Full wwPDB X-ray Structure Validation Report (i)

Jun 24, 2024 – 06:20 PM EDT

PDB ID	:	5R1L
Title	:	PanDDA analysis group deposition – Auto-refined data of Aar2/RNaseH for ground state model 36, DMSO-free
Authors	:	Wollenhaupt, J.; Metz, A.; Barthel, T.; Lima, G.M.A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M.S.
Deposited on	:	2020-02-12
Resolution	:	1.94 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:

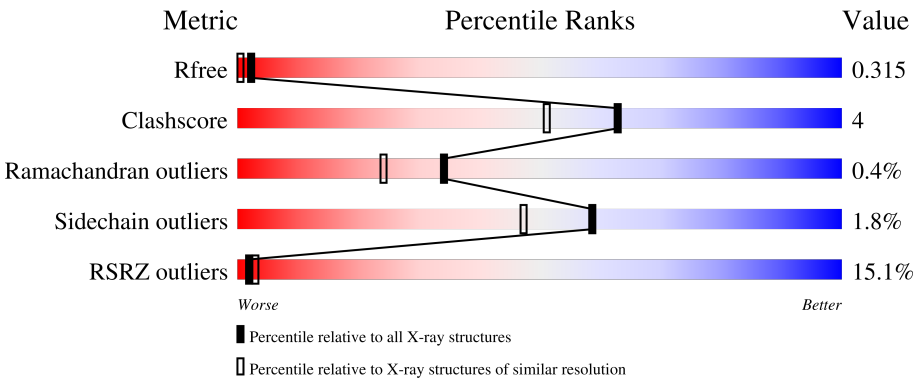
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 1.94 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	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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	10% 79% 11% • 8%
2	B	308	18% 83% 14% • •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4632 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	2568	1645	420	483	20	0	9	0

There are 20 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	LEU	conflict	UNP P32357
B	167	SER	LYS	conflict	UNP P32357
B	170	SER	LEU	conflict	UNP P32357
B	?	-	GLN	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357
B	?	-	ALA	deletion	UNP P32357
B	?	-	GLY	deletion	UNP P32357
B	?	-	SER	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	MET	deletion	UNP P32357
B	?	-	GLU	deletion	UNP P32357
B	?	-	ALA	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357
B	?	-	ASN	deletion	UNP P32357
B	?	-	GLU	deletion	UNP P32357
B	?	-	ASP	deletion	UNP P32357

- Molecule 3 is water.

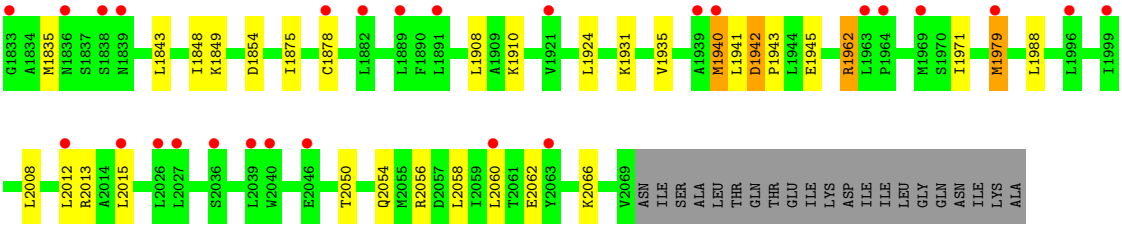
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	39	39	39	0	0
			Total	O		
3	B	23	23	23	0	0

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.

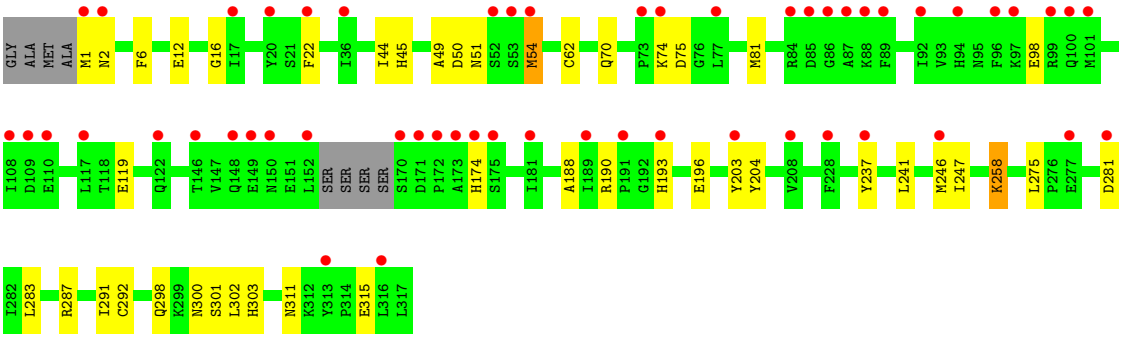
• Molecule 1: Pre-mRNA-splicing factor 8

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	88.67Å 82.67Å 92.19Å 90.00° 107.51° 90.00°	Depositor
Resolution (Å)	26.72 – 1.94 44.88 – 1.94	Depositor EDS
% Data completeness (in resolution range)	99.2 (26.72-1.94) 99.3 (44.88-1.94)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.96 (at 1.94Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.251 , 0.309 0.261 , 0.315	Depositor DCC
R_{free} test set	2101 reflections (4.52%)	wwPDB-VP
Wilson B-factor (Å^2)	50.7	Xtriage
Anisotropy	0.191	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 54.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4632	wwPDB-VP
Average B, all atoms (Å^2)	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.89% 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	0.82	1/2049 (0.0%)	1.00	11/2775 (0.4%)
2	B	0.73	1/2638 (0.0%)	0.78	1/3563 (0.0%)
All	All	0.77	2/4687 (0.0%)	0.88	12/6338 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1878	CYS	CB-SG	-8.81	1.67	1.82
2	B	62	CYS	CB-SG	-5.06	1.73	1.81

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	2012	LEU	CB-CG-CD1	-8.01	97.38	111.00
1	A	1988	LEU	CB-CG-CD2	-7.23	98.72	111.00
1	A	1908	LEU	CB-CG-CD1	-7.04	99.03	111.00
1	A	2015	LEU	CB-CG-CD1	7.01	122.92	111.00
1	A	2008	LEU	CB-CG-CD1	-7.00	99.10	111.00
2	B	246	MET	CG-SD-CE	-6.30	90.12	100.20
1	A	1854	ASP	CB-CG-OD2	6.00	123.70	118.30
1	A	2012	LEU	CB-CG-CD2	5.92	121.06	111.00
1	A	1940	MET	CG-SD-CE	-5.80	90.93	100.20
1	A	1924	LEU	CB-CG-CD2	-5.60	101.48	111.00
1	A	1942	ASP	CB-CG-OD2	-5.41	113.44	118.30
1	A	1935	VAL	CG1-CB-CG2	-5.16	102.65	110.90

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	14	0
2	B	2568	0	2439	25	0
3	A	39	0	0	0	0
3	B	23	0	0	4	0
All	All	4632	0	4468	39	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 4.

All (39) 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:1875:ILE:HD12	1:A:1979[B]:MET:HE3	1.63	0.81
1:A:2062:GLU:O	1:A:2066:LYS:HG2	1.82	0.79
2:B:258:LYS:HD2	2:B:258:LYS:H	1.52	0.74
2:B:74:LYS:NZ	3:B:401:HOH:O	2.13	0.70
1:A:1962:ARG:O	1:A:2013:ARG:NH1	2.26	0.67
2:B:1:MET:N	3:B:403:HOH:O	2.26	0.67
2:B:70:GLN:HB3	2:B:81:MET:HE1	1.78	0.64
1:A:1848:ILE:H	1:A:1931[A]:LYS:HZ2	1.47	0.62
2:B:287:ARG:O	2:B:291:ILE:HD13	2.01	0.60
1:A:2058:LEU:C	1:A:2058:LEU:HD23	2.27	0.55
2:B:300:ASN:O	2:B:303:HIS:NE2	2.39	0.55
2:B:311:ASN:OD1	3:B:402:HOH:O	2.18	0.55
2:B:193:HIS:ND1	2:B:196:GLU:OE1	2.35	0.51
2:B:12:GLU:HG2	2:B:49:ALA:HB2	1.93	0.51
2:B:258:LYS:H	2:B:258:LYS:CD	2.20	0.49
1:A:2062:GLU:HB3	1:A:2066:LYS:HE3	1.95	0.49
2:B:190:ARG:HG3	2:B:203[B]:TYR:CE2	2.48	0.48
2:B:301:SER:O	2:B:302:LEU:HD23	2.14	0.48
2:B:70:GLN:HB3	2:B:81:MET:CE	2.44	0.48
2:B:50:ASP:OD1	2:B:51:ASN:N	2.47	0.47
1:A:1941:LEU:O	1:A:1945:GLU:HB2	2.15	0.47
2:B:275:LEU:CD2	2:B:283:LEU:HD13	2.45	0.45
2:B:298:GLN:HA	2:B:298:GLN:OE1	2.16	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:237:TYR:CE2	2:B:241:LEU:HD11	2.51	0.45
1:A:1910:LYS:HG2	1:A:1940:MET:SD	2.57	0.45
2:B:275:LEU:HD21	2:B:283:LEU:HD13	2.00	0.44
1:A:1942:ASP:HB2	1:A:1943:PRO:HD3	1.99	0.44
1:A:2050:THR:HG22	1:A:2054:GLN:NE2	2.33	0.44
2:B:247:ILE:HG22	2:B:292:CYS:SG	2.58	0.44
1:A:1843:LEU:HA	1:A:1849:LYS:HD2	2.01	0.43
2:B:188:ALA:HA	2:B:204:TYR:CD1	2.54	0.43
1:A:2056[B]:ARG:O	1:A:2060:LEU:HG	2.19	0.43
2:B:54[B]:MET:N	3:B:405:HOH:O	2.38	0.42
2:B:16:GLY:HA3	2:B:45:HIS:CE1	2.54	0.42
2:B:6:PHE:HZ	2:B:44[A]:ILE:HD11	1.85	0.41
1:A:1971:ILE:HG21	1:A:1971:ILE:HD13	1.82	0.41
1:A:1940:MET:HE2	1:A:1940:MET:HB3	1.82	0.41
2:B:281:ASP:OD1	2:B:281:ASP:N	2.47	0.40
2:B:98:GLU:HA	2:B:98:GLU:OE1	2.22	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%)	243 (98%)	5 (2%)	0	100 100
2	B	305/308 (99%)	290 (95%)	12 (4%)	3 (1%)	15 6
All	All	553/566 (98%)	533 (96%)	17 (3%)	3 (0%)	34 17

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	54[A]	MET
2	B	54[B]	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	119	GLU

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%)	221 (98%)	5 (2%)	52	39
2	B	286/284 (101%)	280 (98%)	6 (2%)	53	41
All	All	512/517 (99%)	501 (98%)	11 (2%)	59	41

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1835	MET
1	A	1962	ARG
1	A	1979[A]	MET
1	A	1979[B]	MET
1	A	1979[C]	MET
2	B	2	ASN
2	B	22	PHE
2	B	75	ASP
2	B	174	HIS
2	B	258	LYS
2	B	315	GLU

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

Mol	Chain	Res	Type
1	A	1907	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)

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%)	1.09	27 (11%)	5	7	43, 69, 103, 133	0
2	B	300/308 (97%)	1.21	54 (18%)	1	1	45, 76, 126, 185	0
All	All	537/566 (94%)	1.16	81 (15%)	2	3	43, 72, 121, 185	0

All (81) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	53	SER	7.9
1	A	1833	GLY	6.9
2	B	52	SER	6.6
1	A	2027	LEU	6.2
2	B	109	ASP	6.0
1	A	1878	CYS	5.9
2	B	175	SER	4.8
1	A	1838	SER	4.7
1	A	1969	MET	4.6
2	B	86	GLY	4.4
2	B	172	PRO	4.4
1	A	1889	LEU	4.4
2	B	89	PHE	4.1
1	A	2060	LEU	3.8
2	B	174	HIS	3.7
2	B	101	MET	3.7
2	B	96	PHE	3.7
2	B	92	ILE	3.7
1	A	1940	MET	3.6
2	B	108	ILE	3.6
2	B	170	SER	3.5
2	B	181	ILE	3.5
2	B	193	HIS	3.5
2	B	189	ILE	3.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	148	GLN	3.4
2	B	1	MET	3.4
2	B	100	GLN	3.2
2	B	110	GLU	3.2
1	A	2046	GLU	3.2
2	B	122[A]	GLN	3.2
2	B	277	GLU	3.2
2	B	54[A]	MET	3.1
2	B	237	TYR	3.1
2	B	84	ARG	2.9
2	B	87	ALA	2.9
1	A	2039	LEU	2.9
2	B	171	ASP	2.8
2	B	208	VAL	2.8
2	B	281	ASP	2.8
2	B	313	TYR	2.8
2	B	88	LYS	2.7
2	B	85	ASP	2.7
2	B	99	ARG	2.6
2	B	74	LYS	2.6
1	A	2012	LEU	2.6
2	B	149	GLU	2.6
1	A	1921	VAL	2.6
1	A	1882	LEU	2.5
2	B	152	LEU	2.5
1	A	1979[A]	MET	2.5
2	B	150	ASN	2.5
2	B	17	ILE	2.5
2	B	22	PHE	2.5
2	B	73	PRO	2.4
2	B	146	THR	2.4
2	B	117	LEU	2.4
1	A	1996	LEU	2.3
1	A	2036	SER	2.3
1	A	2015	LEU	2.3
1	A	2040	TRP	2.3
2	B	191	PRO	2.2
1	A	1839	ASN	2.2
1	A	1999	ILE	2.2
1	A	1964	PRO	2.2
1	A	1963	LEU	2.2
2	B	203[A]	TYR	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	20	TYR	2.2
2	B	77	LEU	2.2
2	B	97	LYS	2.2
2	B	36	ILE	2.1
1	A	2026	LEU	2.1
2	B	228	PHE	2.1
1	A	1836	ASN	2.1
1	A	1891	LEU	2.1
1	A	1939	ALA	2.1
2	B	173	ALA	2.1
2	B	316	LEU	2.1
2	B	2	ASN	2.0
2	B	94	HIS	2.0
2	B	246	MET	2.0
1	A	2063	TYR	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.