Full wwPDB X-ray Structure Validation Report (i)

Jun 24, 2024 – 05:22 PM EDT

PDB ID	:	5QYV
Title	:	PanDDA analysis group deposition – Auto-refined data of Aar2/RNaseH for ground state model 11
Authors	:	Weiss, M.S.; Wollenhaupt, J.; Metz, A.; Barthel, T.; Lima, G.M.A.; Heine, A.; Mueller, U.; Klebe, G.
Deposited on	:	2020-02-12
Resolution	:	1.59 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 (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 (1)) were used in the production of this report:

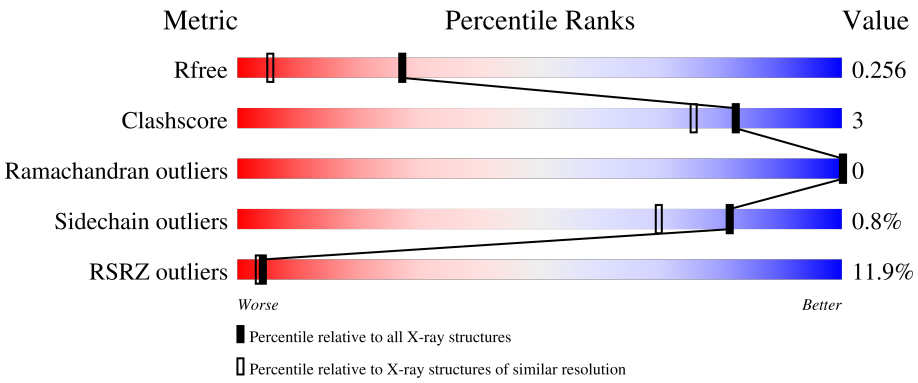
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.59 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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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	258	8%, 85%, 6% • 8%
2	B	308	14%, 89%, 8% •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4612 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 Pre-mRNA-splicing factor 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	237	2002	1283	335	372	12	0	12	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1833	GLY	-	expression tag	UNP P33334
A	1834	ALA	-	expression tag	UNP P33334
A	1835	MET	-	expression tag	UNP P33334

- Molecule 2 is a protein called A1 cistron-splicing factor AAR2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	300	2580	1654	421	485	20	0	9	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	expression tag	UNP P32357
B	-2	ALA	-	expression tag	UNP P32357
B	-1	MET	-	expression tag	UNP P32357
B	0	ALA	-	expression tag	UNP P32357
B	166	SER	-	linker	UNP P32357
B	167	SER	-	linker	UNP P32357
B	168	SER	-	linker	UNP P32357
B	169	SER	-	linker	UNP P32357
B	170	SER	-	linker	UNP P32357

- Molecule 3 is water.

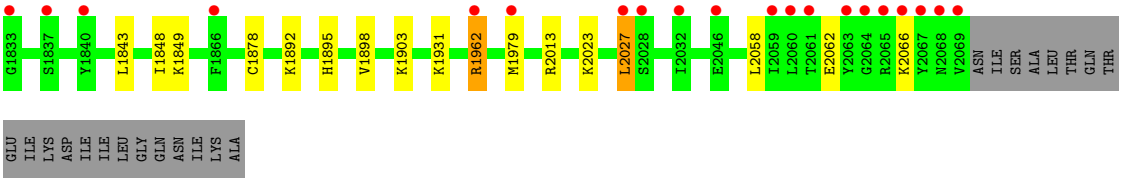
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	16	16	16	0	0
			Total	O		
3	B	14	14	14	0	0

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.

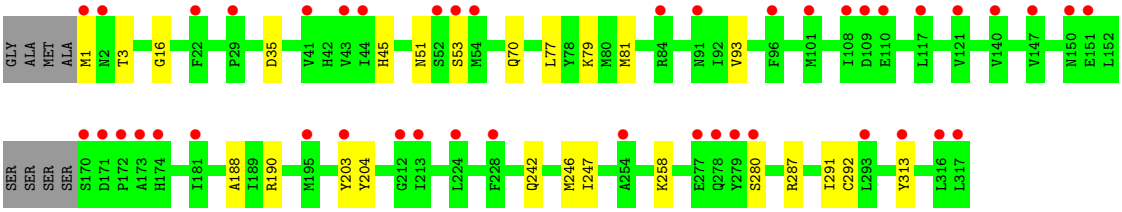
• Molecule 1: Pre-mRNA-splicing factor 8

Chain A:



• Molecule 2: A1 cistron-splicing factor AAR2

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	89.17Å 81.88Å 93.45Å 90.00° 108.56° 90.00°	Depositor
Resolution (Å)	44.01 – 1.59 43.97 – 1.59	Depositor EDS
% Data completeness (in resolution range)	99.5 (44.01-1.59) 99.5 (43.97-1.59)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.94 (at 1.58Å)	Xtriage
Refinement program	REFMAC 5.8.0238, PHENIX 1.16.3549	Depositor
R, R_{free}	0.244 , 0.245 0.253 , 0.256	Depositor DCC
R_{free} test set	4164 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	34.1	Xtriage
Anisotropy	0.167	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 33.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4612	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.58% 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 (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	0.71	0/2049	0.77	0/2775
2	B	0.68	0/2651	0.76	0/3581
All	All	0.69	0/4700	0.77	0/6356

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 (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	2002	0	2029	8	0
2	B	2580	0	2448	15	0
3	A	16	0	0	0	0
3	B	14	0	0	0	0
All	All	4612	0	4477	23	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 3.

All (23) 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:2062:GLU:O	1:A:2066:LYS:HG2	1.88	0.73

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:1:MET:HB3	2:B:35:ASP:HA	1.79	0.62
2:B:280:SER:HB3	2:B:313:TYR:CE1	2.37	0.59
2:B:70:GLN:HB3	2:B:81:MET:HE1	1.84	0.58
2:B:16:GLY:HA3	2:B:45:HIS:CE1	2.39	0.57
1:A:1962:ARG:O	1:A:2013:ARG:NH1	2.39	0.55
2:B:77:LEU:HD21	2:B:79:LYS:HE3	1.88	0.54
2:B:3:THR:HG21	2:B:93:VAL:HG13	1.96	0.48
1:A:1843:LEU:HA	1:A:1849:LYS:HD2	1.96	0.47
1:A:2058:LEU:C	1:A:2058:LEU:HD23	2.34	0.47
2:B:70:GLN:HB3	2:B:81:MET:CE	2.44	0.47
2:B:258:LYS:HD2	2:B:258:LYS:H	1.78	0.47
1:A:1848:ILE:H	1:A:1931[A]:LYS:HZ2	1.62	0.46
2:B:287:ARG:O	2:B:291:ILE:HD13	2.16	0.46
1:A:2023:LYS:O	1:A:2027:LEU:HG	2.17	0.45
2:B:247:ILE:HG22	2:B:292:CYS:SG	2.59	0.43
1:A:1895:HIS:O	1:A:1898[A]:VAL:HG22	2.18	0.43
2:B:291:ILE:HG22	2:B:292:CYS:N	2.33	0.43
2:B:190:ARG:HG3	2:B:203[B]:TYR:CZ	2.55	0.41
2:B:188:ALA:HA	2:B:204:TYR:CD1	2.56	0.41
1:A:1878:CYS:HA	1:A:1892:LYS:O	2.20	0.41
2:B:242:GLN:O	2:B:246:MET:HG3	2.22	0.40
2:B:51:ASN:C	2:B:53:SER:H	2.23	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	A	248/258 (96%)	244 (98%)	4 (2%)	0	100	100
2	B	306/308 (99%)	295 (96%)	11 (4%)	0	100	100
All	All	554/566 (98%)	539 (97%)	15 (3%)	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	226/233 (97%)	220 (97%)	6 (3%)	44	20
2	B	287/284 (101%)	287 (100%)	0	100	100
All	All	513/517 (99%)	507 (99%)	6 (1%)	81	54

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1903	LYS
1	A	1962	ARG
1	A	1979[A]	MET
1	A	1979[B]	MET
1	A	1979[C]	MET
1	A	2027	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	150	ASN

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)

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	237/258 (91%)	0.72	20 (8%)	11	9	23, 33, 65, 91	0
2	B	300/308 (97%)	0.93	44 (14%)	2	2	24, 36, 77, 90	0
All	All	537/566 (94%)	0.84	64 (11%)	4	3	23, 36, 71, 91	0

All (64) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	53	SER	9.6
1	A	2027	LEU	8.6
2	B	52	SER	8.5
2	B	1	MET	7.8
1	A	2063	TYR	5.6
1	A	1833	GLY	5.4
2	B	170	SER	4.9
2	B	173	ALA	4.7
2	B	279	TYR	4.6
2	B	54[A]	MET	4.5
2	B	172	PRO	4.3
1	A	2028	SER	4.3
1	A	2060	LEU	4.2
2	B	22	PHE	4.0
2	B	109	ASP	3.8
1	A	2064	GLY	3.7
2	B	101	MET	3.6
2	B	96	PHE	3.6
2	B	195	MET	3.3
1	A	2066	LYS	3.1
2	B	313	TYR	3.1
2	B	147	VAL	3.0
1	A	1962	ARG	2.8
2	B	213	ILE	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	2061	THR	2.8
1	A	2046	GLU	2.7
2	B	280	SER	2.7
2	B	91	ASN	2.7
2	B	212	GLY	2.6
1	A	2069	VAL	2.6
1	A	2067	TYR	2.6
2	B	277	GLU	2.5
2	B	171	ASP	2.5
2	B	84	ARG	2.5
2	B	150	ASN	2.5
2	B	316	LEU	2.4
2	B	29	PRO	2.4
2	B	43	VAL	2.4
2	B	278	GLN	2.4
2	B	293	LEU	2.4
2	B	121	VAL	2.3
2	B	203[A]	TYR	2.3
2	B	110	GLU	2.3
2	B	181	ILE	2.3
1	A	2068	ASN	2.3
2	B	140	VAL	2.3
1	A	2032	ILE	2.2
2	B	108	ILE	2.2
2	B	224	LEU	2.2
2	B	317	LEU	2.2
2	B	228	PHE	2.2
2	B	254	ALA	2.2
2	B	2	ASN	2.2
1	A	2065	ARG	2.2
1	A	1837	SER	2.1
1	A	1866	PHE	2.1
2	B	117	LEU	2.1
2	B	44[A]	ILE	2.1
2	B	41[A]	VAL	2.1
1	A	2059	ILE	2.1
2	B	151	GLU	2.1
1	A	1979[A]	MET	2.1
1	A	1840	TYR	2.0
2	B	174	HIS	2.0

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 (i)

There are no such residues in this entry.