



# wwPDB X-ray Structure Validation Summary Report

Apr 8, 2014 – 02:14 AM EDT

PDB ID : 4KGC  
Title : Nucleosome Core Particle Containing (ETA6-P-CYMENE)-(1, 2-ETHYLENEDIAMINE)-RUTHENIUM  
Authors : Adhireksan, Z.; Davey, C.A.  
Deposited on : 2013-04-29  
Resolution : 2.69 Å(reported)

This is a wwPDB 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

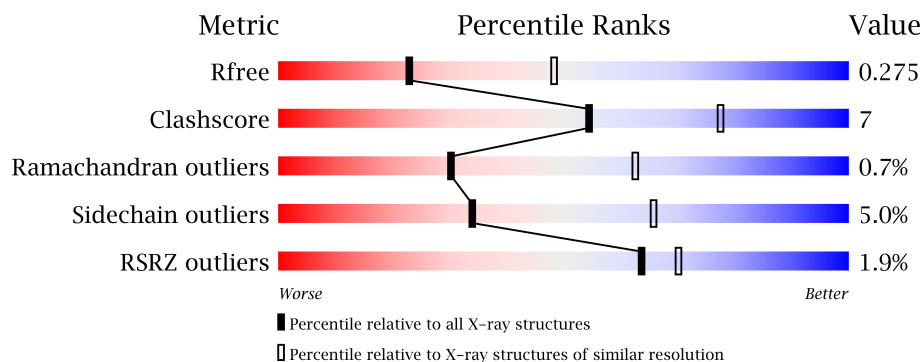
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	<b>FAILED</b>
Xtriage (Phenix)	:	dev-1439
EDS	:	stable22978
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)	:	stable22978

# 1 Overall quality at a glance

The reported resolution of this entry is 2.69 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	1557 (2.70-2.70)
Clashscore	79885	1939 (2.70-2.70)
Ramachandran outliers	78287	1905 (2.70-2.70)
Sidechain outliers	78261	1905 (2.70-2.70)
RSRZ outliers	66119	1559 (2.70-2.70)

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. 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	136	
1	E	136	
2	B	103	
2	F	103	
3	C	130	
3	G	130	
4	D	126	
4	H	126	
5	I	145	
6	J	145	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
7	HRU	G	201	-	X
7	HRU	I	101	-	X
7	HRU	J	101	-	X
8	SO4	H	202	-	X
9	MG	E	201	-	X

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 12091 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 Histone H3.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	97	Total	C	N	O	S	0	0	0
			797	504	152	138	3			
1	E	97	Total	C	N	O	S	0	0	0
			797	504	152	138	3			

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	82	Total	C	N	O	S	0	0	0
			653	412	127	113	1			
2	F	87	Total	C	N	O	S	0	0	0
			703	442	142	118	1			

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	106	Total	C	N	O	0	0	0
			818	516	160	142			
3	G	106	Total	C	N	O	0	0	0
			818	516	160	142			

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	95	Total	C	N	O	S	0	0	0
			745	469	134	140	2			
4	H	95	Total	C	N	O	S	0	0	0
			745	469	134	140	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	29	THR	SER	SEE REMARK 999	UNP P02281
H	29	THR	SER	SEE REMARK 999	UNP P02281

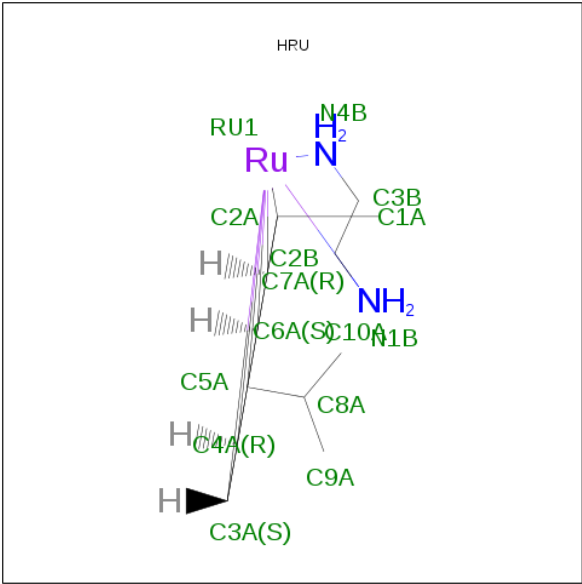
- Molecule 5 is a DNA chain called DNA (145-mer).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	145	Total	C	N	O	P	0	0	0
			2970	1421	538	867	144			

- Molecule 6 is a DNA chain called DNA (145-mer).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	J	145	Total	C	N	O	P	0	0	0
			2969	1421	535	869	144			

- Molecule 7 is (ETHANE-1,2-DIAMINE-KAPPA 2 N,N')[(1,2,3,4,5,6-ETA)-1-METHYL-4-(PROPAN-2-YL)CYCLOHEXANE-1,2,3,4,5,6-HEXAYL]RUTHENIUM (three-letter code: HRU) (formula: C<sub>12</sub>H<sub>22</sub>N<sub>2</sub>Ru).



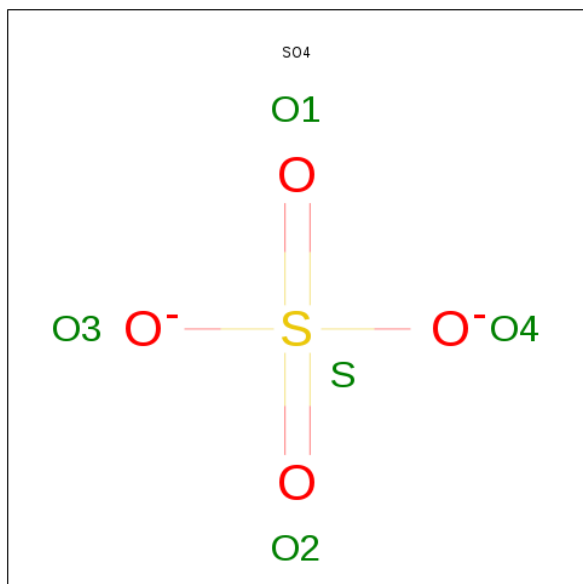
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	C	1	Total	C	N	Ru	0	0
			15	12	2	1		
7	G	1	Total	C	N	Ru	0	0
			15	12	2	1		
7	I	1	Total	C	N	Ru	0	0
			15	12	2	1		

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	J	1	Total	C	N	Ru	0	0
			15	12	2	1		

- Molecule 8 is SULFATE ION (three-letter code: SO<sub>4</sub>) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	D	1	Total	O	S	0	0
			5	4	1		
8	H	1	Total	O	S	0	0
			5	4	1		
8	H	1	Total	O	S	0	0
			5	4	1		

- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

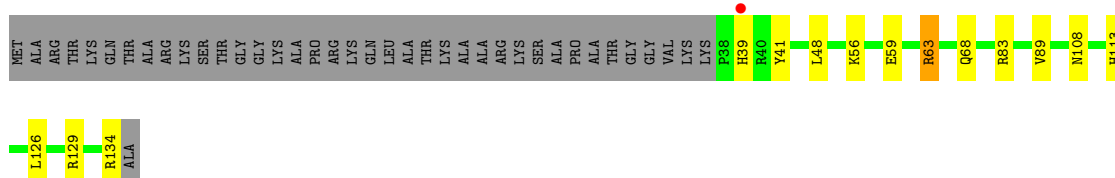
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	E	1	Total	Mg	0	0
			1	1		

### 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.

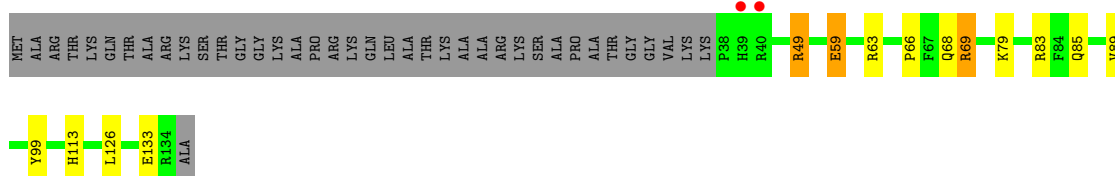
- Molecule 1: Histone H3.2

Chain A: 



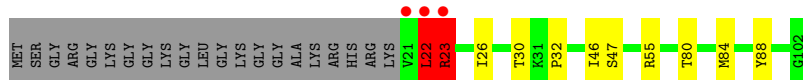
- Molecule 1: Histone H3.2

Chain E: 



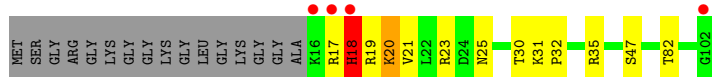
- Molecule 2: Histone H4

Chain B: 



- Molecule 2: Histone H4

Chain F: 

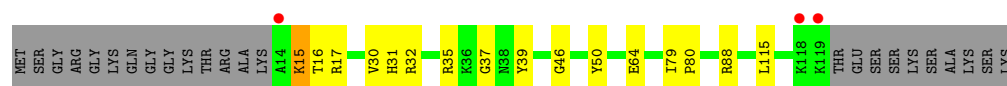


- Molecule 3: Histone H2A

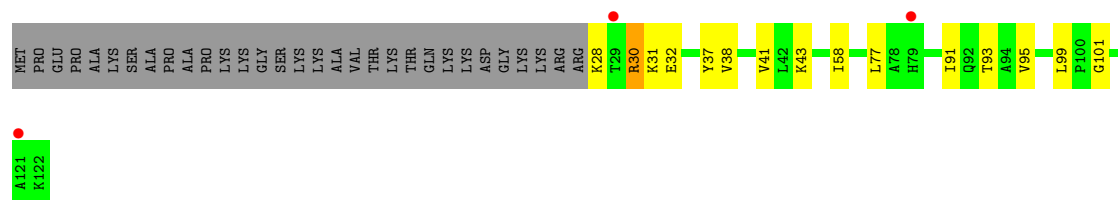
Chain C: 



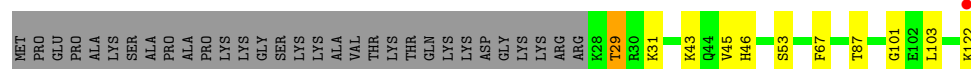
- Molecule 3: Histone H2A

Chain G: 

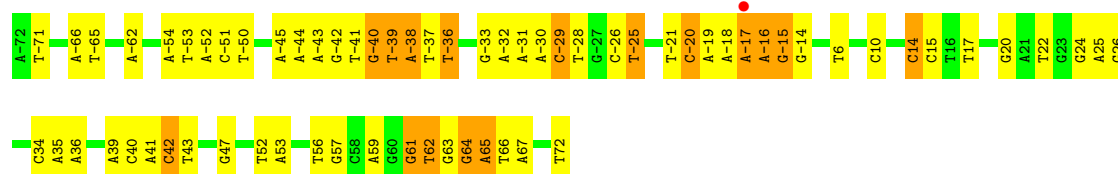
- Molecule 4: Histone H2B 1.1

Chain D: 

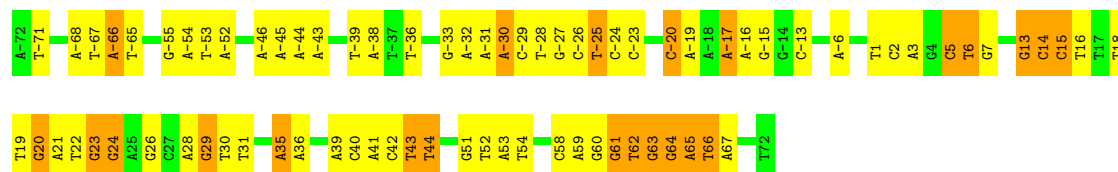
- Molecule 4: Histone H2B 1.1

Chain H: 

- Molecule 5: DNA (145-mer)

Chain I: 

- Molecule 6: DNA (145-mer)

Chain J: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.62Å 109.71Å 181.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	60.60 – 2.69 58.53 – 2.69	Depositor EDS
% Data completeness (in resolution range)	92.9 (60.60-2.69) 92.9 (58.53-2.69)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.43 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.248 , 0.282 0.241 , 0.275	Depositor DCC
$R_{free}$ test set	1143 reflections (2.09%)	DCC
Wilson B-factor (Å <sup>2</sup> )	73.4	Xtriage
Anisotropy	0.087	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 34.1	EDS
Estimated twinning fraction	0.020 for k,h,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 55720 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	12091	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.15% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: HRU, MG, SO4

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	0.50	0/809	0.61	0/1085
1	E	0.51	0/809	0.71	1/1085 (0.1%)
2	B	0.47	0/660	0.61	0/883
2	F	0.58	0/711	0.72	0/948
3	C	0.53	0/828	0.70	1/1117 (0.1%)
3	G	0.44	0/828	0.60	0/1117
4	D	0.52	0/756	0.59	0/1015
4	H	0.47	0/756	0.55	0/1015
5	I	0.78	0/3332	1.47	44/5141 (0.9%)
6	J	0.79	0/3330	1.48	48/5138 (0.9%)
All	All	0.67	0/12819	1.18	94/18544 (0.5%)

There are no bond length outliers.

The worst 5 of 94 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	I	-25	DT	O4'-C1'-N1	9.83	114.88	108.00
6	J	-25	DT	O4'-C1'-N1	9.74	114.82	108.00
6	J	53	DA	O4'-C1'-N9	9.19	114.44	108.00
5	I	-38	DA	P-O3'-C3'	8.20	129.54	119.70
5	I	14	DC	O4'-C1'-N1	-7.99	102.41	108.00

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	797	0	832	9	0
1	E	797	0	832	15	0
2	B	653	0	696	7	0
2	F	703	0	755	10	0
3	C	818	0	877	16	0
3	G	818	0	877	12	0
4	D	745	0	773	11	0
4	H	745	0	773	9	0
5	I	2970	0	1640	38	0
6	J	2969	0	1641	52	0
7	C	15	0	0	0	0
7	G	15	0	0	0	0
7	I	15	0	0	0	0
7	J	15	0	0	0	0
8	D	5	0	0	0	0
8	H	10	0	0	5	0
9	E	1	0	0	0	0
All	All	12091	0	9696	146	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 7.

The worst 5 of 146 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:E:49:ARG:HG3	1:E:49:ARG:HH11	1.34	0.91
6:J:22:DT:H2''	6:J:23:DG:O4'	1.78	0.82
3:C:32:ARG:NH1	5:I:44:DA:OP1	2.16	0.79
6:J:-26:DC:H2''	6:J:-25:DT:C6	2.16	0.79
5:I:-33:DG:H2''	5:I:-32:DA:H5'	1.68	0.76

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	95/136 (70%)	94 (99%)	1 (1%)	0	100	100
1	E	95/136 (70%)	95 (100%)	0	0	100	100
2	B	80/103 (78%)	77 (96%)	1 (1%)	2 (2%)	9	21
2	F	85/103 (82%)	83 (98%)	1 (1%)	1 (1%)	19	45
3	C	104/130 (80%)	98 (94%)	6 (6%)	0	100	100
3	G	104/130 (80%)	101 (97%)	3 (3%)	0	100	100
4	D	93/126 (74%)	92 (99%)	0	1 (1%)	21	49
4	H	93/126 (74%)	91 (98%)	1 (1%)	1 (1%)	21	49
All	All	749/990 (76%)	731 (98%)	13 (2%)	5 (1%)	30	62

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	101	GLY
2	F	18	HIS
2	B	22	LEU
4	H	101	GLY
2	B	23	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	84/111 (76%)	80 (95%)	4 (5%)	35	68
1	E	84/111 (76%)	81 (96%)	3 (4%)	47	79
2	B	67/79 (85%)	64 (96%)	3 (4%)	38	70
2	F	72/79 (91%)	67 (93%)	5 (7%)	22	48
3	C	84/102 (82%)	79 (94%)	5 (6%)	27	56
3	G	84/102 (82%)	80 (95%)	4 (5%)	35	68
4	D	81/106 (76%)	78 (96%)	3 (4%)	45	78

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	H	81/106 (76%)	76 (94%)	5 (6%)	26 54
All	All	637/796 (80%)	605 (95%)	32 (5%)	34 66

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	31	LYS
1	E	79	LYS
4	H	53	SER
1	E	59	GLU
2	F	18	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	31	HIS
4	H	92	GLN
4	D	92	GLN
1	A	108	ASN
3	G	31	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.5 Carbohydrates ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.6 Ligand geometry ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.7 Other polymers ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 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, 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	97/136 (71%)	0.16	1 (1%) 79 83	41, 58, 78, 96	0
1	E	97/136 (71%)	0.17	2 (2%) 60 67	34, 47, 67, 86	0
2	B	82/103 (79%)	0.25	3 (3%) 39 44	45, 53, 75, 96	0
2	F	87/103 (84%)	0.39	4 (4%) 31 35	36, 46, 71, 106	0
3	C	106/130 (81%)	0.18	1 (0%) 81 85	38, 52, 73, 86	0
3	G	106/130 (81%)	0.26	3 (2%) 50 56	44, 62, 85, 95	0
4	D	95/126 (75%)	0.35	3 (3%) 45 50	41, 54, 87, 97	0
4	H	95/126 (75%)	0.37	1 (1%) 77 82	47, 64, 93, 98	0
5	I	145/145 (100%)	-0.19	1 (0%) 84 89	59, 109, 145, 160	0
6	J	145/145 (100%)	-0.17	0 100 100	63, 111, 146, 152	0
All	All	1055/1280 (82%)	0.14	19 (1%) 64 71	34, 61, 136, 160	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	17	ARG	7.0
3	C	119	LYS	7.0
2	B	21	VAL	6.0
2	B	22	LEU	4.8
1	A	39	HIS	4.5

### 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 ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
9	MG	E	201	1/1	0.49	8.99	58,58,58,58	0
7	HRU	G	201	15/15	0.23	8.02	154,154,154,155	0
7	HRU	J	101	15/15	0.21	4.06	137,138,138,138	0
7	HRU	I	101	15/15	0.33	3.50	181,182,183,184	0
8	SO4	H	202	5/5	0.33	2.19	87,89,90,91	0
8	SO4	D	201	5/5	0.23	1.40	84,85,85,86	0
8	SO4	H	201	5/5	0.22	1.33	91,91,93,94	0
7	HRU	C	201	15/15	0.25	1.15	174,175,175,175	0

### 6.5 Other polymers ⓘ

There are no such residues in this entry.