



wwPDB X-ray Structure Validation Summary Report ⓘ

May 8, 2023 – 12:07 PM JST

PDB ID : 8ID2
Title : Crystal structure of the ubiquitin-like domain in the SF3A1 subunit of human U2 snRNP complexed with the stem-loop 4 of U1 snRNA
Authors : Terawaki, S.; Nameki, N.; Kuwasako, K.
Deposited on : 2023-02-12
Resolution : 1.80 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 : 2.32.2
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) : 2.32.2

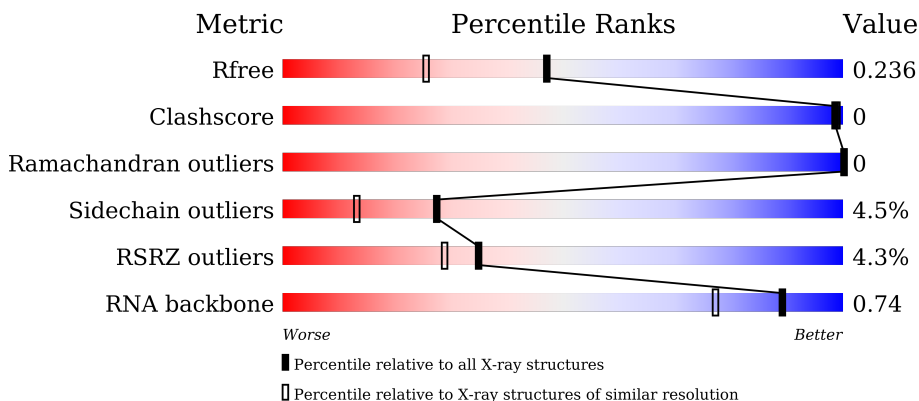
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)
RNA backbone	3102	1060 (2.40-1.20)

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	117	 2% 76% 5% 19%
1	B	117	 7% 74% 1% 24%
2	C	24	 83% 17%
2	D	24	 92% 8%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2540 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 3A subunit 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	95	749	483	129	134	3	0	0	0
1	B	89	685	440	117	125	3	0	0	0

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	677	MET	-	initiating methionine	UNP Q15459
A	678	GLY	-	expression tag	UNP Q15459
A	679	SER	-	expression tag	UNP Q15459
A	680	SER	-	expression tag	UNP Q15459
A	681	HIS	-	expression tag	UNP Q15459
A	682	HIS	-	expression tag	UNP Q15459
A	683	HIS	-	expression tag	UNP Q15459
A	684	HIS	-	expression tag	UNP Q15459
A	685	HIS	-	expression tag	UNP Q15459
A	686	HIS	-	expression tag	UNP Q15459
A	687	SER	-	expression tag	UNP Q15459
A	688	SER	-	expression tag	UNP Q15459
A	689	GLY	-	expression tag	UNP Q15459
A	690	LEU	-	expression tag	UNP Q15459
A	691	VAL	-	expression tag	UNP Q15459
A	692	PRO	-	expression tag	UNP Q15459
A	693	ARG	-	expression tag	UNP Q15459
A	694	GLY	-	expression tag	UNP Q15459
A	695	SER	-	expression tag	UNP Q15459
A	696	HIS	-	expression tag	UNP Q15459
A	697	GLU	-	expression tag	UNP Q15459
A	698	ASN	-	expression tag	UNP Q15459
A	699	LEU	-	expression tag	UNP Q15459
A	700	TYR	-	expression tag	UNP Q15459
A	701	PHE	-	expression tag	UNP Q15459

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	702	GLN	-	expression tag	UNP Q15459
B	677	MET	-	initiating methionine	UNP Q15459
B	678	GLY	-	expression tag	UNP Q15459
B	679	SER	-	expression tag	UNP Q15459
B	680	SER	-	expression tag	UNP Q15459
B	681	HIS	-	expression tag	UNP Q15459
B	682	HIS	-	expression tag	UNP Q15459
B	683	HIS	-	expression tag	UNP Q15459
B	684	HIS	-	expression tag	UNP Q15459
B	685	HIS	-	expression tag	UNP Q15459
B	686	HIS	-	expression tag	UNP Q15459
B	687	SER	-	expression tag	UNP Q15459
B	688	SER	-	expression tag	UNP Q15459
B	689	GLY	-	expression tag	UNP Q15459
B	690	LEU	-	expression tag	UNP Q15459
B	691	VAL	-	expression tag	UNP Q15459
B	692	PRO	-	expression tag	UNP Q15459
B	693	ARG	-	expression tag	UNP Q15459
B	694	GLY	-	expression tag	UNP Q15459
B	695	SER	-	expression tag	UNP Q15459
B	696	HIS	-	expression tag	UNP Q15459
B	697	GLU	-	expression tag	UNP Q15459
B	698	ASN	-	expression tag	UNP Q15459
B	699	LEU	-	expression tag	UNP Q15459
B	700	TYR	-	expression tag	UNP Q15459
B	701	PHE	-	expression tag	UNP Q15459
B	702	GLN	-	expression tag	UNP Q15459

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	24	503	225	84	171	23	0	0	0
2	D	24	503	225	84	171	23	0	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	24	Total	O	0	0
			24	24		

Continued on next page...

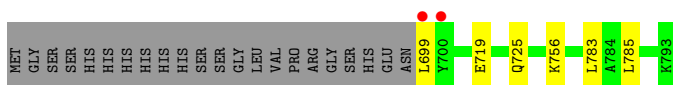
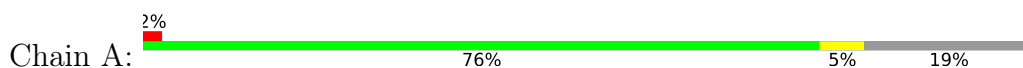
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	30	Total O 30 30	0	0
3	C	27	Total O 27 27	0	0
3	D	19	Total O 19 19	0	0

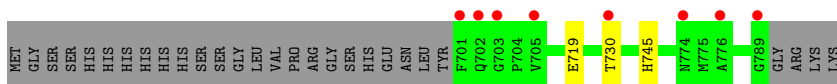
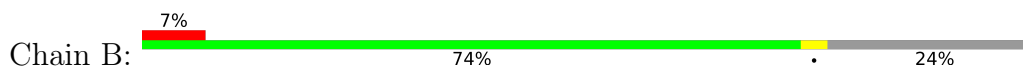
3 Residue-property plots [i](#)

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.

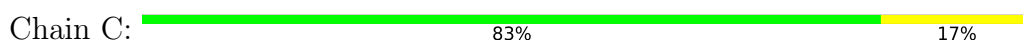
- Molecule 1: Splicing factor 3A subunit 1



- Molecule 1: Splicing factor 3A subunit 1



- Molecule 2: RNA (5'-R(*GP*GP*GP*GP*AP*CP*UP*GP*CP*GP*UP*UP*CP*GP*CP*GP*CP*UP*UP*UP*CP*CP*CP*C)-3')



- Molecule 2: RNA (5'-R(*GP*GP*GP*GP*AP*CP*UP*GP*CP*GP*UP*UP*CP*GP*CP*GP*CP*UP*UP*UP*CP*CP*CP*C)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	54.45Å 60.21Å 103.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.39 – 1.80 40.39 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.1 (40.39-1.80) 99.1 (40.39-1.80)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.18 (at 1.81Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.194 , 0.237 0.192 , 0.236	Depositor DCC
R_{free} test set	1997 reflections (6.23%)	wwPDB-VP
Wilson B-factor (Å ²)	46.7	Xtrriage
Anisotropy	0.452	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 47.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	2540	wwPDB-VP
Average B, all atoms (Å ²)	59.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.53% 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](#)

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	0.54	0/763	0.67	0/1026
1	B	0.49	0/697	0.62	0/942
2	C	0.82	0/559	1.38	3/869 (0.3%)
2	D	0.72	0/559	1.28	2/869 (0.2%)
All	All	0.64	0/2578	1.03	5/3706 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	8	G	O5'-P-OP2	-8.55	98.01	105.70
2	C	10	G	N7-C8-N9	-5.53	110.33	113.10
2	C	11	U	C2-N3-C4	-5.13	123.92	127.00
2	D	9	C	O5'-P-OP2	-5.11	101.10	105.70
2	D	10	G	O5'-P-OP2	-5.11	101.10	105.70

There are no chirality outliers.

There are no planarity outliers.

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	749	0	780	1	0
1	B	685	0	705	1	0
2	C	503	0	260	1	0
2	D	503	0	260	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	24	0	0	0	0
3	B	30	0	0	1	0
3	C	27	0	0	0	0
3	D	19	0	0	0	0
All	All	2540	0	2005	2	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 0.

All (2) 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:745:HIS:HE1	3:B:801:HOH:O	1.85	0.59
1:A:756:LYS:NZ	2:C:13:C:N3	2.66	0.42

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	93/117 (80%)	92 (99%)	1 (1%)	0	100	100
1	B	87/117 (74%)	84 (97%)	3 (3%)	0	100	100
All	All	180/234 (77%)	176 (98%)	4 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	81/100 (81%)	76 (94%)	5 (6%)	18 6
1	B	74/100 (74%)	72 (97%)	2 (3%)	44 31
All	All	155/200 (78%)	148 (96%)	7 (4%)	27 13

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	783	LEU
1	A	785	LEU
1	B	730	THR
1	B	719	GLU
1	A	725	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	23/24 (95%)	0	0
2	D	23/24 (95%)	0	0
All	All	46/48 (95%)	0	0

There are no RNA backbone outliers to report.

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 monosaccharides 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	95/117 (81%)	0.34	2 (2%) 63 59	42, 57, 74, 106	0
1	B	89/117 (76%)	0.43	8 (8%) 9 7	41, 54, 81, 123	0
2	C	24/24 (100%)	-0.35	0 100 100	43, 56, 69, 72	0
2	D	24/24 (100%)	-0.27	0 100 100	48, 64, 76, 80	0
All	All	232/282 (82%)	0.24	10 (4%) 35 29	41, 57, 78, 123	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	701	PHE	11.6
1	A	699	LEU	4.9
1	A	700	TYR	4.0
1	B	730	THR	3.4
1	B	774	ASN	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](#)

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.