



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 03:17 pm BST

PDB ID : 5FJ4  
Title : Structure of the standard kink turn HmKt-7 as stem loop bound with U1A and L7Ae proteins  
Authors : Huang, L.; Lilley, D.M.J.  
Deposited on : 2015-10-06  
Resolution : 2.95 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	:	2.11
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

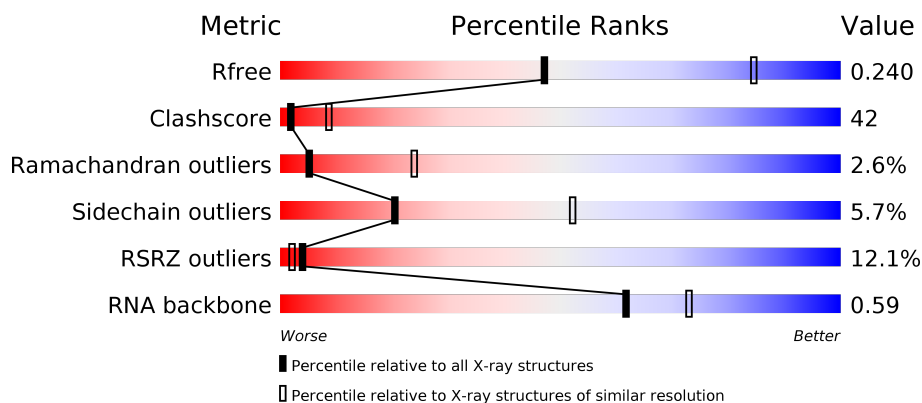
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.95 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)
RNA backbone	3102	1065 (3.22-2.70)

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	102	<div> <div>5%</div> <div> <div>61%</div> <div>32%</div> <div>• 5%</div> </div> </div>
1	B	102	<div> <div>11%</div> <div> <div>39%</div> <div>39%</div> <div>14%</div> <div>• 5%</div> </div> </div>
1	E	102	<div> <div>6%</div> <div> <div>61%</div> <div>28%</div> <div>6%</div> <div>5%</div> </div> </div>
1	F	102	<div> <div>19%</div> <div> <div>39%</div> <div>45%</div> <div>9%</div> <div>• 5%</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	C	123	<div><div></div><div>17%</div><div>39%</div><div>47%</div><div>9%</div><div></div><div></div></div>
2	G	123	<div><div></div><div>18%</div><div>36%</div><div>47%</div><div>12%</div><div></div><div></div></div>
3	D	35	<div><div></div><div>54%</div><div>34%</div><div>11%</div><div></div><div></div></div>
3	H	35	<div><div></div><div>66%</div><div>29%</div><div>6%</div><div></div><div></div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6471 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 U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	97	Total	C	N	O	S	0	0	0
			787	506	139	138	4			
1	B	97	Total	C	N	O	S	0	3	0
			794	512	139	139	4			
1	E	97	Total	C	N	O	S	0	0	0
			787	506	139	138	4			
1	F	97	Total	C	N	O	S	0	2	0
			785	506	137	138	4			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	31	HIS	TYR	conflict	UNP P09012
A	36	ARG	GLN	conflict	UNP P09012
B	31	HIS	TYR	conflict	UNP P09012
B	36	ARG	GLN	conflict	UNP P09012
E	31	HIS	TYR	conflict	UNP P09012
E	36	ARG	GLN	conflict	UNP P09012
F	31	HIS	TYR	conflict	UNP P09012
F	36	ARG	GLN	conflict	UNP P09012

- Molecule 2 is a protein called 50S RIBOSOMAL PROTEIN L7AE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	118	Total	C	N	O	S	0	0	0
			907	576	151	177	3			
2	G	118	Total	C	N	O	S	0	0	0
			907	576	151	177	3			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-3	GLY	-	expression tag	UNP O29494
C	-2	PRO	-	expression tag	UNP O29494
C	-1	GLU	-	expression tag	UNP O29494
C	0	ALA	-	expression tag	UNP O29494
C	1	SER	-	expression tag	UNP O29494
C	2	TYR	-	expression tag	UNP O29494
C	3	VAL	-	expression tag	UNP O29494
C	4	LYS	-	expression tag	UNP O29494
G	-3	GLY	-	expression tag	UNP O29494
G	-2	PRO	-	expression tag	UNP O29494
G	-1	GLU	-	expression tag	UNP O29494
G	0	ALA	-	expression tag	UNP O29494
G	1	SER	-	expression tag	UNP O29494
G	2	TYR	-	expression tag	UNP O29494
G	3	VAL	-	expression tag	UNP O29494
G	4	LYS	-	expression tag	UNP O29494

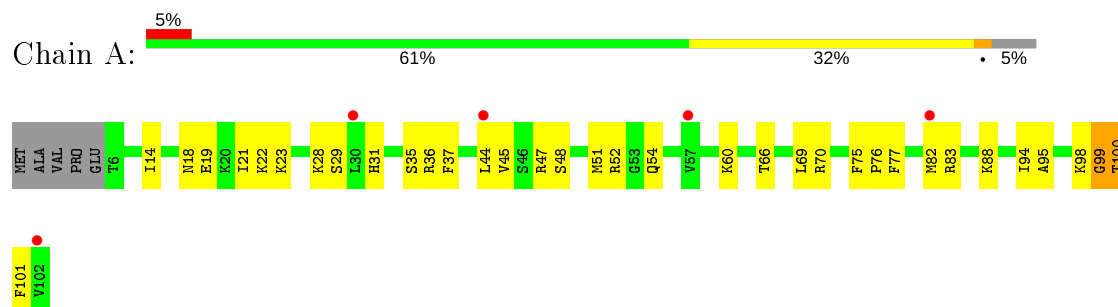
- Molecule 3 is a RNA chain called HMKT-7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	35	Total	C	N	O	P	0	0	0
			752	335	140	242	35			
3	H	35	Total	C	N	O	P	0	0	0
			752	335	140	242	35			

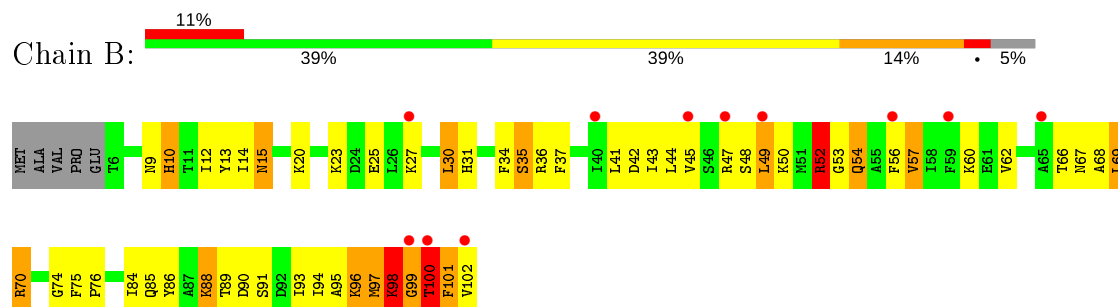
### 3 Residue-property plots [i](#)

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.

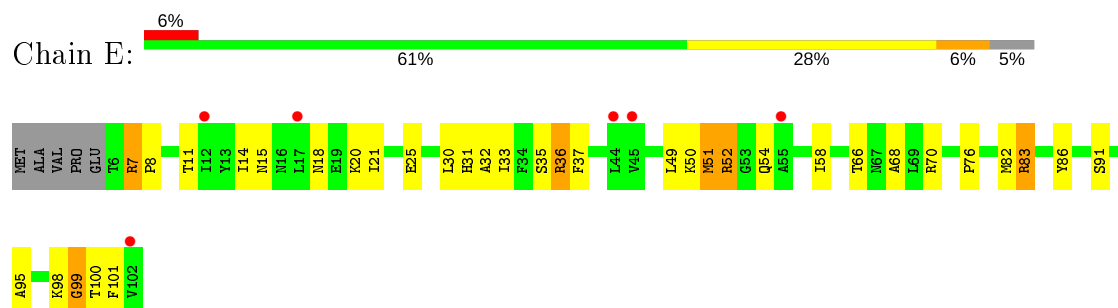
#### • Molecule 1: U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A



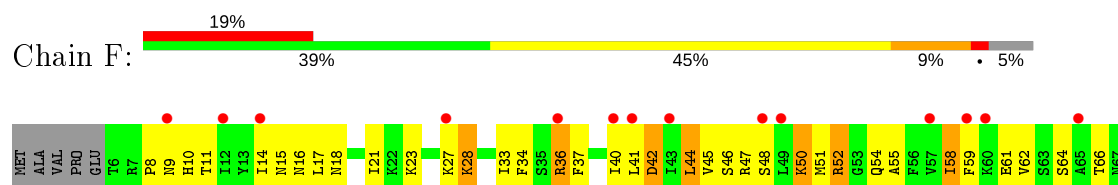
#### • Molecule 1: U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A

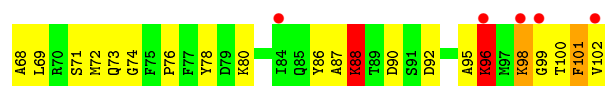


#### • Molecule 1: U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A

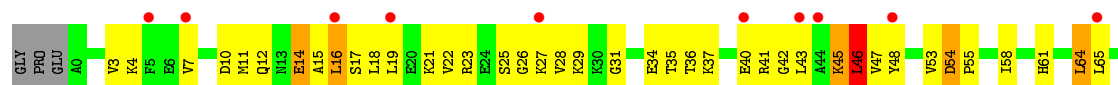


#### • Molecule 1: U1 SMALL NUCLEAR RIBONUCLEOPROTEIN A

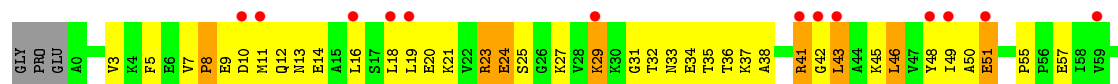




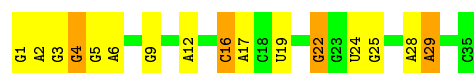
• Molecule 2: 50S RIBOSOMAL PROTEIN L7AE



• Molecule 2: 50S RIBOSOMAL PROTEIN L7AE



• Molecule 3: HMKT-7



• Molecule 3: HMKT-7



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	131.85Å 161.23Å 149.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	80.61 – 2.95 84.37 – 2.95	Depositor EDS
% Data completeness (in resolution range)	99.9 (80.61-2.95) 99.9 (84.37-2.95)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.22 (at 2.96Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.199 , 0.237 0.196 , 0.240	Depositor DCC
$R_{free}$ test set	1655 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	103.4	Xtriage
Anisotropy	0.008	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 102.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6471	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	128.0	wwPDB-VP

Xtriage'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 ⓘ

### 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	0.66	0/801	1.01	2/1071 (0.2%)
1	B	0.95	7/808 (0.9%)	1.09	4/1081 (0.4%)
1	E	0.66	0/801	0.98	5/1071 (0.5%)
1	F	0.72	0/799	1.05	5/1070 (0.5%)
2	C	0.82	3/918 (0.3%)	1.12	8/1238 (0.6%)
2	G	0.82	2/918 (0.2%)	1.35	14/1238 (1.1%)
3	D	0.78	0/841	1.24	6/1310 (0.5%)
3	H	0.63	0/841	1.20	3/1310 (0.2%)
All	All	0.76	12/6727 (0.2%)	1.14	47/9389 (0.5%)

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
1	F	0	1
2	C	0	2
2	G	0	3
All	All	0	6

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	14	GLU	CD-OE1	-7.99	1.16	1.25
2	C	14	GLU	CG-CD	-7.11	1.41	1.51
1	B	52	ARG	NE-CZ	-6.94	1.24	1.33
1	B	52	ARG	CZ-NH2	-6.37	1.24	1.33
2	G	113	LYS	CD-CE	-5.77	1.36	1.51

The worst 5 of 47 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	60	LYS	CD-CE-NZ	-12.64	82.63	111.70
2	G	43	LEU	CA-CB-CG	12.39	143.79	115.30
2	G	41	ARG	NE-CZ-NH2	9.91	125.26	120.30
2	C	64	LEU	CB-CG-CD2	-9.65	94.60	111.00
2	G	41	ARG	NE-CZ-NH1	-9.58	115.51	120.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	53	VAL	Peptide
2	C	54	ASP	Peptide
1	F	88	LYS	Peptide
2	G	76	VAL	Peptide
2	G	8	PRO	Peptide

## 5.2 Too-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 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	787	0	821	40	0
1	B	794	0	813	111	0
1	E	787	0	821	42	0
1	F	785	0	801	90	0
2	C	907	0	948	102	0
2	G	907	0	948	157	0
3	D	752	0	381	14	0
3	H	752	0	381	11	0
All	All	6471	0	5914	518	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 42.

The worst 5 of 518 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:G:111:VAL:O	2:G:115:LYS:NZ	1.68	1.27

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:G:41:ARG:NH1	2:G:98:ILE:HD11	1.53	1.23
1:B:43:ILE:CD1	1:B:57:VAL:HG23	1.74	1.16
2:G:41:ARG:HH12	2:G:98:ILE:CD1	1.59	1.15
2:G:27:LYS:O	2:G:98:ILE:HG22	1.43	1.14

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	95/102 (93%)	91 (96%)	3 (3%)	1 (1%)	14	46
1	B	98/102 (96%)	84 (86%)	7 (7%)	7 (7%)	1	4
1	E	95/102 (93%)	92 (97%)	3 (3%)	0	100	100
1	F	97/102 (95%)	87 (90%)	7 (7%)	3 (3%)	4	19
2	C	116/123 (94%)	103 (89%)	10 (9%)	3 (3%)	5	24
2	G	116/123 (94%)	103 (89%)	11 (10%)	2 (2%)	9	34
All	All	617/654 (94%)	560 (91%)	41 (7%)	16 (3%)	5	24

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	49	LEU
1	B	89	THR
1	B	97	MET
1	B	98	LYS
1	B	100	THR

### 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	87/91 (96%)	87 (100%)	0	100	100
1	B	85/91 (93%)	75 (88%)	10 (12%)	5	20
1	E	87/91 (96%)	82 (94%)	5 (6%)	20	52
1	F	84/91 (92%)	77 (92%)	7 (8%)	11	35
2	C	100/104 (96%)	96 (96%)	4 (4%)	31	64
2	G	100/104 (96%)	95 (95%)	5 (5%)	24	57
All	All	543/572 (95%)	512 (94%)	31 (6%)	20	52

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	7	ARG
1	E	52	ARG
2	G	79	LYS
1	E	36	ARG
1	E	83	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
2	C	61	HIS
2	G	13	ASN
2	C	99	ASN
1	B	54	GLN
1	E	39	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	D	34/35 (97%)	6 (17%)	0
3	H	34/35 (97%)	2 (5%)	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
All	All	68/70 (97%)	8 (11%)	0

5 of 8 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	D	4	G
3	D	5	G
3	D	6	A
3	D	19	U
3	D	22	G

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 ⓘ

### 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, 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	97/102 (95%)	0.68	5 (5%) 27 17	63, 90, 123, 148	0
1	B	97/102 (95%)	0.78	11 (11%) 5 3	81, 126, 191, 236	0
1	E	97/102 (95%)	0.77	6 (6%) 20 12	62, 87, 130, 159	0
1	F	97/102 (95%)	1.15	19 (19%) 1 0	78, 118, 194, 250	0
2	C	118/123 (95%)	1.02	21 (17%) 1 1	98, 145, 229, 278	0
2	G	118/123 (95%)	0.92	22 (18%) 1 0	113, 158, 234, 245	0
3	D	35/35 (100%)	0.12	0 100 100	90, 117, 166, 187	0
3	H	35/35 (100%)	0.07	0 100 100	85, 122, 179, 205	0
All	All	694/724 (95%)	0.81	84 (12%) 4 2	62, 122, 209, 278	0

The worst 5 of 84 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	99	GLY	10.2
1	F	49	LEU	6.4
2	C	101	GLY	5.5
2	C	100	GLU	5.4
1	B	100	THR	5.3

### 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.