

Full wwPDB X-ray Structure Validation Report (i)

Oct 25, 2022 – 05:19 PM EDT

PDB ID	:	7FML
Title	:	PanDDA analysis group deposition – Aar2/RNaseH in complex with fragment
		P06D04 from the F2X-Universal Library
Authors	:	Barthel, T.; Wollenhaupt, J.; Lima, G.M.A.; Wahl, M.C.; Weiss, M.S.
Deposited on	:	2022-08-26
Resolution	:	1.58 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:

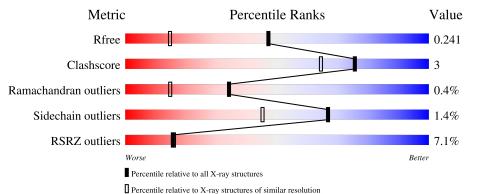
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: $X\text{-}RAY\;DIFFRACTION$

The reported resolution of this entry is 1.58 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R_{free}	130704	5534 (1.60-1.56)
Clashscore	141614	5861 (1.60-1.56)
Ramachandran outliers	138981	5708 (1.60-1.56)
Sidechain outliers	138945	5703 (1.60-1.56)
RSRZ outliers	127900	5431 (1.60-1.56)

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	А	258	8% 84% 7%	• 8%				
2	В	308	<u>6%</u> 90%	7% •				

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:



Μ	l Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	VR6	А	2101	-	-	-	Х



$7 \mathrm{FML}$

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 9244 atoms, of which 4524 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	
1	А	237	Total 4068	C 1287	Н 2060	N 336	O 373	S 12	0	21	0

There are 3 discrepancies between the modelled and reference sequences:

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

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace		
2	В	300	Total 5044	C 1654	Н 2464	N 421	O 485	S 20	0	17	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
В	-3	GLY	-	expression tag	UNP P32357
В	-2	ALA	-	expression tag	UNP P32357
В	-1	MET	-	expression tag	UNP P32357
В	0	ALA	-	expression tag	UNP P32357
В	166	SER	LEU	conflict	UNP P32357
В	167	SER	LYS	conflict	UNP P32357
В	?	-	LEU	deletion	UNP P32357
В	?	-	GLN	deletion	UNP P32357
В	?	-	LYS	deletion	UNP P32357
В	?	-	ALA	deletion	UNP P32357
В	?	-	GLY	deletion	UNP P32357
В	?	-	SER	deletion	UNP P32357
В	?	-	LYS	deletion	UNP P32357

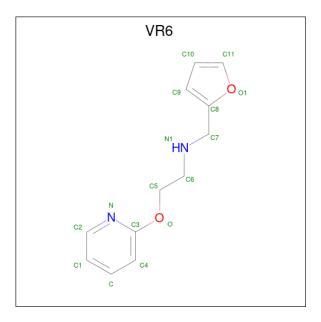
Continued on next page...



Contentia	ca from pre	erous page			
Chain	Residue	Modelled	Actual	Comment	Reference
В	?	-	MET	deletion	UNP P32357
В	?	-	GLU	deletion	UNP P32357
В	?	-	ALA	deletion	UNP P32357
В	?	-	LYS	deletion	UNP P32357
В	?	-	ASN	deletion	UNP P32357
В	?	-	GLU	deletion	UNP P32357
В	170	SER	ASP	conflict	UNP P32357

Continued from previous page..

• Molecule 3 is N-[(furan-2-yl)methyl]-2-[(pyridin-2-yl)oxy]ethan-1-amine (three-letter code: VR6) (formula: $C_{12}H_{14}N_2O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	А	1	Total 16	C 12	N 2	O 2	0	0

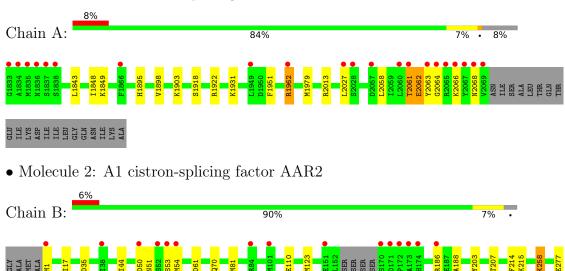
• Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	55	Total O 55 55	0	0
4	В	61	Total O 61 61	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



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants	88.10Å 82.11Å 93.46Å	Depositor
a, b, c, α , β , γ	90.00° 108.12° 90.00°	Depositor
Resolution (Å)	44.66 - 1.58	Depositor
Resolution (A)	44.66 - 1.58	EDS
% Data completeness	99.6 (44.66-1.58)	Depositor
(in resolution range)	99.7 (44.66-1.58)	EDS
R _{merge}	(Not available)	Depositor
R _{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$0.96 (at 1.58 \text{\AA})$	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
D D.	0.205 , 0.239	Depositor
R, R_{free}	0.212 , 0.241	DCC
R_{free} test set	2099 reflections $(2.42%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	34.8	Xtriage
Anisotropy	0.368	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.40 , 48.0	EDS
L-test for twinning ²	$ \langle L \rangle = 0.50, \langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	9244	wwPDB-VP
Average B, all atoms $(Å^2)$	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.56% of the height of the origin peak. No significant pseudotranslation is detected.

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.



¹Intensities estimated from amplitudes.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: $\mathrm{VR6}$

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		nd lengths	Bond	angles
Moi Chain		RMSZ	# Z > 5	RMSZ	# Z > 5
1	А	0.60 1/2149 (0.0%		0.67	0/2911
2	В	0.54	0/2739	0.67	0/3699
All	All	0.57	1/4888~(0.0%)	0.67	0/6610

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	Observed(Å)	Ideal(Å)
1	А	1951	PHE	C-N	9.55	1.52	1.34

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	А	2008	2060	1974	12	0
2	В	2580	2464	2398	16	0
3	А	16	0	0	0	0
4	А	55	0	0	0	0
4	В	61	0	0	1	0
All	All	4720	4524	4372	28	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 (28) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic	Clash
Atom-1	Atom-2	distance (Å)	overlap (Å)
1:A:2061:THR:O	1:A:2064:GLY:N	2.11	0.82
1:A:2064:GLY:O	1:A:2068:ASN:N	2.32	0.62
1:A:1918[A]:SER:O	1:A:1922:ARG:HG3	1.98	0.62
1:A:1918[B]:SER:O	1:A:1922:ARG:HG3	1.99	0.62
2:B:1:MET:HB3	2:B:35:ASP:HA	1.87	0.56
2:B:70:GLN:HB3	2:B:81:MET:HE1	1.89	0.54
2:B:258:LYS:HD2	2:B:258:LYS:H	1.73	0.54
2:B:287:ARG:O	2:B:291:ILE:HD13	2.09	0.52
2:B:50:ASP:OD1	2:B:51:ASN:N	2.43	0.52
1:A:1848:ILE:H	1:A:1931[A]:LYS:HZ2	1.57	0.52
2:B:277:GLU:CD	2:B:277:GLU:H	2.18	0.47
1:A:1962:ARG:O	1:A:2013:ARG:NH1	2.47	0.46
2:B:70:GLN:HB3	2:B:81:MET:CE	2.45	0.46
2:B:17:ILE:HD13	2:B:44[B]:ILE:CG1	2.45	0.45
1:A:2061:THR:O	1:A:2063:TYR:N	2.49	0.45
2:B:54[A]:MET:HE2	4:B:460:HOH:O	2.17	0.44
1:A:2058:LEU:C	1:A:2058:LEU:HD23	2.38	0.44
1:A:1895:HIS:O	1:A:1898[A]:VAL:HG22	2.18	0.43
2:B:214:PHE:O	2:B:215:LYS:HB2	2.18	0.43
1:A:1843:LEU:HA	1:A:1849:LYS:HD2	2.01	0.43
1:A:2061:THR:O	1:A:2062:GLU:C	2.57	0.42
1:A:2062:GLU:O	1:A:2066:LYS:HB2	2.20	0.41
2:B:51:ASN:ND2	2:B:53:SER:O	2.54	0.41
2:B:203[A]:TYR:CZ	2:B:207:THR:HG21	2.56	0.40
2:B:51:ASN:OD1	2:B:53:SER:HB2	2.22	0.40
2:B:110:GLU:O	2:B:110:GLU:HG2	2.20	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



Mol	Chain	Analysed	Favoured	Allowed	Outliers	Perce	ntiles
1	А	258/258~(100%)	250~(97%)	6~(2%)	2(1%)	19	5
2	В	315/308~(102%)	305~(97%)	10 (3%)	0	100	100
All	All	573/566~(101%)	555~(97%)	16 (3%)	2(0%)	34	21

analysed, and the total number of residues.

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	А	2061	THR
1	А	2062	GLU

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent side chain 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	Analysed Rotameric Outliers		Percentiles		
1	А	237/233~(102%)	231~(98%)	6~(2%)	47 20		
2	В	294/284~(104%)	290~(99%)	4 (1%)	67 45		
All	All	531/517~(103%)	521 (98%)	10 (2%)	67 31		

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	1903	LYS
1	А	1962	ARG
1	А	1979[A]	MET
1	А	1979[B]	MET
1	А	1979[C]	MET
1	А	2027	LEU
2	В	123[A]	MET
2	В	123[B]	MET
2	В	186	ARG
2	В	258	LYS

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



sidechains are listed below:

Mol	Chain	Res	Type
2	В	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)

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).

Mal	Mol Type Chain Res	e Chain Res Link Bond lengths		3						
INIOI		Unain	Cham Res	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	VR6	А	2101	-	14,17,17	1.89	3 (21%)	16,20,20	1.00	1 (6%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	VR6	А	2101	-	-	2/6/8/8	0/2/2/2

All (3) bond length outliers are listed below:



Mol	Chain	Res	Type	Atoms	Ζ	Observed(Å)	Ideal(Å)
3	А	2101	VR6	C7-C8	-4.58	1.46	1.51
3	А	2101	VR6	C3-N	3.23	1.37	1.32
3	А	2101	VR6	C6-N1	2.47	1.55	1.46

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
3	А	2101	VR6	C2-N-C3	2.15	117.53	115.71

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	А	2101	VR6	O-C5-C6-N1
3	А	2101	VR6	C5-C6-N1-C7

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 (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, 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	$\langle RSRZ \rangle$	#RSRZ>2	$OWAB(Å^2)$	Q<0.9
1	А	237/258~(91%)	0.43	21 (8%) 9 9	33, 42, 80, 103	0
2	В	300/308~(97%)	0.43	17 (5%) 23 23	31, 44, 77, 114	0
All	All	537/566~(94%)	0.43	38 (7%) 16 16	31, 43, 80, 114	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	1833	GLY	10.1
2	В	170	SER	7.4
1	А	2069	VAL	7.1
2	В	172	PRO	6.6
2	В	1	MET	5.7
1	А	2060	LEU	5.7
1	А	2065	ARG	5.6
2	В	53	SER	5.1
1	А	2068	ASN	4.6
1	А	2063	TYR	4.6
2	В	174	HIS	4.1
2	В	279	TYR	4.0
1	А	2027	LEU	3.9
2	В	316	LEU	3.9
2	В	171	ASP	3.8
1	А	1837	SER	3.4
2	В	173	ALA	3.2
1	А	1838	SER	3.1
1	А	1836	ASN	3.1
1	А	2028	SER	3.1
2	В	54[A]	MET	3.0
2	В	52	SER	2.9
1	А	1866	PHE	2.8
1	А	1835	MET	2.8

Continued on next page...



Mol	Chain	Res	Type	RSRZ
1	А	2057[A]	ASP	2.7
1	А	2061	THR	2.7
2	В	186	ARG	2.7
2	В	50	ASP	2.7
2	В	84	ARG	2.5
2	В	151	GLU	2.5
1	А	1834	ALA	2.5
1	А	2064	GLY	2.5
1	А	2067	TYR	2.4
1	А	1962	ARG	2.3
2	В	101	MET	2.2
1	А	1949	LEU	2.2
2	В	38	ILE	2.2
1	А	2066	LYS	2.0

Continued from previous page...

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)

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	$\mathbf{B} ext{-factors}(\mathbf{A}^2)$	Q<0.9
3	VR6	А	2101	16/16	0.43	0.43	20,20,20,20	16

6.5 Other polymers (i)

There are no such residues in this entry.

