



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 27, 2016 – 10:35 PM EDT

PDB ID : 5E5A  
Title : Crystal structure of the chromatin-tethering domain of Human cytomegalovirus IE1 protein bound to the nucleosome core particle  
Authors : Fang, Q.; Chen, P.; Wang, M.; Fang, J.; Yang, N.; Li, G.; Xu, R.M.  
Deposited on : 2015-10-08  
Resolution : 2.81 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  
<http://wwpdb.org/validation/2016/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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027107  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027107

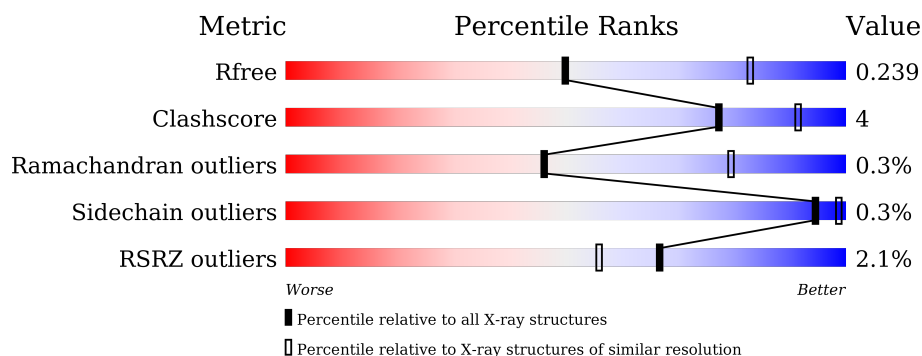
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.81 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))
$R_{free}$	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (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. 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	I	146	<div> <div>5%</div> <div> <div></div> <div>75%</div> <div>25%</div> </div> </div>
1	J	146	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>25%</div> </div> </div>
2	A	136	<div> <div>2%</div> <div> <div></div> <div>67%</div> <div>6%</div> <div>27%</div> </div> </div>
2	E	136	<div> <div>%</div> <div> <div></div> <div>65%</div> <div>7%</div> <div>28%</div> </div> </div>
3	B	103	<div> <div></div> <div> <div></div> <div>73%</div> <div></div> <div>23%</div> </div> </div>
3	F	103	<div> <div>%</div> <div> <div></div> <div>75%</div> <div>8%</div> <div>17%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	C	130	<div><div><div>%</div><div><div></div><div></div><div></div></div><div>75%6%19%</div></div></div>
4	G	130	<div><div><div></div><div></div><div></div></div><div>77%19%</div></div>
5	D	123	<div><div><div>3%</div><div><div></div><div></div><div></div></div><div>73%7%20%</div></div></div>
5	H	123	<div><div><div>%</div><div><div></div><div></div><div></div></div><div>67%9%24%</div></div></div>
6	K	16	<div><div><div></div><div></div><div></div></div><div>75%13%13%</div></div>

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 12407 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 DNA chain called DNA (146-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	I	146	Total	C	N	O	P	0	0	0
			2990	1430	541	874	145			
1	J	146	Total	C	N	O	P	0	0	0
			2993	1430	541	876	146			

- Molecule 2 is a protein called Histone H3.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	99	Total	C	N	O	S	0	0	0
			817	515	158	141	3			
2	E	98	Total	C	N	O	S	0	1	0
			816	517	156	140	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	ALA	GLY	conflict	UNP P84233
E	102	ALA	GLY	conflict	UNP P84233

- Molecule 3 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	79	Total	C	N	O	S	0	0	0
			627	395	121	110	1			
3	F	86	Total	C	N	O	S	0	1	0
			702	441	143	117	1			

- Molecule 4 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	C	105	Total	C	N	O	0	0	0
			809	510	158	141			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	G	105	Total	C	N	O	0	0	0
			809	510	158	141			

- Molecule 5 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	D	98	Total	C	N	O	S	0	0	0
			775	487	144	142	2			
5	H	94	Total	C	N	O	S	0	0	0
			731	460	131	138	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	MET	-	expression tag	UNP P02281
D	29	THR	SER	conflict	UNP P02281
H	0	MET	-	expression tag	UNP P02281
H	29	THR	SER	conflict	UNP P02281

- Molecule 6 is a protein called C-terminal domain of Regulatory protein IE1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	K	14	Total	C	N	O	S	0	0	0
			104	62	21	20	1			

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	I	1	Total	Mg	0	0
			1	1		
7	D	1	Total	Mg	0	0
			1	1		
7	C	1	Total	Mg	0	0
			1	1		
7	E	1	Total	Mg	0	0
			1	1		

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	I	15	Total	O	0	0
			15	15		

*Continued on next page...*

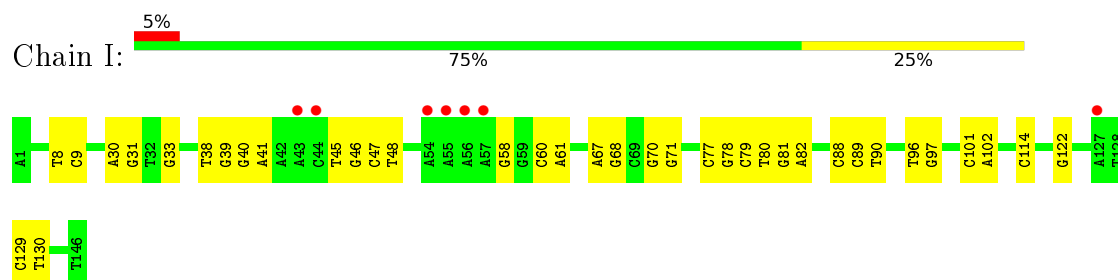
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	J	15	Total 15	O 15	0	0
8	A	24	Total 24	O 24	0	0
8	B	18	Total 18	O 18	0	0
8	C	33	Total 33	O 33	0	0
8	D	22	Total 22	O 22	0	0
8	E	37	Total 37	O 37	0	0
8	F	34	Total 34	O 34	0	0
8	G	19	Total 19	O 19	0	0
8	H	11	Total 11	O 11	0	0
8	K	2	Total 2	O 2	0	0

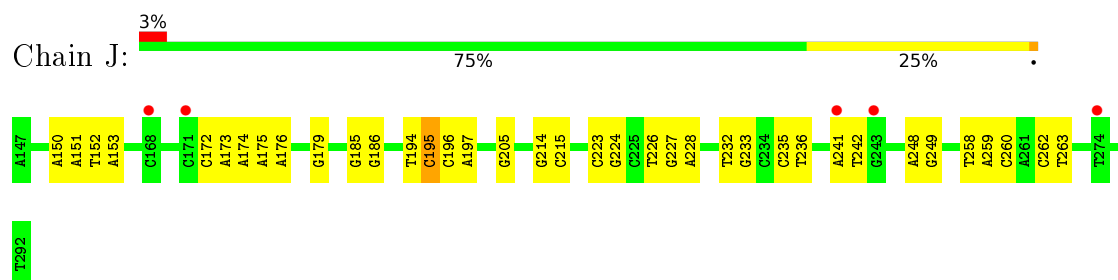
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

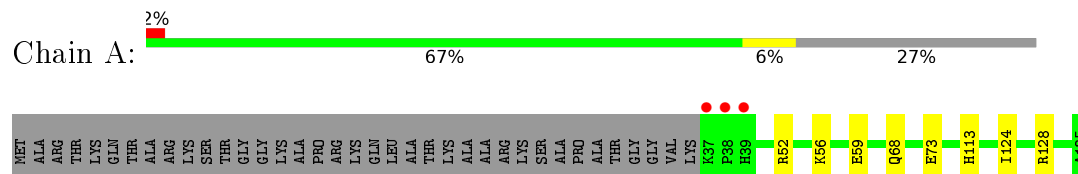
- Molecule 1: DNA (146-MER)



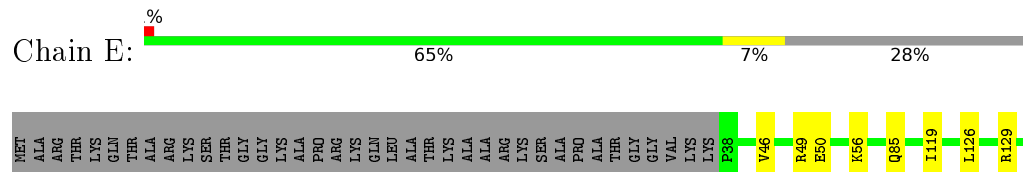
- Molecule 1: DNA (146-MER)



- Molecule 2: Histone H3.2

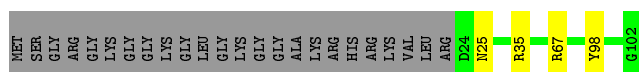


- Molecule 2: Histone H3.2

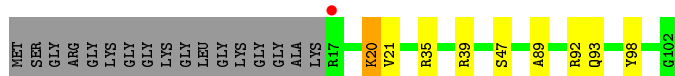
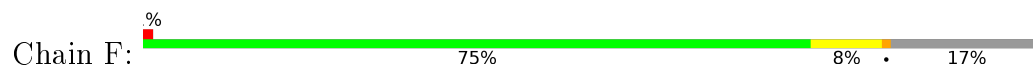


- Molecule 3: Histone H4

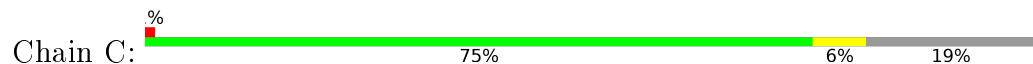




- Molecule 3: Histone H4



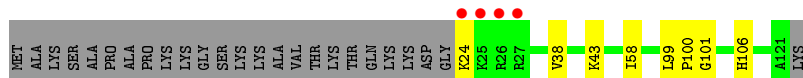
- Molecule 4: Histone H2A



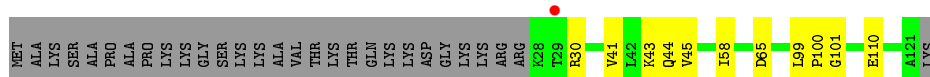
- Molecule 4: Histone H2A



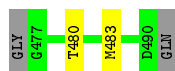
- Molecule 5: Histone H2B 1.1



- Molecule 5: Histone H2B 1.1



- Molecule 6: C-terminal domain of Regulatory protein IE1





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.70Å 109.47Å 181.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.81 30.12 – 2.81	Depositor EDS
% Data completeness (in resolution range)	99.5 (30.00-2.81) 99.5 (30.12-2.81)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.63 (at 2.80Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.196 , 0.242 0.197 , 0.239	Depositor DCC
$R_{free}$ test set	4979 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	54.7	Xtriage
Anisotropy	0.355	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 47.6	EDS
Estimated twinning fraction	0.015 for k,h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 52391 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	12407	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.88% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: MG

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	I	0.75	1/3354 (0.0%)	0.97	0/5175
1	J	0.78	1/3357 (0.0%)	0.97	1/5179 (0.0%)
2	A	0.41	0/829	0.50	0/1111
2	E	0.46	0/832	0.52	0/1115
3	B	0.41	0/634	0.54	0/848
3	F	0.45	0/713	0.59	0/951
4	C	0.43	0/819	0.53	0/1106
4	G	0.36	0/819	0.47	0/1106
5	D	0.45	0/786	0.54	0/1054
5	H	0.39	0/742	0.53	0/1000
6	K	0.26	0/105	0.48	0/139
All	All	0.62	2/12990 (0.0%)	0.80	1/18784 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	J	195	DC	C3'-O3'	-5.64	1.36	1.44
1	I	58	DG	C3'-O3'	-5.23	1.37	1.44

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	J	197	DA	O5'-P-OP2	-5.38	100.86	105.70

There are no chirality outliers.

There are no planarity outliers.

## 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	I	2990	0	1651	26	0
1	J	2993	0	1650	25	0
2	A	817	0	858	7	0
2	E	816	0	855	8	0
3	B	627	0	663	3	0
3	F	702	0	755	7	0
4	C	809	0	864	7	0
4	G	809	0	864	3	0
5	D	775	0	812	6	0
5	H	731	0	749	8	0
6	K	104	0	106	2	0
7	C	1	0	0	0	0
7	D	1	0	0	0	0
7	E	1	0	0	0	0
7	I	1	0	0	0	0
8	A	24	0	0	1	0
8	B	18	0	0	0	0
8	C	33	0	0	1	0
8	D	22	0	0	1	0
8	E	37	0	0	0	0
8	F	34	0	0	0	0
8	G	19	0	0	0	0
8	H	11	0	0	0	0
8	I	15	0	0	1	0
8	J	15	0	0	1	0
8	K	2	0	0	0	0
All	All	12407	0	9827	80	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 4.

The worst 5 of 80 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:I:38:DT:H2"	1:I:39:DG:C8	2.22	0.75

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:I:80:DT:OP2	8:I:301:HOH:O	2.08	0.71
1:J:205:DG:OP2	8:J:301:HOH:O	2.12	0.67
1:I:45:DT:H2"	1:I:46:DG:C8	2.31	0.65
1:J:226:DT:H2"	1:J:227:DG:C8	2.32	0.65

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	
2	A	97/136 (71%)	96 (99%)	1 (1%)	0	100	100
2	E	97/136 (71%)	96 (99%)	1 (1%)	0	100	100
3	B	77/103 (75%)	76 (99%)	1 (1%)	0	100	100
3	F	85/103 (82%)	83 (98%)	2 (2%)	0	100	100
4	C	103/130 (79%)	102 (99%)	1 (1%)	0	100	100
4	G	103/130 (79%)	101 (98%)	2 (2%)	0	100	100
5	D	96/123 (78%)	95 (99%)	0	1 (1%)	19	52
5	H	92/123 (75%)	91 (99%)	0	1 (1%)	17	50
6	K	12/16 (75%)	10 (83%)	2 (17%)	0	100	100
All	All	762/1000 (76%)	750 (98%)	10 (1%)	2 (0%)	46	79

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	D	101	GLY
5	H	101	GLY

### 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	
2	A	86/111 (78%)	86 (100%)	0	100	100
2	E	86/111 (78%)	86 (100%)	0	100	100
3	B	64/79 (81%)	63 (98%)	1 (2%)	70	93
3	F	72/79 (91%)	71 (99%)	1 (1%)	74	94
4	C	83/102 (81%)	83 (100%)	0	100	100
4	G	83/102 (81%)	83 (100%)	0	100	100
5	D	84/103 (82%)	84 (100%)	0	100	100
5	H	79/103 (77%)	79 (100%)	0	100	100
6	K	12/13 (92%)	12 (100%)	0	100	100
All	All	649/803 (81%)	647 (100%)	2 (0%)	94	99

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	67	ARG
3	F	20	LYS

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

Mol	Chain	Res	Type
2	A	68	GLN
6	K	481	HIS

### 5.3.3 RNA ⓘ

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 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 ⓘ

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	I	146/146 (100%)	0.37	7 (4%) 34 23	42, 85, 132, 146	0
1	J	146/146 (100%)	0.32	5 (3%) 49 36	40, 85, 123, 146	0
2	A	99/136 (72%)	-0.18	3 (3%) 54 41	25, 35, 61, 106	0
2	E	98/136 (72%)	-0.24	1 (1%) 84 77	18, 27, 47, 83	0
3	B	79/103 (76%)	-0.39	0 100 100	25, 33, 50, 58	0
3	F	86/103 (83%)	-0.37	1 (1%) 81 73	18, 25, 40, 95	0
4	C	105/130 (80%)	-0.34	1 (0%) 84 77	18, 29, 50, 75	0
4	G	105/130 (80%)	-0.25	0 100 100	24, 36, 62, 90	0
5	D	98/123 (79%)	-0.26	4 (4%) 41 29	19, 31, 85, 99	0
5	H	94/123 (76%)	-0.20	1 (1%) 82 74	24, 37, 65, 101	0
6	K	14/16 (87%)	-0.23	0 100 100	43, 55, 74, 75	0
All	All	1070/1292 (82%)	-0.11	23 (2%) 67 56	18, 37, 111, 146	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	D	25	LYS	4.3
1	I	44	DC	3.8
2	A	38	PRO	3.8
1	I	56	DA	3.4
1	I	57	DA	3.3

### 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 [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [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. 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	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
7	MG	E	201	1/1	0.99	0.15	-0.21	23,23,23,23	0
7	MG	D	201	1/1	0.96	0.08	-	51,51,51,51	0
7	MG	I	201	1/1	0.95	0.12	-	52,52,52,52	0
7	MG	C	201	1/1	0.93	0.21	-	35,35,35,35	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.