



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2018 – 10:55 pm GMT

PDB ID : 2F9D  
Title : 2.5 angstrom resolution structure of the spliceosomal protein p14 bound to region of SF3b155  
Authors : Schellenberg, M.J.; Edwards, R.A.; Ritchie, D.B.; Glover, J.N.M.; Macmillan, A.M.  
Deposited on : 2005-12-05  
Resolution : 2.50 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](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	trunk30967
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30967

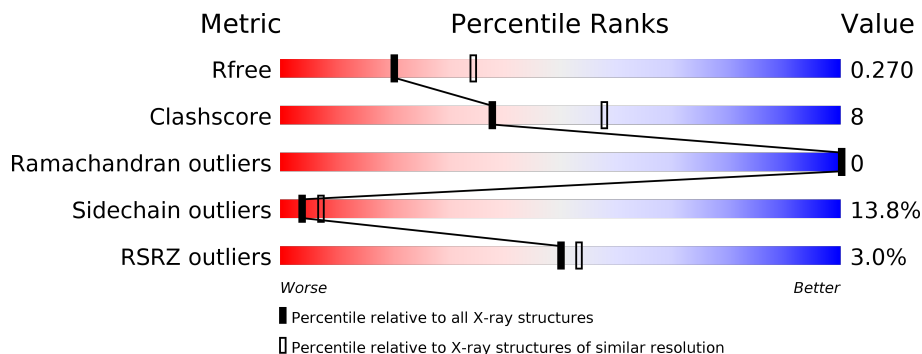
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
$R_{free}$	111664	4155 (2.50-2.50)
Clashscore	122126	4827 (2.50-2.50)
Ramachandran outliers	120053	4735 (2.50-2.50)
Sidechain outliers	120020	4737 (2.50-2.50)
RSRZ outliers	108989	4058 (2.50-2.50)

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  $\geq 3$ , 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	A	125	<div> <div>2%</div> <div> <div></div> <div>67%</div> <div>18%</div> <div>6%</div> <div>9%</div> </div> </div>
1	B	125	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div>22%</div> <div>6%</div> <div>9%</div> </div> </div>
2	P	43	<div> <div>2%</div> <div> <div></div> <div>74%</div> <div>12%</div> <div>5%</div> <div>9%</div> </div> </div>
2	Q	43	<div> <div>2%</div> <div> <div></div> <div>63%</div> <div>23%</div> <div>5%</div> <div>9%</div> </div> </div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2627 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 Pre-mRNA branch site protein p14.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	114	Total	C	N	O	S	0	0	0
			945	606	163	172	4			
1	B	114	Total	C	N	O	S	0	0	0
			945	606	163	172	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	39	Total	C	N	O	Se	0	0	0
			335	209	57	67	2			
2	Q	39	Total	C	N	O	Se	0	0	0
			335	209	57	67	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	376	MSE	MET	MODIFIED RESIDUE	UNP O75533
P	378	MSE	MET	MODIFIED RESIDUE	UNP O75533
P	407	MSE	MET	MODIFIED RESIDUE	UNP O75533
Q	376	MSE	MET	MODIFIED RESIDUE	UNP O75533
Q	378	MSE	MET	MODIFIED RESIDUE	UNP O75533
Q	407	MSE	MET	MODIFIED RESIDUE	UNP O75533

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	27	Total	O	0	0
			27	27		
3	B	28	Total	O	0	0
			28	28		
3	P	8	Total	O	0	0
			8	8		

*Continued on next page...*

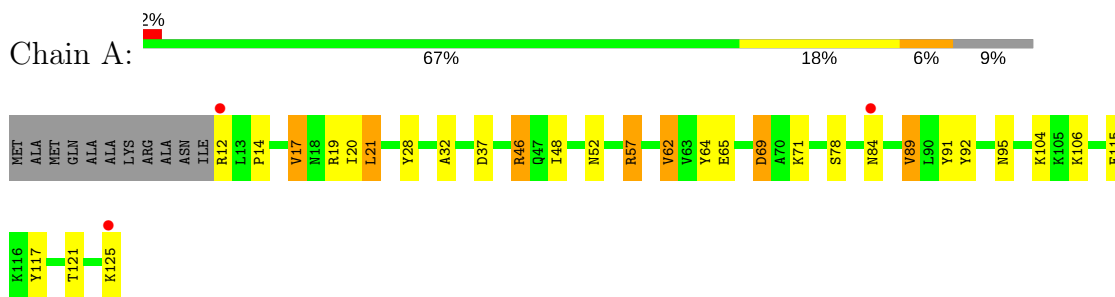
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	Q	4	Total	O	0	0
			4	4		

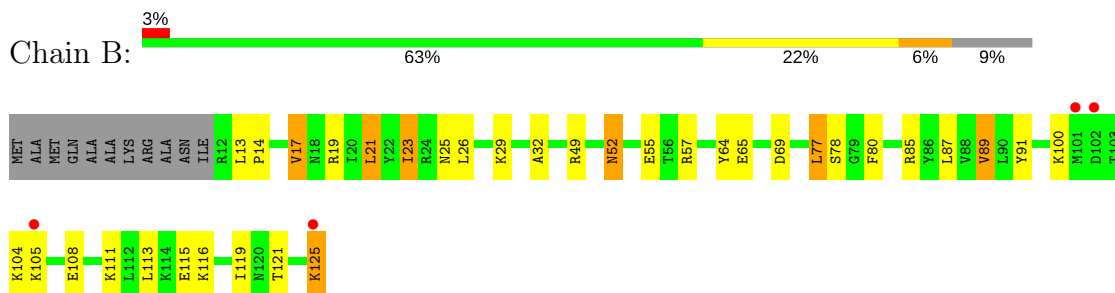
### 3 Residue-property plots [i](#)

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.

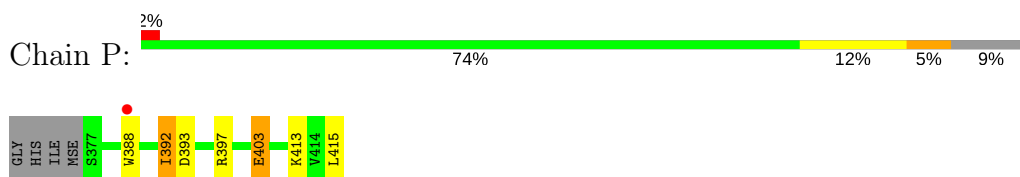
- Molecule 1: Pre-mRNA branch site protein p14



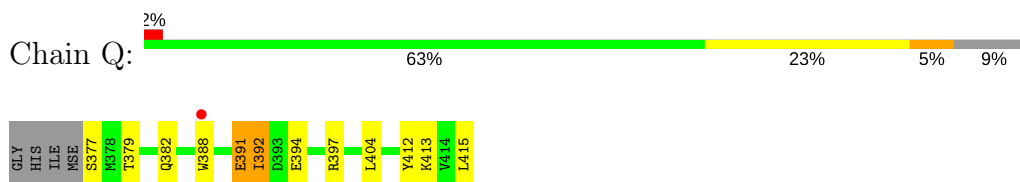
- Molecule 1: Pre-mRNA branch site protein p14



- Molecule 2: Splicing factor 3B subunit 1



- Molecule 2: Splicing factor 3B subunit 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.87Å 115.23Å 82.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	76.25 – 2.50 56.04 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.6 (76.25-2.50) 99.6 (56.04-2.50)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.18 (at 2.51Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.221 , 0.276 0.218 , 0.270	Depositor DCC
$R_{free}$ test set	858 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	43.0	Xtriage
Anisotropy	0.870	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 37.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	2627	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.66% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

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	0.71	0/966	0.82	1/1303 (0.1%)
1	B	0.67	0/966	0.82	1/1303 (0.1%)
2	P	0.83	1/341 (0.3%)	0.86	0/456
2	Q	0.75	0/341	0.74	0/456
All	All	0.72	1/2614 (0.0%)	0.81	2/3518 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	388	TRP	CB-CG	6.44	1.61	1.50

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	69	ASP	CB-CG-OD1	5.62	123.36	118.30
1	A	69	ASP	CB-CG-OD1	5.36	123.12	118.30

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	A	945	0	951	18	0
1	B	945	0	951	20	0
2	P	335	0	311	3	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	Q	335	0	311	7	0
3	A	27	0	0	2	0
3	B	28	0	0	1	0
3	P	8	0	0	0	0
3	Q	4	0	0	0	0
All	All	2627	0	2524	42	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 8.

All (42) 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:57:ARG:HD2	3:A:137:HOH:O	1.77	0.84
1:B:111:LYS:O	1:B:115:GLU:HG2	1.91	0.71
2:Q:379:THR:HB	2:Q:382:GLN:HB2	1.73	0.70
1:B:113:LEU:HB3	1:B:119:ILE:HD11	1.74	0.68
1:A:19:ARG:HD3	3:A:146:HOH:O	1.98	0.63
1:B:57:ARG:HD2	3:B:150:HOH:O	2.02	0.58
1:B:25:ASN:O	1:B:85:ARG:HD3	2.04	0.57
1:B:14:PRO:O	1:B:17:VAL:HG13	2.08	0.54
2:Q:388:TRP:CZ2	2:Q:392:ILE:HG12	2.42	0.54
1:B:17:VAL:HB	1:B:91:TYR:CZ	2.43	0.54
1:A:17:VAL:HB	1:A:91:TYR:CZ	2.44	0.52
1:B:19:ARG:NH1	2:Q:397:ARG:O	2.38	0.51
2:Q:379:THR:HB	2:Q:382:GLN:CB	2.42	0.50
2:Q:388:TRP:CZ3	2:Q:391:GLU:HG2	2.45	0.50
1:A:117:TYR:CE1	2:P:403:GLU:HG2	2.47	0.49
1:A:78:SER:HA	1:A:89:VAL:HG13	1.95	0.49
1:A:14:PRO:O	1:A:17:VAL:HG13	2.13	0.49
1:A:125:LYS:HE2	1:A:125:LYS:HA	1.93	0.49
1:B:23:ILE:HD11	1:B:87:LEU:HD22	1.95	0.49
1:A:28:TYR:CZ	1:A:57:ARG:HD3	2.49	0.48
1:B:21:LEU:HD22	1:B:64:TYR:HE1	1.78	0.48
1:A:46:ARG:HG2	1:A:65:GLU:CD	2.34	0.48
1:B:115:GLU:HB2	1:B:116:LYS:HE2	1.95	0.48
1:B:52:ASN:H	1:B:52:ASN:HD22	1.60	0.47
1:B:32:ALA:HA	2:Q:415:LEU:HD13	1.98	0.46
1:B:125:LYS:H	1:B:125:LYS:HD3	1.81	0.46
1:A:48:ILE:HG12	1:A:62:VAL:HB	1.98	0.46
1:B:49:ARG:HB3	2:Q:412:TYR:HB3	1.98	0.45

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:P:392:ILE:HD13	2:P:393:ASP:N	2.31	0.44
1:A:20:ILE:HG22	1:A:92:TYR:HB2	1.99	0.44
1:A:84:ASN:O	1:A:84:ASN:OD1	2.36	0.44
1:A:71:LYS:HA	1:A:91:TYR:OH	2.18	0.44
1:A:46:ARG:HG2	1:A:65:GLU:OE2	2.18	0.43
1:A:32:ALA:HA	2:P:415:LEU:HD13	2.00	0.43
1:A:21:LEU:HD22	1:A:64:TYR:HE1	1.84	0.43
1:B:52:ASN:N	1:B:52:ASN:HD22	2.17	0.43
1:B:13:LEU:HA	1:B:14:PRO:HD3	1.88	0.42
1:B:125:LYS:N	1:B:125:LYS:HD3	2.35	0.42
1:B:77:LEU:O	1:B:80:PHE:HB2	2.19	0.42
1:A:95:ASN:OD1	1:A:106:LYS:NZ	2.46	0.41
1:A:69:ASP:OD2	1:B:85:ARG:NH2	2.54	0.41
1:B:78:SER:HA	1:B:89:VAL:HG13	2.03	0.41

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	112/125 (90%)	109 (97%)	3 (3%)	0	100	100
1	B	112/125 (90%)	107 (96%)	5 (4%)	0	100	100
2	P	37/43 (86%)	36 (97%)	1 (3%)	0	100	100
2	Q	37/43 (86%)	37 (100%)	0	0	100	100
All	All	298/336 (89%)	289 (97%)	9 (3%)	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	102/109 (94%)	90 (88%)	12 (12%)	6	11
1	B	102/109 (94%)	86 (84%)	16 (16%)	3	5
2	P	36/36 (100%)	32 (89%)	4 (11%)	7	13
2	Q	36/36 (100%)	30 (83%)	6 (17%)	2	4
All	All	276/290 (95%)	238 (86%)	38 (14%)	4	7

All (38) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	12	ARG
1	A	17	VAL
1	A	21	LEU
1	A	37	ASP
1	A	46	ARG
1	A	52	ASN
1	A	57	ARG
1	A	62	VAL
1	A	89	VAL
1	A	104	LYS
1	A	115	GLU
1	A	121	THR
1	B	17	VAL
1	B	21	LEU
1	B	23	ILE
1	B	26	LEU
1	B	29	LYS
1	B	52	ASN
1	B	55	GLU
1	B	65	GLU
1	B	77	LEU
1	B	89	VAL
1	B	100	LYS
1	B	104	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	105	LYS
1	B	108	GLU
1	B	121	THR
1	B	125	LYS
2	P	392	ILE
2	P	397	ARG
2	P	403	GLU
2	P	413	LYS
2	Q	377	SER
2	Q	391	GLU
2	Q	392	ILE
2	Q	394	GLU
2	Q	404	LEU
2	Q	413	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	52	ASN
1	B	52	ASN
1	B	72	ASN
2	Q	384	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](#)

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

### 5.5 Carbohydrates [i](#)

There are no carbohydrates 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	114/125 (91%)	0.15	3 (2%) 56 59	32, 39, 56, 71	0
1	B	114/125 (91%)	0.34	4 (3%) 44 47	32, 39, 56, 61	0
2	P	37/43 (86%)	-0.05	1 (2%) 54 58	35, 49, 56, 58	0
2	Q	37/43 (86%)	0.19	1 (2%) 54 58	39, 54, 92, 94	0
All	All	302/336 (89%)	0.20	9 (2%) 50 53	32, 42, 59, 94	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	125	LYS	3.5
1	B	125	LYS	3.3
2	P	388	TRP	3.2
1	B	102	ASP	2.8
1	B	105	LYS	2.2
1	A	84	ASN	2.2
2	Q	388	TRP	2.1
1	A	12	ARG	2.1
1	B	101	MET	2.0

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.