



Full wwPDB X-ray Structure Validation Report ⓘ

Oct 25, 2022 – 05:24 pm BST

PDB ID : 7PJH
Title : Crystal structure of the human spliceosomal maturation factor AAR2 bound to the RNase H domain of PRPF8
Authors : Preussner, M.; Santos, K.; Heroven, A.C.; Alles, J.; Heyd, F.; Wahl, M.C.; Weber, G.
Deposited on : 2021-08-24
Resolution : 2.35 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 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 : 2.31.2
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.31.2

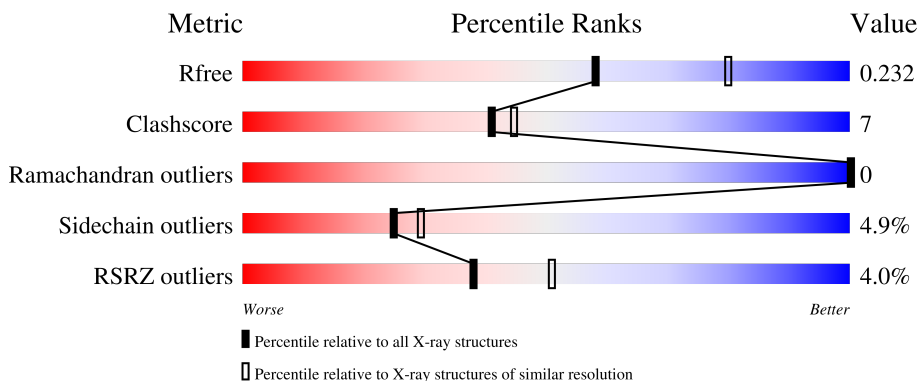
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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 $\leq 5\%$. 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	401	 2% 62% 16% 21%
2	B	259	 5% 83% 15% ..

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4790 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 Protein AAR2 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	317	2554	1655	424	462	13	0	3	0

There are 57 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-27	MET	-	initiating methionine	UNP Q9Y312
A	-26	LYS	-	expression tag	UNP Q9Y312
A	-25	HIS	-	expression tag	UNP Q9Y312
A	-24	HIS	-	expression tag	UNP Q9Y312
A	-23	HIS	-	expression tag	UNP Q9Y312
A	-22	HIS	-	expression tag	UNP Q9Y312
A	-21	HIS	-	expression tag	UNP Q9Y312
A	-20	HIS	-	expression tag	UNP Q9Y312
A	-19	PRO	-	expression tag	UNP Q9Y312
A	-18	MET	-	expression tag	UNP Q9Y312
A	-17	SER	-	expression tag	UNP Q9Y312
A	-16	ASP	-	expression tag	UNP Q9Y312
A	-15	TYR	-	expression tag	UNP Q9Y312
A	-14	ASP	-	expression tag	UNP Q9Y312
A	-13	ILE	-	expression tag	UNP Q9Y312
A	-12	PRO	-	expression tag	UNP Q9Y312
A	-11	THR	-	expression tag	UNP Q9Y312
A	-10	THR	-	expression tag	UNP Q9Y312
A	-9	GLU	-	expression tag	UNP Q9Y312
A	-8	ASN	-	expression tag	UNP Q9Y312
A	-7	LEU	-	expression tag	UNP Q9Y312
A	-6	TYR	-	expression tag	UNP Q9Y312
A	-5	PHE	-	expression tag	UNP Q9Y312
A	-4	GLN	-	expression tag	UNP Q9Y312
A	-3	GLY	-	expression tag	UNP Q9Y312
A	-2	ALA	-	expression tag	UNP Q9Y312
A	-1	GLU	-	expression tag	UNP Q9Y312

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	PHE	-	expression tag	UNP Q9Y312
A	?	-	VAL	deletion	UNP Q9Y312
A	180	SER	THR	conflict	UNP Q9Y312
A	181	SER	LYS	conflict	UNP Q9Y312
A	182	SER	ASP	conflict	UNP Q9Y312
A	184	ALA	VAL	conflict	UNP Q9Y312
A	186	THR	GLN	conflict	UNP Q9Y312
A	187	GLU	ASN	conflict	UNP Q9Y312
A	188	ILE	LEU	conflict	UNP Q9Y312
A	189	ARG	PRO	conflict	UNP Q9Y312
A	190	PHE	ARG	conflict	UNP Q9Y312
A	191	SER	CYS	conflict	UNP Q9Y312
A	192	GLU	GLY	conflict	UNP Q9Y312
A	193	LEU	ILE	conflict	UNP Q9Y312
A	194	PRO	GLU	conflict	UNP Q9Y312
A	195	THR	CYS	conflict	UNP Q9Y312
A	196	GLN	LYS	conflict	UNP Q9Y312
A	197	MET	SER	conflict	UNP Q9Y312
A	198	PHE	TYR	conflict	UNP Q9Y312
A	199	PRO	GLN	conflict	UNP Q9Y312
A	?	-	GLY	deletion	UNP Q9Y312
A	?	-	LEU	deletion	UNP Q9Y312
A	?	-	ALA	deletion	UNP Q9Y312
A	?	-	ARG	deletion	UNP Q9Y312
A	?	-	LEU	deletion	UNP Q9Y312
A	?	-	PRO	deletion	UNP Q9Y312
A	?	-	GLU	deletion	UNP Q9Y312
A	?	-	MET	deletion	UNP Q9Y312
A	?	-	LYS	deletion	UNP Q9Y312
A	?	-	PRO	deletion	UNP Q9Y312

- Molecule 2 is a protein called Pre-mRNA-processing-splicing factor 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	257	2088	1358	350	376	4	0	1	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	78	78	78	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	69	Total	O	0	1
			70	70		

4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	145.28Å 57.26Å 111.23Å 90.00° 112.74° 90.00°	Depositor
Resolution (Å)	48.21 – 2.35 48.21 – 2.35	Depositor EDS
% Data completeness (in resolution range)	97.7 (48.21-2.35) 97.8 (48.21-2.35)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.15 (at 2.34Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158, PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.189 , 0.235 0.186 , 0.232	Depositor DCC
R_{free} test set	1723 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å ²)	55.4	Xtrriage
Anisotropy	0.686	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	4790	wwPDB-VP
Average B, all atoms (Å ²)	77.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.90% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.46	0/2627	0.61	0/3561
2	B	0.44	0/2138	0.61	0/2905
All	All	0.45	0/4765	0.61	0/6466

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	2554	0	2541	40	0
2	B	2088	0	2156	26	0
3	A	78	0	0	4	0
3	B	70	0	0	1	0
All	All	4790	0	4697	66	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 7.

All (66) 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:1957:ASP:HB3	2:B:1960:THR:HG23	1.75	0.68

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:36:TYR:HB2	1:A:129:THR:HG21	1.78	0.66
1:A:74:ARG:HB3	3:A:401:HOH:O	1.96	0.65
1:A:14:LEU:HD21	1:A:52:PRO:HG3	1.78	0.65
1:A:309:ILE:HG22	1:A:311:ALA:H	1.62	0.64
1:A:23:ILE:HG22	1:A:89:LEU:HB2	1.79	0.63
1:A:282:CYS:HB3	1:A:331:PHE:HB2	1.79	0.62
1:A:252:LEU:HG	1:A:301:ILE:HD12	1.81	0.62
1:A:232:LEU:HD22	1:A:283:ARG:HB2	1.82	0.61
1:A:95:ARG:HB3	1:A:97:GLU:HG2	1.83	0.61
1:A:90:ARG:NH2	1:A:107:GLU:OE1	2.35	0.60
2:B:1953:ILE:HD13	2:B:1982:GLN:HB3	1.83	0.59
2:B:1991:TYR:CE2	2:B:2010:ILE:HD12	2.38	0.59
2:B:1835:GLN:OE1	2:B:1835:GLN:N	2.34	0.58
1:A:71:VAL:HG22	1:A:72:GLY:H	1.69	0.58
1:A:323:PHE:O	1:A:327:THR:HG23	2.05	0.57
2:B:1999:VAL:HB	2:B:2001:SER:H	1.70	0.56
1:A:157:ALA:HB2	3:A:401:HOH:O	2.05	0.56
1:A:15:PHE:HA	1:A:50:MET:HG2	1.87	0.56
1:A:23:ILE:HD12	1:A:23:ILE:O	2.06	0.56
2:B:1949:ARG:HG2	2:B:1986:LEU:HD13	1.87	0.56
1:A:24:LEU:HD23	1:A:46[B]:ARG:HH22	1.70	0.56
1:A:24:LEU:HD23	1:A:46[B]:ARG:NH2	2.22	0.53
1:A:31:GLU:O	1:A:60:TYR:HA	2.08	0.53
2:B:1865[A]:ARG:NH2	3:B:2101:HOH:O	2.34	0.53
2:B:1782:ASP:O	2:B:1785:VAL:HG13	2.09	0.52
2:B:1991:TYR:HD2	2:B:2010:ILE:HG23	1.73	0.52
1:A:60:TYR:OH	1:A:97:GLU:HB2	2.10	0.52
1:A:278:LEU:HD11	1:A:302:LEU:HD11	1.92	0.51
2:B:1905:LEU:HD21	2:B:1915:VAL:HG21	1.91	0.51
1:A:133:TRP:CZ2	1:A:137:THR:HG21	2.46	0.51
1:A:296:ILE:O	1:A:347:LYS:NZ	2.41	0.51
2:B:1957:ASP:O	2:B:1960:THR:OG1	2.25	0.50
2:B:1782:ASP:OD1	2:B:1865[B]:ARG:NE	2.40	0.50
2:B:2007:ILE:O	2:B:2011:ILE:HG12	2.13	0.49
1:A:23:ILE:HG13	1:A:46[B]:ARG:O	2.13	0.48
2:B:1939:ILE:HG21	2:B:1968:TRP:CZ3	2.47	0.48
1:A:311:ALA:HB2	1:A:359:PHE:CE2	2.48	0.48
1:A:23:ILE:HD13	1:A:26:MET:SD	2.53	0.48
1:A:266:ASN:HB2	3:A:401:HOH:O	2.13	0.48
1:A:135:SER:HB2	1:A:304:HIS:CG	2.48	0.48
2:B:2009:ASP:N	2:B:2009:ASP:OD1	2.46	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:1869:LEU:O	2:B:1873:GLU:HG3	2.13	0.48
1:A:90:ARG:HG2	1:A:99:ASP:HB3	1.96	0.47
1:A:337:SER:HB2	3:A:469:HOH:O	2.14	0.47
1:A:262:PHE:CD1	1:A:271:PHE:HB2	2.49	0.47
2:B:1956:PRO:HG2	2:B:1960:THR:HG21	1.96	0.47
1:A:301:ILE:O	1:A:305:GLN:HG3	2.15	0.46
2:B:1809:ILE:HB	2:B:1818:PHE:HB2	1.98	0.46
1:A:283:ARG:NH1	1:A:330:VAL:HG21	2.31	0.45
2:B:1782:ASP:OD1	2:B:1865[A]:ARG:HD3	2.16	0.45
1:A:46[A]:ARG:CZ	1:A:118:LEU:HD21	2.46	0.45
2:B:1760:GLU:HG2	2:B:1760:GLU:O	2.17	0.45
2:B:1850:ARG:HG2	2:B:1879:PHE:HZ	1.82	0.45
1:A:342:ALA:O	1:A:346:LYS:HG3	2.16	0.45
2:B:1870:ASP:HB3	2:B:1871:PRO:HD3	1.99	0.44
2:B:1995:ASN:O	2:B:1997:VAL:HG23	2.17	0.44
2:B:1991:TYR:CD2	2:B:2010:ILE:HG23	2.53	0.44
2:B:1892:PRO:HD3	2:B:1941:ARG:NH2	2.32	0.43
1:A:58:LEU:O	1:A:76:GLY:HA2	2.19	0.43
1:A:149:GLN:O	1:A:207:ARG:NH1	2.45	0.42
1:A:133:TRP:CH2	1:A:137:THR:HG21	2.54	0.42
1:A:269:GLU:CD	1:A:269:GLU:H	2.22	0.42
2:B:1783:THR:OG1	2:B:1865[A]:ARG:HD2	2.19	0.42
1:A:311:ALA:HB2	1:A:359:PHE:HE2	1.85	0.41
1:A:252:LEU:HG	1:A:301:ILE:CD1	2.50	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	314/401 (78%)	301 (96%)	13 (4%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	256/259 (99%)	244 (95%)	12 (5%)	0	100	100
All	All	570/660 (86%)	545 (96%)	25 (4%)	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	279/352 (79%)	261 (94%)	18 (6%)	17	18
2	B	234/235 (100%)	227 (97%)	7 (3%)	41	50
All	All	513/587 (87%)	488 (95%)	25 (5%)	25	29

All (25) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	13	ARG
1	A	14	LEU
1	A	38	SER
1	A	61	SER
1	A	79	LEU
1	A	90	ARG
1	A	114	ASN
1	A	122	LEU
1	A	129	THR
1	A	146	GLU
1	A	155	ILE
1	A	210	GLU
1	A	232	LEU
1	A	252	LEU
1	A	264	LEU
1	A	280	LEU
1	A	355	LEU
1	A	361	TRP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	1819	LEU
2	B	1872	LEU
2	B	1919	LEU
2	B	1962	THR
2	B	1991	TYR
2	B	2006	GLU
2	B	2009	ASP

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, 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	317/401 (79%)	0.37	10 (3%) 47 59	48, 68, 119, 194	0
2	B	257/259 (99%)	0.47	13 (5%) 28 40	45, 73, 142, 171	0
All	All	574/660 (86%)	0.42	23 (4%) 38 51	45, 70, 136, 194	0

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	66	ALA	5.0
2	B	2007	ILE	4.5
2	B	1998	ASN	4.4
2	B	2001	SER	3.5
2	B	2011	ILE	3.4
1	A	70	GLU	3.4
2	B	2003	THR	3.4
2	B	1994	LYS	3.4
2	B	2014	MET	3.3
2	B	1997	VAL	3.3
2	B	1996	ASN	3.2
2	B	2002	LEU	3.2
2	B	1798	LEU	3.1
1	A	202	ALA	2.9
1	A	324	LEU	2.8
1	A	127	TYR	2.7
1	A	67	ASN	2.5
1	A	71	VAL	2.4
2	B	2008	ARG	2.4
1	A	323	PHE	2.3
2	B	1954	LEU	2.2
1	A	311	ALA	2.0
1	A	69	LYS	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.