



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 15, 2022 – 02:16 PM EDT

PDB ID : 7UK1
Title : Complex Structure of Human Polypyrimidine Splicing Factor (PSF/SFPQ) with Murine Virus-like 30S Transcript-1 (VS30-1) Reveals Cooperative Binding of RNA
Authors : Lomakin, I.B.; Wang, J.
Deposited on : 2022-03-31
Resolution : 2.70 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.29
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)	:	2.29

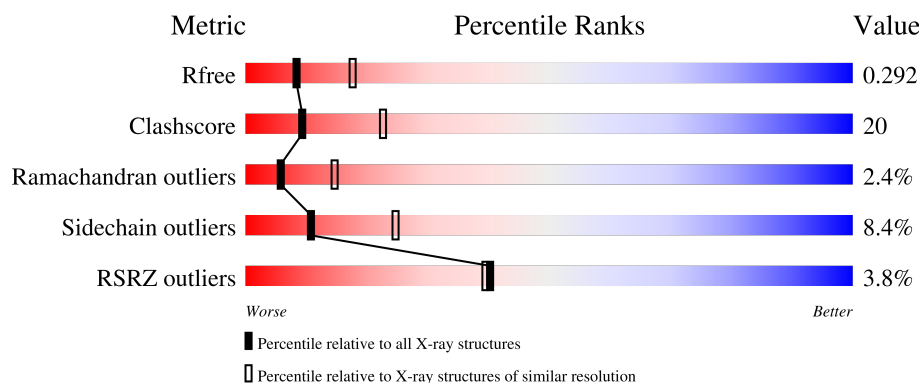
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

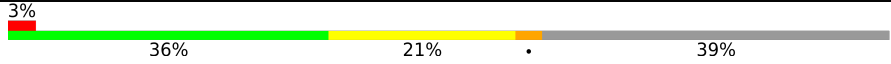
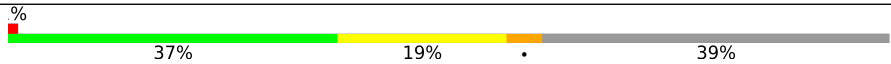
The reported resolution of this entry is 2.70 Å.

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}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 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 $\leq 5\%$. 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	412	
1	B	412	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	MG	A	602	-	-	-	X

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4098 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	250	Total	C	N	O	S	0	1	0
			2030	1274	363	386	7			
1	B	251	Total	C	N	O	S	0	1	0
			2035	1278	361	389	7			

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 MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	3	Total Mg 3 3	0	0
2	B	2	Total Mg 2 2	0	0

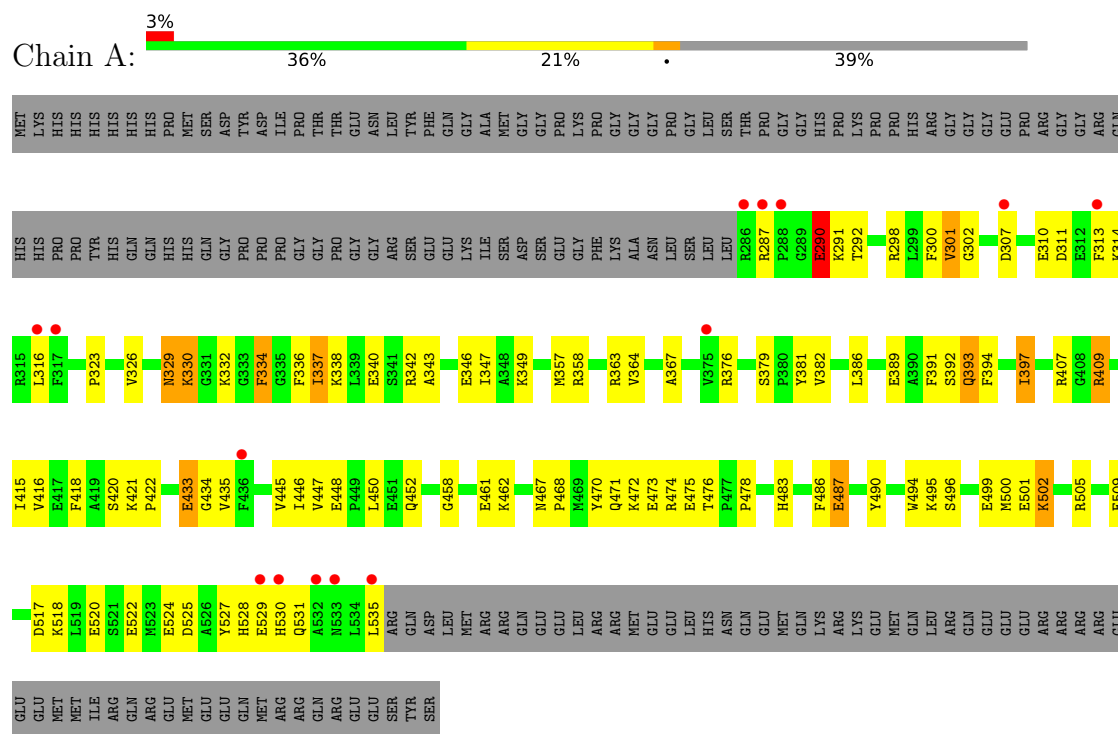
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	15	Total 15	O 15	0	0
3	B	13	Total 13	O 13	0	0

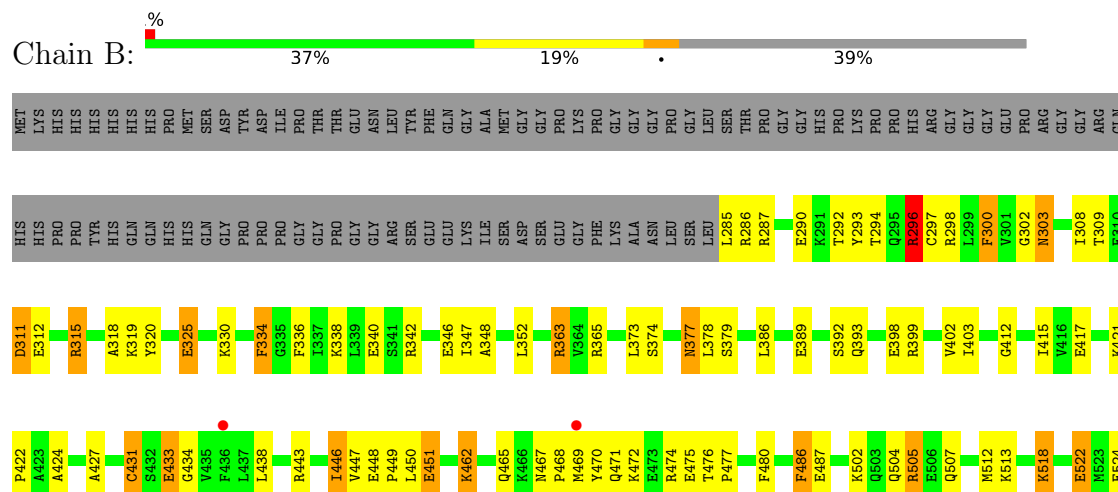
3 Residue-property plots [i](#)

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



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



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

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	61.45Å 64.66Å 68.84Å 90.00° 97.79° 90.00°	Depositor
Resolution (Å)	48.87 – 2.70 48.83 – 2.70	Depositor EDS
% Data completeness (in resolution range)	98.4 (48.87-2.70) 98.4 (48.83-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.39 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.200 , 0.291 0.198 , 0.292	Depositor DCC
R_{free} test set	762 reflections (5.21%)	wwPDB-VP
Wilson B-factor (Å ²)	90.6	Xtriage
Anisotropy	0.538	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 71.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4098	wwPDB-VP
Average B, all atoms (Å ²)	112.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.25% 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 [i](#)

5.1 Standard geometry [i](#)

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

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	0.39	0/2073	0.88	1/2786 (0.0%)
1	B	0.67	3/2078 (0.1%)	1.04	6/2794 (0.2%)
All	All	0.55	3/4151 (0.1%)	0.96	7/5580 (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	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	363	ARG	CZ-NH1	19.36	1.58	1.33
1	B	363	ARG	NE-CZ	-13.10	1.16	1.33
1	B	363	ARG	CZ-NH2	5.09	1.39	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	363	ARG	NE-CZ-NH2	-24.46	108.07	120.30
1	B	363	ARG	NE-CZ-NH1	12.79	126.69	120.30
1	B	311	ASP	CB-CA-C	6.94	124.28	110.40
1	B	365	ARG	CG-CD-NE	-6.69	97.75	111.80
1	B	296	ARG	CG-CD-NE	6.25	124.92	111.80

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	290	GLU	Peptide
1	B	363	ARG	Sidechain

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	2030	0	2006	84	0
1	B	2035	0	2008	89	0
2	A	3	0	0	0	0
2	B	2	0	0	0	0
3	A	15	0	0	0	0
3	B	13	0	0	0	0
All	All	4098	0	4014	160	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 20.

The worst 5 of 160 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:329:ASN:HB3	1:A:334:PHE:CE1	1.92	1.04
1:B:462:LYS:HD2	1:B:462:LYS:H	1.26	0.99
1:A:329:ASN:HB3	1:A:334:PHE:HE1	1.26	0.98
1:B:532:ALA:HA	1:B:535:LEU:HG	1.44	0.96
1:B:312:GLU:HA	1:B:315:ARG:HD2	1.51	0.92

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	249/412 (60%)	215 (86%)	29 (12%)	5 (2%)	7	19
1	B	250/412 (61%)	215 (86%)	28 (11%)	7 (3%)	5	11
All	All	499/824 (61%)	430 (86%)	57 (11%)	12 (2%)	6	15

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	290	GLU
1	A	393	GLN
1	A	487	GLU
1	B	377	ASN
1	B	393	GLN

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	215/354 (61%)	198 (92%)	17 (8%)	12	28
1	B	216/354 (61%)	197 (91%)	19 (9%)	10	23
All	All	431/708 (61%)	395 (92%)	36 (8%)	11	25

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	446	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	524	GLU
1	B	451	GLU
1	B	505	ARG
1	A	483	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	471	GLN
1	A	507	GLN
1	A	511	ASN
1	B	452	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 5 ligands modelled in this entry, 5 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 [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	250/412 (60%)	-0.01	14 (5%) 24 23	63, 104, 174, 312	0
1	B	251/412 (60%)	-0.07	5 (1%) 65 67	69, 102, 166, 256	0
All	All	501/824 (60%)	-0.04	19 (3%) 40 39	63, 103, 170, 312	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	533	ASN	5.8
1	A	533	ASN	4.2
1	A	530	HIS	3.2
1	A	286	ARG	3.1
1	A	532	ALA	3.1

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(\AA^2)	Q<0.9
2	MG	B	601	1/1	0.42	0.19	152,152,152,152	0
2	MG	A	602	1/1	0.61	0.63	124,124,124,124	0
2	MG	B	602	1/1	0.73	0.13	105,105,105,105	0
2	MG	A	603	1/1	0.87	0.34	136,136,136,136	0
2	MG	A	601	1/1	0.91	0.30	141,141,141,141	0

6.5 Other polymers [i](#)

There are no such residues in this entry.