wwPDB X-ray Structure Validation Summary Report (i)

Oct 30, 2023 – 02:01 PM JST

PDB ID	:	5B31
Title	:	The crystal structure of the heterotypic H2AZ/H2A nucleosome with H3.1.
Authors	:	Horikoshi, N.; Taguchi, H.; Arimura, Y.; Kurumizaka, H.
Deposited on	:	2016-02-08
Resolution	:	2.20 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:

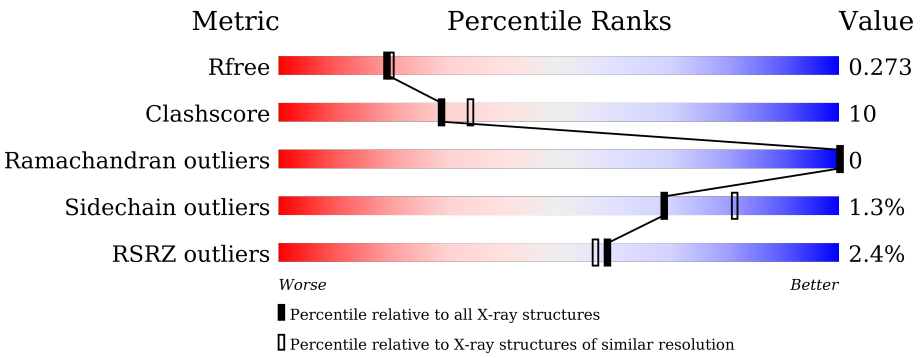
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	139	.% 60% 9% 30%
1	E	139	58% 12% 30%
2	B	106	.% 68% 6% 26%
2	F	106	.% 71% 8% 22%
3	C	133	2% 61% 18% • 20%
4	D	129	63% 9% 28%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	H	129	60% 11% 29%
5	G	131	2% 60% 19% 21%
6	I	146	8% 52% 47%
6	J	146	3% 49% 50%

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 12219 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	97	801	505	155	137	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	391	120	107	1	0	0	0
2	F	83	668	422	132	113	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	107	828	523	162	143	0	0	0

There are 3 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

- 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	93	725	456	130	137	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 protein called Histone H2A.Z.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
5	G	104	785	492	153	140	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-3	GLY	-	expression tag	UNP P0C0S5
G	-2	SER	-	expression tag	UNP P0C0S5
G	-1	HIS	-	expression tag	UNP P0C0S5

- Molecule 6 is a DNA chain called DNA (146-MER).

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

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mn		
7	A	1	1	1	0	0
7	E	1	1	1	0	0
7	I	2	2	2	0	0
7	J	4	4	4	0	0

- Molecule 8 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
8	A	1	1	1	0	0
8	D	1	1	1	0	0
8	E	1	1	1	0	0
8	G	1	1	1	0	0

- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
9	A	33	33	33	0	0
9	B	27	27	27	0	0
9	C	51	51	51	0	0
9	D	19	19	19	0	0
9	E	58	58	58	0	0

Continued on next page...

Continued from previous page...

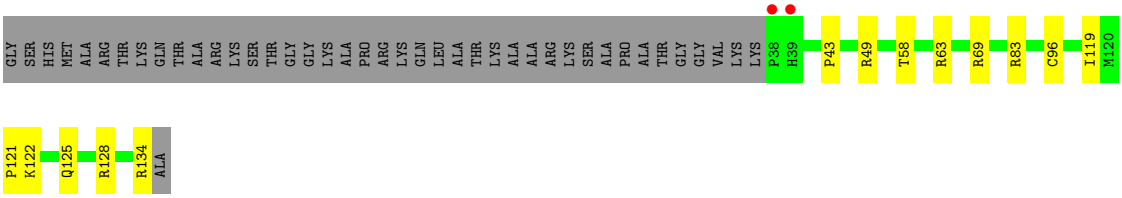
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
9	F	46	46	46	0	0
			Total	O		
9	G	23	23	23	0	0
			Total	O		
9	H	15	15	15	0	0
			Total	O		
9	I	11	11	11	0	0
			Total	O		
9	J	9	9	9	0	0

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.

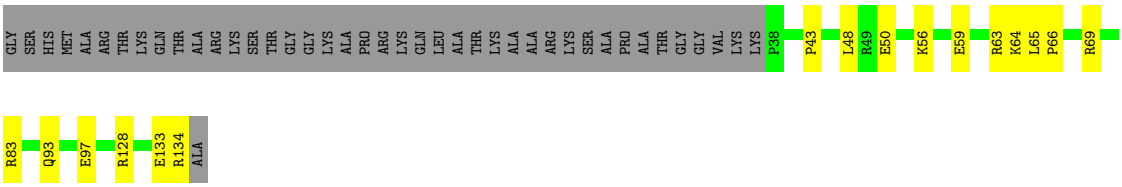
• Molecule 1: Histone H3.1

Chain A:



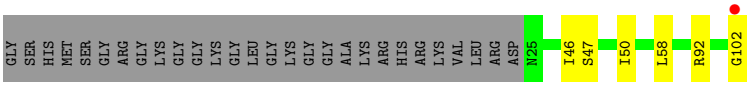
• Molecule 1: Histone H3.1

Chain E:



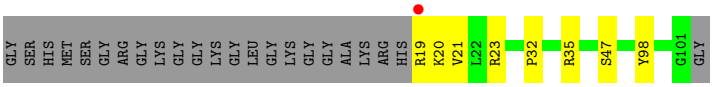
• Molecule 2: Histone H4

Chain B:



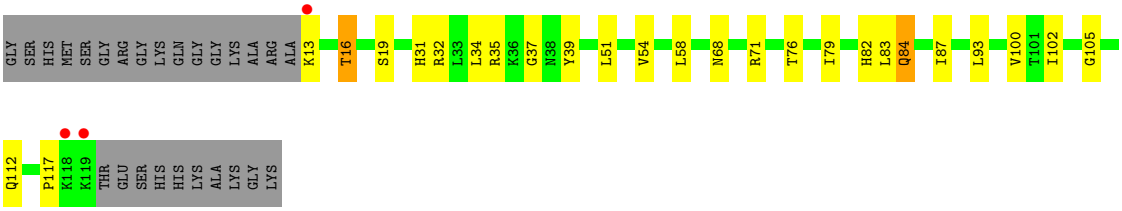
• Molecule 2: Histone H4

Chain F:



• Molecule 3: Histone H2A type 1-B/E

Chain C:

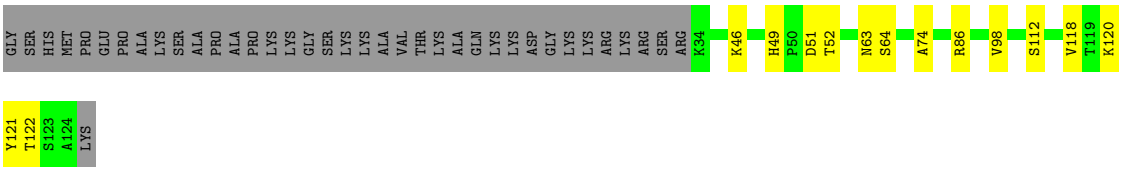


• Molecule 4: Histone H2B type 1-J

Chain D:

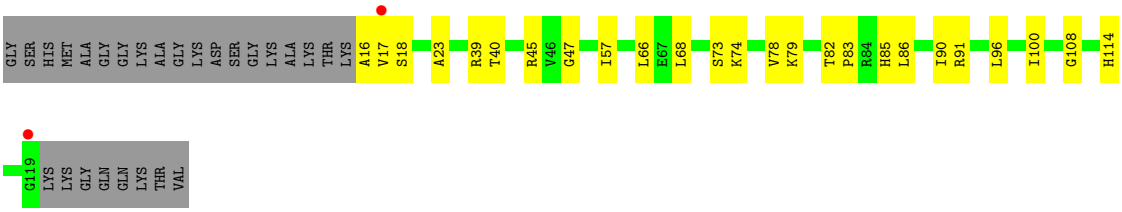
• Molecule 4: Histone H2B type 1-J

Chain H:



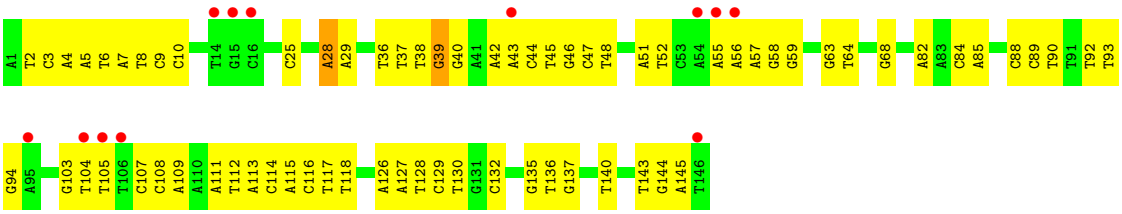
• Molecule 5: Histone H2A.Z

Chain G:



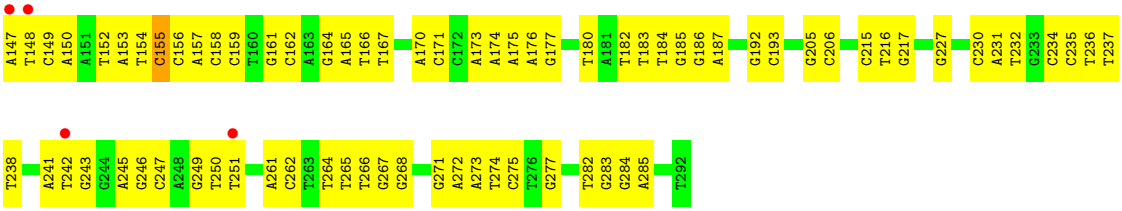
• Molecule 6: DNA (146-MER)

Chain I:



• Molecule 6: DNA (146-MER)

Chain J:


A147
T148
C149
A150
A151
T152
A153
T154
C155
C156
A157
C158
C159
T160
G161
C162
A163
G164
A165
T166
T167
A170
C171
C172
A173
A174
A175
A176
G177
T180
A181
T182
T183
T184
G185
G186
A187
G192
C193
G205
C206
C215
T216
G217
G227
C230
A231
T232
G233
C234
C235
T236
T237
T238
A241
T242
G243
G244
A245
G246
C247
A248
G249
T250
T251
A261
C262
T263
T264
T265
T266
G267
G268
G271
A272
A273
T274
C275
T276
G277
T282
G283
G284
A285
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	105.11Å 109.69Å 181.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.62 – 2.20 48.62 – 2.19	Depositor EDS
% Data completeness (in resolution range)	98.0 (48.62-2.20) 97.1 (48.62-2.19)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.03 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.225 , 0.271 0.227 , 0.273	Depositor DCC
R_{free} test set	5266 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	34.7	Xtriage
Anisotropy	0.175	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 49.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.024 for k,h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	12219	wwPDB-VP
Average B, all atoms (Å^2)	65.0	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.*

[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)

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	0.54	1/813 (0.1%)	0.61	0/1090
1	E	0.60	0/813	0.66	0/1090
2	B	0.55	0/626	0.64	0/837
2	F	0.63	0/675	0.73	0/903
3	C	0.54	0/838	0.63	0/1129
4	D	0.55	0/736	0.67	0/990
4	H	0.51	0/719	0.61	0/968
5	G	0.47	0/796	0.60	0/1073
6	I	0.77	0/3354	1.01	4/5175 (0.1%)
6	J	0.74	0/3354	0.99	1/5175 (0.0%)
All	All	0.67	1/12724 (0.0%)	0.86	5/18430 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	96	CYS	CB-SG	-5.50	1.72	1.81

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	J	155	DC	O5'-P-OP1	-8.85	97.73	105.70
6	I	39	DG	O4'-C1'-N9	6.54	112.58	108.00
6	I	39	DG	O4'-C1'-C2'	-5.67	101.36	105.90
6	I	28	DA	O4'-C1'-N9	5.43	111.80	108.00
6	I	107	DC	O4'-C4'-C3'	-5.03	102.49	104.50

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	13	0
1	E	801	0	839	17	0
2	B	619	0	659	4	0
2	F	668	0	719	5	0
3	C	828	0	892	23	0
4	D	725	0	745	11	0
4	H	708	0	727	14	0
5	G	785	0	827	22	0
6	I	2990	0	1652	61	0
6	J	2990	0	1652	74	0
7	A	1	0	0	0	0
7	E	1	0	0	0	0
7	I	2	0	0	0	0
7	J	4	0	0	0	0
8	A	1	0	0	1	0
8	D	1	0	0	0	0
8	E	1	0	0	0	0
8	G	1	0	0	1	0
9	A	33	0	0	2	0
9	B	27	0	0	0	0
9	C	51	0	0	0	0
9	D	19	0	0	2	0
9	E	58	0	0	2	0
9	F	46	0	0	2	0
9	G	23	0	0	1	0
9	H	15	0	0	1	0
9	I	11	0	0	1	0
9	J	9	0	0	0	0
All	All	12219	0	9551	197	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 10.

The worst 5 of 197 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:E:133:GLU:O	1:E:134:ARG:NH1	1.95	0.97
6:I:145:DA:N1	6:J:147:DA:N6	2.19	0.90
6:J:154:DT:H2"	6:J:155:DC:H5"	1.58	0.86
3:C:16:THR:HG22	3:C:19:SER:H	1.42	0.82
5:G:79:LYS:NZ	4:H:51:ASP:O	2.14	0.80

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	95/139 (68%)	94 (99%)	1 (1%)	0	100	100
2	B	76/106 (72%)	75 (99%)	1 (1%)	0	100	100
2	F	81/106 (76%)	77 (95%)	4 (5%)	0	100	100
3	C	105/133 (79%)	104 (99%)	1 (1%)	0	100	100
4	D	91/129 (70%)	90 (99%)	1 (1%)	0	100	100
4	H	89/129 (69%)	88 (99%)	1 (1%)	0	100	100
5	G	102/131 (78%)	98 (96%)	4 (4%)	0	100	100
All	All	734/1012 (72%)	720 (98%)	14 (2%)	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	85/113 (75%)	85 (100%)	0	100	100
1	E	85/113 (75%)	85 (100%)	0	100	100
2	B	63/81 (78%)	61 (97%)	2 (3%)	39	50
2	F	69/81 (85%)	67 (97%)	2 (3%)	42	54
3	C	85/102 (83%)	83 (98%)	2 (2%)	49	62
4	D	79/107 (74%)	79 (100%)	0	100	100
4	H	77/107 (72%)	76 (99%)	1 (1%)	69	81
5	G	81/99 (82%)	80 (99%)	1 (1%)	71	83
All	All	624/803 (78%)	616 (99%)	8 (1%)	69	81

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	H	112	SER
5	G	40	THR
2	F	32	PRO
3	C	84	GLN
2	F	47	SER

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)

Of 12 ligands modelled in this entry, 12 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, 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	97/139 (69%)	-0.02	2 (2%)	63	61	27, 39, 60, 94	0
1	E	97/139 (69%)	-0.03	0	100	100	22, 31, 48, 71	0
2	B	78/106 (73%)	0.02	1 (1%)	77	75	30, 38, 50, 57	0
2	F	83/106 (78%)	0.25	1 (1%)	79	77	22, 30, 42, 80	0
3	C	107/133 (80%)	0.06	3 (2%)	53	51	23, 36, 60, 95	0
4	D	93/129 (72%)	0.02	0	100	100	25, 37, 61, 81	0
4	H	91/129 (70%)	0.19	0	100	100	29, 41, 60, 74	0
5	G	104/131 (79%)	0.01	2 (1%)	66	65	31, 45, 63, 87	0
6	I	146/146 (100%)	0.18	12 (8%)	11	10	42, 90, 127, 143	0
6	J	146/146 (100%)	0.01	4 (2%)	54	52	46, 90, 128, 138	0
All	All	1042/1304 (79%)	0.07	25 (2%)	59	56	22, 43, 112, 143	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
6	I	146	DT	9.4
6	J	148	DT	7.0
3	C	118	LYS	5.5
6	J	147	DA	5.1
6	I	15	DG	4.4

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)

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. The B-factors column lists the minimum, median, 95th 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	B-factors(Å^2)	Q<0.9
7	MN	J	402	1/1	0.74	0.10	102,102,102,102	0
7	MN	J	403	1/1	0.78	0.08	101,101,101,101	0
7	MN	I	301	1/1	0.81	0.23	87,87,87,87	0
7	MN	I	302	1/1	0.88	0.06	111,111,111,111	0
7	MN	J	401	1/1	0.92	0.16	81,81,81,81	0
8	CL	G	301	1/1	0.96	0.09	49,49,49,49	0
8	CL	D	301	1/1	0.97	0.07	50,50,50,50	0
7	MN	A	301	1/1	0.98	0.08	92,92,92,92	0
8	CL	A	302	1/1	0.98	0.09	54,54,54,54	0
8	CL	E	302	1/1	0.99	0.05	49,49,49,49	0
7	MN	J	404	1/1	0.99	0.08	70,70,70,70	0
7	MN	E	301	1/1	1.00	0.13	29,29,29,29	0

6.5 Other polymers (i)

There are no such residues in this entry.