

Full wwPDB X-ray Structure Validation Report (i)

May 23, 2020 – 08:59 pm BST

PDB ID : 5EV4

Title: Structure IV of Intact U2AF65 Recognizing the 3' Splice Site Signal

Authors : Agrawal, A.A.; Jenkins, J.L.; Kielkopf, C.L.

Deposited on : 2015-11-19

Resolution : 1.57 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 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 : 2.11

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)

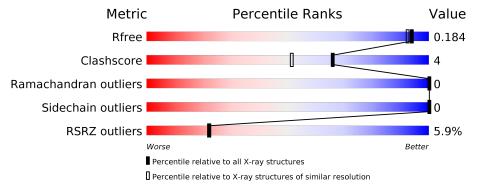
Validation Pipeline (wwPDB-VP) : 2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: X-RAY DIFFRACTION

The reported resolution of this entry is 1.57 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	$\begin{array}{c} \text{Whole archive} \\ (\#\text{Entries}) \end{array}$	$egin{aligned} ext{Similar resolution} \ (\# ext{Entries}, ext{resolution range}(\AA)) \end{aligned}$		
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 on 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	A	201	92%	6% •				
2	В	9	67% 33%					



2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3410 atoms, of which 1549 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.

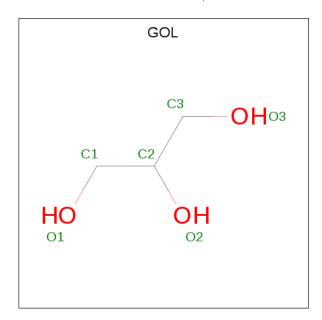
• Molecule 1 is a protein called Splicing factor U2AF 65 kDa subunit.

Mol	Chain	Residues		Atoms					ZeroOcc	AltConf	Trace
1	A	197	Total 2978	C 954	H 1479	N 257	O 281	S 7	0	1	0

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace			
2	В	9	Total 241	Br 1	C 81	H 62	N 19	O 69	P 9	0	0	0

• Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
9	Α	1	Total	С	Н	О	0	0
3	A	1	14	3	8	3	U	U

• Molecule 4 is water.



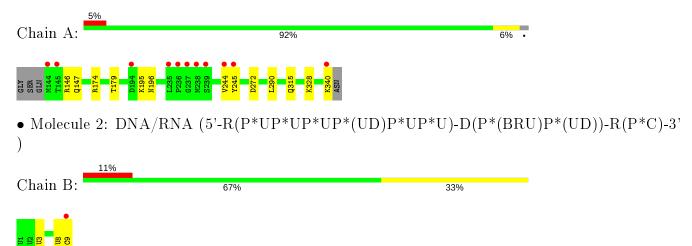
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	154	Total O 154 154	0	0
4	В	23	Total O 23 23	0	0



3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	43.50Å 63.41Å 77.66Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.83 - 1.57	Depositor
Resolution (A)	38.83 - 1.57	EDS
% Data completeness	97.7 (38.83-1.57)	Depositor
(in resolution range)	98.2 (38.83-1.57)	EDS
R_{merge}	0.02	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	4.09 (at 1.57Å)	Xtriage
Refinement program	PHENIX	Depositor
P. P.	0.154 , 0.176	Depositor
R, R_{free}	0.161 , 0.184	DCC
R_{free} test set	2000 reflections (6.63%)	wwPDB-VP
Wilson B-factor (Å ²)	18.8	Xtriage
Anisotropy	0.202	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.42 , 46.0	EDS
L-test for twinning ²	$ < L >=0.48, < L^2>=0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	3410	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 8.60% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical values of <|L|>, $< L^2>$ 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: GOL, 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		
IVIOI	Chain	RMSZ	# Z >5	RMSZ	# Z > 5	
1	A	0.44	0/1531	0.63	0/2076	
2	В	0.75	0/173	1.04	0/260	
All	All	0.48	0/1704	0.69	0/2336	

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	1499	1479	1484	12	0
2	В	179	62	91	6	0
3	A	6	8	8	0	0
4	A	154	0	0	2	1
4	В	23	0	0	0	0
All	All	1861	1549	1583	13	1

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 4.

All (13) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.



Atom-1	Atom-2	$egin{array}{ll} ext{Interatomic} \ ext{distance} & (ext{Å}) \end{array}$	Clash overlap (Å)
1:A:315:GLN:OE1	4:A:501:HOH:O	2.03	0.76
2:B:8:DU:H2"	2:B:9:C:C6	2.21	0.75
1:A:328:LYS:NZ	4:A:504:HOH:O	2.36	0.58
1:A:174:ARG:CZ	1:A:179:THR:HG21	2.37	0.55
1:A:146:ARG:HH12	2:B:9:C:H1'	1.74	0.53
1:A:174:ARG:NH1	1:A:179:THR:HG21	2.26	0.50
1:A:340:LYS:HE2	2:B:3:U:H2'	1.98	0.44
1:A:272:ASP:OD1	1:A:290:LEU:HD23	2.19	0.43
1:A:147:GLN:HE22	2:B:8:DU:H3	1.66	0.43
1:A:195:LYS:O	1:A:196:ASN:HB3	2.20	0.42
1:A:340:LYS:HG2	2:B:3:U:O2'	2.20	0.42
1:A:147:GLN:OE1	2:B:9:C:O2	2.39	0.41
1:A:244:VAL:HG12	1:A:245:TYR:N	2.36	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	$\begin{array}{c} \text{Interatomic} \\ \text{distance (Å)} \end{array}$	$egin{aligned} ext{Clash} \ ext{overlap } (ext{Å}) \end{aligned}$
4:A:520:HOH:O	4:A:606:HOH:O[4_545]	2.13	0.07

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	Perce	Percentiles	
1	A	196/201 (98%)	194 (99%)	2 (1%)	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	158/166 (95%)	158 (100%)	0	100 100		

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	315	GLN

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	Dog	Res Link	Bond lengths			$ \hspace{.05cm} {f B}$	ond ang	les
				nes		Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
	2	BRU	В	7	2	15,21,22	1.86	1 (6%)	17,30,33	1.62	3 (17%)

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	В	7	2	-	0/4/21/22	0/2/2/2

All (1) bond length outliers are listed below:



Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\operatorname{\AA})$	$Ideal(\AA)$
2	В	7	BRU	C4-C5	6.73	1.47	1.38

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\mathbf{Ideal}(^{o})$
2	В	7	BRU	C4-N3-C2	4.99	119.36	115.14
2	В	7	BRU	C5-C4-N3	-3.04	120.01	123.64
2	В	7	BRU	O3'-C3'-C2'	-2.02	103.66	110.90

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 carbohydrates 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 L	Link	Bond lengths			Bond angles		
10101	Type			LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	GOL	A	401	-	5,5,5	0.42	0	5,5,5	1.04	0

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.

Mo	l Type	Chain	${f Res}$	Link	Chirals	Torsions	Rings
3	GOL	A	401	-	-	3/4/4/4	-



There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	401	GOL	C1-C2-C3-O3
3	A	401	GOL	O2-C2-C3-O3
3	A	401	GOL	O1-C1-C2-O2

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, 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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(Å^2)$	Q < 0.9
1	A	197/201 (98%)	0.13	11 (5%) 24 24	13, 22, 44, 59	0
2	В	6/9 (66%)	0.85	1 (16%) 1 1	16, 20, 46, 57	0
All	All	203/210 (96%)	0.15	12 (5%) 22 22	13, 22, 44, 59	0

All (12) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	236	PRO	5.9
2	В	9	С	5.7
1	A	237	GLY	5.3
1	A	244	VAL	4.6
1	A	235	LEU	4.6
1	A	238	MET	4.5
1	A	144	MET	3.4
1	A	239	SER	3.4
1	A	145	THR	3.2
1	A	194	ASP	3.1
1	A	340	LYS	2.8
1	A	245	TYR	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, 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	${ m Res}$	Atoms	RSCC	RSR	${f B\text{-factors}}({f A}^2)$	Q<0.9
2	BRU	В	7	20/21	0.99	0.07	14,17,20,22	0



6.3 Carbohydrates (i)

There are no carbohydrates 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	\mathbf{Type}	Chain	Res	Atoms	RSCC	RSR	${f B-factors}({f A}^2)$	Q<0.9
3	GOL	A	401	6/6	0.90	0.10	23,28,35,42	0

6.5 Other polymers (i)

There are no such residues in this entry.

