



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 29, 2021 – 04:10 pm GMT

PDB ID : 6TFL  
Title : Lsm protein (SmAP) from Halobacterium salinarum  
Authors : Nikulin, A.D.; Fando, M.S.; Lekontseva, N.V.  
Deposited on : 2019-11-14  
Resolution : 2.40 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  
Mogul : 1.8.4 (270009), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.23.2  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0267  
CCP4 : 7.1.010 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.23.2

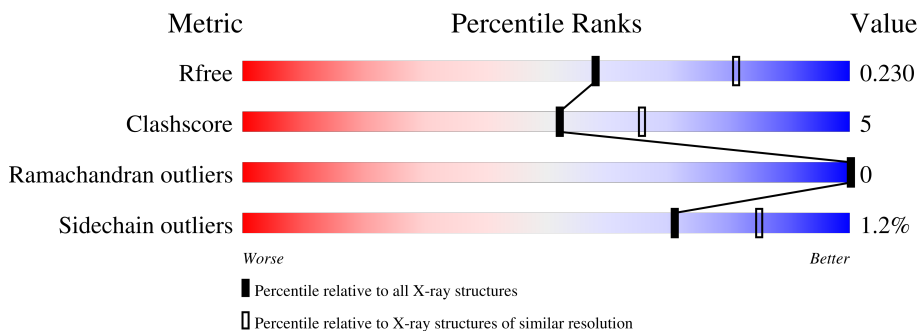
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)

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 of 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	62	90% (Green), 10% (Yellow)
1	B	62	89% (Green), 11% (Yellow)
1	C	62	90% (Green), 10% (Yellow)
1	D	62	82% (Green), 18% (Yellow)
1	E	62	82% (Green), 18% (Yellow)
1	F	62	87% (Green), 13% (Yellow)
1	G	62	95% (Green), 5% (Yellow)

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	H	62	 92% 8%
1	I	62	 87% 13%
1	J	62	 90% 8%
1	K	62	 89% 11%
1	L	62	 85% 15%
1	M	62	 94% 6%
1	N	62	 94% 6%

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 7184 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 RNA-binding protein Lsm.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	62	471	290	78	101	2	0	0	0
1	B	62	471	290	78	101	2	0	0	0
1	C	62	471	290	78	101	2	0	0	0
1	D	62	471	290	78	101	2	0	0	0
1	E	62	471	290	78	101	2	0	0	0
1	F	62	471	290	78	101	2	0	0	0
1	G	62	471	290	78	101	2	0	0	0
1	H	62	471	290	78	101	2	0	0	0
1	I	62	471	290	78	101	2	0	0	0
1	J	62	471	290	78	101	2	0	0	0
1	K	62	471	290	78	101	2	0	0	0
1	L	62	471	290	78	101	2	0	0	0
1	M	62	480	295	79	104	2	0	1	0
1	N	62	471	290	78	101	2	0	0	0

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP B0R5R2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	ALA	-	expression tag	UNP B0R5R2
B	-1	GLY	-	expression tag	UNP B0R5R2
B	0	ALA	-	expression tag	UNP B0R5R2
C	-1	GLY	-	expression tag	UNP B0R5R2
C	0	ALA	-	expression tag	UNP B0R5R2
D	-1	GLY	-	expression tag	UNP B0R5R2
D	0	ALA	-	expression tag	UNP B0R5R2
E	-1	GLY	-	expression tag	UNP B0R5R2
E	0	ALA	-	expression tag	UNP B0R5R2
F	-1	GLY	-	expression tag	UNP B0R5R2
F	0	ALA	-	expression tag	UNP B0R5R2
G	-1	GLY	-	expression tag	UNP B0R5R2
G	0	ALA	-	expression tag	UNP B0R5R2
H	-1	GLY	-	expression tag	UNP B0R5R2
H	0	ALA	-	expression tag	UNP B0R5R2
I	-1	GLY	-	expression tag	UNP B0R5R2
I	0	ALA	-	expression tag	UNP B0R5R2
J	-1	GLY	-	expression tag	UNP B0R5R2
J	0	ALA	-	expression tag	UNP B0R5R2
K	-1	GLY	-	expression tag	UNP B0R5R2
K	0	ALA	-	expression tag	UNP B0R5R2
L	-1	GLY	-	expression tag	UNP B0R5R2
L	0	ALA	-	expression tag	UNP B0R5R2
M	-1	GLY	-	expression tag	UNP B0R5R2
M	0	ALA	-	expression tag	UNP B0R5R2
N	-1	GLY	-	expression tag	UNP B0R5R2
N	0	ALA	-	expression tag	UNP B0R5R2

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

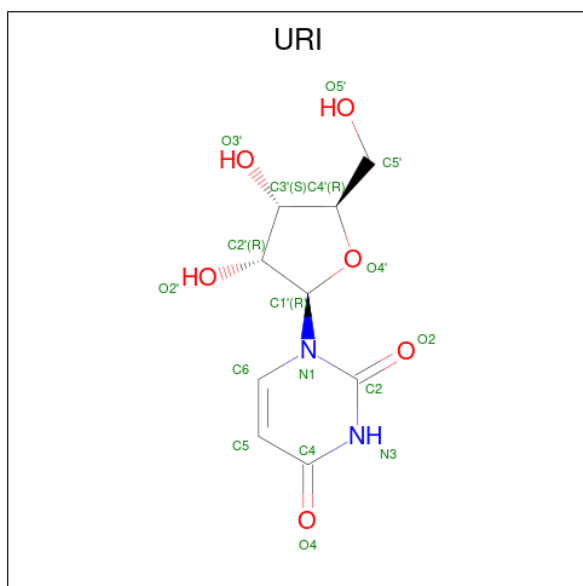
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total Ca 2 2	0	0
2	B	2	Total Ca 2 2	0	0
2	D	4	Total Ca 4 4	0	0
2	E	2	Total Ca 2 2	0	0
2	F	1	Total Ca 1 1	0	0
2	H	3	Total Ca 3 3	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	I	3	Total Ca 3 3	0	0
2	J	1	Total Ca 1 1	0	0
2	K	2	Total Ca 2 2	0	0
2	L	3	Total Ca 3 3	0	0
2	M	1	Total Ca 1 1	0	0
2	N	4	Total Ca 4 4	0	0

- Molecule 3 is URIDINE (three-letter code: URI) (formula: C<sub>9</sub>H<sub>12</sub>N<sub>2</sub>O<sub>6</sub>).



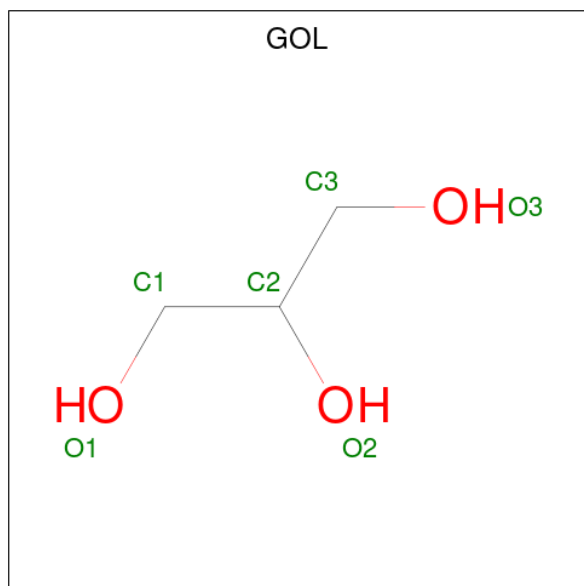
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C N O 17 9 2 6	0	0
3	B	1	Total C N O 17 9 2 6	0	0
3	C	1	Total C N O 17 9 2 6	0	0
3	D	1	Total C N O 17 9 2 6	0	0
3	E	1	Total C N O 17 9 2 6	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	F	1	Total 17	C 9	N 2	O 6	0	0
3	G	1	Total 17	C 9	N 2	O 6	0	0
3	H	1	Total 17	C 9	N 2	O 6	0	0
3	I	1	Total 17	C 9	N 2	O 6	0	0
3	J	1	Total 17	C 9	N 2	O 6	0	0
3	K	1	Total 17	C 9	N 2	O 6	0	0
3	L	1	Total 17	C 9	N 2	O 6	0	0
3	M	1	Total 17	C 9	N 2	O 6	0	0
3	N	1	Total 17	C 9	N 2	O 6	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	G	1	Total 6	C 3	O 3	0	0

- Molecule 5 is water.


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	20	Total O 20 20	0	0
5	B	22	Total O 22 22	0	0
5	C	15	Total O 15 15	0	0
5	D	29	Total O 29 29	0	0
5	E	22	Total O 22 22	0	0
5	F	21	Total O 21 21	0	0
5	G	26	Total O 26 26	0	0
5	H	26	Total O 26 26	0	0
5	I	17	Total O 17 17	0	0
5	J	13	Total O 13 13	0	0
5	K	22	Total O 22 22	0	0
5	L	28	Total O 28 28	0	0
5	M	19	Total O 19 19	0	0
5	N	29	Total O 29 29	0	0



### 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: RNA-binding protein Lsm

Chain A:  90% 10%




- Molecule 1: RNA-binding protein Lsm

Chain B:  89% 11%




- Molecule 1: RNA-binding protein Lsm

Chain C:  90% 10%




- Molecule 1: RNA-binding protein Lsm

Chain D:  82% 18%



- Molecule 1: RNA-binding protein Lsm

Chain E:  82% 18%



- Molecule 1: RNA-binding protein Lsm

Chain F:  87% 13%



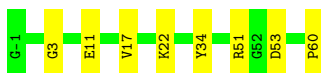
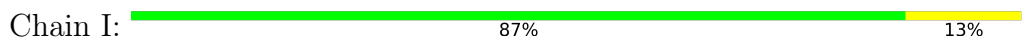
- Molecule 1: RNA-binding protein Lsm



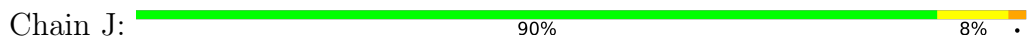
- Molecule 1: RNA-binding protein Lsm



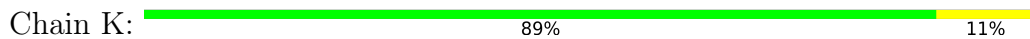
- Molecule 1: RNA-binding protein Lsm



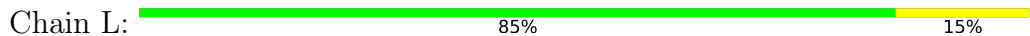
- Molecule 1: RNA-binding protein Lsm



- Molecule 1: RNA-binding protein Lsm



- Molecule 1: RNA-binding protein Lsm



- Molecule 1: RNA-binding protein Lsm





- Molecule 1: RNA-binding protein Lsm

Chain N:  94% 6%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	156.45Å 156.45Å 106.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.40 49.47 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.8 (50.00-2.40) 99.9 (49.47-2.40)	Depositor EDS
$R_{merge}$	0.26	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.47 (at 2.39Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, $R_{free}$	0.178 , 0.231 0.181 , 0.230	Depositor DCC
$R_{free}$ test set	2702 reflections (5.17%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.9	Xtriage
Anisotropy	0.660	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	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	0.94	EDS
Total number of atoms	7184	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.48% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, GOL, URI

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	0.39	0/475	0.56	0/642
1	B	0.40	0/475	0.58	0/642
1	C	0.41	0/475	0.58	0/642
1	D	0.38	0/475	0.58	0/642
1	E	0.41	0/475	0.56	0/642
1	F	0.38	0/475	0.55	0/642
1	G	0.39	0/475	0.58	0/642
1	H	0.36	0/475	0.57	0/642
1	I	0.40	0/475	0.56	0/642
1	J	0.43	0/475	0.54	0/642
1	K	0.40	0/475	0.55	0/642
1	L	0.42	0/475	0.58	0/642
1	M	0.38	0/484	0.58	0/654
1	N	0.39	0/475	0.57	0/642
All	All	0.40	0/6659	0.57	0/9000

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 [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	471	0	462	5	0
1	B	471	0	462	4	0
1	C	471	0	462	6	0
1	D	471	0	462	11	0
1	E	471	0	462	10	0
1	F	471	0	462	10	0
1	G	471	0	462	2	0
1	H	471	0	462	6	0
1	I	471	0	462	7	0
1	J	471	0	462	5	0
1	K	471	0	462	6	0
1	L	471	0	462	6	0
1	M	480	0	467	2	0
1	N	471	0	462	5	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	D	4	0	0	0	0
2	E	2	0	0	0	0
2	F	1	0	0	0	0
2	H	3	0	0	0	0
2	I	3	0	0	0	0
2	J	1	0	0	0	0
2	K	2	0	0	0	0
2	L	3	0	0	0	0
2	M	1	0	0	0	0
2	N	4	0	0	0	0
3	A	17	0	12	1	0
3	B	17	0	12	2	0
3	C	17	0	12	4	0
3	D	17	0	12	3	0
3	E	17	0	12	3	0
3	F	17	0	12	4	0
3	G	17	0	12	2	0
3	H	17	0	12	0	0
3	I	17	0	12	1	0
3	J	17	0	12	1	0
3	K	17	0	12	2	0
3	L	17	0	12	4	0
3	M	17	0	12	2	0
3	N	17	0	12	1	0
4	G	6	0	8	0	0
5	A	20	0	0	0	0
5	B	22	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	15	0	0	2	0
5	D	29	0	0	0	0
5	E	22	0	0	0	0
5	F	21	0	0	0	0
5	G	26	0	0	0	0
5	H	26	0	0	0	0
5	I	17	0	0	0	0
5	J	13	0	0	0	0
5	K	22	0	0	0	0
5	L	28	0	0	0	0
5	M	19	0	0	0	0
5	N	29	0	0	0	0
All	All	7184	0	6649	74	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 5.

The worst 5 of 74 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:F:51:ARG:HD3	1:F:53:ASP:OD1	1.88	0.73
3:D:105:URI:H4'	3:E:103:URI:H5'1	1.73	0.71
1:F:51:ARG:NH2	1:F:53:ASP:OD1	2.21	0.70
3:A:103:URI:H5'1	3:G:102:URI:H3'	1.72	0.70
1:J:51:ARG:NH1	1:J:53:ASP:OD1	2.32	0.62

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	60/62 (97%)	58 (97%)	2 (3%)	0	100	100
1	B	60/62 (97%)	58 (97%)	2 (3%)	0	100	100
1	C	60/62 (97%)	57 (95%)	3 (5%)	0	100	100
1	D	60/62 (97%)	57 (95%)	3 (5%)	0	100	100
1	E	60/62 (97%)	58 (97%)	2 (3%)	0	100	100
1	F	60/62 (97%)	58 (97%)	2 (3%)	0	100	100
1	G	60/62 (97%)	59 (98%)	1 (2%)	0	100	100
1	H	60/62 (97%)	57 (95%)	3 (5%)	0	100	100
1	I	60/62 (97%)	59 (98%)	1 (2%)	0	100	100
1	J	60/62 (97%)	59 (98%)	1 (2%)	0	100	100
1	K	60/62 (97%)	58 (97%)	2 (3%)	0	100	100
1	L	60/62 (97%)	59 (98%)	1 (2%)	0	100	100
1	M	61/62 (98%)	59 (97%)	2 (3%)	0	100	100
1	N	60/62 (97%)	59 (98%)	1 (2%)	0	100	100
All	All	841/868 (97%)	815 (97%)	26 (3%)	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	53/54 (98%)	53 (100%)	0	100	100
1	B	53/54 (98%)	53 (100%)	0	100	100
1	C	53/54 (98%)	52 (98%)	1 (2%)	57	75
1	D	53/54 (98%)	51 (96%)	2 (4%)	33	51
1	E	53/54 (98%)	51 (96%)	2 (4%)	33	51
1	F	53/54 (98%)	53 (100%)	0	100	100
1	G	53/54 (98%)	52 (98%)	1 (2%)	57	75

*Continued on next page...*



Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	H	53/54 (98%)	53 (100%)	0	100	100
1	I	53/54 (98%)	53 (100%)	0	100	100
1	J	53/54 (98%)	52 (98%)	1 (2%)	57	75
1	K	53/54 (98%)	53 (100%)	0	100	100
1	L	53/54 (98%)	52 (98%)	1 (2%)	57	75
1	M	54/54 (100%)	53 (98%)	1 (2%)	57	75
1	N	53/54 (98%)	53 (100%)	0	100	100
All	All	743/756 (98%)	734 (99%)	9 (1%)	71	85

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	L	45	GLU
1	M	53	ASP
1	E	45	GLU
1	E	46	ASP
1	G	53	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	36	GLN
1	N	54	ASN

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 43 ligands modelled in this entry, 28 are monoatomic - leaving 15 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

### 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.