



wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 11:08 AM GMT

PDB ID : 4M7D
Title : Crystal structure of Lsm2-8 complex bound to the RNA fragment CGUUU
Authors : Zhou, L.; Hang, J.; Zhou, Y.; Wan, R.; Lu, G.; Yan, C.; Shi, Y.
Deposited on : 2013-08-12
Resolution : 2.60 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

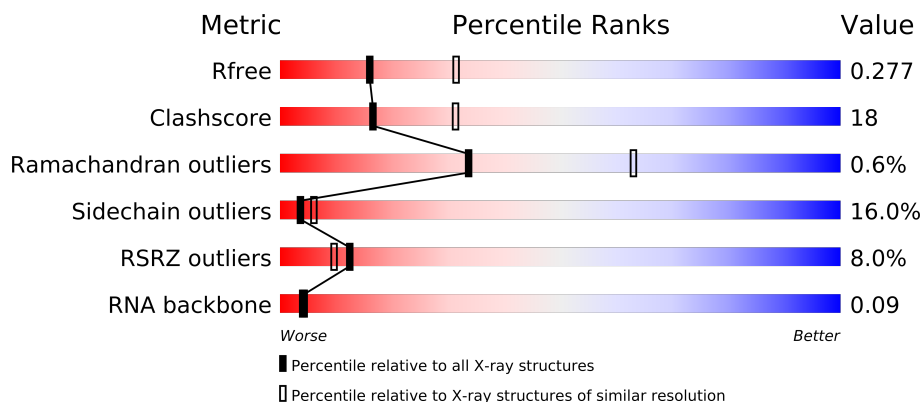
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.60 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.




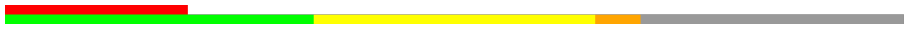


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	1718 (2.60-2.60)
Clashscore	79885	2154 (2.60-2.60)
Ramachandran outliers	78287	2113 (2.60-2.60)
Sidechain outliers	78261	2113 (2.60-2.60)
RSRZ outliers	66119	1718 (2.60-2.60)
RNA backbone	1838	1002 (3.12-2.08)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	96	
1	H	96	
2	B	95	
2	I	95	
3	C	89	
3	J	89	
4	D	86	
4	K	86	
5	E	93	
5	L	93	
6	F	115	
6	M	115	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	G	93	
7	N	93	
8	O	8	
8	P	8	

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 8455 atoms, of which 0 are hydrogen and 0 are deuterium.

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	65	Total	C	N	O	0	0	0
			508	329	86	93			
1	H	64	Total	C	N	O	0	0	0
			494	317	85	92			

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	92	Total	C	N	O	S	0	0	0
			750	478	126	143	3			
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	C	N	O	S	0	0	0
			611	382	105	123	1			
3	J	78	Total	C	N	O	S	0	0	0
			616	385	106	124	1			

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	C	N	O	S	0	0	0
			577	364	95	116	2			
4	K	74	Total	C	N	O	S	0	0	0
			556	351	93	110	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	77	Total	C	N	O	S	0	0	0
			603	387	100	114	2			
5	L	76	Total	C	N	O	S	0	0	0
			594	382	99	111	2			

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	77	Total	C	N	O	S	0	0	0
			613	394	99	117	3			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	N	66	Total	C	N	O	S	0	0	0
			536	349	87	97	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	14	38	5			
8	P	4	Total	C	N	O	P	0	0	0
			83	37	11	31	4			

- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	2	Total	O	0	0
			2	2		
9	B	7	Total	O	0	0
			7	7		
9	C	13	Total	O	0	0
			13	13		
9	D	10	Total	O	0	0
			10	10		
9	E	8	Total	O	0	0
			8	8		
9	F	6	Total	O	0	0
			6	6		
9	G	6	Total	O	0	0
			6	6		
9	H	4	Total	O	0	0
			4	4		
9	I	5	Total	O	0	0
			5	5		
9	J	5	Total	O	0	0
			5	5		
9	K	9	Total	O	0	0
			9	9		
9	L	5	Total	O	0	0
			5	5		
9	M	5	Total	O	0	0
			5	5		
9	N	3	Total	O	0	0
			3	3		

Continued on next page...

Continued from previous page...

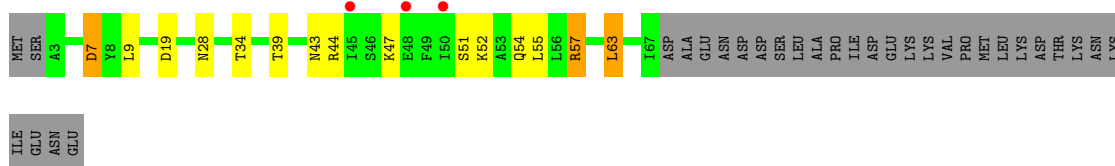
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	O	2	Total	O	0	0
			2	2		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

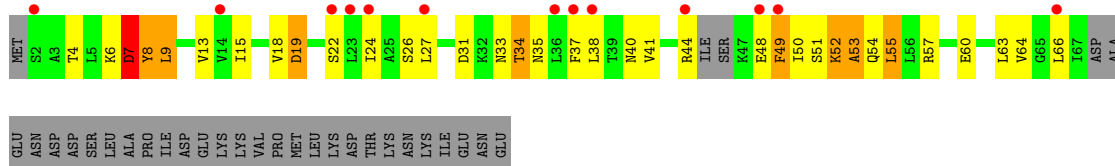
- Molecule 1: U6 snRNA-associated Sm-like protein LSm8

Chain A: 



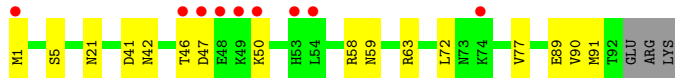
- Molecule 1: U6 snRNA-associated Sm-like protein LSm8

Chain H: 



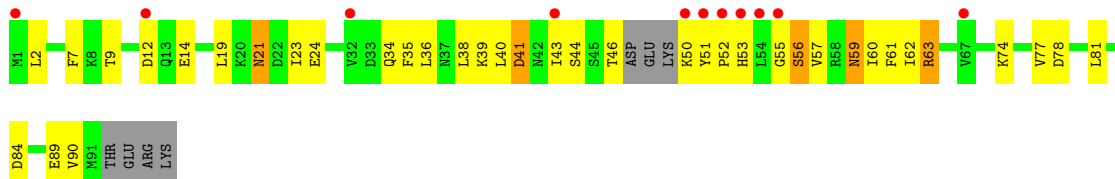
- Molecule 2: U6 snRNA-associated Sm-like protein LSm2

Chain B: 



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

Chain I: 



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

Chain C: 



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

Chain J:



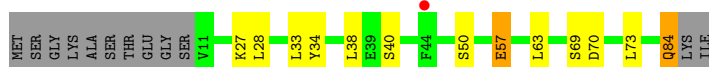
- Molecule 4: U6 snRNA-associated Sm-like protein LSm6

Chain D:



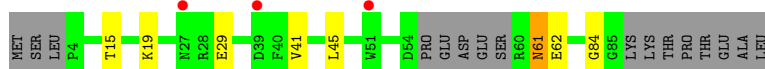
- Molecule 4: U6 snRNA-associated Sm-like protein LSm6

Chain K:



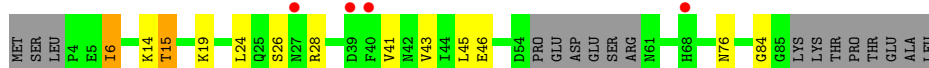
- Molecule 5: U6 snRNA-associated Sm-like protein LSm5

Chain E:



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

Chain L:



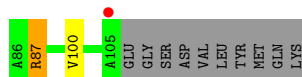
- Molecule 6: U6 snRNA-associated Sm-like protein LSm7

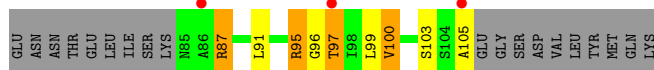
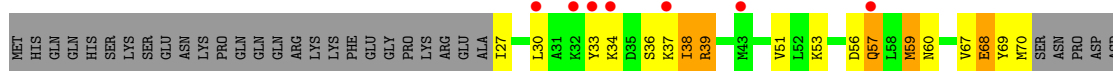
Chain F:



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

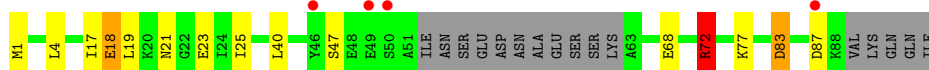
Chain M:





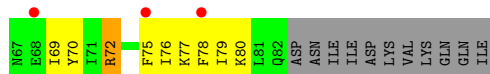
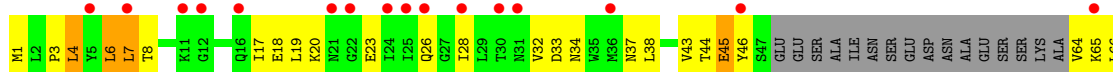
- Molecule 7: U6 snRNA-associated Sm-like protein LSm4

Chain G:



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

Chain N:



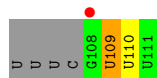
- Molecule 8: U6 snRNA

Chain O:



- Molecule 8: U6 snRNA

Chain P:



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	125.54Å 78.48Å 142.99Å 90.00° 103.17° 90.00°	Depositor
Resolution (Å)	34.18 – 2.60 34.18 – 2.59	Depositor EDS
% Data completeness (in resolution range)	98.3 (34.18-2.60) 98.4 (34.18-2.59)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.18 (at 2.61Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.240 , 0.283 0.238 , 0.277	Depositor DCC
R_{free} test set	2096 reflections (5.06%)	DCC
Wilson B-factor (Å ²)	66.8	Xtriage
Anisotropy	0.293	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 51.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 41446 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	8455	wwPDB-VP
Average B, all atoms (Å ²)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.76% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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	0.56	0/512	0.86	2/689 (0.3%)
1	H	0.39	0/496	0.57	0/666
2	B	0.60	0/760	0.76	0/1025
2	I	0.34	0/730	0.53	0/982
3	C	0.69	0/617	0.89	0/836
3	J	0.64	1/622 (0.2%)	0.86	1/843 (0.1%)
4	D	0.64	0/584	0.78	0/787
4	K	0.67	0/563	0.82	0/760
5	E	0.58	0/611	0.76	0/827
5	L	0.53	0/602	0.74	0/815
6	F	0.58	0/505	0.76	0/675
6	M	0.35	0/497	0.55	0/664
7	G	0.57	0/619	0.86	2/835 (0.2%)
7	N	0.34	0/542	0.56	0/730
8	O	0.22	0/113	0.63	0/173
8	P	0.28	0/91	0.67	0/139
All	All	0.55	1/8464 (0.0%)	0.75	5/11446 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	J	4	PRO	N-CD	5.31	1.55	1.47

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	57	ARG	NE-CZ-NH1	-6.24	117.18	120.30
7	G	72	ARG	NE-CZ-NH1	-5.82	117.39	120.30
1	A	63	LEU	CB-CG-CD2	-5.74	101.25	111.00
3	J	3	THR	C-N-CD	5.56	140.07	128.40
7	G	40	LEU	CA-CB-CG	5.02	126.85	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	508	0	0	4	0
1	H	494	0	4	37	0
2	B	750	0	0	3	0
2	I	721	0	0	28	0
3	C	611	0	0	6	0
3	J	616	0	1	11	0
4	D	577	0	0	3	0
4	K	556	0	0	6	0
5	E	603	0	0	5	0
5	L	594	0	0	4	0
6	F	504	0	0	9	0
6	M	496	0	0	19	0
7	G	613	0	0	7	0
7	N	536	0	0	25	0
8	O	103	0	0	11	0
8	P	83	0	0	1	0
9	A	2	0	0	0	0
9	B	7	0	0	1	0
9	C	13	0	0	4	0
9	D	10	0	0	1	0
9	E	8	0	0	1	0
9	F	6	0	0	1	0
9	G	6	0	0	2	0
9	H	4	0	0	4	0
9	I	5	0	0	2	0
9	J	5	0	0	2	0
9	K	9	0	0	2	0
9	L	5	0	0	0	0
9	M	5	0	0	0	0
9	N	3	0	0	1	0
9	O	2	0	0	0	0
All	All	8455	0	5	151	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 18.

The worst 5 of 151 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
6:M:37:LYS:CG	6:M:105:ALA:CB	2.05	1.34
6:M:67:VAL:CG2	6:M:69:TYR:CE1	2.16	1.27
2:I:43:ILE:CD1	2:I:60:ILE:CG2	2.32	1.07
3:C:69:ARG:NH1	9:C:102:HOH:O	1.92	1.01
2:I:41:ASP:OD1	2:I:59:ASN:ND2	1.95	1.00

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	63/96 (66%)	61 (97%)	2 (3%)	0	100	100
1	H	60/96 (62%)	53 (88%)	5 (8%)	2 (3%)	6	9
2	B	90/95 (95%)	83 (92%)	7 (8%)	0	100	100
2	I	84/95 (88%)	74 (88%)	10 (12%)	0	100	100
3	C	75/89 (84%)	72 (96%)	3 (4%)	0	100	100
3	J	76/89 (85%)	73 (96%)	3 (4%)	0	100	100
4	D	72/86 (84%)	70 (97%)	2 (3%)	0	100	100
4	K	72/86 (84%)	65 (90%)	7 (10%)	0	100	100
5	E	73/93 (78%)	69 (94%)	3 (4%)	1 (1%)	16	32
5	L	72/93 (77%)	63 (88%)	8 (11%)	1 (1%)	16	32
6	F	62/115 (54%)	62 (100%)	0	0	100	100
6	M	61/115 (53%)	54 (88%)	7 (12%)	0	100	100
7	G	73/93 (78%)	72 (99%)	1 (1%)	0	100	100
7	N	62/93 (67%)	53 (86%)	7 (11%)	2 (3%)	6	9
All	All	995/1334 (75%)	924 (93%)	65 (6%)	6 (1%)	33	63

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	L	41	VAL
1	H	7	ASP
1	H	53	ALA
7	N	34	ASN
7	N	69	ILE

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	57/86 (66%)	47 (82%)	10 (18%)	3	4
1	H	55/86 (64%)	39 (71%)	16 (29%)	0	1
2	B	87/91 (96%)	75 (86%)	12 (14%)	5	9
2	I	84/91 (92%)	71 (84%)	13 (16%)	4	6
3	C	71/81 (88%)	55 (78%)	16 (22%)	1	2
3	J	71/81 (88%)	58 (82%)	13 (18%)	2	4
4	D	66/75 (88%)	58 (88%)	8 (12%)	7	13
4	K	58/75 (77%)	49 (84%)	9 (16%)	4	6
5	E	68/84 (81%)	65 (96%)	3 (4%)	39	68
5	L	67/84 (80%)	58 (87%)	9 (13%)	6	10
6	F	56/103 (54%)	52 (93%)	4 (7%)	21	40
6	M	55/103 (53%)	42 (76%)	13 (24%)	1	2
7	G	68/85 (80%)	59 (87%)	9 (13%)	6	10
7	N	60/85 (71%)	47 (78%)	13 (22%)	1	2
All	All	923/1210 (76%)	775 (84%)	148 (16%)	3	6

5 of 148 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	15	ILE
2	I	21	ASN
7	N	4	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	H	19	ASP
1	H	52	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	O	4/8 (50%)	4 (100%)	0
8	P	3/8 (37%)	2 (66%)	0
All	All	7/16 (43%)	6 (85%)	0

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	O	108	G
8	O	109	U
8	O	110	U
8	O	111	U
8	P	109	U

There are no RNA pucker outliers to report.

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 ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues

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	65/96 (67%)	0.28	3 (4%) 31 27	40, 68, 124, 144	0
1	H	64/96 (66%)	1.12	13 (20%) 1 1	27, 97, 152, 201	0
2	B	92/95 (96%)	0.49	9 (9%) 8 6	44, 76, 145, 193	0
2	I	88/95 (92%)	0.90	11 (12%) 5 3	50, 87, 156, 199	0
3	C	77/89 (86%)	0.09	1 (1%) 74 75	41, 66, 93, 119	0
3	J	78/89 (87%)	0.28	1 (1%) 74 75	30, 64, 100, 132	0
4	D	74/86 (86%)	0.45	1 (1%) 72 72	44, 71, 116, 157	0
4	K	74/86 (86%)	0.07	1 (1%) 72 72	36, 60, 97, 115	0
5	E	77/93 (82%)	0.25	3 (3%) 37 33	42, 61, 109, 123	0
5	L	76/93 (81%)	0.31	4 (5%) 25 22	46, 73, 117, 140	0
6	F	66/115 (57%)	0.09	3 (4%) 32 28	39, 67, 114, 144	0
6	M	65/115 (56%)	0.99	10 (15%) 3 2	60, 94, 139, 160	0
7	G	77/93 (82%)	0.29	4 (5%) 26 22	40, 65, 124, 180	0
7	N	66/93 (70%)	1.51	19 (28%) 1 1	75, 112, 149, 166	0
8	O	5/8 (62%)	0.57	0 100 100	81, 87, 154, 184	0
8	P	4/8 (50%)	1.23	1 (25%) 1 1	96, 111, 117, 180	0
All	All	1048/1350 (77%)	0.50	84 (8%) 12 9	27, 75, 138, 201	0

The worst 5 of 84 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	I	53	HIS	9.0
2	B	53	HIS	8.2
2	I	50	LYS	6.7
5	L	40	PHE	5.0
2	I	52	PRO	4.9

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 carbohydrates in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.