



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 10:48 pm GMT

PDB ID : 3AZN  
Title : Crystal Structure of Human Nucleosome Core Particle Containing H4K91Q mutation  
Authors : Iwasaki, W.; Tachiwana, H.; Kawaguchi, K.; Shibata, T.; Kagawa, W.; Kurumizaka, H.  
Deposited on : 2011-05-25  
Resolution : 3.00 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  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

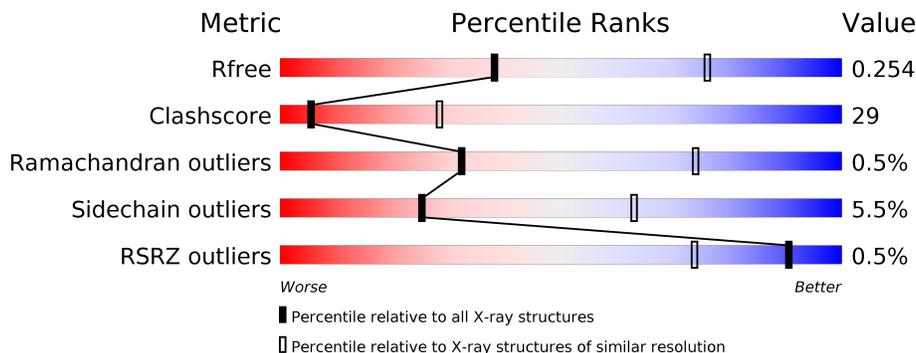
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
$R_{free}$	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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	A	139	
1	E	139	
2	B	106	
2	F	106	
3	C	133	
3	G	133	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	129	 43% 26% 5% 26%
4	H	129	 45% 26% 28%
5	I	146	 % 15% 84%
5	J	146	 11% 88%

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 11973 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 H3.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	97	801	505	155	137	4	0	0	0
1	E	99	816	514	158	140	4	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	EXPRESSION TAG	UNP P68431
A	-2	SER	-	EXPRESSION TAG	UNP P68431
A	-1	HIS	-	EXPRESSION TAG	UNP P68431
E	-3	GLY	-	EXPRESSION TAG	UNP P68431
E	-2	SER	-	EXPRESSION TAG	UNP P68431
E	-1	HIS	-	EXPRESSION TAG	UNP P68431

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	78	619	390	120	108	1	0	0	0
2	F	84	673	423	133	116	1	0	0	0

There are 8 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
B	91	GLN	LYS	ENGINEERED MUTATION	UNP P62805
F	-3	GLY	-	EXPRESSION TAG	UNP P62805
F	-2	SER	-	EXPRESSION TAG	UNP P62805

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	HIS	-	EXPRESSION TAG	UNP P62805
F	91	GLN	LYS	ENGINEERED MUTATION	UNP P62805

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	108	Total	C	N	O	0	0	0
			835	526	165	144			
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	95	Total	C	N	O	S	0	0	0
			745	468	136	139	2			
4	H	93	Total	C	N	O	S	0	0	0
			725	456	130	137	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
			Total	C	N	O	P			
5	I	145	2970	1421	538	867	144	0	0	0
5	J	145	2969	1421	535	869	144	0	0	0

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	G	1	Total 1	Cl 1	0	0
6	A	1	Total 1	Cl 1	0	0
6	C	1	Total 1	Cl 1	0	0
6	E	1	Total 1	Cl 1	0	0

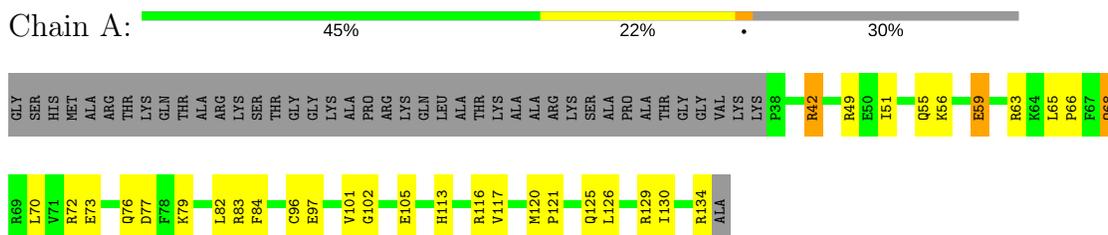
- Molecule 7 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	4	Total 4	Mn 4	0	0
7	I	6	Total 6	Mn 6	0	0
7	D	1	Total 1	Mn 1	0	0

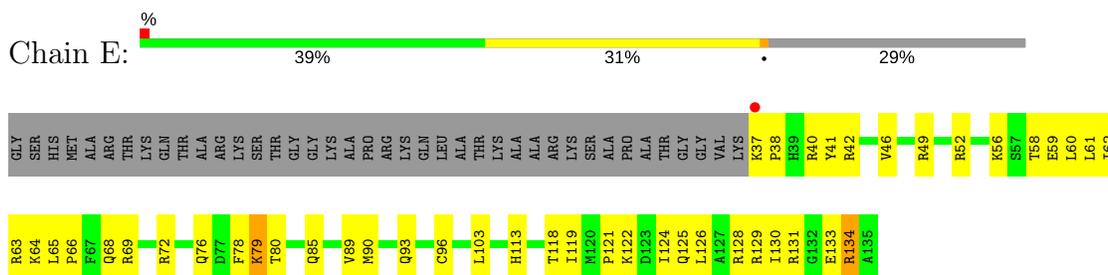
### 3 Residue-property plots [i](#)

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

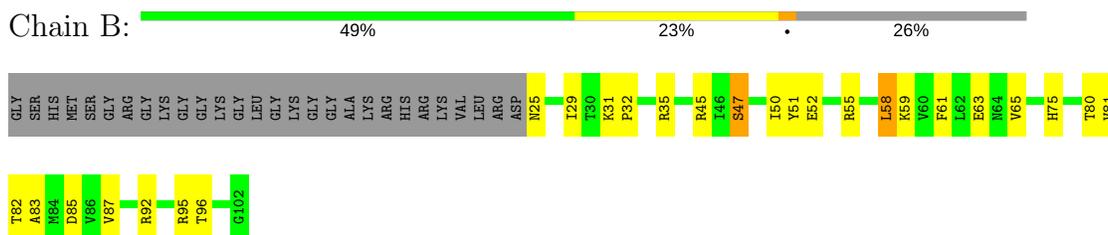
- Molecule 1: Histone H3.1



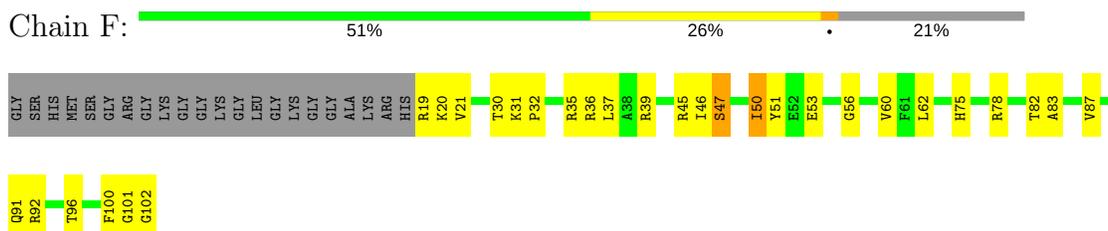
- Molecule 1: Histone H3.1



- Molecule 2: Histone H4



- Molecule 2: Histone H4





C129  
T130  
G131  
C132  
A133  
G134  
G135  
T136  
G137  
A138  
A139  
T140  
A141  
T142  
T143  
G144  
A145  
DT

● Molecule 5: 146-MER DNA

Chain J: 11% 88%

DA  
T148  
C149  
A150  
A151  
T152  
A153  
T154  
C155  
C156  
A157  
C158  
G161  
C162  
A163  
G164  
A165  
T166  
T167  
C168  
T169  
A170  
C171  
C172  
A173  
A174  
A175  
T176  
G177  
T178  
T179  
T180  
A181  
T182  
T183  
G186  
A187  
A188  
A189  
C190  
T191  
G192  
C193  
T194  
C195  
C196  
A197  
T198  
C199  
A200  
A201  
A202  
G205  
C206  
A207  
T208  
G209

T210  
T211  
C212  
C215  
T216  
G217  
A218  
A219  
T220  
T221  
C222  
A223  
G224  
C225  
T226  
A228  
A229  
C230  
A231  
T232  
G233  
C234  
C235  
T236  
T237  
T238  
T239  
G240  
A241  
T242  
G243  
G244  
A245  
G246  
C247  
A248  
G249  
T250  
T251  
T252  
C253  
C254  
A255  
A256  
A257  
T258  
A259  
C260  
A261  
C262  
T263  
T264  
T265  
T266  
G267  
G268  
T269  
A270

G271  
A272  
A273  
T274  
C275  
T276  
G277  
C278  
A279  
G280  
G281  
T282  
G283  
G284  
A285  
T286  
A287  
T288  
T289  
G290  
A291  
T292

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.38Å 109.69Å 182.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.75 – 3.00 48.75 – 3.01	Depositor EDS
% Data completeness (in resolution range)	99.3 (48.75-3.00) 99.3 (48.75-3.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.15 (at 3.01Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.197 , 0.254 0.198 , 0.254	Depositor DCC
$R_{free}$ test set	2161 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	54.9	Xtrriage
Anisotropy	0.319	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 55.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.019 for k,h,-l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	11973	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	69.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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: MN, CL

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	0.41	0/813	0.65	0/1090
1	E	0.43	0/828	0.62	0/1109
2	B	0.38	0/626	0.64	0/838
2	F	0.44	0/680	0.69	0/909
3	C	0.38	0/845	0.60	0/1139
3	G	0.36	0/815	0.60	0/1100
4	D	0.39	0/756	0.60	0/1015
4	H	0.40	0/736	0.62	0/990
5	I	0.38	0/3332	0.78	0/5141
5	J	0.36	0/3330	0.78	0/5138
All	All	0.39	0/12761	0.72	0/18469

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	801	0	839	38	0
1	E	816	0	856	39	0
2	B	619	0	654	31	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	F	673	0	717	35	0
3	C	835	0	897	48	0
3	G	805	0	861	34	0
4	D	745	0	771	48	0
4	H	725	0	745	35	0
5	I	2970	0	1640	170	0
5	J	2969	0	1641	229	0
6	A	1	0	0	1	0
6	C	1	0	0	0	0
6	E	1	0	0	1	0
6	G	1	0	0	1	0
7	D	1	0	0	0	0
7	I	6	0	0	0	0
7	J	4	0	0	0	0
All	All	11973	0	9621	618	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 29.

The worst 5 of 618 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:J:242:DT:H2''	5:J:243:DG:H5''	1.20	1.17
5:J:252:DT:H2''	5:J:253:DC:H5''	1.22	1.17
5:J:285:DA:H2''	5:J:286:DT:H5''	1.30	1.11
5:I:101:DC:H2''	5:I:102:DA:H5'	1.27	1.09
5:I:136:DT:H2''	5:I:137:DG:H5'	1.27	1.09

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%)	92 (97%)	3 (3%)	0	100	100
1	E	97/139 (70%)	93 (96%)	4 (4%)	0	100	100
2	B	76/106 (72%)	74 (97%)	2 (3%)	0	100	100
2	F	82/106 (77%)	79 (96%)	2 (2%)	1 (1%)	15	53
3	C	106/133 (80%)	103 (97%)	3 (3%)	0	100	100
3	G	102/133 (77%)	95 (93%)	7 (7%)	0	100	100
4	D	93/129 (72%)	87 (94%)	4 (4%)	2 (2%)	8	36
4	H	91/129 (70%)	81 (89%)	9 (10%)	1 (1%)	17	56
All	All	742/1014 (73%)	704 (95%)	34 (5%)	4 (0%)	32	74

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	104	GLY
4	H	104	GLY
2	F	101	GLY
4	D	123	SER

### 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	85/113 (75%)	80 (94%)	5 (6%)	23	60
1	E	86/113 (76%)	79 (92%)	7 (8%)	14	45
2	B	63/81 (78%)	61 (97%)	2 (3%)	44	79
2	F	69/81 (85%)	66 (96%)	3 (4%)	33	72
3	C	85/102 (83%)	78 (92%)	7 (8%)	13	44
3	G	83/102 (81%)	80 (96%)	3 (4%)	40	77
4	D	81/107 (76%)	74 (91%)	7 (9%)	12	42
4	H	79/107 (74%)	78 (99%)	1 (1%)	73	92
All	All	631/806 (78%)	596 (94%)	35 (6%)	25	63

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	34	LYS
4	D	106	LEU
3	G	81	ARG
4	D	63	ASN
4	D	70	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	84	GLN
3	C	110	ASN
3	G	31	HIS
3	C	73	ASN
3	G	73	ASN

### 5.3.3 RNA [i](#)

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 15 ligands modelled in this entry, 15 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 [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	97/139 (69%)	-0.56	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	16, 40, 70, 98	0
1	E	99/139 (71%)	-0.51	1 (1%) <span style="border: 1px solid blue; padding: 2px;">82</span> <span style="border: 1px solid blue; padding: 2px;">58</span>	17, 34, 79, 107	0
2	B	78/106 (73%)	-0.61	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	23, 37, 56, 105	0
2	F	84/106 (79%)	-0.54	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	17, 33, 52, 94	0
3	C	108/133 (81%)	-0.45	3 (2%) <span style="border: 1px solid gray; padding: 2px;">53</span> <span style="border: 1px solid red; padding: 2px;">25</span>	16, 39, 71, 137	0
3	G	104/133 (78%)	-0.55	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	22, 43, 71, 92	0
4	D	95/129 (73%)	-0.49	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	21, 41, 81, 110	0
4	H	93/129 (72%)	-0.50	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	20, 44, 74, 98	0
5	I	145/146 (99%)	-0.18	1 (0%) <span style="border: 1px solid blue; padding: 2px;">87</span> <span style="border: 1px solid blue; padding: 2px;">67</span>	41, 93, 140, 164	0
5	J	145/146 (99%)	-0.24	0 <span style="border: 1px solid blue; padding: 2px;">100</span> <span style="border: 1px solid blue; padding: 2px;">100</span>	44, 97, 141, 152	0
All	All	1048/1306 (80%)	-0.44	5 (0%) <span style="border: 1px solid blue; padding: 2px;">90</span> <span style="border: 1px solid blue; padding: 2px;">74</span>	16, 45, 123, 164	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	13	LYS	4.5
3	C	12	ALA	4.3
5	I	55	DA	3.1
3	C	11	ARG	2.7
1	E	37	LYS	2.5

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

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	MN	D	201	1/1	0.99	0.22	1.91	28,28,28,28	0
6	CL	C	1001	1/1	0.97	0.15	-0.11	52,52,52,52	0
6	CL	G	1001	1/1	0.98	0.15	-0.26	48,48,48,48	0
7	MN	J	1002	1/1	0.91	0.16	-	92,92,92,92	0
7	MN	I	1005	1/1	0.87	0.08	-	116,116,116,116	0
7	MN	J	1001	1/1	0.82	0.12	-	116,116,116,116	0
6	CL	A	1001	1/1	0.96	0.12	-	62,62,62,62	0
7	MN	I	1004	1/1	0.95	0.13	-	68,68,68,68	0
7	MN	I	1001	1/1	0.70	0.13	-	111,111,111,111	0
7	MN	I	1002	1/1	0.96	0.07	-	96,96,96,96	0
7	MN	J	1004	1/1	0.97	0.20	-	65,65,65,65	0
7	MN	I	1006	1/1	0.65	0.27	-	119,119,119,119	0
6	CL	E	1001	1/1	0.97	0.06	-	46,46,46,46	0
7	MN	J	1003	1/1	0.97	0.16	-	63,63,63,63	0
7	MN	I	1003	1/1	0.95	0.32	-	75,75,75,75	0

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

There are no such residues in this entry.