



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 25, 2022 – 05:58 PM EDT

PDB ID : 5STW
Title : PanDDA analysis group deposition – Aar2/RNaseH in complex with fragment P03C12 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 : 1.89 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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
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

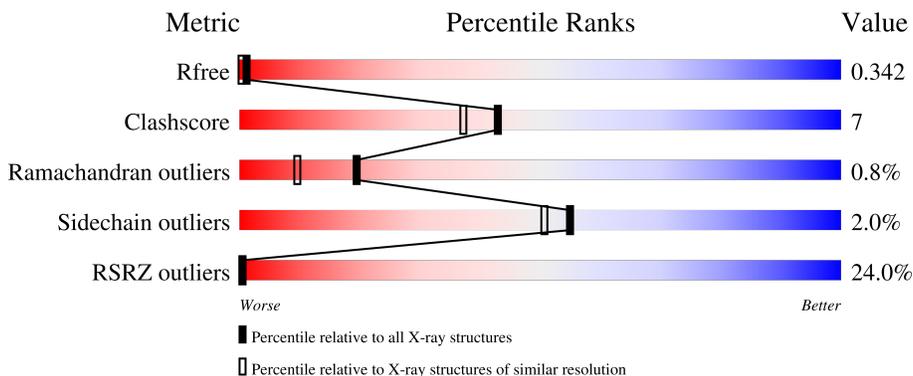
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.89 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	258	
2	B	308	

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:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	V8L	B	401	-	X	-	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 9173 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	
			Total	C	H	N	O				S
1	A	237	4068	1287	2060	336	373	12	18	21	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	H	N	O				S
2	B	300	5044	1654	2464	421	485	20	0	17	0

There are 20 discrepancies between the modelled and reference sequences:

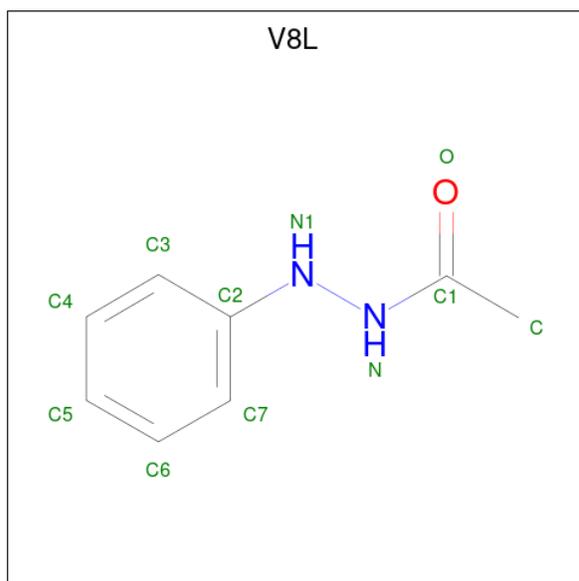
Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	expression tag	UNP P32357
B	-2	ALA	-	expression tag	UNP P32357
B	-1	MET	-	expression tag	UNP P32357
B	0	ALA	-	expression tag	UNP P32357
B	166	SER	LEU	conflict	UNP P32357
B	167	SER	LYS	conflict	UNP P32357
B	?	-	LEU	deletion	UNP P32357
B	?	-	GLN	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	170	SER	ASP	conflict	UNP P32357

- Molecule 3 is N¹-phenylacetohydrazide (three-letter code: V8L) (formula: C₈H₁₀N₂O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	B	1	11	8	2	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	21	21	21	0	0
4	B	29	29	29	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	89.10Å 82.20Å 93.53Å 90.00° 108.52° 90.00°	Depositor
Resolution (Å)	44.34 – 1.89 44.72 – 1.89	Depositor EDS
% Data completeness (in resolution range)	99.2 (44.34-1.89) 99.3 (44.72-1.89)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.97 (at 1.89Å)	Xtrriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.283 , 0.341 0.292 , 0.342	Depositor DCC
R_{free} test set	2100 reflections (4.15%)	wwPDB-VP
Wilson B-factor (Å ²)	51.6	Xtrriage
Anisotropy	0.385	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 70.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	9173	wwPDB-VP
Average B, all atoms (Å ²)	107.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.00% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: V8L

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	2/2149 (0.1%)	0.95	4/2911 (0.1%)
2	B	0.91	5/2739 (0.2%)	0.99	4/3699 (0.1%)
All	All	0.87	7/4888 (0.1%)	0.97	8/6610 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	B	0	4

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	71	TYR	CD1-CE1	-8.78	1.26	1.39
2	B	115	TYR	CB-CG	-7.18	1.40	1.51
2	B	56	TYR	CD1-CE1	5.81	1.48	1.39
1	A	1890	PHE	CE2-CZ	5.77	1.48	1.37
1	A	1858	TYR	CG-CD1	5.24	1.46	1.39

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	249	LEU	CB-CG-CD2	-7.94	97.50	111.00
1	A	1850	LEU	CB-CG-CD2	-7.84	97.67	111.00
1	A	2008	LEU	CB-CG-CD1	-6.97	99.15	111.00
2	B	77	LEU	CA-CB-CG	6.35	129.91	115.30
2	B	115	TYR	CB-CG-CD1	-6.26	117.24	121.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	122[A]	GLN	Mainchain
2	B	122[B]	GLN	Mainchain
2	B	176	LEU	Mainchain

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	2008	2060	1974	20	0
2	B	2580	2464	2396	42	0
3	B	11	0	0	0	0
4	A	21	0	0	1	0
4	B	29	0	0	2	0
All	All	4649	4524	4370	62	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.

The worst 5 of 62 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:1962:ARG:O	1:A:2013:ARG:NH1	1.97	0.97
2:B:129:ILE:HG23	2:B:177[A]:ASN:C	2.01	0.81
2:B:129:ILE:HG23	2:B:177[B]:ASN:C	2.02	0.81
2:B:70:GLN:HB3	2:B:81:MET:HE2	1.72	0.70
2:B:287:ARG:O	2:B:291:ILE:HD13	1.95	0.67

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	258/258 (100%)	247 (96%)	9 (4%)	2 (1%)	19	9
2	B	315/308 (102%)	290 (92%)	22 (7%)	3 (1%)	15	6
All	All	573/566 (101%)	537 (94%)	31 (5%)	5 (1%)	19	7

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1896	THR
2	B	177[A]	ASN
2	B	177[B]	ASN
1	A	2061	THR
2	B	215	LYS

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	237/233 (102%)	230 (97%)	7 (3%)	41	33
2	B	294/284 (104%)	289 (98%)	5 (2%)	60	57
All	All	531/517 (103%)	519 (98%)	12 (2%)	55	45

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	77	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	124	ASP
2	B	317	LEU
2	B	258	LYS
1	A	1979[A]	MET

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	290	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).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	V8L	B	401	-	11,11,11	2.28	5 (45%)	13,13,13	4.47	4 (30%)

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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	V8L	B	401	-	-	4/5/5/5	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	401	V8L	O-C1	-3.68	1.14	1.23
3	B	401	V8L	N1-N	-3.52	1.32	1.39
3	B	401	V8L	C6-C7	-3.23	1.32	1.38
3	B	401	V8L	C3-C2	-2.94	1.34	1.39
3	B	401	V8L	C4-C3	-2.16	1.34	1.38

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	401	V8L	C-C1-N	10.97	124.52	114.97
3	B	401	V8L	O-C1-N	-7.59	113.95	121.97
3	B	401	V8L	C2-N1-N	-6.68	105.83	116.76
3	B	401	V8L	C1-N-N1	5.58	131.01	119.53

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	401	V8L	C7-C2-N1-N
3	B	401	V8L	C3-C2-N1-N
3	B	401	V8L	O-C1-N-N1
3	B	401	V8L	C-C1-N-N1

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

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%)	1.44	53 (22%) 0 0	52, 91, 157, 204	0
2	B	300/308 (97%)	1.56	76 (25%) 0 0	53, 89, 153, 268	0
All	All	537/566 (94%)	1.51	129 (24%) 0 0	52, 91, 153, 268	0

The worst 5 of 129 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	109	ASP	10.3
2	B	1	MET	8.9
2	B	53	SER	8.4
1	A	2068	ASN	7.5
2	B	52	SER	7.1

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](#)

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, 95th 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	B-factors(\AA^2)	Q<0.9
3	V8L	B	401	11/11	0.86	0.39	20,20,20,20	11

6.5 Other polymers [i](#)

There are no such residues in this entry.