



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 17, 2024 – 08:36 PM EDT

PDB ID : 5R1P  
Title : PanDDA analysis group deposition – Auto-refined data of Aar2/RNaseH for ground state model 40, 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 : 1.95 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](mailto: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 : 2.37.1  
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

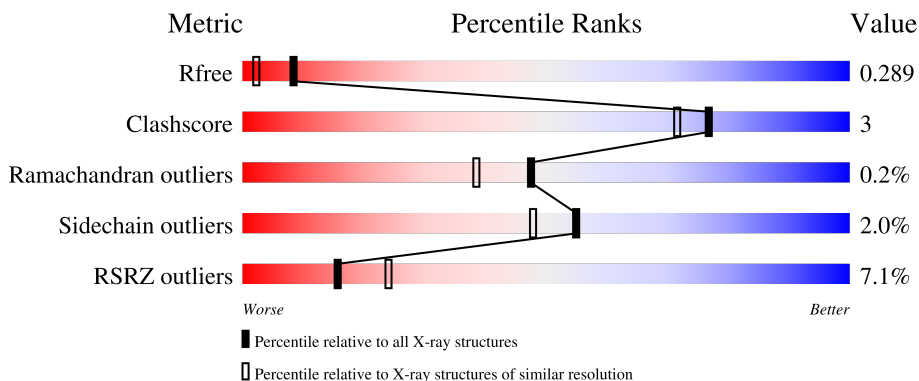
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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  $\geq 3$ , 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	258	
2	B	308	

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 4655 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	2575	1651	420	484	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

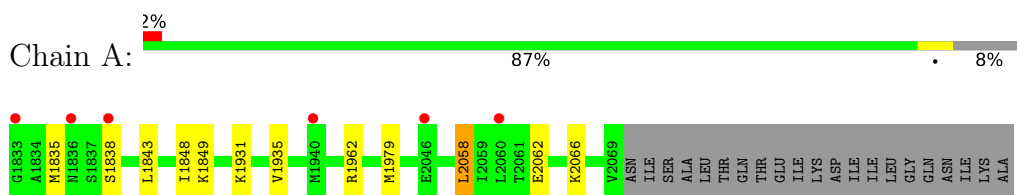
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	57	Total O 57 57	0	0
3	B	21	Total O 21 21	0	0

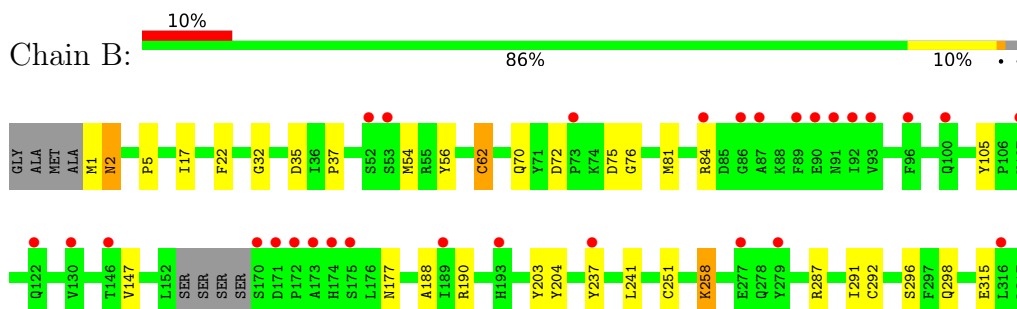
### 3 Residue-property plots [i](#)

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



- Molecule 2: A1 cistron-splicing factor AAR2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	87.21Å 82.47Å 92.22Å 90.00° 107.44° 90.00°	Depositor
Resolution (Å)	21.75 – 1.95 44.61 – 1.95	Depositor EDS
% Data completeness (in resolution range)	99.8 (21.75-1.95) 99.8 (44.61-1.95)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.01 (at 1.95Å)	Xtrriage
Refinement program	REFMAC 5.8.0238	Depositor
R, $R_{free}$	0.218 , 0.285 0.231 , 0.289	Depositor DCC
$R_{free}$ test set	2100 reflections (4.61%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	47.3	Xtrriage
Anisotropy	0.211	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 50.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4655	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.33% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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	0.69	1/2049 (0.0%)	0.78	0/2775
2	B	0.67	3/2643 (0.1%)	0.70	0/3570
All	All	0.68	4/4692 (0.1%)	0.73	0/6345

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	119	GLU	CB-CG	5.55	1.62	1.52
2	B	56	TYR	CG-CD1	5.49	1.46	1.39
2	B	62	CYS	CB-SG	-5.34	1.73	1.81
1	A	1935	VAL	CB-CG2	-5.13	1.42	1.52

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	5	0
2	B	2575	0	2444	22	0
3	A	57	0	0	0	0
3	B	21	0	0	1	0
All	All	4655	0	4473	27	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 (27) 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)
2:B:1:MET:N	3:B:401:HOH:O	2.12	0.82
1:A:1848:ILE:H	1:A:1931[A]:LYS:HZ2	1.32	0.75
1:A:2062:GLU:O	1:A:2066:LYS:HG2	2.02	0.59
1:A:2058:LEU:C	1:A:2058:LEU:HD23	2.24	0.58
2:B:251:CYS:SG	2:B:292:CYS:SG	3.01	0.57
2:B:251:CYS:CB	2:B:292:CYS:HG	2.18	0.55
2:B:298:GLN:HA	2:B:298:GLN:OE1	2.08	0.54
2:B:251:CYS:HG	2:B:292:CYS:HG	1.49	0.53
2:B:287:ARG:O	2:B:291:ILE:HD13	2.09	0.51
2:B:2:ASN:HB3	2:B:62:CYS:SG	2.53	0.49
2:B:70:GLN:HB3	2:B:81:MET:HE1	1.94	0.48
2:B:258:LYS:HD2	2:B:258:LYS:H	1.78	0.48
2:B:1:MET:HB3	2:B:35:ASP:HA	1.95	0.48
2:B:37:PRO:HD3	2:B:105:TYR:CD1	2.50	0.47
2:B:72:ASP:O	2:B:76:GLY:N	2.47	0.46
1:A:2062:GLU:HB3	1:A:2066:LYS:HE3	1.98	0.46
2:B:147:VAL:HG21	2:B:177[B]:ASN:HD22	1.81	0.44
2:B:251:CYS:HB2	2:B:292:CYS:HG	1.83	0.44
2:B:188:ALA:HA	2:B:204:TYR:CD1	2.53	0.44
2:B:2:ASN:CB	2:B:62:CYS:SG	3.07	0.43
2:B:237:TYR:CE2	2:B:241:LEU:HD11	2.54	0.43
1:A:1843:LEU:HA	1:A:1849:LYS:HD2	2.01	0.43
2:B:190:ARG:HG3	2:B:203[B]:TYR:CE2	2.53	0.43
2:B:72:ASP:OD2	2:B:75:ASP:HB2	2.19	0.43
2:B:251:CYS:O	2:B:296:SER:HB2	2.20	0.42
2:B:5:PRO:HA	2:B:32:GLY:HA3	2.03	0.40
2:B:17:ILE:O	2:B:17:ILE:HG23	2.21	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%)	241 (97%)	7 (3%)	0	100	100
2	B	305/308 (99%)	295 (97%)	8 (3%)	2 (1%)	22	11
All	All	553/566 (98%)	536 (97%)	15 (3%)	2 (0%)	47	22

All (2) Ramachandran outliers are listed below:

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

### 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%)	219 (97%)	7 (3%)	40	28
2	B	286/284 (101%)	281 (98%)	5 (2%)	60	55
All	All	512/517 (99%)	500 (98%)	12 (2%)	55	42

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1835	MET
1	A	1838	SER
1	A	1962	ARG
1	A	1979[A]	MET
1	A	1979[B]	MET
1	A	1979[C]	MET
1	A	2058	LEU
2	B	2	ASN
2	B	22	PHE
2	B	84	ARG
2	B	258	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	315	GLU

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

### 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

### 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	237/258 (91%)	0.52	6 (2%) 57 66	39, 58, 100, 117	0
2	B	300/308 (97%)	0.69	32 (10%) 6 9	42, 67, 120, 157	0
All	All	537/566 (94%)	0.62	38 (7%) 16 24	39, 64, 108, 157	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1833	GLY	5.0
2	B	53	SER	5.0
2	B	87	ALA	4.2
2	B	109	ASP	3.8
2	B	86	GLY	3.7
2	B	52	SER	3.7
2	B	175	SER	3.6
1	A	1940	MET	3.5
2	B	90	GLU	3.3
2	B	92	ILE	3.3
1	A	2060	LEU	3.3
2	B	279	TYR	3.3
2	B	96	PHE	3.2
2	B	172	PRO	3.2
1	A	1838	SER	3.2
2	B	73	PRO	3.0
2	B	193	HIS	2.9
2	B	171	ASP	2.9
2	B	84	ARG	2.8
2	B	91	ASN	2.8
2	B	110	GLU	2.8
2	B	316	LEU	2.7
2	B	93	VAL	2.6
2	B	130	VAL	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	146	THR	2.5
2	B	173	ALA	2.5
1	A	1836	ASN	2.4
2	B	237	TYR	2.3
2	B	108	ILE	2.2
2	B	107	LYS	2.2
2	B	89	PHE	2.2
1	A	2046	GLU	2.2
2	B	100	GLN	2.2
2	B	174	HIS	2.2
2	B	277	GLU	2.1
2	B	189	ILE	2.1
2	B	122[A]	GLN	2.1
2	B	170	SER	2.1

## 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 [i](#)

There are no such residues in this entry.