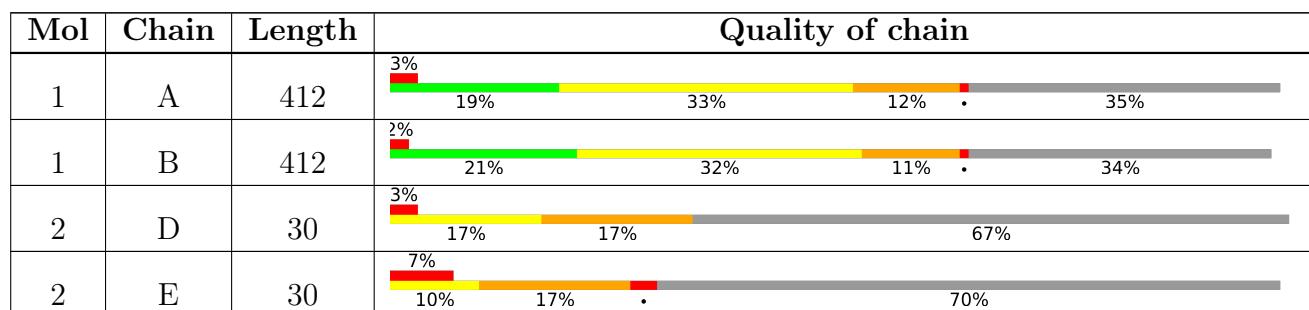
2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 4729 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 Splicing factor, proline- and glutamine-rich.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	268	Total	C	N	O	S	0	0	0
			2155	1351	381	415	8			
1	B	270	Total	C	N	O	S	0	0	0
			2184	1367	392	417	8			

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	187	MET	-	initiating methionine	UNP P23246
A	188	LYS	-	expression tag	UNP P23246
A	189	HIS	-	expression tag	UNP P23246
A	190	HIS	-	expression tag	UNP P23246
A	191	HIS	-	expression tag	UNP P23246
A	192	HIS	-	expression tag	UNP P23246
A	193	HIS	-	expression tag	UNP P23246
A	194	HIS	-	expression tag	UNP P23246
A	195	PRO	-	expression tag	UNP P23246
A	196	MET	-	expression tag	UNP P23246
A	197	SER	-	expression tag	UNP P23246
A	198	ASP	-	expression tag	UNP P23246
A	199	TYR	-	expression tag	UNP P23246
A	200	ASP	-	expression tag	UNP P23246
A	201	ILE	-	expression tag	UNP P23246
A	202	PRO	-	expression tag	UNP P23246
A	203	THR	-	expression tag	UNP P23246
A	204	THR	-	expression tag	UNP P23246
A	205	GLU	-	expression tag	UNP P23246
A	206	ASN	-	expression tag	UNP P23246
A	207	LEU	-	expression tag	UNP P23246
A	208	TYR	-	expression tag	UNP P23246
A	209	PHE	-	expression tag	UNP P23246
A	210	GLN	-	expression tag	UNP P23246
A	211	GLY	-	expression tag	UNP P23246

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	212	ALA	-	expression tag	UNP P23246
A	213	MET	-	expression tag	UNP P23246
B	187	MET	-	initiating methionine	UNP P23246
B	188	LYS	-	expression tag	UNP P23246
B	189	HIS	-	expression tag	UNP P23246
B	190	HIS	-	expression tag	UNP P23246
B	191	HIS	-	expression tag	UNP P23246
B	192	HIS	-	expression tag	UNP P23246
B	193	HIS	-	expression tag	UNP P23246
B	194	HIS	-	expression tag	UNP P23246
B	195	PRO	-	expression tag	UNP P23246
B	196	MET	-	expression tag	UNP P23246
B	197	SER	-	expression tag	UNP P23246
B	198	ASP	-	expression tag	UNP P23246
B	199	TYR	-	expression tag	UNP P23246
B	200	ASP	-	expression tag	UNP P23246
B	201	ILE	-	expression tag	UNP P23246
B	202	PRO	-	expression tag	UNP P23246
B	203	THR	-	expression tag	UNP P23246
B	204	THR	-	expression tag	UNP P23246
B	205	GLU	-	expression tag	UNP P23246
B	206	ASN	-	expression tag	UNP P23246
B	207	LEU	-	expression tag	UNP P23246
B	208	TYR	-	expression tag	UNP P23246
B	209	PHE	-	expression tag	UNP P23246
B	210	GLN	-	expression tag	UNP P23246
B	211	GLY	-	expression tag	UNP P23246
B	212	ALA	-	expression tag	UNP P23246
B	213	MET	-	expression tag	UNP P23246

- Molecule 2 is a RNA chain called RNA (30-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	10	Total	C	N	O	P	0	0	0
			200	90	20	80	10			
2	E	9	Total	C	N	O	P	0	0	0
			180	81	18	72	9			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	6	Total 6	O 6	0	0
3	B	4	Total 4	O 4	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: Splicing factor, proline- and glutamine-rich

Chain A:

• Molecule 1: Splicing factor, proline- and glutamine-rich

Chain B:

L378 S379 P380 Y381 V382 S383 N384 E385 L386 L387 E388 E389 A390 F391 S392 Q393 F394 G395 P396 I397 E398 R399 A400 V401 V402 I403 V404 D405 D406 R407 G408 R409 G412 K413 G414 I415 V416 E417 F418 A419 S420 K421 P422 A423 A424 R425 K426 E429 R430 C431 S432 E433 L437 L438 T439 T440 T441

P442 R443 P444 V445 I446 V447 E448 P449 L450 E451 Q452 L453 D457 G458 L459 P460 E461 K462 L463 A464 Q465 K466 N467 P468 M469 Y470 Q471 K472 E473 R474 E475 T476 P477 P478 R479 F480 A481 Q482 H483 G484 T485 F486 E487 Y488 E489 Y490 S496 L497 M500 E501 K502 Q503 Q504 R505 E506 E509

K510 N511 M512 K513 K516 D517 K518 L519 E520 M523 E524 D525 A526 Y527 Q531 A532 N533 L534 L535 R536 Q537 D538 L539 M540 R541 R542 GLN GLU GLU LEU ARG ARG MET GLU LEU HIS ASN GLN GLU MET GLN LYS ARG LYS GLU MET GLN LEU ARG GLN GLU GLU GLU ARG ARG ARG

ARG GLU GLU GLU MET MET ILE ARG GLN GLU ARG GLU MET GLU GLU GLN MET ARG ARG GLN GLN ARG GLU GLU SER TYR SER

• Molecule 2: RNA (30-MER)

Chain D:

3%

17% 17% 67%

U10 U11 U12 U13 U14 U15 U16 U17 U18 U19 U U U U U U U U U U U U U U U U

• Molecule 2: RNA (30-MER)

Chain E:

7%

10% 17% • 70%

U15 U16 U17 U18 U19 U20 U21 U22 U23 U U U U U U U U U U U U U U U U U U

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 3 2 1	Depositor
Cell constants a, b, c, α, β, γ	186.74Å 186.74Å 55.37Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.67 – 3.50 47.63 – 3.50	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.67-3.50) 100.0 (47.63-3.50)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.48 (at 3.48Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.165 , 0.241 0.166 , 0.243	Depositor DCC
R_{free} test set	711 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	136.5	Xtriage
Anisotropy	0.162	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 151.0	EDS
L-test for twinning [2]	$< |L| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.088 for -h,-k,l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4729	wwPDB-VP
Average B, all atoms (Å^2)	171.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.71% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[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 ⓘ

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	0.44	1/2195 (0.0%)	0.98	4/2948 (0.1%)
1	B	0.47	0/2223	1.08	3/2981 (0.1%)
2	D	0.57	0/219	1.02	0/336
2	E	0.74	1/197 (0.5%)	0.94	0/302
All	All	0.47	2/4834 (0.0%)	1.03	7/6567 (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	1
1	B	0	5
All	All	0	6

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	456	GLU	CD-OE2	7.26	1.33	1.25
2	E	18	U	O3'-P	5.29	1.67	1.61

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	275	SER	CB-CA-C	8.43	126.12	110.10
1	B	288	PRO	CA-N-CD	-8.20	100.02	111.50
1	B	281	ASN	CB-CA-C	8.11	126.61	110.40
1	A	493	ARG	CB-CG-CD	5.31	125.41	111.60
1	A	485	THR	CB-CA-C	-5.26	97.39	111.60

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	437	LEU	Peptide
1	B	273	SER	Peptide
1	B	274	ASP	Peptide
1	B	285	LEU	Peptide
1	B	437	LEU	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	2155	0	2119	207	0
1	B	2184	0	2157	206	0
2	D	200	0	101	15	0
2	E	180	0	91	11	0
3	A	6	0	0	2	0
3	B	4	0	0	0	0
All	All	4729	0	4468	391	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 43.

The worst 5 of 391 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:B:286:ARG:HA	1:B:287:ARG:N	1.47	1.28
1:B:502:LYS:NZ	1:B:505:ARG:HD2	1.63	1.12
1:A:303:ASN:OD1	2:E:18:U:H5'	1.46	1.12
1:B:287:ARG:HB3	1:B:288:PRO:N	1.62	1.11
1:A:305:PRO:CG	1:A:360:ARG:HD2	1.81	1.09

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	264/412 (64%)	218 (83%)	33 (12%)	13 (5%)	2	19
1	B	265/412 (64%)	214 (81%)	38 (14%)	13 (5%)	2	19
All	All	529/824 (64%)	432 (82%)	71 (13%)	26 (5%)	2	19

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	313	PHE
1	A	314	LYS
1	A	485	THR
1	B	419	ALA
1	B	483	HIS

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	229/354 (65%)	158 (69%)	71 (31%)	0	2
1	B	232/354 (66%)	169 (73%)	63 (27%)	0	3
All	All	461/708 (65%)	327 (71%)	134 (29%)	0	2

5 of 134 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	487	GLU

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type
1	B	502	LYS
1	B	538	ASP
1	A	467	ASN
1	A	466	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	537	GLN
1	B	492	GLN
1	B	504	GLN
1	A	504	GLN
1	A	467	ASN

5.3.3   RNA   ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	D	9/30 (30%)	6 (66%)	0
2	E	9/30 (30%)	8 (88%)	4 (44%)
All	All	18/60 (30%)	14 (77%)	4 (22%)

5 of 14 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	D	11	U
2	D	12	U
2	D	14	U
2	D	16	U
2	D	18	U

All (4) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	E	15	U
2	E	19	U
2	E	20	U
2	E	21	U

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 ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	2
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	288:PRO	C	289:GLY	N	3.23
1	B	287:ARG	C	288:PRO	N	2.89
1	B	286:ARG	C	287:ARG	N	2.55

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	268/412 (65%)	0.07	13 (4%)	29	26	111, 166, 243, 441	0
1	B	270/412 (65%)	-0.08	10 (3%)	41	37	82, 139, 241, 297	0
2	D	10/30 (33%)	0.20	1 (10%)	7	8	148, 231, 307, 324	0
2	E	9/30 (30%)	1.35	2 (22%)	0	0	195, 308, 368, 370	0
All	All	557/884 (63%)	0.02	26 (4%)	31	28	82, 154, 260, 441	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	539	LEU	5.7
2	E	18	U	4.8
1	A	535	LEU	4.8
1	A	527	TYR	4.0
1	A	539	LEU	3.5

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.