wwPDB X-ray Structure Validation Summary Report (i)

Nov 16, 2023 – 04:53 AM JST

PDB ID	:	6KXV
Title	:	Crystal structure of a nucleosome containing Leishmania histone H3
Authors	:	Dacher, M.; Taguchi, H.; Kujirai, T.; Kurumizaka, H.
Deposited on	:	2019-09-13
Resolution	:	3.63 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

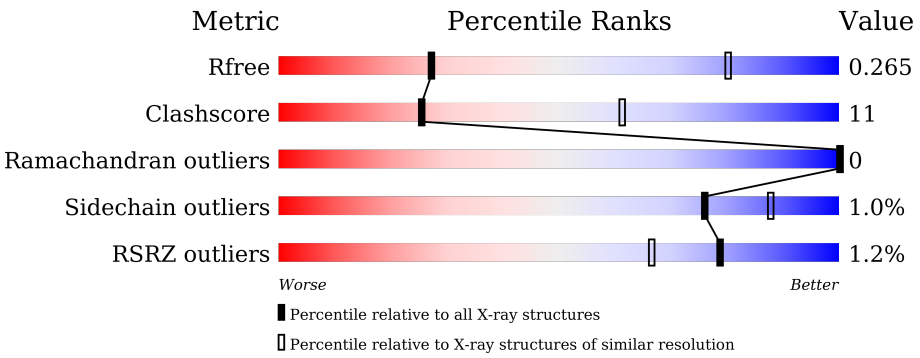
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.63 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	1341 (3.78-3.50)
Clashscore	141614	1439 (3.78-3.50)
Ramachandran outliers	138981	1391 (3.78-3.50)
Sidechain outliers	138945	1391 (3.78-3.50)
RSRZ outliers	127900	1242 (3.78-3.50)

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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	133	58% 13% 29%
1	E	133	55% 17% 29%
2	B	106	% 63% 10% 26%
2	F	106	4% 61% 18% 21%
3	C	133	3% 67% 14% 19%
3	G	133	68% 11% 22%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	129	2% 60% 13% 26%
4	H	129	60% 11% 29%
5	I	146	% 38% 62%
5	J	146	45% 55%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 11876 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	94	745	467	141	132	5	0	0	0
1	E	95	756	473	145	133	5	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 Q4QHB5
A	-2	SER	-	expression tag	UNP Q4QHB5
A	-1	HIS	-	expression tag	UNP Q4QHB5
E	-3	GLY	-	expression tag	UNP Q4QHB5
E	-2	SER	-	expression tag	UNP Q4QHB5
E	-1	HIS	-	expression tag	UNP Q4QHB5

- Molecule 2 is a protein called Histone H4.

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

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
			Total	C	N	O			
3	C	108	835	526	165	144	0	0	0
3	G	104	805	508	157	140	0	0	0

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
			Total	C	N	O	S			
4	D	96	755	474	138	141	2	0	0	0
4	H	91	708	447	125	134	2	0	0	0

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 DNA (146-MER).

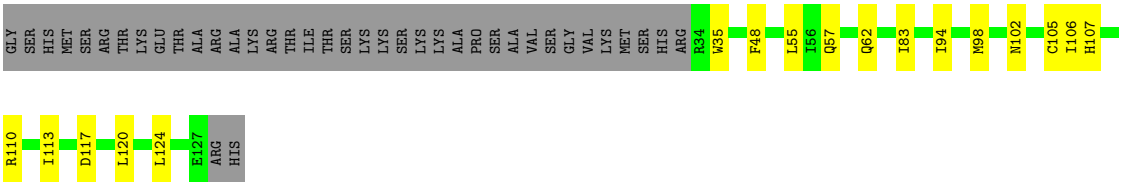
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	I	146	2990	1431	540	874	145	0	0	0
5	J	146	2990	1431	540	874	145	0	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

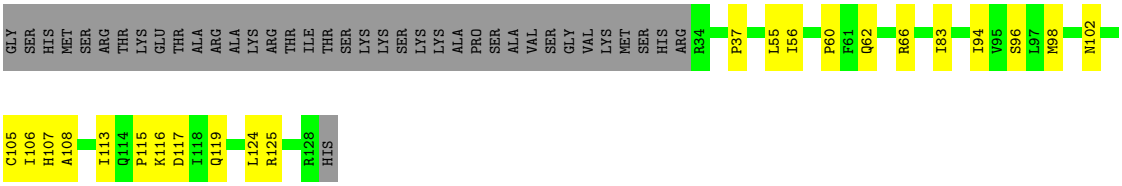
- Molecule 1: Histone H3

Chain A:



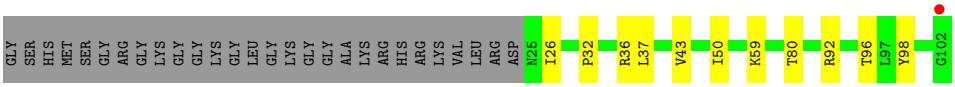
- Molecule 1: Histone H3

Chain E:



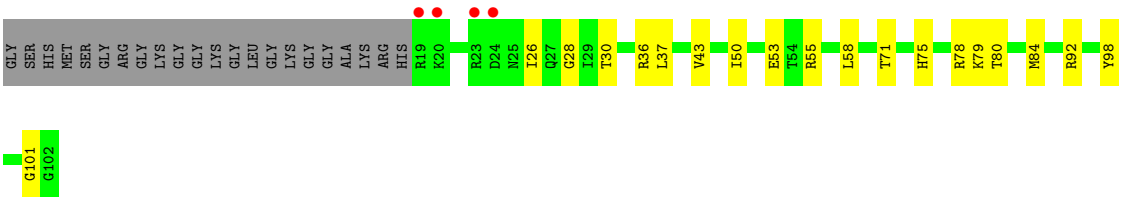
- Molecule 2: Histone H4

Chain B:



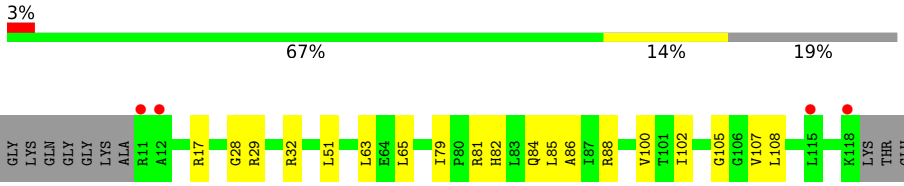
- Molecule 2: Histone H4

Chain F:



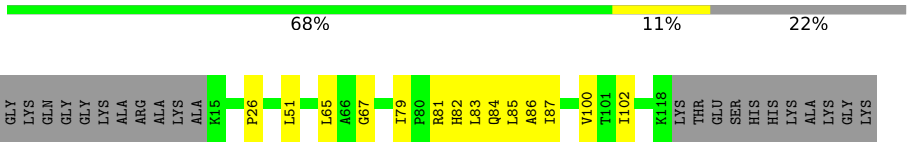
- Molecule 3: Histone H2A type 1-B/E

Chain C:



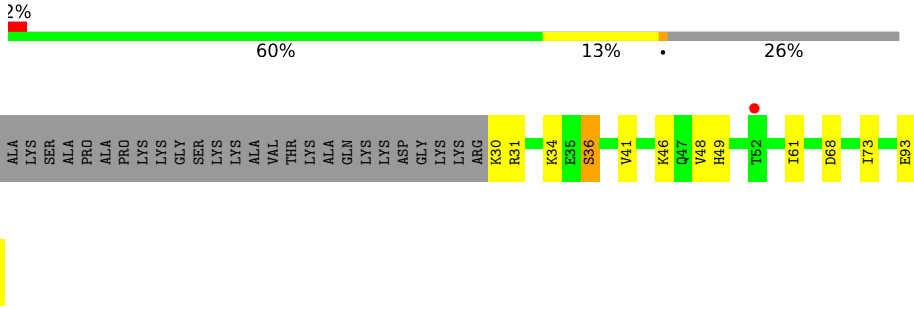
• Molecule 3: Histone H2A type 1-B/E

Chain G:



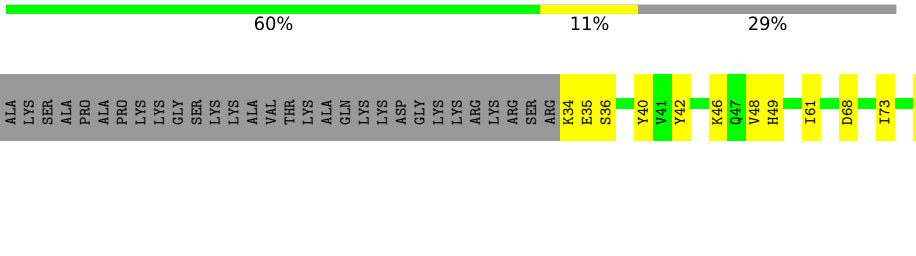
• Molecule 4: Histone H2B type 1-J

Chain D:



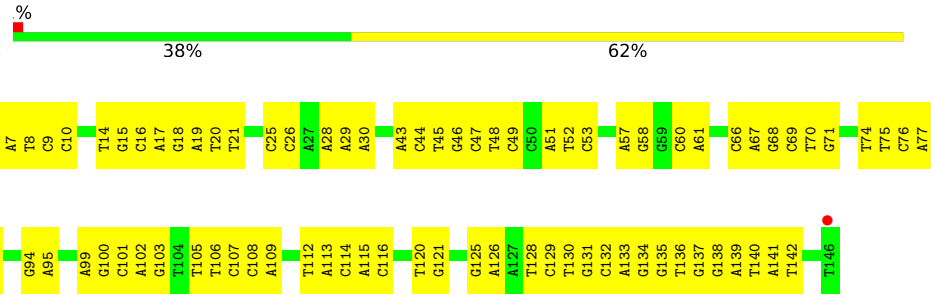
• Molecule 4: Histone H2B type 1-J

Chain H:



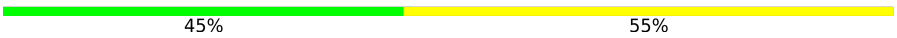
• Molecule 5: DNA (146-MER)

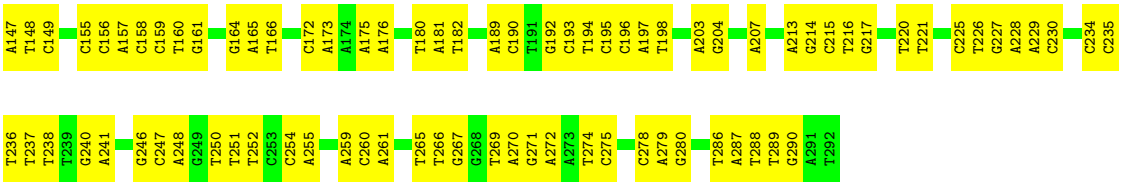
Chain I:



• Molecule 5: DNA (146-MER)

Chain J:




A147 T148 C149
C155 C156 A157 C158 C159 T160 G161
G164 A165 T166
C172 A173 A174 A175 A176
T180 A181 T182
A189 C190 T191 G192 C193 T194 C195 C196 A197 T198
A203 G204
A207
A213 G214 C215 T216 G217
T220 T221
C225 T226 G227 A228 A229 C230
C234 C235
T236 T237 T238 T239 G240 A241
G246 C247 A248 G249 T250 T251 T252 C253 C254 A255
A259 C260 A261
T265 T266 G267 G268 T269 A270 G271 A272 A273 T274 C275
C278 A279 G280
T286 A287 T288 T289 G290 A291 T292

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	101.90Å 101.10Å 175.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.93 – 3.63 48.93 – 3.63	Depositor EDS
% Data completeness (in resolution range)	99.6 (48.93-3.63) 99.6 (48.93-3.63)	Depositor EDS
R_{merge}	0.23	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.43 (at 3.67Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.220 , 0.264 0.220 , 0.265	Depositor DCC
R_{free} test set	1080 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å^2)	135.0	Xtriage
Anisotropy	0.128	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 53.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.043 for k,h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	11876	wwPDB-VP
Average B, all atoms (Å^2)	126.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.11% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.25	0/754	0.39	0/1013
1	E	0.25	0/765	0.40	0/1027
2	B	0.25	0/626	0.42	0/837
2	F	0.25	0/680	0.43	0/908
3	C	0.24	0/845	0.42	0/1139
3	G	0.23	0/815	0.41	0/1100
4	D	0.27	0/766	0.40	0/1026
4	H	0.29	0/719	0.40	0/968
5	I	0.52	0/3354	0.96	0/5175
5	J	0.51	0/3354	0.94	0/5175
All	All	0.41	0/12678	0.76	0/18368

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	745	0	783	17	0
1	E	756	0	796	23	0
2	B	619	0	659	10	0
2	F	673	0	722	20	0
3	C	835	0	897	15	0
3	G	805	0	861	11	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	755	0	784	14	0
4	H	708	0	727	11	0
5	I	2990	0	1652	81	0
5	J	2990	0	1652	70	0
All	All	11876	0	9533	227	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 11.

The worst 5 of 227 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:259:DA:H2"	5:J:260:DC:H5"	1.66	0.76
5:I:101:DC:H2"	5:I:102:DA:H5"	1.67	0.75
5:I:108:DC:H2"	5:I:109:DA:N7	2.04	0.72
5:I:14:DT:O4	5:J:278:DC:N4	2.24	0.71
5:I:120:DT:H2"	5:I:121:DG:N7	2.06	0.70

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	92/133 (69%)	90 (98%)	2 (2%)	0	100	100
1	E	93/133 (70%)	90 (97%)	3 (3%)	0	100	100
2	B	76/106 (72%)	74 (97%)	2 (3%)	0	100	100
2	F	82/106 (77%)	78 (95%)	4 (5%)	0	100	100
3	C	106/133 (80%)	104 (98%)	2 (2%)	0	100	100
3	G	102/133 (77%)	101 (99%)	1 (1%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	D	94/129 (73%)	91 (97%)	3 (3%)	0	100	100
4	H	89/129 (69%)	87 (98%)	2 (2%)	0	100	100
All	All	734/1002 (73%)	715 (97%)	19 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	80/113 (71%)	80 (100%)	0	100	100
1	E	81/113 (72%)	81 (100%)	0	100	100
2	B	63/81 (78%)	62 (98%)	1 (2%)	62	82
2	F	69/81 (85%)	68 (99%)	1 (1%)	67	84
3	C	85/102 (83%)	85 (100%)	0	100	100
3	G	83/102 (81%)	82 (99%)	1 (1%)	71	86
4	D	82/107 (77%)	81 (99%)	1 (1%)	71	86
4	H	77/107 (72%)	75 (97%)	2 (3%)	46	73
All	All	620/806 (77%)	614 (99%)	6 (1%)	76	88

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	84	GLN
4	H	36	SER
4	H	42	TYR
4	D	36	SER
2	B	92	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 monosaccharides 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 (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, 95th 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^2)	Q<0.9
1	A	94/133 (70%)	-0.24	0	100	100	78, 98, 121, 131	0
1	E	95/133 (71%)	-0.15	0	100	100	80, 100, 124, 144	0
2	B	78/106 (73%)	-0.11	1 (1%)	77	64	79, 96, 115, 137	0
2	F	84/106 (79%)	0.00	4 (4%)	30	20	76, 94, 151, 163	0
3	C	108/133 (81%)	0.01	4 (3%)	41	27	73, 100, 131, 169	0
3	G	104/133 (78%)	-0.11	0	100	100	77, 99, 129, 140	0
4	D	96/129 (74%)	-0.08	2 (2%)	63	48	72, 96, 134, 164	0
4	H	91/129 (70%)	-0.18	0	100	100	72, 96, 116, 123	0
5	I	146/146 (100%)	-0.51	1 (0%)	87	79	105, 150, 185, 210	0
5	J	146/146 (100%)	-0.50	0	100	100	109, 148, 193, 199	0
All	All	1042/1294 (80%)	-0.22	12 (1%)	79	66	72, 104, 173, 210	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	12	ALA	5.1
2	F	19	ARG	4.1
3	C	11	ARG	3.7
2	F	23	ARG	3.3
2	F	24	ASP	3.2

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 monosaccharides 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.