



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 02:32 am BST

PDB ID : 3AV2  
Title : The human nucleosome structure containing the histone variant H3.3  
Authors : Tachiwana, H.; Osakabe, A.; Shiga, T.; Miya, M.; Kimura, H.; Kagawa, W.; Kurumizaka, H.  
Deposited on : 2011-02-18  
Resolution : 2.80 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



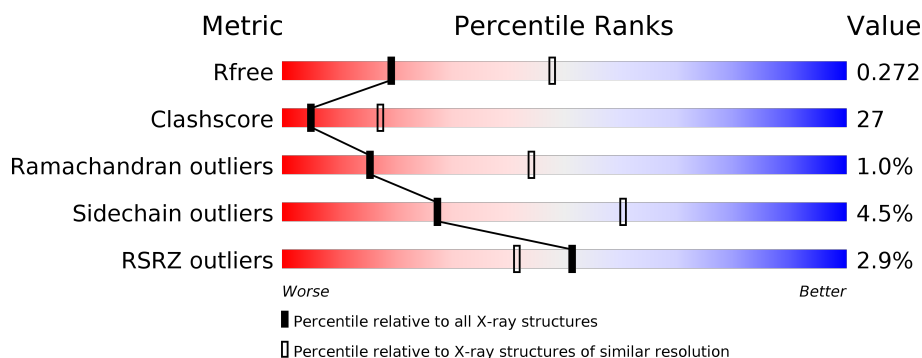
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	139	<div> <div>%</div> <div> <div></div> <div>45%</div> <div>24%</div> <div>30%</div> </div> </div>
1	E	139	<div> <div>%</div> <div> <div></div> <div>46%</div> <div>23%</div> <div>•</div> <div>29%</div> </div> </div>
2	B	106	<div> <div>%</div> <div> <div></div> <div>46%</div> <div>26%</div> <div>•</div> <div>26%</div> </div> </div>
2	F	106	<div> <div>%</div> <div> <div></div> <div>61%</div> <div>18%</div> <div>21%</div> </div> </div>
3	C	133	<div> <div>%</div> <div> <div></div> <div>53%</div> <div>23%</div> <div>•</div> <div>21%</div> </div> </div>
3	G	133	<div> <div>%</div> <div> <div></div> <div>56%</div> <div>20%</div> <div>•</div> <div>22%</div> </div> </div>

Continued on next page...



Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	129	<div><div><div></div><div></div><div></div><div></div></div><div>2%46%26%•27%</div></div>
4	H	129	<div><div><div></div><div></div><div></div><div></div></div><div>%47%22%••29%</div></div>
5	I	146	<div><div><div></div><div></div><div></div><div></div></div><div>8%17%83%</div></div>
5	J	146	<div><div><div></div><div></div><div></div><div></div></div><div>6%8%92%</div></div>



## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 12026 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 Histone H3.3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	97	Total	C	N	O	S	0	0	0
			797	503	155	137	2			
1	E	98	Total	C	N	O	S	0	0	0
			803	506	156	139	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	EXPRESSION TAG	UNP P84243
A	-2	SER	-	EXPRESSION TAG	UNP P84243
A	-1	HIS	-	EXPRESSION TAG	UNP P84243
E	-3	GLY	-	EXPRESSION TAG	UNP P84243
E	-2	SER	-	EXPRESSION TAG	UNP P84243
E	-1	HIS	-	EXPRESSION TAG	UNP P84243

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	78	Total	C	N	O	S	0	0	0
			619	391	120	107	1			
2	F	84	Total	C	N	O	S	0	0	0
			673	424	133	115	1			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	EXPRESSION TAG	UNP P62805
B	-2	SER	-	EXPRESSION TAG	UNP P62805
B	-1	HIS	-	EXPRESSION TAG	UNP P62805
F	-3	GLY	-	EXPRESSION TAG	UNP P62805
F	-2	SER	-	EXPRESSION TAG	UNP P62805
F	-1	HIS	-	EXPRESSION TAG	UNP P62805



- Molecule 3 is a protein called Histone H2A type 1-B/E.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	105	Total	C	N	O	0	0	0
			810	511	158	141			
3	G	104	Total	C	N	O	0	0	0
			805	508	157	140			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-3	GLY	-	EXPRESSION TAG	UNP P04908
C	-2	SER	-	EXPRESSION TAG	UNP P04908
C	-1	HIS	-	EXPRESSION TAG	UNP P04908
G	-3	GLY	-	EXPRESSION TAG	UNP P04908
G	-2	SER	-	EXPRESSION TAG	UNP P04908
G	-1	HIS	-	EXPRESSION TAG	UNP P04908

- Molecule 4 is a protein called Histone H2B type 1-J.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	94	Total	C	N	O	S	0	0	0
			736	462	134	138	2			
4	H	92	Total	C	N	O	S	0	0	0
			719	453	129	135	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	GLY	-	EXPRESSION TAG	UNP P06899
D	-2	SER	-	EXPRESSION TAG	UNP P06899
D	-1	HIS	-	EXPRESSION TAG	UNP P06899
H	-3	GLY	-	EXPRESSION TAG	UNP P06899
H	-2	SER	-	EXPRESSION TAG	UNP P06899
H	-1	HIS	-	EXPRESSION TAG	UNP P06899

- Molecule 5 is a DNA chain called 146-MER DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	146	Total	C	N	O	P	0	0	0
			2990	1431	540	874	145			
5	J	146	Total	C	N	O	P	0	0	0
			2990	1431	540	874	145			



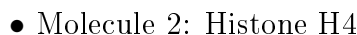
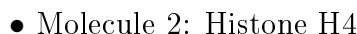
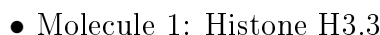
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	6	Total O 6 6	0	0
6	B	5	Total O 5 5	0	0
6	C	10	Total O 10 10	0	0
6	D	4	Total O 4 4	0	0
6	E	18	Total O 18 18	0	0
6	F	14	Total O 14 14	0	0
6	G	5	Total O 5 5	0	0
6	H	1	Total O 1 1	0	0
6	I	11	Total O 11 11	0	0
6	J	10	Total O 10 10	0	0



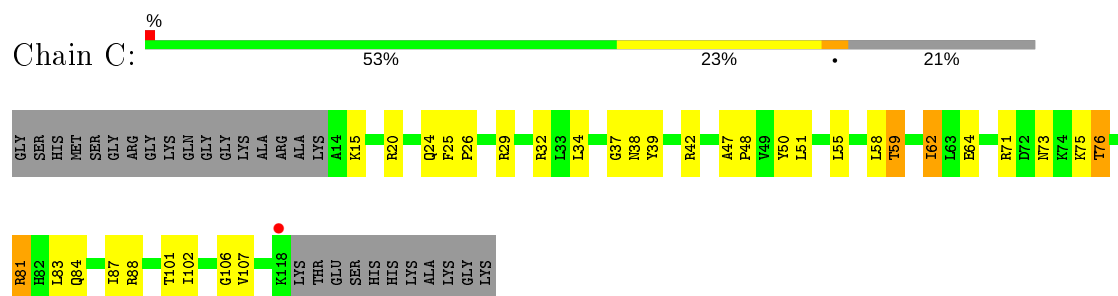


- Molecule 1: Histone H3.3

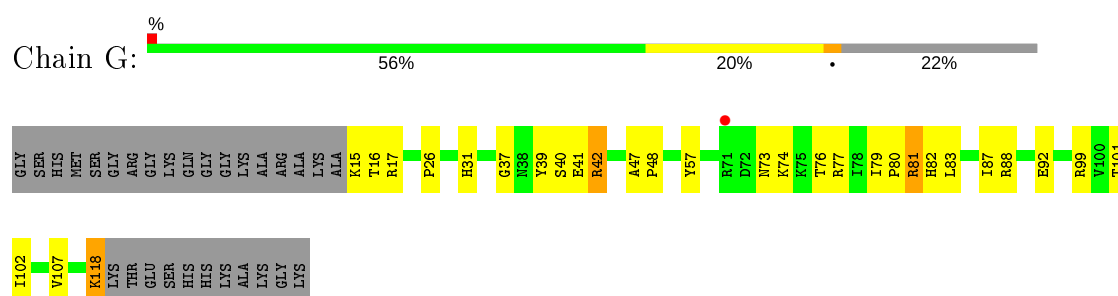




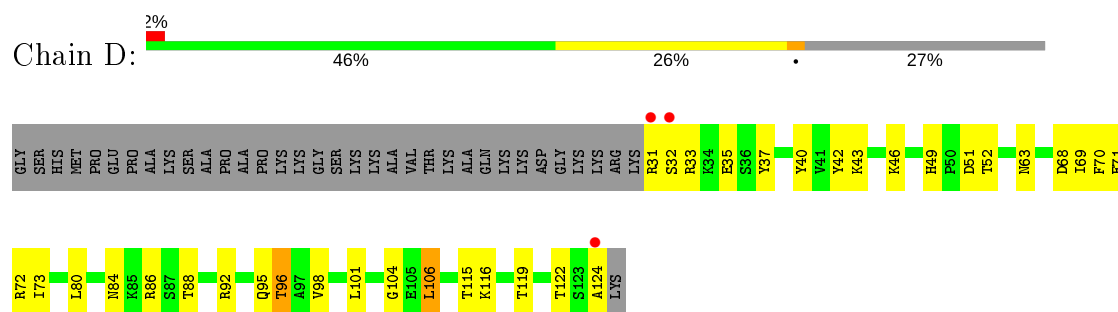
- Molecule 3: Histone H2A type 1-B/E



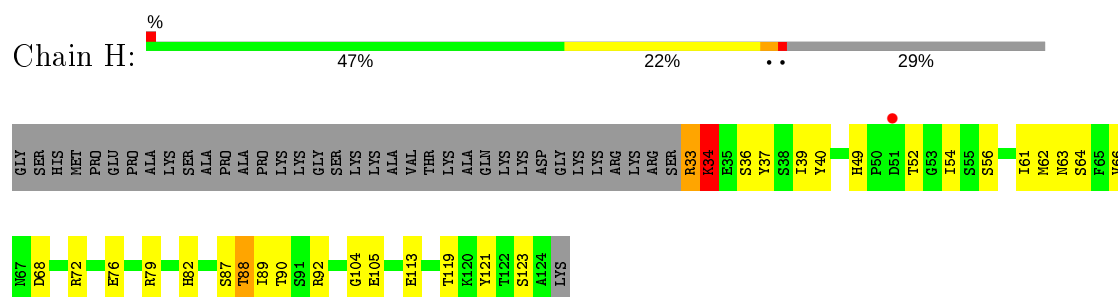
- Molecule 3: Histone H2A type 1-B/E



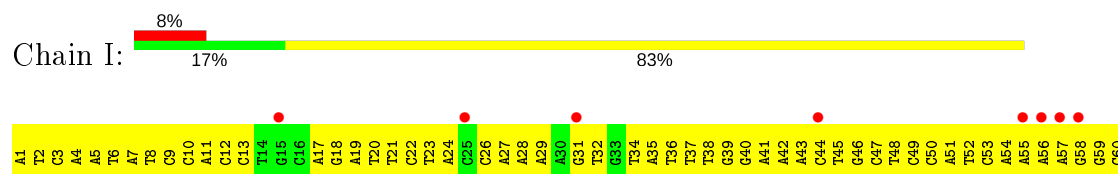
- Molecule 4: Histone H2B type 1-J



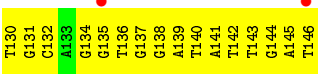
- Molecule 4: Histone H2B type 1-J



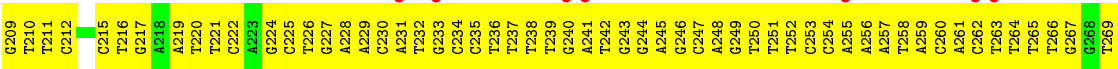
- Molecule 5: 146-MER DNA







● Molecule 5: 146-MER DNA





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.84Å 109.69Å 181.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 38.93 – 2.79	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.80) 99.6 (38.93-2.79)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.38 (at 2.81Å)	Xtriage
Refinement program	CNS 1.2.1	Depositor
R, $R_{free}$	0.230 , 0.272 0.230 , 0.272	Depositor DCC
$R_{free}$ test set	2674 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	55.7	Xtriage
Anisotropy	0.426	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 72.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.025 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	12026	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.46% 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	0/809	0.64	0/1085
1	E	0.51	0/815	0.69	0/1092
2	B	0.43	0/626	0.68	0/837
2	F	0.50	0/680	0.72	0/908
3	C	0.44	0/820	0.67	0/1107
3	G	0.41	0/815	0.60	0/1100
4	D	0.46	0/747	0.67	0/1004
4	H	0.47	0/730	0.62	0/982
5	I	0.40	0/3354	0.81	0/5175
5	J	0.38	0/3354	0.78	0/5175
All	All	0.43	0/12750	0.74	0/18465

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	797	0	835	33	0
1	E	803	0	840	32	0
2	B	619	0	659	37	0
2	F	673	0	722	17	0
3	C	810	0	866	37	0
3	G	805	0	861	34	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	736	0	758	39	0
4	H	719	0	740	30	0
5	I	2990	0	1652	182	0
5	J	2990	0	1652	216	0
6	A	6	0	0	0	0
6	B	5	0	0	1	0
6	C	10	0	0	1	0
6	D	4	0	0	1	0
6	E	18	0	0	1	0
6	F	14	0	0	0	0
6	G	5	0	0	0	0
6	H	1	0	0	0	0
6	I	11	0	0	0	0
6	J	10	0	0	3	0
All	All	12026	0	9585	579	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 27.

The worst 5 of 579 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)
5:I:36:DT:H2''	5:I:37:DT:H5''	1.22	1.14
1:E:134:ARG:HH11	1:E:134:ARG:HB3	1.12	1.14
5:I:51:DA:H2''	5:I:52:DT:H5''	1.28	1.12
5:I:20:DT:H2''	5:I:21:DT:H5''	1.32	1.12
5:I:52:DT:H2''	5:I:53:DC:H5'	1.34	1.04

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	95/139 (68%)	94 (99%)	1 (1%)	0	100	100
1	E	96/139 (69%)	91 (95%)	4 (4%)	1 (1%)	15	44
2	B	76/106 (72%)	74 (97%)	2 (3%)	0	100	100
2	F	82/106 (77%)	78 (95%)	3 (4%)	1 (1%)	13	39
3	C	103/133 (77%)	98 (95%)	5 (5%)	0	100	100
3	G	102/133 (77%)	96 (94%)	6 (6%)	0	100	100
4	D	92/129 (71%)	86 (94%)	5 (5%)	1 (1%)	14	41
4	H	90/129 (70%)	84 (93%)	2 (2%)	4 (4%)	2	8
All	All	736/1014 (73%)	701 (95%)	28 (4%)	7 (1%)	15	44

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	H	34	LYS
4	H	104	GLY
4	H	123	SER
4	D	104	GLY
1	E	134	ARG

### 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	83/112 (74%)	80 (96%)	3 (4%)	35	69
1	E	83/112 (74%)	82 (99%)	1 (1%)	71	92
2	B	63/81 (78%)	62 (98%)	1 (2%)	62	88
2	F	69/81 (85%)	67 (97%)	2 (3%)	42	76
3	C	83/102 (81%)	77 (93%)	6 (7%)	14	38
3	G	83/102 (81%)	78 (94%)	5 (6%)	19	48
4	D	80/107 (75%)	76 (95%)	4 (5%)	24	56
4	H	78/107 (73%)	72 (92%)	6 (8%)	13	35

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	622/804 (77%)	594 (96%)	28 (4%)	27 60

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	101	LEU
2	F	24	ASP
4	H	56	SER
4	D	106	LEU
1	E	122	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	49	HIS
4	D	63	ASN
2	F	93	GLN
3	C	73	ASN
3	C	84	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 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 [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	97/139 (69%)	-0.06	1 (1%) 82 77	32, 46, 68, 93	0
1	E	98/139 (70%)	-0.16	1 (1%) 82 77	25, 38, 61, 77	0
2	B	78/106 (73%)	-0.25	1 (1%) 77 72	31, 46, 62, 71	0
2	F	84/106 (79%)	-0.12	1 (1%) 79 73	23, 38, 53, 84	0
3	C	105/133 (78%)	-0.20	1 (0%) 82 77	24, 43, 67, 82	0
3	G	104/133 (78%)	-0.18	1 (0%) 82 77	33, 49, 73, 86	0
4	D	94/129 (72%)	-0.04	3 (3%) 47 37	29, 46, 71, 122	0
4	H	92/129 (71%)	-0.14	1 (1%) 80 75	34, 49, 74, 89	0
5	I	146/146 (100%)	0.52	11 (7%) 14 8	47, 104, 143, 168	0
5	J	146/146 (100%)	0.34	9 (6%) 20 13	54, 104, 145, 163	0
All	All	1044/1306 (79%)	0.02	30 (2%) 51 41	23, 50, 132, 168	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	146	DT	10.8
5	J	148	DT	3.9
5	I	56	DA	3.5
5	J	147	DA	3.5
5	J	262	DC	3.4

### 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 [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.