



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 24, 2016 – 06:14 PM EDT

PDB ID : 5IFE
Title : Crystal structure of the human SF3b core complex
Authors : Cretu, C.; Dybkov, O.; De Laurentiis, E.; Will, C.L.; Luhrmann, R.; Pena, V.
Deposited on : 2016-02-25
Resolution : 3.10 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

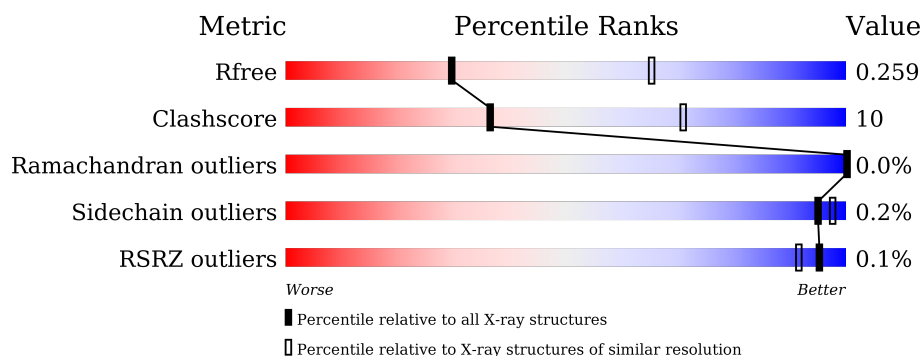
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 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}	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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. 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	B	86	<div> <div>67%</div> <div>9%</div> <div>23%</div> </div>
2	C	1304	<div> <div>49%</div> <div>15%</div> <div>36%</div> </div>
3	D	120	<div> <div>62%</div> <div>13%</div> <div>26%</div> </div>
4	A	1235	<div> <div>70%</div> <div>25%</div> <div>.</div> </div>

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 17103 atoms, of which 0 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 3B subunit 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	66	Total	C	N	O	S	0	0	0
			540	343	94	98	5			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	829	Total	C	N	O	S	0	0	0
			6602	4242	1132	1189	39			

- Molecule 3 is a protein called PHD finger-like domain-containing protein 5A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	89	Total	C	N	O	S	0	0	0
			670	410	119	128	13			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-9	GLY	-	expression tag	UNP Q7RTV0
D	-8	PRO	-	expression tag	UNP Q7RTV0
D	-7	LEU	-	expression tag	UNP Q7RTV0
D	-6	GLY	-	expression tag	UNP Q7RTV0
D	-5	SER	-	expression tag	UNP Q7RTV0
D	-4	PRO	-	expression tag	UNP Q7RTV0
D	-3	GLY	-	expression tag	UNP Q7RTV0
D	-2	SER	-	expression tag	UNP Q7RTV0
D	-1	ARG	-	expression tag	UNP Q7RTV0
D	0	ALA	-	expression tag	UNP Q7RTV0

- Molecule 4 is a protein called Splicing factor 3B subunit 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	1184	Total	C	N	O	S	0	0	0
			9287	5894	1576	1772	45			

There are 18 discrepancies between the modelled and reference sequences:

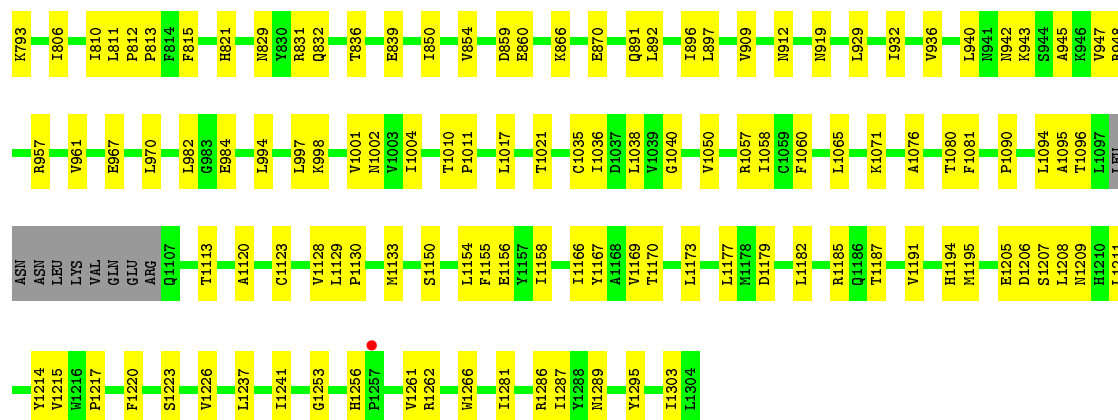
Chain	Residue	Modelled	Actual	Comment	Reference
A	-9	GLY	-	expression tag	UNP Q15393
A	-8	ALA	-	expression tag	UNP Q15393
A	-7	GLU	-	expression tag	UNP Q15393
A	-6	PHE	-	expression tag	UNP Q15393
A	-5	LYS	-	expression tag	UNP Q15393
A	-4	GLY	-	expression tag	UNP Q15393
A	-3	LEU	-	expression tag	UNP Q15393
A	-2	ARG	-	expression tag	UNP Q15393
A	-1	VAL	-	expression tag	UNP Q15393
A	0	ASP	-	expression tag	UNP Q15393
A	1218	ASP	-	expression tag	UNP Q15393
A	1219	TYR	-	expression tag	UNP Q15393
A	1220	LYS	-	expression tag	UNP Q15393
A	1221	ASP	-	expression tag	UNP Q15393
A	1222	ASP	-	expression tag	UNP Q15393
A	1223	ASP	-	expression tag	UNP Q15393
A	1224	ASP	-	expression tag	UNP Q15393
A	1225	LYS	-	expression tag	UNP Q15393

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	3	Total	Zn	0	0
			3	3		

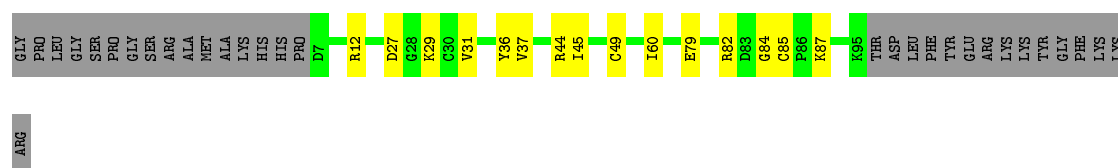
- Molecule 6 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	K	0	0
			1	1		



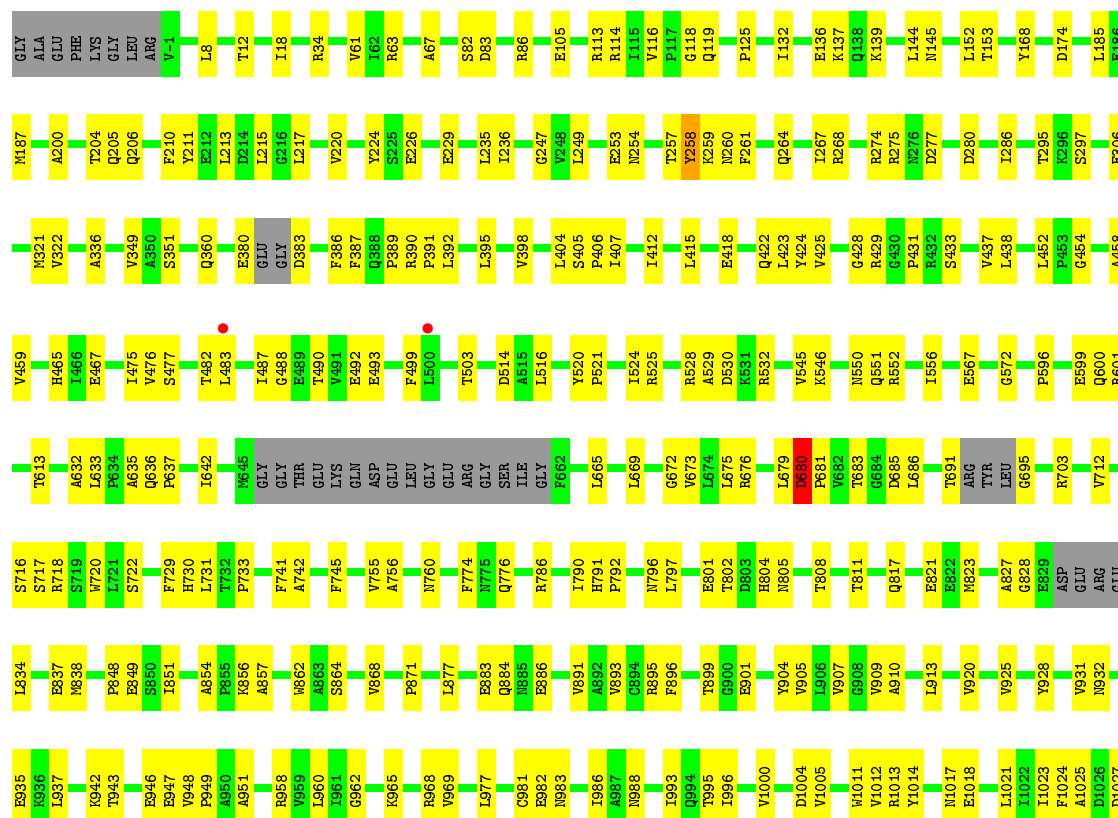
• Molecule 3: PHD finger-like domain-containing protein 5A

Chain D: 62% 13% 26%



• Molecule 4: Splicing factor 3B subunit 3

Chain A: 70% 25%





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	105.06Å 154.44Å 210.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.04 – 3.10 47.04 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.8 (47.04-3.10) 99.9 (47.04-3.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.70 (at 3.12Å)	Xtriage
Refinement program	PHENIX (1.10.1 _2155: ???)	Depositor
R, R_{free}	0.231 , 0.260 0.226 , 0.259	Depositor DCC
R_{free} test set	3105 reflections (4.95%)	DCC
Wilson B-factor (Å ²)	79.9	Xtriage
Anisotropy	0.313	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 39.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	17103	wwPDB-VP
Average B, all atoms (Å ²)	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.62% 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

5.1 Standard geometry

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

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	B	0.24	0/556	0.42	0/751
2	C	0.24	0/6727	0.40	0/9106
3	D	0.25	0/678	0.43	0/909
4	A	0.25	0/9476	0.46	0/12858
All	All	0.25	0/17437	0.43	0/23624

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

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	B	540	0	509	9	0
2	C	6602	0	6800	126	0
3	D	670	0	653	9	0
4	A	9287	0	9201	197	0
5	D	3	0	0	0	0
6	A	1	0	0	0	0
All	All	17103	0	17163	327	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 10.

The worst 5 of 327 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)
2:C:1173:LEU:O	2:C:1177:LEU:HB2	1.64	0.98
2:C:1289:ASN:HB3	2:C:1295:TYR:H	1.39	0.85
2:C:665:ILE:HG23	2:C:690:ILE:HD11	1.63	0.80
2:C:672:ALA:HA	2:C:679:ILE:HD11	1.63	0.80
4:A:1103:SER:OG	4:A:1120:THR:HG22	1.86	0.76

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	B	64/86 (74%)	59 (92%)	5 (8%)	0	100	100
2	C	823/1304 (63%)	806 (98%)	17 (2%)	0	100	100
3	D	87/120 (72%)	79 (91%)	8 (9%)	0	100	100
4	A	1172/1235 (95%)	1104 (94%)	67 (6%)	1 (0%)	56	88
All	All	2146/2745 (78%)	2048 (95%)	97 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	680	ASP

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	B	57/77 (74%)	57 (100%)	0	100	100
2	C	715/1104 (65%)	714 (100%)	1 (0%)	95	98
3	D	76/101 (75%)	76 (100%)	0	100	100
4	A	1027/1066 (96%)	1024 (100%)	3 (0%)	94	97
All	All	1875/2348 (80%)	1871 (100%)	4 (0%)	95	97

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	1081	PHE
4	A	258	TYR
4	A	680	ASP
4	A	1166	TYR

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

Mol	Chain	Res	Type
4	A	760	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

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 [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	B	66/86 (76%)	-0.02	0 100 100	50, 69, 92, 106	0
2	C	829/1304 (63%)	-0.25	1 (0%) 95 91	51, 79, 110, 135	0
3	D	89/120 (74%)	-0.36	0 100 100	45, 61, 88, 106	0
4	A	1184/1235 (95%)	-0.18	2 (0%) 95 91	36, 74, 114, 149	0
All	All	2168/2745 (78%)	-0.21	3 (0%) 95 91	36, 75, 112, 149	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	500	LEU	2.7
2	C	1257	PRO	2.6
4	A	483	LEU	2.4

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 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
5	ZN	D	202	1/1	0.98	0.15	-0.56	52,52,52,52	0
5	ZN	D	201	1/1	0.98	0.13	-0.60	68,68,68,68	0
5	ZN	D	203	1/1	1.00	0.14	-0.61	59,59,59,59	0
6	K	A	1301	1/1	0.84	0.18	-	109,109,109,109	0

6.5 Other polymers [i](#)

There are no such residues in this entry.