Full wwPDB X-ray Structure Validation Report (i)

Oct 10, 2023 – 03:28 AM EDT

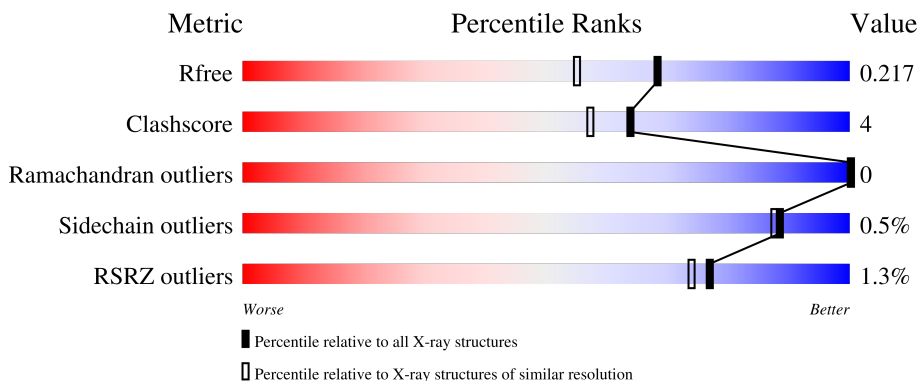|		
PDB ID	:	7SN6
Title	:	U2AF65 UHM BOUND TO SF3B155 ULM5
Authors	:	Loerch, S.; Jenkins, J.L.; Kielkopf, C.L.
Deposited on	:	2021-10-27
Resolution	:	1.80 Å(reported)

1 Overall quality at a glance ⓘ

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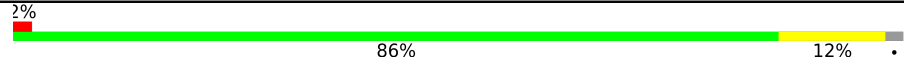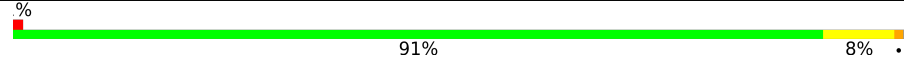*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	Percentile Ranks	Value
Rfree		0.217
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		0.5%
RSRZ outliers		1.3%

Worse   Better

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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	A	106	2% / 86% / 12% / •
1	B	106	% / 91% / 8% / ••
2	C	19	58% / 5% / 37%
2	D	19	47% / 53%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 3578 atoms, of which 1695 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	104	Total	C	H	N	O	S	0	1	0
			1573	517	761	133	153	9			
1	B	105	Total	C	H	N	O	S	0	1	0
			1584	521	768	133	154	8			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	370	GLY	-	expression tag	UNP P26368
A	371	PRO	-	expression tag	UNP P26368
A	372	LEU	-	expression tag	UNP P26368
A	373	GLY	-	expression tag	UNP P26368
A	374	SER	-	expression tag	UNP P26368
B	370	GLY	-	expression tag	UNP P26368
B	371	PRO	-	expression tag	UNP P26368
B	372	LEU	-	expression tag	UNP P26368
B	373	GLY	-	expression tag	UNP P26368
B	374	SER	-	expression tag	UNP P26368

- Molecule 2 is a protein called Splicing factor 3B subunit 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	12	Total	C	H	N	O	S	0	1	0
			192	62	91	20	18	1			
2	D	9	Total	C	H	N	O		0	1	0
			146	48	67	17	14				

- Molecule 3 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C_3H_8O).

IPA

C3

C2

C1

OH

O2
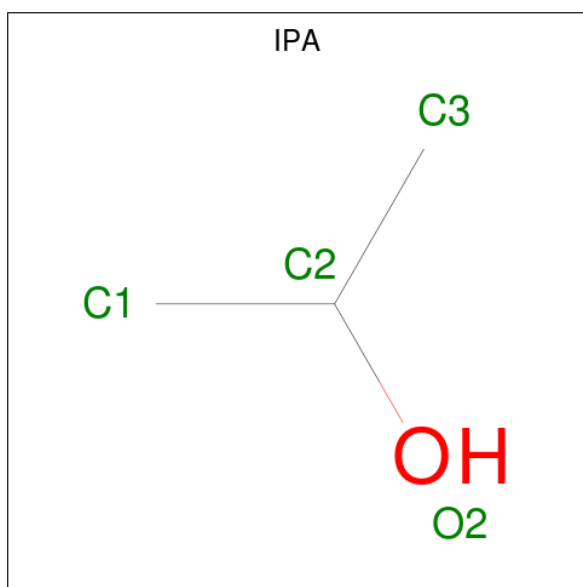
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	B	1	Total 12	C 3	H 8	O 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	38	Total 38	O 38	0	0
4	B	31	Total 31	O 31	0	0
4	C	2	Total 2	O 2	0	0

3 Residue-property plots ⓘ

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 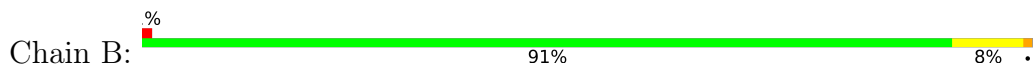(RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.
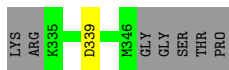
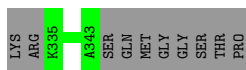• Molecule 1: Splicing factor U2AF 65 kDa subunit
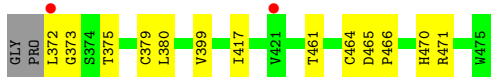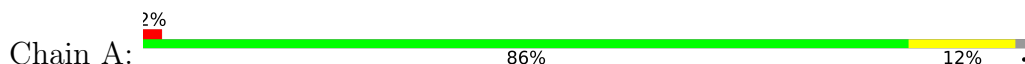
Chain A:

• Molecule 1: Splicing factor U2AF 65 kDa subunit

Chain B:

• Molecule 2: Splicing factor 3B subunit 1

Chain C:

• Molecule 2: Splicing factor 3B subunit 1

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	76.16Å 63.16Å 49.39Å 90.00° 94.37° 90.00°	Depositor
Resolution (Å)	37.97 – 1.80 37.97 – 1.80	Depositor EDS
% Data completeness (in resolution range)	98.1 (37.97-1.80) 98.1 (37.97-1.80)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.72 (at 1.81Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.180 , 0.217 0.180 , 0.217	Depositor DCC
R_{free} test set	2008 reflections (9.43%)	wwPDB-VP
Wilson B-factor (Å^2)	33.3	Xtriage
Anisotropy	0.521	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 52.2	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3578	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.73% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical 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.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

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

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	0.57	0/828	0.68	0/1119
1	B	0.50	0/836	0.63	0/1133
2	C	0.46	0/106	0.55	0/142
2	D	0.40	0/81	0.50	0/110
All	All	0.53	0/1851	0.64	0/2504

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
1	B	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	474	PHE	Peptide

5.2 Too-close contacts ⓘ

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.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	812	761	779	8	0
1	B	816	768	784	6	0
2	C	101	91	93	1	0
2	D	79	67	66	0	0
3	B	4	8	8	0	0
4	A	38	0	0	0	0
4	B	31	0	0	0	0
4	C	2	0	0	0	0
All	All	1883	1695	1730	15	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 4.

All (15) 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:380:LEU:HD23	1:A:461:THR:HG22	1.63	0.81
1:B:462:LYS:CE	1:B:475:TRP:HE1	2.11	0.64
1:A:466:PRO:O	1:A:470:HIS:N	2.31	0.58
2:C:339:ASP:OD1	2:C:339:ASP:C	2.47	0.53
1:A:466:PRO:O	1:A:470:HIS:CG	2.63	0.52
1:B:371:PRO:HG2	1:B:374:SER:OG	2.11	0.51
1:B:462:LYS:HE3	1:B:475:TRP:HE1	1.75	0.51
1:A:399:VAL:HG22	1:A:417:ILE:HD12	1.95	0.49
1:A:379:CYS:HB2	1:A:464[A]:CYS:SG	2.52	0.49
1:B:426:VAL:HB	1:B:427:PRO:HD2	1.95	0.48
1:A:373:GLY:HA2	1:A:375:THR:HG23	1.98	0.45
1:A:372:LEU:HD21	1:A:465:ASP:N	2.32	0.44
1:B:474:PHE:HB3	1:B:475:TRP:HB2	1.99	0.43
1:B:446:MET:O	1:B:450:THR:OG1	2.30	0.42
1:A:466:PRO:O	1:A:470:HIS:CD2	2.73	0.42

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	103/106 (97%)	103 (100%)	0	0	100	100
1	B	104/106 (98%)	103 (99%)	1 (1%)	0	100	100
2	C	11/19 (58%)	11 (100%)	0	0	100	100
2	D	8/19 (42%)	8 (100%)	0	0	100	100
All	All	226/250 (90%)	225 (100%)	1 (0%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2   Protein sidechains   ⓘ

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	88/96 (92%)	87 (99%)	1 (1%)	73	68
1	B	89/96 (93%)	89 (100%)	0	100	100
2	C	10/16 (62%)	10 (100%)	0	100	100
2	D	7/16 (44%)	7 (100%)	0	100	100
All	All	194/224 (87%)	193 (100%)	1 (0%)	88	87

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	471	ARG

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

Mol	Chain	Res	Type
1	A	470	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	IPA	B	501	-	3,3,3	0.52	0	3,3,3	0.44	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

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	104/106 (98%)	-0.01	2 (1%)	66	63	27, 41, 69, 88	0
1	B	105/106 (99%)	-0.10	1 (0%)	82	80	28, 43, 65, 91	0
2	C	12/19 (63%)	-0.09	0	100	100	39, 47, 53, 71	0
2	D	9/19 (47%)	0.42	0	100	100	46, 68, 81, 93	0
All	All	230/250 (92%)	-0.04	3 (1%)	77	74	27, 44, 71, 93	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	475	TRP	5.4
1	A	372	LEU	3.3
1	A	421	VAL	2.0

6.2   Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3   Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4   Ligands ⓘ

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
3	IPA	B	501	4/4	0.82	0.24	65,79,107,107	0

6.5 Other polymers ⓘ

There are no such residues in this entry.