



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 25, 2024 – 04:04 PM EDT

PDB ID : 9C7B
Title : Crystal structure of R150H neurodevelopmental disease-associated U2AF2 variant
Authors : Maji, D.; Jenkins, J.L.; Kielkopf, C.L.
Deposited on : 2024-06-10
Resolution : 1.40 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 : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 1.20.1
EDS : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.002 (Gargrove)
Density-Fitness : 1.0.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.38.3

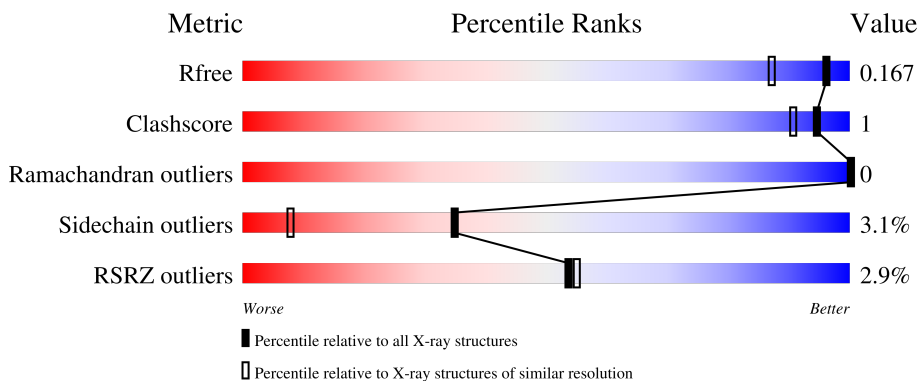
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

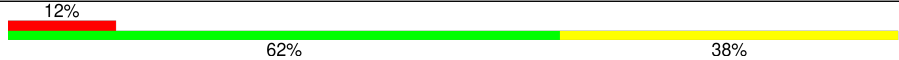
The reported resolution of this entry is 1.40 Å.

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}	164625	2247 (1.40-1.40)
Clashscore	180529	2446 (1.40-1.40)
Ramachandran outliers	177936	2398 (1.40-1.40)
Sidechain outliers	177891	2397 (1.40-1.40)
RSRZ outliers	164620	2246 (1.40-1.40)

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	204	
2	B	8	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 3521 atoms, of which 1578 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 Splicing factor U2AF 65 kDa subunit.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	201	3041	977	1506	258	290	10	0	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	138	GLY	-	expression tag	UNP P26368
A	139	PRO	-	expression tag	UNP P26368
A	140	LEU	-	expression tag	UNP P26368
A	150	HIS	ARG	engineered mutation	UNP P26368

- Molecule 2 is DNA/RNA hybrid called DNA/RNA (5'-R(P*UP*U)-D(P*U)-R(P*UP*U)-D(P*(BRU))-R(P*CP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace	
			Total	Br	C	H	N	O				P
2	B	8	232	1	72	72	18	61	8	0	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	206	Total	O	0	11
			217	217		
3	B	31	Total	O	0	0
			31	31		

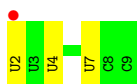
3 Residue-property plots [i](#)

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: Splicing factor U2AF 65 kDa subunit



- Molecule 2: DNA/RNA (5'-R(P*UP*U)-D(P*U)-R(P*UP*U)-D(P*(BRU))-R(P*CP*C)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	43.72Å 55.79Å 76.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.98 – 1.40 37.98 – 1.40	Depositor EDS
% Data completeness (in resolution range)	97.5 (37.98-1.40) 97.5 (37.98-1.40)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.19 (at 1.40Å)	Xtrriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.134 , 0.167 0.135 , 0.167	Depositor DCC
R_{free} test set	35171 reflections (6.65%)	wwPDB-VP
Wilson B-factor (Å ²)	14.8	Xtrriage
Anisotropy	0.284	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.42 , 51.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	3521	wwPDB-VP
Average B, all atoms (Å ²)	23.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.76% 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: BRU

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	0.50	0/1578	0.73	0/2141
2	B	2.97	11/152 (7.2%)	1.31	2/228 (0.9%)
All	All	1.00	11/1730 (0.6%)	0.80	2/2369 (0.1%)

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	4	DU	C3'-C2'	-22.82	1.24	1.52
2	B	4	DU	C5-C6	12.14	1.45	1.34
2	B	2	U	OP3-P	-10.66	1.48	1.61
2	B	4	DU	C2-N3	9.47	1.44	1.37
2	B	4	DU	C2'-C1'	8.75	1.61	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	4	DU	O4'-C4'-C3'	-5.75	102.20	104.50
2	B	4	DU	C4'-C3'-C2'	5.01	107.61	103.10

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	1535	1506	1506	4	0
2	B	160	72	81	0	0
3	A	217	0	0	1	0
3	B	31	0	0	0	0
All	All	1943	1578	1587	4	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 1.

All (4) 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:291:VAL:HG12	1:A:300:LYS:HD2	1.86	0.57
1:A:160:ILE:HD13	1:A:165[A]:MET:HG2	1.94	0.50
1:A:160:ILE:HD13	1:A:165[B]:MET:SD	2.52	0.49
1:A:210:GLN:OE1	3:A:401:HOH:O	2.20	0.46

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	203/204 (100%)	203 (100%)	0	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	164/168 (98%)	158 (96%)	6 (4%)	29 5

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	166	MET
1	A	233	GLN
1	A	328	LYS
1	A	165[A]	MET
1	A	150	HIS

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](#)

1 non-standard protein/DNA/RNA residue 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
2	BRU	B	7	2	18,21,22	1.51	4 (22%)	25,30,33	1.70	7 (28%)

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
2	BRU	B	7	2	-	0/7/21/22	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	7	BRU	C4-N3	-3.37	1.32	1.38
2	B	7	BRU	C6-C5	2.89	1.40	1.34
2	B	7	BRU	C6-N1	-2.73	1.33	1.38
2	B	7	BRU	C2-N1	2.20	1.41	1.38

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	7	BRU	C5-C4-N3	3.78	117.69	113.34
2	B	7	BRU	N3-C2-N1	3.36	119.27	114.89
2	B	7	BRU	O4-C4-C5	-3.25	121.65	125.80
2	B	7	BRU	C4-N3-C2	-2.84	123.62	127.34
2	B	7	BRU	C6-C5-C4	-2.49	118.14	120.67

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no oligosaccharides 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	201/204 (98%)	-0.27	5 (2%) 58 59	9, 18, 46, 65	4 (1%)
2	B	6/8 (75%)	0.12	1 (16%) 5 4	14, 19, 59, 67	0
All	All	207/212 (97%)	-0.26	6 (2%) 54 55	9, 18, 46, 67	4 (1%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	341	ASN	3.2
2	B	2	U	2.5
1	A	239	SER	2.4
1	A	340	LYS	2.2
1	A	298	LEU	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [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(Å ²)	Q<0.9
2	BRU	B	7	20/21	0.99	0.06	12,14,17,18	0

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.