



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 09:24 am BST

PDB ID : 5VSU  
Title : Structure of yeast U6 snRNP with 2'-phosphate terminated U6 RNA  
Authors : Montemayor, E.J.  
Deposited on : 2017-05-12  
Resolution : 3.10 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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

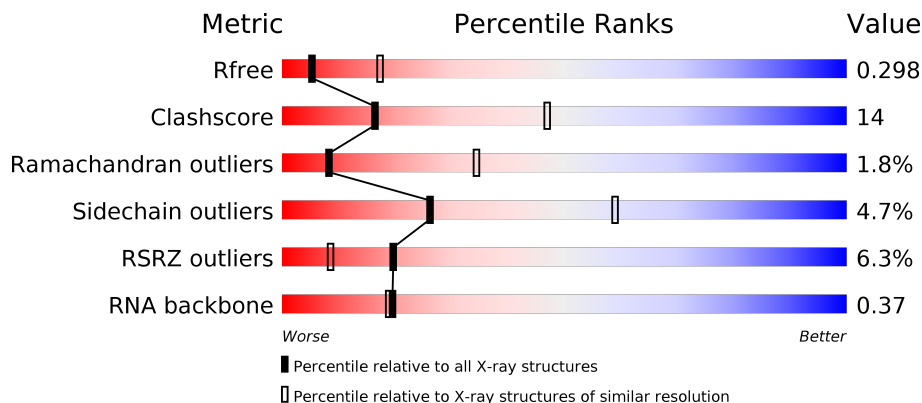
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)
RNA backbone	3102	1116 (3.40-2.80)

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  $\leq 5\%$ . 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	452	 4% 62% 23% 14%
2	B	98	 8% 48% 48% 14%
3	C	92	 4% 50% 37% 12%
4	D	96	 5% 45% 21% 32%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	E	96	
6	F	88	
7	G	118	
8	H	111	
9	I	83	

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 9240 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 U4/U6 snRNA-associated-splicing factor PRP24.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	387	3120	1971	555	579	15	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	445	LEU	-	expression tag	UNP P49960
A	446	GLU	-	expression tag	UNP P49960
A	447	HIS	-	expression tag	UNP P49960
A	448	HIS	-	expression tag	UNP P49960
A	449	HIS	-	expression tag	UNP P49960
A	450	HIS	-	expression tag	UNP P49960
A	451	HIS	-	expression tag	UNP P49960
A	452	HIS	-	expression tag	UNP P49960

- 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	98	798	507	137	149	5	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	MET	-	initiating methionine	UNP P38203
B	-1	GLY	-	expression tag	UNP P38203
B	0	SER	-	expression tag	UNP P38203

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	81	Total	C	N	O	S	0	0	0
			638	397	109	128	4			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-2	MET	-	initiating methionine	UNP P57743
C	-1	GLY	-	expression tag	UNP P57743
C	0	SER	-	expression tag	UNP P57743

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	65	Total	C	N	O	S	0	0	0
			510	330	85	93	2			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-2	MET	-	initiating methionine	UNP P40070
D	-1	GLY	-	expression tag	UNP P40070
D	0	SER	-	expression tag	UNP P40070

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	87	Total	C	N	O	S	0	0	0
			688	438	115	132	3			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	-2	MET	-	initiating methionine	UNP P40089
E	-1	GLY	-	expression tag	UNP P40089
E	0	SER	-	expression tag	UNP P40089

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	77	Total	C	N	O	S	0	0	0
			600	379	99	120	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	GLY	-	expression tag	UNP Q06406
F	0	SER	-	expression tag	UNP Q06406

- 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	70	533	343	90	97	3	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-2	MET	-	initiating methionine	UNP P53905
G	-1	GLY	-	expression tag	UNP P53905
G	0	SER	-	expression tag	UNP P53905

- Molecule 8 is a protein called U6 snRNA-associated Sm-like protein LSm8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	H	100	794	510	133	148	3	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	-1	GLY	-	expression tag	UNP P47093
H	0	SER	-	expression tag	UNP P47093

- Molecule 9 is a RNA chain called Saccharomyces cerevisiae strain T8 chromosome XII sequence.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
9	I	73	1559	697	278	510	74	0	0	0

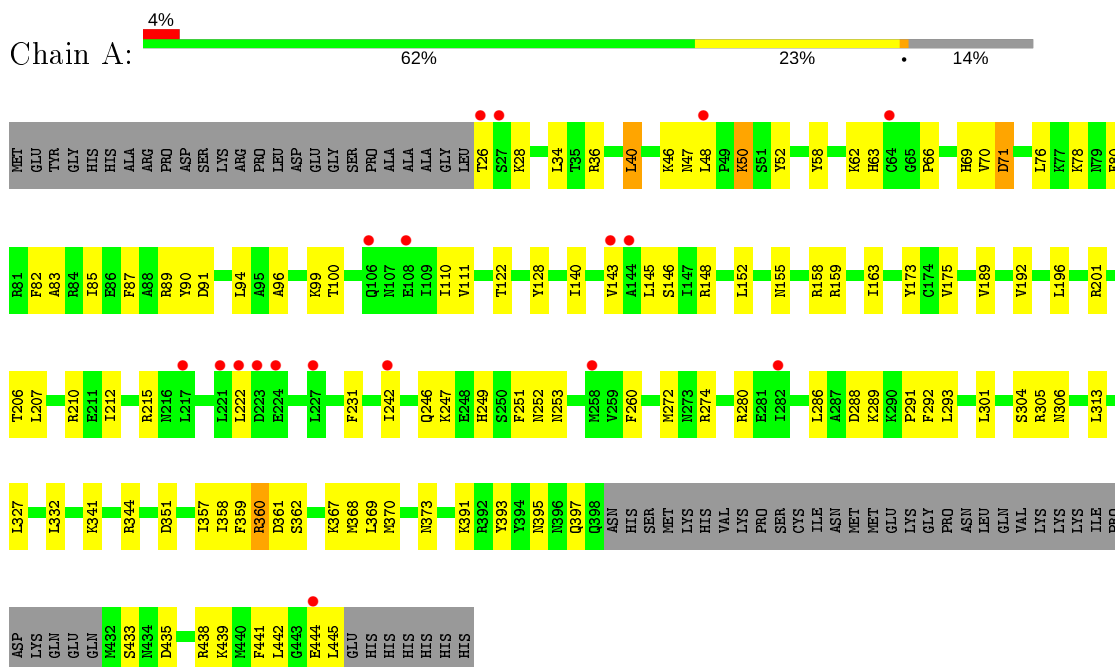
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	62	G	A	conflict	GB 1039023528

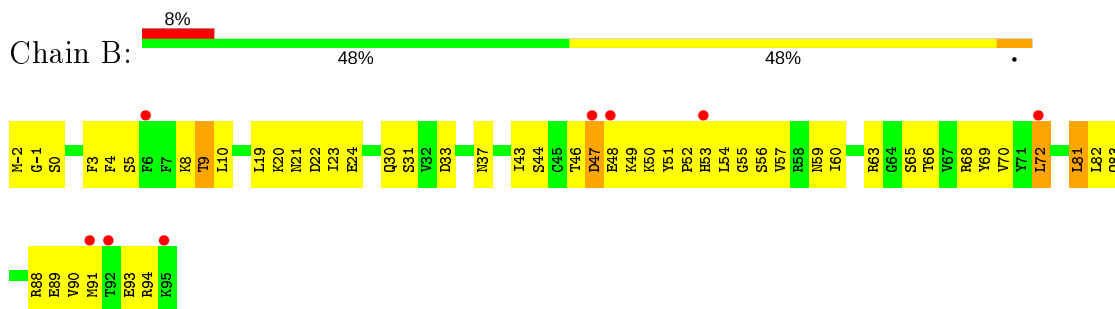
### 3 Residue-property plots [i](#)

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.

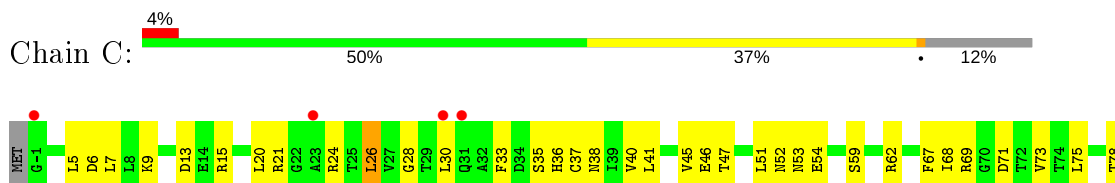
- Molecule 1: U4/U6 snRNA-associated-splicing factor PRP24



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

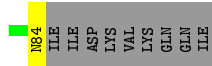
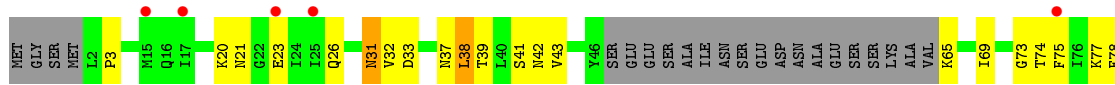


- Molecule 3: U6 snRNA-associated Sm-like protein LSm3

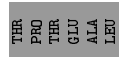




• Molecule 4: U6 snRNA-associated Sm-like protein LSm4



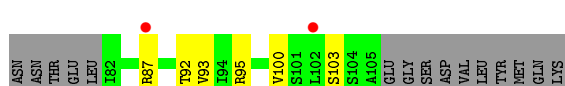
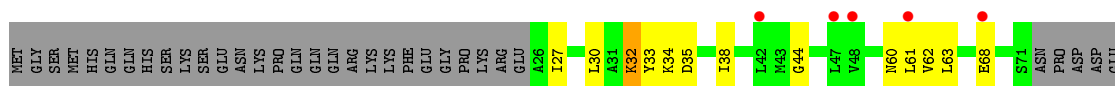
• Molecule 5: U6 snRNA-associated Sm-like protein LSm5



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



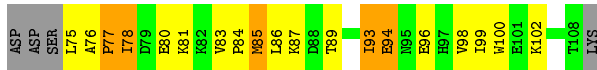
• Molecule 7: U6 snRNA-associated Sm-like protein LSm7



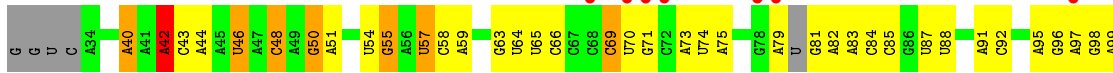
• Molecule 8: U6 snRNA-associated Sm-like protein LSm8







- Molecule 9: *Saccharomyces cerevisiae* strain T8 chromosome XII sequence



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.16Å 114.73Å 179.84Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	96.72 – 3.10 96.72 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.8 (96.72-3.10) 99.8 (96.72-3.10)	Depositor EDS
$R_{merge}$	0.25	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.62 (at 3.13Å)	Xtrriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, $R_{free}$	0.234 , 0.298 0.234 , 0.298	Depositor DCC
$R_{free}$ test set	2000 reflections (7.39%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	110.6	Xtrriage
Anisotropy	0.154	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 72.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	9240	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	133.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.32% 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: 9QV

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	0.57	0/3167	0.89	4/4256 (0.1%)
2	B	0.56	0/808	0.99	2/1085 (0.2%)
3	C	0.58	0/644	0.98	1/871 (0.1%)
4	D	0.49	0/515	0.67	0/696
5	E	0.53	0/698	0.87	2/944 (0.2%)
6	F	0.52	0/607	0.89	2/817 (0.2%)
7	G	0.54	0/534	1.03	0/713
8	H	0.60	0/802	0.97	1/1077 (0.1%)
9	I	0.67	0/1716	1.28	14/2666 (0.5%)
All	All	0.58	0/9491	1.00	26/13125 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
8	H	0	1

There are no bond length outliers.

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	I	108	G	N3-C4-N9	-7.54	121.48	126.00
1	A	40	LEU	CB-CG-CD1	-7.41	98.41	111.00
8	H	76	ALA	C-N-CD	-6.67	105.93	120.60
9	I	48	C	N1-C2-O2	6.38	122.73	118.90
9	I	92	C	C6-N1-C2	-6.37	117.75	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
8	H	77	PRO	Peptide

## 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	3120	0	3182	72	0
2	B	798	0	818	52	0
3	C	638	0	646	33	0
4	D	510	0	519	26	0
5	E	688	0	702	18	0
6	F	600	0	601	16	0
7	G	533	0	591	16	0
8	H	794	0	826	50	0
9	I	1559	0	774	38	0
All	All	9240	0	8659	255	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 14.

The worst 5 of 255 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:305:ARG:NH1	9:I:101:U:OP2	1.81	1.14
4:D:21:ASN:HB2	4:D:75:PHE:HZ	1.32	0.94
2:B:48:GLU:HG3	2:B:50:LYS:HB3	1.49	0.92
8:H:43:ASN:OD1	8:H:44:ARG:N	2.07	0.86
5:E:75:GLY:HA2	5:E:78:ILE:HD12	1.57	0.86

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	383/452 (85%)	371 (97%)	11 (3%)	1 (0%)	41	73
2	B	96/98 (98%)	84 (88%)	8 (8%)	4 (4%)	3	16
3	C	79/92 (86%)	75 (95%)	3 (4%)	1 (1%)	12	42
4	D	61/96 (64%)	54 (88%)	7 (12%)	0	100	100
5	E	85/96 (88%)	79 (93%)	5 (6%)	1 (1%)	13	44
6	F	75/88 (85%)	72 (96%)	3 (4%)	0	100	100
7	G	66/118 (56%)	64 (97%)	0	2 (3%)	4	23
8	H	94/111 (85%)	75 (80%)	11 (12%)	8 (8%)	1	5
All	All	939/1151 (82%)	874 (93%)	48 (5%)	17 (2%)	8	34

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	433	SER
2	B	53	HIS
2	B	55	GLY
5	E	86	LYS
8	H	78	ILE

### 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	349/408 (86%)	340 (97%)	9 (3%)	46	74

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	92/93 (99%)	88 (96%)	4 (4%)	29	62
3	C	74/83 (89%)	71 (96%)	3 (4%)	30	64
4	D	55/87 (63%)	52 (94%)	3 (6%)	21	53
5	E	79/86 (92%)	75 (95%)	4 (5%)	24	56
6	F	69/76 (91%)	68 (99%)	1 (1%)	67	86
7	G	60/105 (57%)	59 (98%)	1 (2%)	60	83
8	H	89/100 (89%)	73 (82%)	16 (18%)	1	7
All	All	867/1038 (84%)	826 (95%)	41 (5%)	26	59

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	84	ASN
5	E	87	LYS
8	H	52	LYS
5	E	10	GLU
5	E	74	SER

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

Mol	Chain	Res	Type
1	A	306	ASN
5	E	66	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
9	I	69/83 (83%)	13 (18%)	1 (1%)

5 of 13 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
9	I	40	A
9	I	42	A
9	I	46	U
9	I	50	G
9	I	51	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
9	I	110	U

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
9	9QV	I	112	9	17,25,26	0.88	1 (5%)	20,37,40	1.86	6 (30%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
9	9QV	I	112	9	-	2/10/30/31	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	I	112	9QV	C6-C5	-2.08	1.33	1.38

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	I	112	9QV	C5-C4-N3	-4.25	113.96	123.31
9	I	112	9QV	O2P-P1-O1P	-3.58	96.67	110.68
9	I	112	9QV	O2P-P1-O2'	2.84	118.74	105.99
9	I	112	9QV	O3P-P1-O2'	2.79	118.52	105.99
9	I	112	9QV	O3P-P1-O1P	-2.76	99.88	110.68

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
9	I	112	9QV	C2'-O2'-P1-O2P
9	I	112	9QV	C2'-O2'-P1-O3P

There are no ring outliers.

No monomer is involved in short contacts.

## 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	387/452 (85%)	0.15	18 (4%) 31 15	70, 112, 157, 201	0
2	B	98/98 (100%)	0.36	8 (8%) 11 4	72, 120, 189, 250	0
3	C	81/92 (88%)	0.22	4 (4%) 29 14	80, 110, 169, 183	0
4	D	65/96 (67%)	0.39	5 (7%) 13 5	92, 125, 151, 174	0
5	E	87/96 (90%)	0.34	8 (9%) 9 3	89, 115, 164, 190	0
6	F	77/88 (87%)	0.20	4 (5%) 27 12	81, 116, 154, 215	0
7	G	70/118 (59%)	0.61	7 (10%) 7 2	78, 112, 184, 229	0
8	H	100/111 (90%)	0.13	4 (4%) 38 19	83, 127, 182, 265	0
9	I	72/83 (86%)	0.06	7 (9%) 7 2	88, 192, 317, 328	0
All	All	1037/1234 (84%)	0.23	65 (6%) 20 8	70, 117, 190, 328	0

The worst 5 of 65 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
6	F	86	ILE	10.9
1	A	26	THR	9.9
2	B	95	LYS	6.7
2	B	47	ASP	5.2
8	H	2	SER	4.9

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
9	9QV	I	112	24/25	0.94	0.16	84,112,151,172	0

### 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.