



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 9, 2020 – 12:12 AM BST

PDB ID : 6PPQ  
Title : Structure of *S. pombe* Lsm1-7 with RNA, polyuridine with 3' adenosine  
Authors : Montemayor, E.J.; Butcher, S.E.  
Deposited on : 2019-07-08  
Resolution : 1.81 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](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 : 2.13.1  
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) : 2.13.1

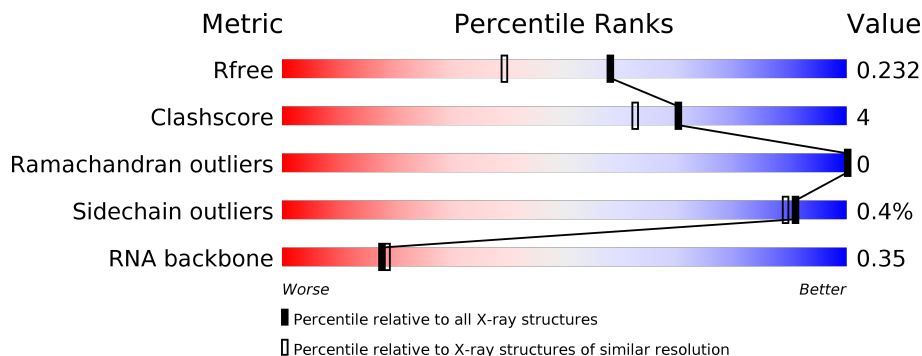
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.81 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))
$R_{free}$	130704	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RNA backbone	3102	1047 (2.40-1.24)

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 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	86	78% (green), 20% (grey), 2% (yellow), 0% (orange), 0% (red)
2	B	96	85% (green), 10% (yellow), 5% (orange), 0% (red), 0% (grey)
3	C	95	72% (green), 7% (yellow), 21% (grey), 0% (orange), 0% (red)
4	D	129	51% (green), 12% (yellow), 36% (grey), 0% (orange), 0% (red)
5	E	80	93% (green), 6% (yellow), 1% (orange), 0% (red), 0% (grey)
6	F	77	86% (green), 6% (yellow), 8% (grey), 0% (orange), 0% (red)

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
7	G	119	 66% 5% 29%
8	H	6	 67% 17% 17%

## 2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 4677 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 U6 snRNA-associated Sm-like protein LSm1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	69	556	360	91	103	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP P87173
A	0	SER	-	expression tag	UNP P87173

- Molecule 2 is a protein called U6 snRNA-associated Sm-like protein LSm2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	93	727	469	121	132	5	0	0	0

- Molecule 3 is a protein called Probable U6 snRNA-associated Sm-like protein LSm3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	75	599	383	104	110	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	GLY	-	expression tag	UNP Q9Y7M4
C	0	SER	-	expression tag	UNP Q9Y7M4

- Molecule 4 is a protein called Probable U6 snRNA-associated Sm-like protein LSm4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	82	630	397	111	117	5	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	122	TRP	-	expression tag	UNP O14352
D	123	SER	-	expression tag	UNP O14352
D	124	HIS	-	expression tag	UNP O14352
D	125	PRO	-	expression tag	UNP O14352
D	126	GLN	-	expression tag	UNP O14352
D	127	PHE	-	expression tag	UNP O14352
D	128	GLU	-	expression tag	UNP O14352
D	129	LYS	-	expression tag	UNP O14352

- Molecule 5 is a protein called U6 snRNA-associated Sm-like protein LSm5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	75	572	366	89	111	6	0	0	0

- Molecule 6 is a protein called U6 snRNA-associated Sm-like protein LSm6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	71	544	346	91	105	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	GLY	-	expression tag	UNP Q9UUI1
F	0	SER	-	expression tag	UNP Q9UUI1

- Molecule 7 is a protein called U6 snRNA-associated Sm-like protein LSm7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	G	85	645	406	115	122	2	0	1	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	114	HIS	-	expression tag	UNP O74499
G	115	HIS	-	expression tag	UNP O74499
G	116	HIS	-	expression tag	UNP O74499
G	117	HIS	-	expression tag	UNP O74499
G	118	HIS	-	expression tag	UNP O74499

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
G	119	HIS	-	expression tag	UNP O74499

- Molecule 8 is a RNA chain called RNA (5'-R(\*UP\*UP\*UP\*UP\*UP\*A)-3').

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
			Total	C	N	O				P
8	H	5	102	46	13	38	5	0	0	0

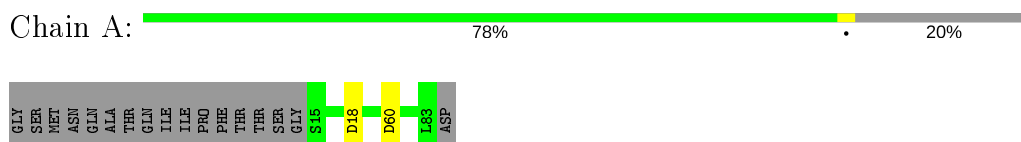
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	41	Total	O	0	0
			41	41		
9	B	34	Total	O	0	0
			34	34		
9	C	51	Total	O	0	0
			51	51		
9	D	36	Total	O	0	0
			36	36		
9	E	51	Total	O	0	0
			51	51		
9	F	31	Total	O	0	0
			31	31		
9	G	46	Total	O	0	0
			46	46		
9	H	12	Total	O	0	0
			12	12		

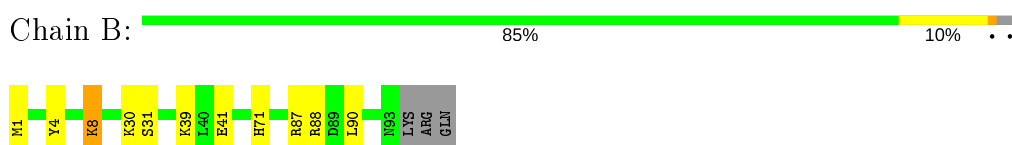
### 3 Residue-property plots [i](#)

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. 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. 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.

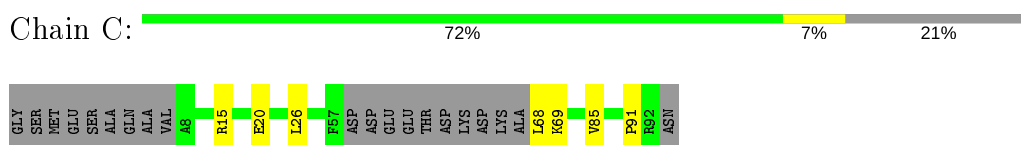
- Molecule 1: U6 snRNA-associated Sm-like protein LSm1



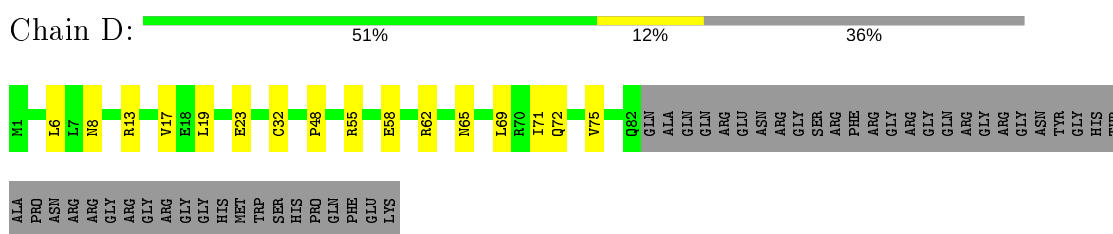
- Molecule 2: U6 snRNA-associated Sm-like protein LSm2



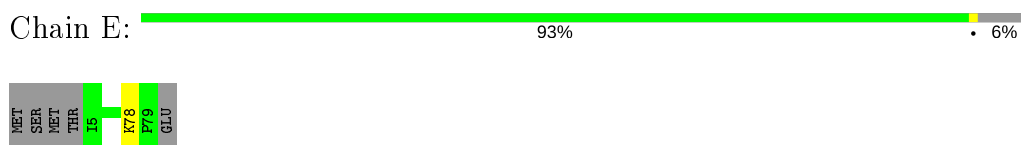
- Molecule 3: Probable U6 snRNA-associated Sm-like protein LSm3




- Molecule 4: Probable U6 snRNA-associated Sm-like protein LSm4



- Molecule 5: U6 snRNA-associated Sm-like protein LSm5



- Molecule 6: U6 snRNA-associated Sm-like protein LSm6

Chain F:  86% 6% 8%



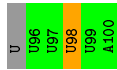
- Molecule 7: U6 snRNA-associated Sm-like protein LSm7

Chain G:  66% 5% 29%



- Molecule 8: RNA (5'-R(\*UP\*UP\*UP\*UP\*UP\*A)-3')

Chain H:  67% 17% 17%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.86 Å 68.86 Å 296.30 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	34.50 – 1.81 98.77 – 1.81	Depositor EDS
% Data completeness (in resolution range)	100.0 (34.50-1.81) 99.9 (98.77-1.81)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.93 (at 1.81 Å)	Xtrriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	(Not available) , (Not available) 0.207 , 0.232	Depositor DCC
$R_{free}$ test set	1995 reflections (2.63%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.3	Xtrriage
Anisotropy	0.349	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 58.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.026 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	4677	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.56% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.35	0/562	0.62	0/758
2	B	0.33	0/738	0.56	0/998
3	C	0.32	0/607	0.59	0/820
4	D	0.38	0/639	0.56	0/867
5	E	0.36	0/581	0.62	0/787
6	F	0.37	0/551	0.60	0/744
7	G	0.33	0/654	0.55	0/885
8	H	0.58	0/112	1.19	1/171 (0.6%)
All	All	0.36	0/4444	0.61	1/6030 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	H	98	U	C5-C6-N1	-5.77	119.82	122.70

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	556	0	570	2	0
2	B	727	0	736	10	0
3	C	599	0	610	5	0
4	D	630	0	615	11	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	E	572	0	573	1	0
6	F	544	0	542	2	0
7	G	645	0	661	5	0
8	H	102	0	52	0	0
9	A	41	0	0	0	0
9	B	34	0	0	3	0
9	C	51	0	0	0	0
9	D	36	0	0	1	0
9	E	51	0	0	1	0
9	F	31	0	0	0	0
9	G	46	0	0	0	0
9	H	12	0	0	0	0
All	All	4677	0	4359	31	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.

The worst 5 of 31 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:B:30:LYS:NZ	2:B:41:GLU:OE1	2.21	0.74
5:E:78:LYS:NZ	9:E:101:HOH:O	2.21	0.73
7:G:93:LEU:HD13	7:G:96:ILE:HD11	1.81	0.61
2:B:88:ARG:NH2	9:B:102:HOH:O	2.29	0.61
3:C:68:LEU:HG	3:C:69:LYS:HD3	1.88	0.55

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	67/86 (78%)	65 (97%)	2 (3%)	0	100	100
2	B	91/96 (95%)	91 (100%)	0	0	100	100
3	C	71/95 (75%)	71 (100%)	0	0	100	100
4	D	80/129 (62%)	78 (98%)	2 (2%)	0	100	100
5	E	73/80 (91%)	71 (97%)	2 (3%)	0	100	100
6	F	69/77 (90%)	68 (99%)	1 (1%)	0	100	100
7	G	84/119 (71%)	82 (98%)	2 (2%)	0	100	100
All	All	535/682 (78%)	526 (98%)	9 (2%)	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	61/76 (80%)	61 (100%)	0	100	100
2	B	79/87 (91%)	78 (99%)	1 (1%)	69	61
3	C	65/83 (78%)	65 (100%)	0	100	100
4	D	66/110 (60%)	66 (100%)	0	100	100
5	E	65/71 (92%)	65 (100%)	0	100	100
6	F	59/66 (89%)	58 (98%)	1 (2%)	60	50
7	G	69/104 (66%)	69 (100%)	0	100	100
All	All	464/597 (78%)	462 (100%)	2 (0%)	91	89

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	8	LYS
6	F	44	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	65	ASN
7	G	45	GLN
7	G	32	GLN
4	D	8	ASN
7	G	34	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	H	4/6 (66%)	1 (25%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	H	98	U

There are no RNA pucker outliers to report.

## 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 monosaccharides 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

### 6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.