Full wwPDB X-ray Structure Validation Report ⓘ

May 28, 2020 – 08:29 pm BST

PDB ID	:	6SQQ
Title	:	Structure of the U1A variant A1-98 Y31H/Q36R/F56W triple mutant in complex with RNA obtained by soaking
Authors	:	Rosenbach, H.; Span, I.
Deposited on	:	2019-09-04
Resolution	:	2.37 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references ⓘ) were used in the production of this report:

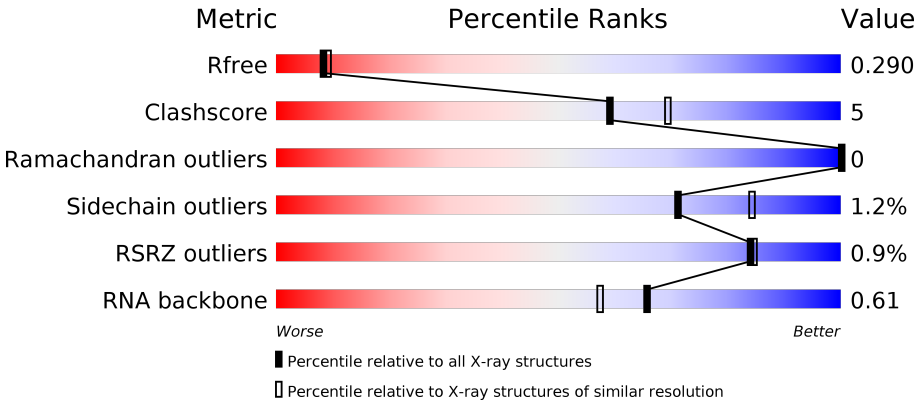
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.37 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.



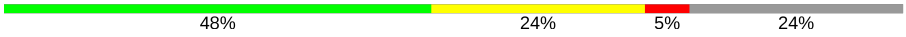
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5509 (2.40-2.36)
Clashscore	141614	6082 (2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)
RSRZ outliers	127900	5397 (2.40-2.36)
RNA backbone	3102	1017 (2.76-2.00)

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 on 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	AAA	98	% 87% 10% •
1	BBB	98	% 82% 11% • 6%
1	CCC	98	% 84% 9% 7%
2	XXX	21	62% 14% 24%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	YYY	21	
2	ZZZ	21	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6194 atoms, of which 2882 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 U1 small nuclear ribonucleoprotein A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	95	Total	C	H	N	O	S	37	0	0
			1574	495	802	137	137	3			
1	BBB	92	Total	C	H	N	O	S	37	1	0
			1554	487	793	137	134	3			
1	CCC	91	Total	C	H	N	O	S	37	0	0
			1512	476	768	132	133	3			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	31	HIS	TYR	engineered mutation	UNP P09012
AAA	36	ARG	GLN	engineered mutation	UNP P09012
AAA	56	TRP	PHE	engineered mutation	UNP P09012
BBB	31	HIS	TYR	engineered mutation	UNP P09012
BBB	36	ARG	GLN	engineered mutation	UNP P09012
BBB	56	TRP	PHE	engineered mutation	UNP P09012
CCC	31	HIS	TYR	engineered mutation	UNP P09012
CCC	36	ARG	GLN	engineered mutation	UNP P09012
CCC	56	TRP	PHE	engineered mutation	UNP P09012

- Molecule 2 is a RNA chain called RNA hairpin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	XXX	16	Total	C	H	N	O	P	16	0	0
			509	152	173	60	109	15			
2	YYY	16	Total	C	H	N	O	P	16	0	0
			509	152	173	60	109	15			
2	ZZZ	16	Total	C	H	N	O	P	16	0	0
			509	152	173	60	109	15			

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	XXX	1	Total Mg 1 1	0	0

- Molecule 4 is water.

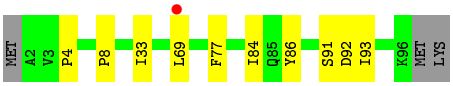
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	10	Total O 10 10	0	0
4	BBB	9	Total O 9 9	0	0
4	CCC	3	Total O 3 3	0	0
4	XXX	2	Total O 2 2	0	0
4	YYY	1	Total O 1 1	0	0
4	ZZZ	1	Total O 1 1	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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: U1 small nuclear ribonucleoprotein A

Chain AAA:

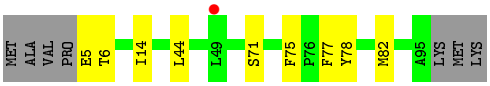


- Molecule 1: U1 small nuclear ribonucleoprotein A

Chain BBB:

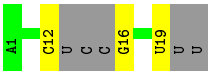
- Molecule 1: U1 small nuclear ribonucleoprotein A

Chain CCC:



- Molecule 2: RNA hairpin

Chain XXX:

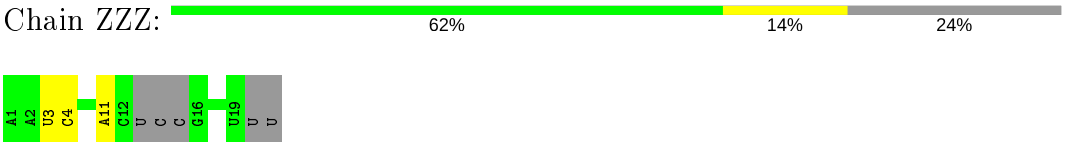


- Molecule 2: RNA hairpin

Chain YYY:

- Molecule 2: RNA hairpin

Chain ZZZ:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	96.08Å 96.08Å 258.01Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	51.03 – 2.37 50.98 – 2.37	Depositor EDS
% Data completeness (in resolution range)	99.9 (51.03-2.37) 100.0 (50.98-2.37)	Depositor EDS
R_{merge}	0.29	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.13 (at 2.37Å)	Xtriage
Refinement program	REFMAC 5.8.0253	Depositor
R, R_{free}	0.245 , 0.294 0.248 , 0.290	Depositor DCC
R_{free} test set	1498 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	46.3	Xtriage
Anisotropy	0.005	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 37.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	6194	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.99% 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

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	AAA	0.67	0/787	0.85	0/1057
1	BBB	0.72	0/778	0.84	0/1042
1	CCC	0.73	1/758 (0.1%)	0.87	0/1017
2	XXX	0.64	0/374	1.16	1/578 (0.2%)
2	YYY	0.71	2/374 (0.5%)	1.02	3/578 (0.5%)
2	ZZZ	0.50	0/374	0.81	0/578
All	All	0.68	3/3445 (0.1%)	0.91	4/4850 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	CCC	5	GLU	CD-OE2	5.50	1.31	1.25
2	YYY	19	U	C3'-O3'	5.33	1.49	1.42
2	YYY	12	C	C3'-O3'	5.06	1.49	1.42

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	XXX	16	G	O5'-P-OP2	-19.23	87.63	110.70
2	YYY	16	G	O5'-P-OP2	-7.83	98.66	105.70
2	YYY	19	U	C2'-C3'-O3'	6.84	124.65	113.70
2	YYY	3	U	O5'-P-OP2	-6.64	99.72	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	AAA	772	802	799	15	0
1	BBB	761	793	791	7	0
1	CCC	744	768	765	8	0
2	XXX	336	173	175	7	0
2	YYY	336	173	175	2	0
2	ZZZ	336	173	175	2	0
3	XXX	1	0	0	0	0
4	AAA	10	0	0	0	0
4	BBB	9	0	0	0	0
4	CCC	3	0	0	0	0
4	XXX	2	0	0	0	0
4	YYY	1	0	0	0	0
4	ZZZ	1	0	0	0	0
All	All	3312	2882	2880	28	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 5.

All (28) 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:AAA:91:SER:HA	2:XXX:12:C:H5	1.46	0.79
1:AAA:91:SER:HA	2:XXX:12:C:C5	2.23	0.74
1:CCC:14:ILE:HD13	1:CCC:82:MET:HE1	1.85	0.57
1:CCC:77:PHE:CZ	1:CCC:78:TYR:HE2	2.25	0.54
1:AAA:92:ASP:H	2:XXX:12:C:H5	1.56	0.54
1:CCC:44:LEU:HD12	2:ZZZ:11:A:H2	1.74	0.51
1:AAA:33:ILE:CD1	1:AAA:77:PHE:CE1	2.93	0.51
1:AAA:69:LEU:HD12	1:AAA:84:ILE:HG22	1.94	0.50
1:AAA:93:ILE:HD12	1:AAA:93:ILE:H	1.77	0.50
2:ZZZ:3:U:H2'	2:ZZZ:4:C:C6	2.47	0.49
1:BBB:31:HIS:O	1:BBB:35:SER:HB3	2.13	0.49
1:AAA:91:SER:CA	2:XXX:12:C:H5	2.19	0.48
1:AAA:92:ASP:N	2:XXX:12:C:H5	2.11	0.48
1:AAA:4:PRO:HB3	1:CCC:75:PHE:CE1	2.49	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:BBB:39:GLN:OE1	1:BBB:60:LYS:HE3	2.14	0.47
2:YYY:18:A:H2'	2:YYY:19:U:C6	2.50	0.47
1:AAA:33:ILE:HD11	1:BBB:6:THR:HG21	1.95	0.47
1:AAA:92:ASP:N	2:XXX:12:C:C5	2.81	0.46
1:CCC:77:PHE:CZ	1:CCC:78:TYR:CE2	3.03	0.46
1:AAA:69:LEU:HD21	1:CCC:71:SER:HB3	2.00	0.43
1:BBB:33:ILE:HG12	1:CCC:6:THR:HG21	2.01	0.42
2:XXX:12:C:O2	2:XXX:12:C:O5'	2.37	0.42
1:AAA:33:ILE:HD11	1:AAA:77:PHE:CE1	2.56	0.41
1:BBB:75:PHE:CD1	1:BBB:76:PRO:HD2	2.56	0.41
1:AAA:69:LEU:CD2	1:CCC:71:SER:HB3	2.51	0.41
1:BBB:33:ILE:HD12	1:BBB:77:PHE:CE1	2.56	0.40
1:BBB:52:ARG:NH1	2:YYY:6:A:N1	2.70	0.40
1:AAA:8:PRO:HA	1:AAA:86:TYR:CE1	2.56	0.40

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	AAA	93/98 (95%)	91 (98%)	2 (2%)	0	100	100
1	BBB	91/98 (93%)	89 (98%)	2 (2%)	0	100	100
1	CCC	89/98 (91%)	87 (98%)	2 (2%)	0	100	100
All	All	273/294 (93%)	267 (98%)	6 (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	AAA	85/88 (97%)	85 (100%)	0	100 100
1	BBB	84/88 (96%)	81 (96%)	3 (4%)	35 51
1	CCC	82/88 (93%)	82 (100%)	0	100 100
All	All	251/264 (95%)	248 (99%)	3 (1%)	71 84

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	35	SER
1	BBB	67	ASN
1	BBB	96	LYS

Some 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	XXX	14/21 (66%)	1 (7%)	0
2	YYY	14/21 (66%)	1 (7%)	0
2	ZZZ	14/21 (66%)	0	0
All	All	42/63 (66%)	2 (4%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	XXX	19	U
2	YYY	19	U

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is 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

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, 95^{th} 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	AAA	95/98 (96%)	0.18	1 (1%)	80	81	33, 41, 63, 88	0
1	BBB	92/98 (93%)	0.18	1 (1%)	80	81	29, 41, 60, 97	0
1	CCC	91/98 (92%)	0.18	1 (1%)	80	81	32, 42, 60, 67	0
2	XXX	16/21 (76%)	-0.37	0	100	100	42, 44, 55, 57	0
2	YYY	16/21 (76%)	-0.31	0	100	100	41, 46, 54, 55	0
2	ZZZ	16/21 (76%)	0.20	0	100	100	46, 61, 85, 95	0
All	All	326/357 (91%)	0.13	3 (0%)	84	84	29, 42, 65, 97	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	69	LEU	2.2
1	CCC	49	LEU	2.1
1	BBB	95	ALA	2.0

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 carbohydrates in this entry.

6.4 Ligands

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, 95^{th} 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	MG	XXX	101	1/1	0.97	0.21	49,49,49,49	0

6.5 Other polymers (i)

There are no such residues in this entry.