



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 11:33 pm GMT

PDB ID : 4M7A  
Title : Crystal structure of Lsm2-8 complex bound to the 3' end sequence of U6 snRNA  
Authors : Zhou, L.; Hang, J.; Zhou, Y.; Wan, R.; Lu, G.; Yan, C.; Shi, Y.  
Deposited on : 2013-08-12  
Resolution : 2.78 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

<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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

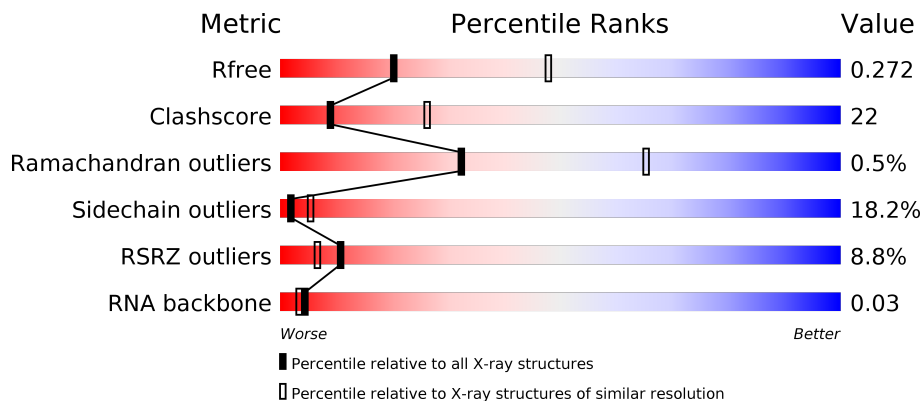
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.78 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))
$R_{free}$	100719	3276 (2.80-2.76)
Clashscore	112137	3771 (2.80-2.76)
Ramachandran outliers	110173	3707 (2.80-2.76)
Sidechain outliers	110143	3709 (2.80-2.76)
RSRZ outliers	101464	3307 (2.80-2.76)
RNA backbone	2435	1019 (3.10-2.46)

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	96	<div> <div>5%</div> <div> <div>33%</div> <div>26%</div> <div>7%</div> <div>33%</div> </div> </div>
1	H	96	<div> <div>22%</div> <div> <div>22%</div> <div>31%</div> <div>10%</div> <div>35%</div> </div> </div>
2	B	95	<div> <div>11%</div> <div> <div>51%</div> <div>37%</div> <div>7%</div> <div>5%</div> </div> </div>
2	I	95	<div> <div>12%</div> <div> <div>35%</div> <div>45%</div> <div>13%</div> <div>7%</div> </div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	C	89	
3	J	89	
4	D	86	
4	K	86	
5	E	93	
5	L	93	
6	F	115	
6	M	115	
7	G	93	
7	N	93	
8	O	8	
8	P	8	

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 8295 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 LSm8.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	64	Total	C	N	O	0	0	0
			497	320	85	92			
1	H	62	Total	C	N	O	0	0	0
			483	311	83	89			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	17	LEU	LYS	ENGINEERED MUTATION	UNP P47093
A	22	SER	CYS	ENGINEERED MUTATION	UNP P47093
A	38	LEU	ILE	ENGINEERED MUTATION	UNP P47093
A	51	SER	CYS	ENGINEERED MUTATION	UNP P47093
H	17	LEU	LYS	ENGINEERED MUTATION	UNP P47093
H	22	SER	CYS	ENGINEERED MUTATION	UNP P47093
H	38	LEU	ILE	ENGINEERED MUTATION	UNP P47093
H	51	SER	CYS	ENGINEERED MUTATION	UNP P47093

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	90	Total	C	N	O	S	0	0	0
			735	469	124	140	2			
2	I	88	Total	C	N	O	S	0	0	0
			721	462	122	134	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	45	SER	CYS	ENGINEERED MUTATION	UNP P38203
I	45	SER	CYS	ENGINEERED MUTATION	UNP P38203

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	77	Total 611	C 382	N 105	O 123	S 1	0	0	0
3	J	77	Total 611	C 382	N 105	O 123	S 1	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	37	SER	CYS	ENGINEERED MUTATION	UNP P57743
C	63	SER	CYS	ENGINEERED MUTATION	UNP P57743
J	37	SER	CYS	ENGINEERED MUTATION	UNP P57743
J	63	SER	CYS	ENGINEERED MUTATION	UNP P57743

- Molecule 4 is a protein called U6 snRNA-associated Sm-like protein LSm6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	74	Total 577	C 364	N 95	O 116	S 2	0	0	0
4	K	74	Total 556	C 351	N 93	O 110	S 2	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	75	Total 588	C 378	N 98	O 110	S 2	0	0	0
5	L	73	Total 578	C 372	N 96	O 108	S 2	0	0	0

- Molecule 6 is a protein called U6 snRNA-associated Sm-like protein LSm7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	66	Total 504	C 325	N 85	O 91	S 3	0	0	0
6	M	65	Total 496	C 321	N 83	O 89	S 3	0	0	0

- Molecule 7 is a protein called U6 snRNA-associated Sm-like protein LSm4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	74	Total 588	C 381	N 96	O 108	S 3	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	N	67	Total	C	N	O	S	0	0	0
			544	353	88	100	3			

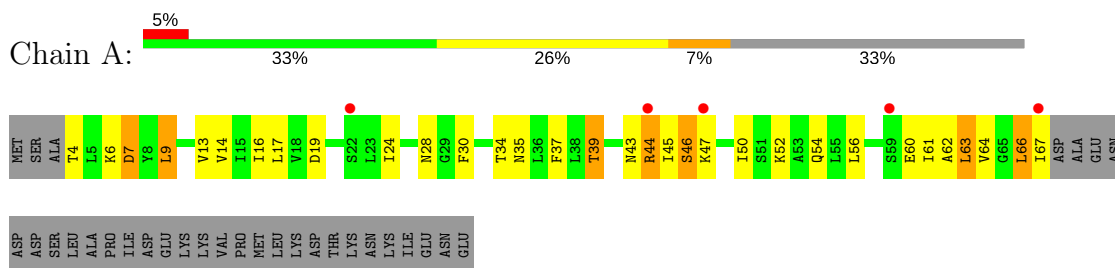
- Molecule 8 is a RNA chain called U6 snRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	O	5	Total	C	N	O	P	0	0	0
			103	46	13	39	5			
8	P	5	Total	C	N	O	P	0	0	0
			103	46	13	39	5			

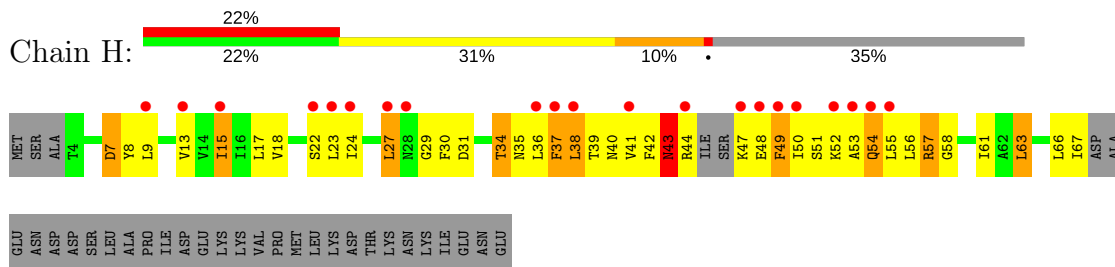
### 3 Residue-property plots [i](#)

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.

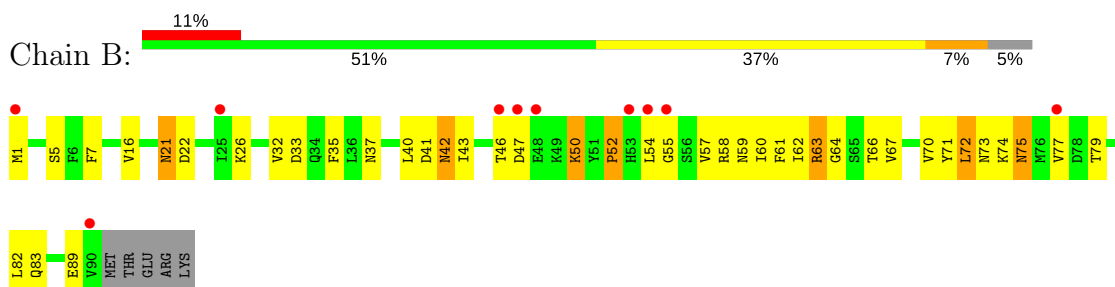
- Molecule 1: U6 snRNA-associated Sm-like protein LSm8



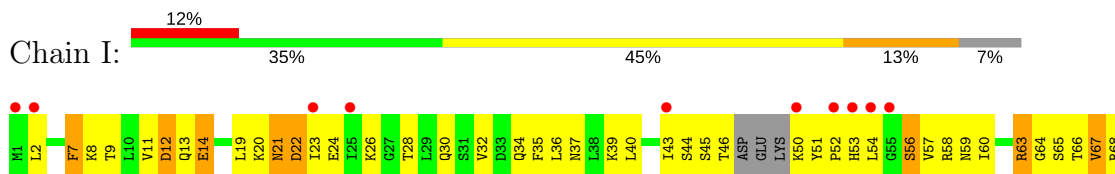
- Molecule 1: U6 snRNA-associated Sm-like protein LSm8



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

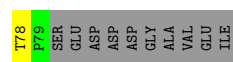


- Molecule 2: U6 snRNA-associated Sm-like protein LSm2

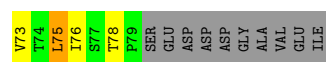




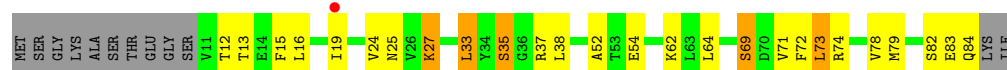
- Molecule 3: U6 snRNA-associated Sm-like protein LSm3



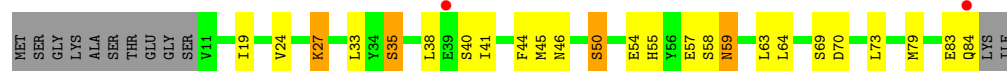
- Molecule 3: U6 snRNA-associated Sm-like protein LSm3



- Molecule 4: U6 snRNA-associated Sm-like protein LSm6



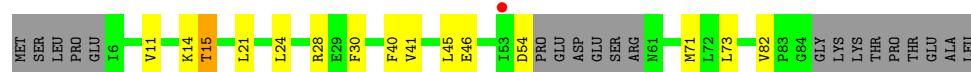
- Molecule 4: U6 snRNA-associated Sm-like protein LSm6



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

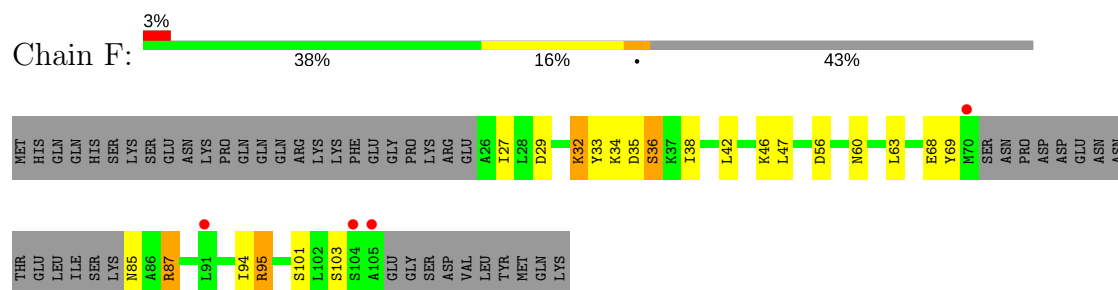


- Molecule 5: U6 snRNA-associated Sm-like protein LSm5

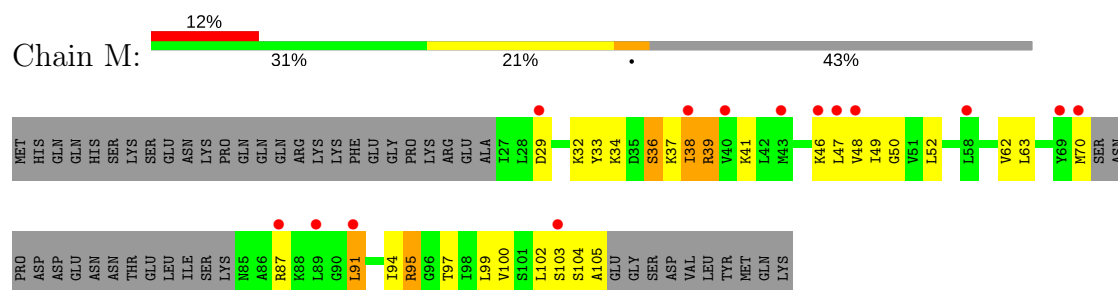




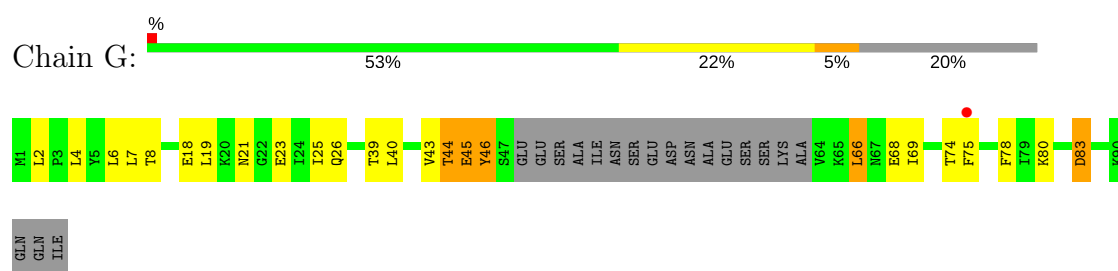
- Molecule 6: U6 snRNA-associated Sm-like protein LSm7



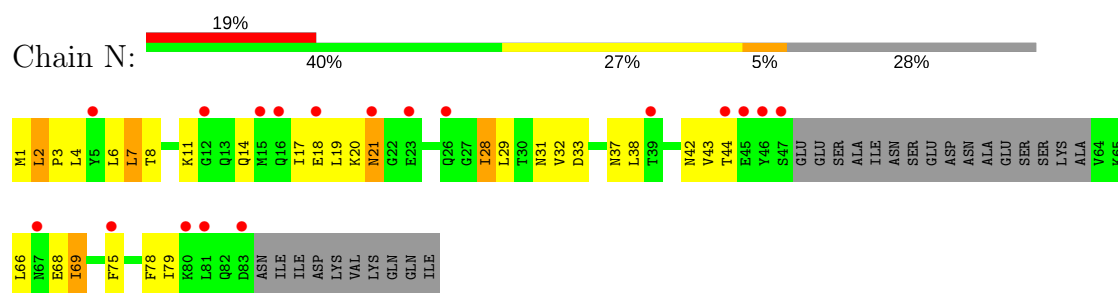
- Molecule 6: U6 snRNA-associated Sm-like protein LSm7



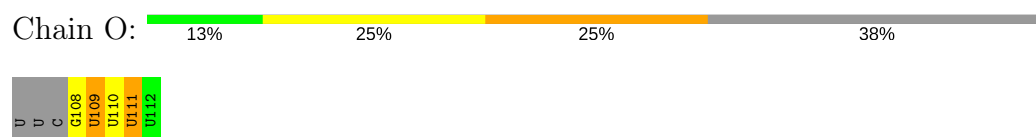
- Molecule 7: U6 snRNA-associated Sm-like protein LSm4



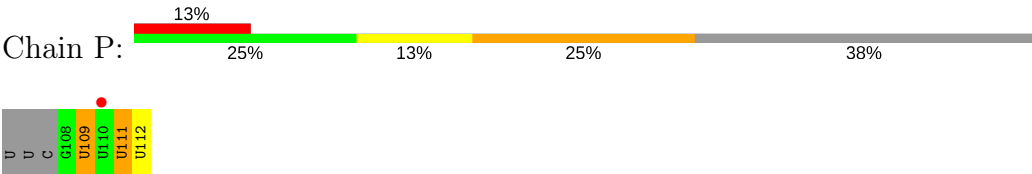
- Molecule 7: U6 snRNA-associated Sm-like protein LSm4



- Molecule 8: U6 snRNA



- Molecule 8: U6 snRNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	117.62Å 78.38Å 140.56Å 90.00° 100.58° 90.00°	Depositor
Resolution (Å)	39.19 – 2.78 39.53 – 2.78	Depositor EDS
% Data completeness (in resolution range)	97.9 (39.19-2.78) 97.9 (39.53-2.78)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.91 (at 2.77Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, $R_{free}$	0.245 , 0.278 0.239 , 0.272	Depositor DCC
$R_{free}$ test set	1576 reflections (5.06%)	DCC
Wilson B-factor (Å <sup>2</sup> )	74.1	Xtriage
Anisotropy	0.243	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 67.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	8295	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.31% 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	0.44	0/500	0.69	0/673
1	H	0.42	0/485	0.63	0/651
2	B	0.44	0/745	0.68	0/1005
2	I	0.38	0/730	0.58	0/982
3	C	0.47	0/617	0.69	0/836
3	J	0.47	0/617	0.66	0/836
4	D	0.53	0/584	0.68	0/787
4	K	0.48	0/563	0.61	0/760
5	E	0.50	0/595	0.67	1/806 (0.1%)
5	L	0.45	0/585	0.63	0/792
6	F	0.42	0/505	0.66	0/675
6	M	0.43	0/497	0.68	0/664
7	G	0.42	0/594	0.68	0/802
7	N	0.40	0/550	0.62	0/741
8	O	0.27	0/113	0.70	0/173
8	P	0.27	0/113	0.68	0/173
All	All	0.44	0/8393	0.66	1/11356 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	E	4	PRO	N-CA-CB	5.95	110.44	103.30

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	497	0	531	34	0
1	H	483	0	514	55	0
2	B	735	0	744	37	0
2	I	721	0	740	78	0
3	C	611	0	620	45	0
3	J	611	0	620	30	0
4	D	577	0	572	24	0
4	K	556	0	524	20	0
5	E	588	0	594	14	0
5	L	578	0	590	15	0
6	F	504	0	557	23	0
6	M	496	0	548	26	0
7	G	588	0	602	15	0
7	N	544	0	566	24	0
8	O	103	0	52	10	0
8	P	103	0	52	3	0
All	All	8295	0	8426	374	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 22.

The worst 5 of 374 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)
6:M:37:LYS:HG2	6:M:105:ALA:CB	1.25	1.57
6:M:37:LYS:CG	6:M:105:ALA:HB3	1.30	1.54
1:A:66:LEU:O	2:B:57:VAL:HG13	1.37	1.19
7:N:21:ASN:OD1	7:N:75:PHE:CZ	1.95	1.18
7:N:21:ASN:ND2	7:N:75:PHE:HE2	1.39	1.17

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	62/96 (65%)	60 (97%)	2 (3%)	0	100	100
1	H	58/96 (60%)	50 (86%)	6 (10%)	2 (3%)	4	13
2	B	88/95 (93%)	78 (89%)	8 (9%)	2 (2%)	7	22
2	I	84/95 (88%)	75 (89%)	9 (11%)	0	100	100
3	C	75/89 (84%)	72 (96%)	3 (4%)	0	100	100
3	J	75/89 (84%)	72 (96%)	3 (4%)	0	100	100
4	D	72/86 (84%)	70 (97%)	2 (3%)	0	100	100
4	K	72/86 (84%)	69 (96%)	3 (4%)	0	100	100
5	E	71/93 (76%)	66 (93%)	5 (7%)	0	100	100
5	L	69/93 (74%)	63 (91%)	6 (9%)	0	100	100
6	F	62/115 (54%)	61 (98%)	1 (2%)	0	100	100
6	M	61/115 (53%)	54 (88%)	7 (12%)	0	100	100
7	G	70/93 (75%)	66 (94%)	4 (6%)	0	100	100
7	N	63/93 (68%)	54 (86%)	8 (13%)	1 (2%)	11	33
All	All	982/1334 (74%)	910 (93%)	67 (7%)	5 (0%)	32	65

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	74	LYS
1	H	43	ASN
1	H	49	PHE
7	N	69	ILE
2	B	52	PRO

### 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	56/86 (65%)	44 (79%)	12 (21%)	1	3
1	H	54/86 (63%)	39 (72%)	15 (28%)	0	1
2	B	85/91 (93%)	74 (87%)	11 (13%)	5	14
2	I	84/91 (92%)	70 (83%)	14 (17%)	2	7
3	C	71/81 (88%)	51 (72%)	20 (28%)	0	1
3	J	71/81 (88%)	51 (72%)	20 (28%)	0	1
4	D	66/75 (88%)	56 (85%)	10 (15%)	3	9
4	K	58/75 (77%)	49 (84%)	9 (16%)	3	8
5	E	66/84 (79%)	60 (91%)	6 (9%)	11	29
5	L	66/84 (79%)	60 (91%)	6 (9%)	11	29
6	F	56/103 (54%)	50 (89%)	6 (11%)	8	21
6	M	55/103 (53%)	44 (80%)	11 (20%)	1	4
7	G	64/85 (75%)	51 (80%)	13 (20%)	1	4
7	N	61/85 (72%)	48 (79%)	13 (21%)	1	3
All	All	913/1210 (76%)	747 (82%)	166 (18%)	2	5

5 of 166 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	G	66	LEU
1	H	63	LEU
7	N	1	MET
7	G	83	ASP
1	H	27	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	33	ASN
1	H	40	ASN
2	I	53	HIS
6	F	85	ASN
2	I	21	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	O	4/8 (50%)	2 (50%)	0
8	P	4/8 (50%)	2 (50%)	0
All	All	8/16 (50%)	4 (50%)	0

All (4) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	O	109	U
8	O	111	U
8	P	109	U
8	P	111	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 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	64/96 (66%)	0.43	5 (7%) 14 9	44, 73, 122, 145	0
1	H	62/96 (64%)	1.79	21 (33%) 0 0	65, 116, 185, 226	0
2	B	90/95 (94%)	0.72	10 (11%) 6 4	40, 77, 144, 189	0
2	I	88/95 (92%)	1.04	11 (12%) 4 3	50, 95, 143, 209	0
3	C	77/89 (86%)	0.52	1 (1%) 77 74	43, 66, 107, 139	0
3	J	77/89 (86%)	0.44	1 (1%) 77 74	38, 68, 104, 115	0
4	D	74/86 (86%)	0.28	1 (1%) 75 71	41, 67, 104, 132	0
4	K	74/86 (86%)	0.24	2 (2%) 55 48	39, 70, 125, 139	0
5	E	75/93 (80%)	0.31	0 100 100	36, 63, 96, 113	0
5	L	73/93 (78%)	0.48	1 (1%) 75 71	51, 72, 102, 122	0
6	F	66/115 (57%)	0.53	4 (6%) 22 16	48, 77, 124, 141	0
6	M	65/115 (56%)	1.26	14 (21%) 1 1	61, 104, 156, 210	0
7	G	74/93 (79%)	0.37	1 (1%) 75 71	50, 81, 115, 138	0
7	N	67/93 (72%)	1.46	18 (26%) 1 0	70, 114, 157, 169	0
8	O	5/8 (62%)	1.35	0 100 100	68, 84, 110, 135	0
8	P	5/8 (62%)	1.58	1 (20%) 1 1	78, 78, 120, 120	0
All	All	1036/1350 (76%)	0.69	91 (8%) 11 7	36, 80, 138, 226	0

The worst 5 of 91 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	52	LYS	9.8
2	I	53	HIS	8.3
2	I	54	LEU	7.6
1	H	53	ALA	7.1
2	B	53	HIS	7.0

## 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.