Full wwPDB X-ray Structure Validation Report ⓘ

May 28, 2020 – 08:10 pm BST

PDB ID	:	6SR7
Title	:	Structure of the U1A variant A1-98 Y31H/Q36R/K98W
Authors	:	Rosenbach, H.; Span, I.
Deposited on	:	2019-09-05
Resolution	:	1.86 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:

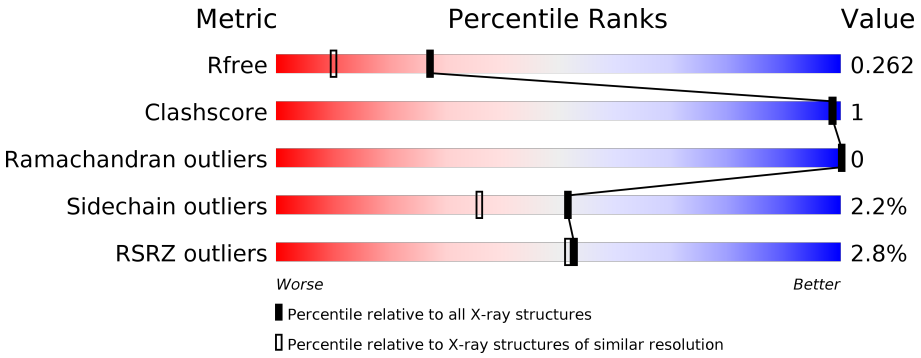
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.86 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	5% 92% • 7%
1	BBB	98	% 91% •• 7%
1	CCC	98	% 88% 5% 7%
1	DDD	98	3% 78% 5% 17%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6015 atoms, of which 3020 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.

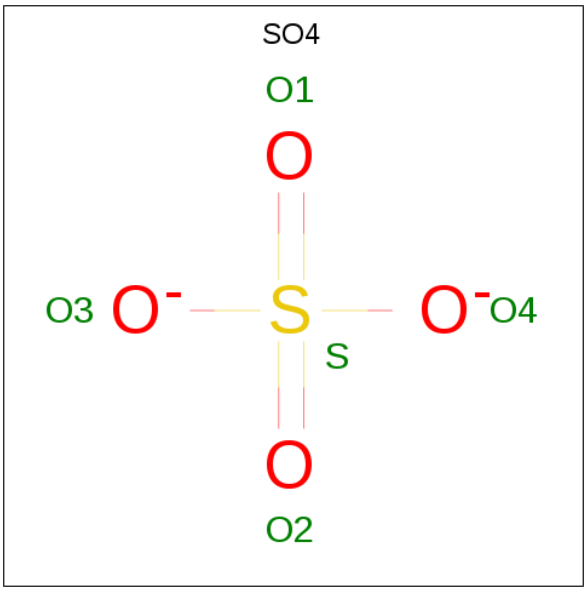
- Molecule 1 is a protein called U1 small nuclear ribonucleoprotein A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	91	Total	C	H	N	O	S	38	1	0
			1533	486	780	132	131	4			
1	BBB	91	Total	C	H	N	O	S	36	0	0
			1519	481	773	130	131	4			
1	CCC	91	Total	C	H	N	O	S	36	0	0
			1519	481	773	130	131	4			
1	DDD	81	Total	C	H	N	O	S	34	0	0
			1358	429	694	118	114	3			

There are 12 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	98	TRP	LYS	engineered mutation	UNP P09012
BBB	31	HIS	TYR	engineered mutation	UNP P09012
BBB	36	ARG	GLN	engineered mutation	UNP P09012
BBB	98	TRP	LYS	engineered mutation	UNP P09012
CCC	31	HIS	TYR	engineered mutation	UNP P09012
CCC	36	ARG	GLN	engineered mutation	UNP P09012
CCC	98	TRP	LYS	engineered mutation	UNP P09012
DDD	31	HIS	TYR	engineered mutation	UNP P09012
DDD	36	ARG	GLN	engineered mutation	UNP P09012
DDD	98	TRP	LYS	engineered mutation	UNP P09012

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	CCC	1	Total O S 5 4 1	0	0

- Molecule 3 is water.

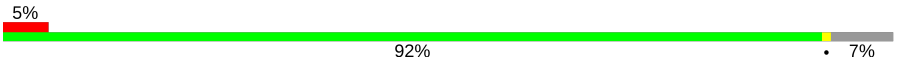
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	24	Total O 24 24	0	0
3	BBB	19	Total O 19 19	0	0
3	CCC	23	Total O 23 23	0	0
3	DDD	15	Total O 15 15	0	0

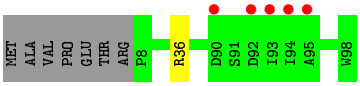
3 Residue-property plots (i)

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

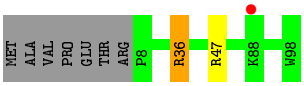
Chain AAA:





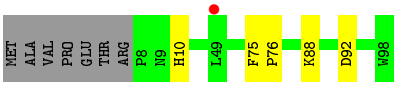
- Molecule 1: U1 small nuclear ribonucleoprotein A

Chain BBB:



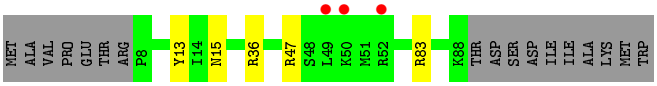
- Molecule 1: U1 small nuclear ribonucleoprotein A

Chain CCC:



- Molecule 1: U1 small nuclear ribonucleoprotein A

Chain DDD:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	76.26Å 76.26Å 151.48Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	68.11 – 1.86 68.11 – 1.86	Depositor EDS
% Data completeness (in resolution range)	99.9 (68.11-1.86) 100.0 (68.11-1.86)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.28 (at 1.86Å)	Xtriage
Refinement program	REFMAC 5.8.0253	Depositor
R, R_{free}	0.221 , 0.256 0.229 , 0.262	Depositor DCC
R_{free} test set	1920 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	25.4	Xtriage
Anisotropy	0.103	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.43 , 30.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6015	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.04% 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: SO4

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	0.74	0/772	0.86	0/1032
1	BBB	0.72	0/761	0.83	0/1017
1	CCC	0.70	0/761	0.87	0/1017
1	DDD	0.67	0/677	0.87	0/904
All	All	0.71	0/2971	0.86	0/3970

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	AAA	753	780	778	0	0
1	BBB	746	773	771	0	1
1	CCC	746	773	771	2	1
1	DDD	664	694	692	1	0
2	CCC	5	0	0	0	0
3	AAA	24	0	0	0	0
3	BBB	19	0	0	0	0
3	CCC	23	0	0	0	0
3	DDD	15	0	0	0	0
All	All	2995	3020	3012	3	1

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 1.

All (3) 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:CCC:88:LYS:HB2	1:CCC:92:ASP:OD2	2.19	0.43
1:DDD:13:TYR:CZ	1:DDD:15:ASN:HB3	2.55	0.42
1:CCC:75:PHE:HA	1:CCC:76:PRO:HD3	1.94	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:BBB:36:ARG:HD2	1:CCC:76:PRO:HB2[8_665]	1.29	0.31

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	AAA	90/98 (92%)	90 (100%)	0	0	100	100
1	BBB	89/98 (91%)	87 (98%)	2 (2%)	0	100	100
1	CCC	89/98 (91%)	88 (99%)	1 (1%)	0	100	100
1	DDD	79/98 (81%)	79 (100%)	0	0	100	100
All	All	347/392 (88%)	344 (99%)	3 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	AAA	83/88 (94%)	82 (99%)	1 (1%)	71	62
1	BBB	82/88 (93%)	80 (98%)	2 (2%)	49	33
1	CCC	82/88 (93%)	81 (99%)	1 (1%)	71	62
1	DDD	73/88 (83%)	70 (96%)	3 (4%)	30	13
All	All	320/352 (91%)	313 (98%)	7 (2%)	52	36

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	36	ARG
1	BBB	36	ARG
1	BBB	47	ARG
1	CCC	10	HIS
1	DDD	36	ARG
1	DDD	47	ARG
1	DDD	83	ARG

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

5.3.3 RNA

There are no RNA molecules in this entry.

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 carbohydrates in this entry.

5.6 Ligand geometry

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	SO4	CCC	101	-	4,4,4	0.21	0	6,6,6	0.16	0

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	91/98 (92%)	0.55	5 (5%)	25	24	20, 28, 47, 49	0
1	BBB	91/98 (92%)	0.40	1 (1%)	80	81	18, 27, 45, 64	0
1	CCC	91/98 (92%)	0.30	1 (1%)	80	81	17, 24, 43, 55	0
1	DDD	81/98 (82%)	0.43	3 (3%)	41	39	19, 27, 61, 83	0
All	All	354/392 (90%)	0.42	10 (2%)	53	52	17, 27, 48, 83	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	93	ILE	5.2
1	DDD	49	LEU	4.4
1	AAA	94	ILE	4.4
1	DDD	50	LYS	4.1
1	AAA	92	ASP	4.0
1	CCC	49	LEU	3.6
1	DDD	52	ARG	3.1
1	AAA	95	ALA	3.1
1	AAA	90	ASP	2.6
1	BBB	88	LYS	2.5

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 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
2	SO4	CCC	101	5/5	0.95	0.11	37,39,40,42	0

6.5 Other polymers

There are no such residues in this entry.