wwPDB X-ray Structure Validation Summary Report (i)

Oct 22, 2023 – 04:16 AM EDT

PDB ID	:	3AZM
Title	:	Crystal Structure of Human Nucleosome Core Particle Containing H4K79Q mutation
Authors	:	Iwasaki, W.; Tachiwana, H.; Kawaguchi, K.; Shibata, T.; Kagawa, W.; Kurumizaka, H.
Deposited on	:	2011-05-25
Resolution	:	2.89 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:

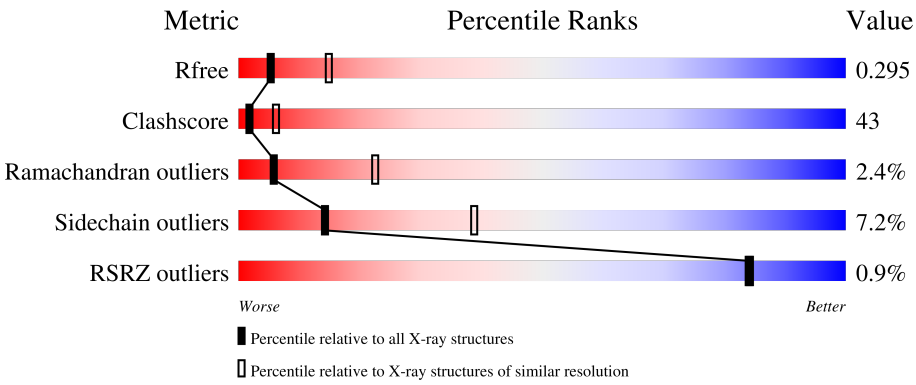
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.89 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	.% 27% 40% • 29%
1	E	139	35% 32% • 29%
2	B	106	2% 34% 35% • 27%
2	F	106	26% 46% 7% 21%
3	C	133	.% 36% 38% 6% 20%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	133	% 39% 35% 5% 22%
4	D	129	% 35% 33% 6% 26%
4	H	129	2% 29% 39% 29%
5	I	146	8% 91%
5	J	146	97%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 11947 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	98	807	508	156	139	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	77	614	388	119	106	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	79	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	79	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
			Total	C	N	O			
3	C	106	819	517	160	142	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	95	745	468	136	139	2	0	0	0
4	H	92	719	453	129	135	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 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
			Total	Cl		
6	A	1	1	1	0	0
6	C	1	1	1	0	0
6	G	1	1	1	0	0

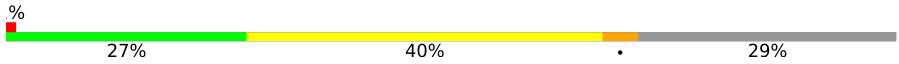
- Molecule 7 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

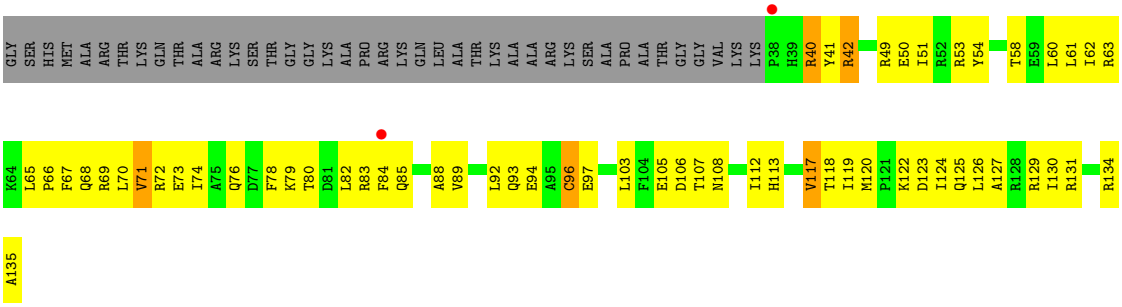
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mn		
7	E	1	1	1	0	0
7	I	3	3	3	0	0
7	J	3	3	3	0	0

3 Residue-property plots (i)

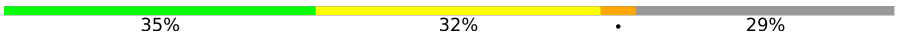
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.

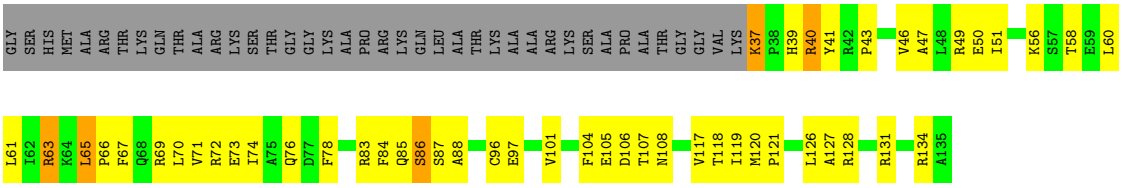
• Molecule 1: Histone H3.1

Chain A: 

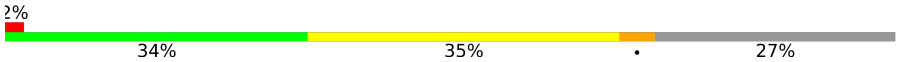


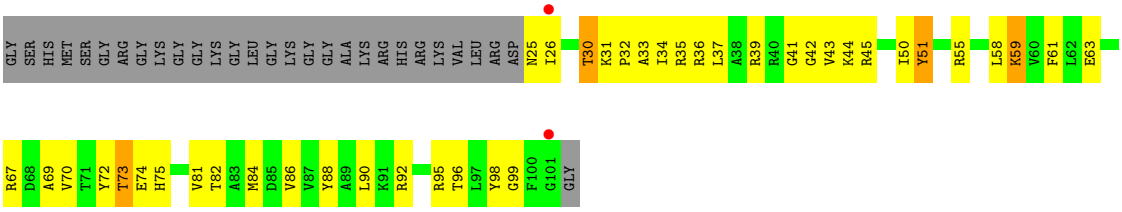
• Molecule 1: Histone H3.1

Chain E: 

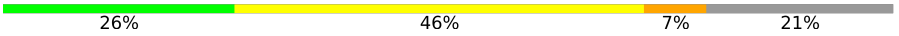


• Molecule 2: Histone H4

Chain B: 

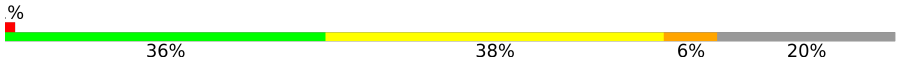


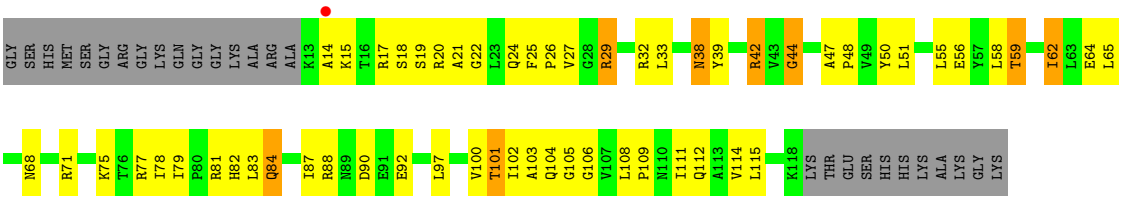
• Molecule 2: Histone H4

Chain F: 

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

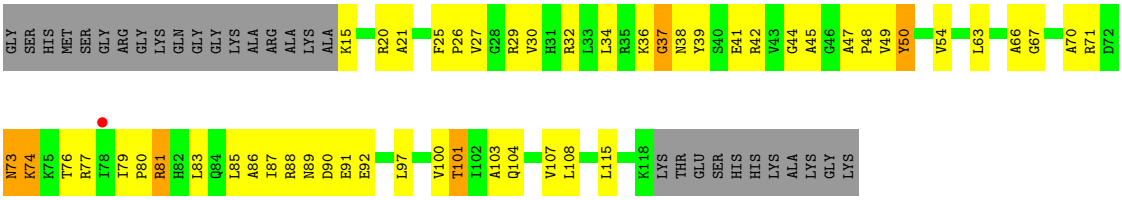
Chain C:





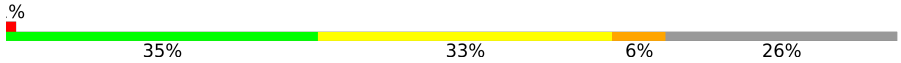
- Molecule 3: Histone H2A type 1-B/E

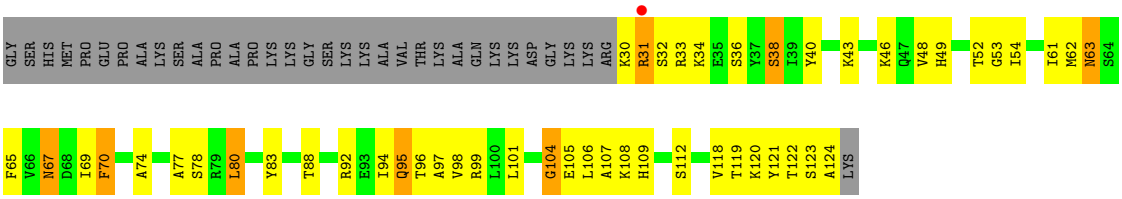
Chain G:



- Molecule 4: Histone H2B type 1-J

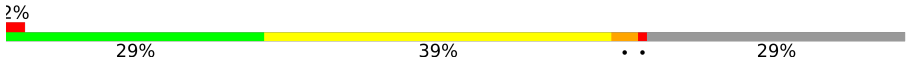
Chain D:

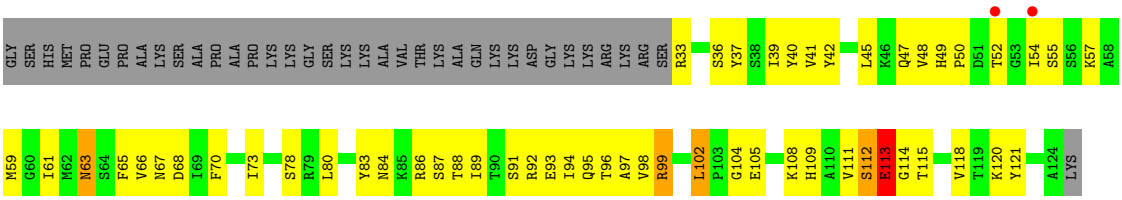




- Molecule 4: Histone H2B type 1-J

Chain H:





- Molecule 5: 146-MER DNA

Chain I: 8% 91%

A1 T2 C3 A4 A5 T6 A7 T8 C9 C10 A11 C12 C13 T14 G15 C16 A17 G18 A19 T20 T21 C22 T23 A24 C25 C26 A27 A28 A29 A30 G31 T32 G33 T34 A35 T36 T37 T38 G39 G40 A41 A42 A43 C44 T45 G46 C47 T48 C49 C50 A51 T52 C53 A54 A55 A56 A57 G58 G59 C60

A61 T62 G63 T64 G68 C69 T70 G71 A72 A73 T74 T75 C76 A77 G78 C79 T80 G81 C84 A85 T86 G87 C88 C89 T90 T91 T92 T93 G94 A95 T96 G97 G98 A99 G100 C101 A102 G103 T104 T105 T106 C107 C108 A109 A110 A111 T112 A113 C114 A115 C116 T117 T118 T119 T120 G121 G122 T123

A124 G125 A126 A127 T128 C129 T130 G131 C132 A133 G134 G135 T136 G137 G138 A139 T140 A141 T142 T143 G144 A145 DT

- Molecule 5: 146-MER DNA

Chain J: 97%

DA T148 C149 A150 A151 T152 A153 T154 C155 C156 A157 C158 C159 T160 G161 C162 A163 G164 A165 T166 T167 C168 T169 A170 C171 C172 A173 A174 A175 A176 G177 T178 G179 T180 A181 T182 T183 T184 G185 G186 A187 A188 A189 C190 T191 G192 C193 T194 C195 C196 A197 T198 C199 A200 A201 A202 A203 G204 G205 C206

A207 T208 G209 T210 T211 C212 A213 G214 C215 T216 G217 A218 A219 T220 T221 C222 A223 G224 C225 T226 G227 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 (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	104.79Å 108.77Å 174.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.52 – 2.89 41.52 – 2.89	Depositor EDS
% Data completeness (in resolution range)	99.3 (41.52-2.89) 99.2 (41.52-2.89)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	3.03 (at 2.90Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.241 , 0.296 0.240 , 0.295	Depositor DCC
R_{free} test set	2262 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	69.4	Xtriage
Anisotropy	0.827	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 66.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.023 for k,h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11947	wwPDB-VP
Average B, all atoms (Å^2)	104.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.81% 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: CL, MN

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	0.45	0/819	0.61	0/1097
1	E	0.52	0/828	0.71	0/1109
2	B	0.44	0/621	0.65	0/833
2	F	0.52	0/680	0.79	0/909
3	C	0.47	0/829	0.70	0/1118
3	G	0.42	0/815	0.66	0/1100
4	D	0.49	0/756	0.68	0/1015
4	H	0.44	0/730	0.64	0/982
5	I	0.41	0/3332	0.78	0/5141
5	J	0.38	0/3330	0.78	0/5138
All	All	0.43	0/12740	0.74	0/18442

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	807	0	844	78	0
1	E	816	0	856	73	0
2	B	614	0	651	71	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	84	0
3	C	819	0	879	73	0
3	G	805	0	861	60	0
4	D	745	0	771	62	0
4	H	719	0	740	66	0
5	I	2970	0	1640	240	0
5	J	2969	0	1641	291	0
6	A	1	0	0	0	0
6	C	1	0	0	1	0
6	G	1	0	0	0	0
7	E	1	0	0	0	0
7	I	3	0	0	0	0
7	J	3	0	0	0	0
All	All	11947	0	9600	923	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 43.

The worst 5 of 923 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)
2:F:19:ARG:HA	2:F:19:ARG:NH1	1.57	1.19
3:C:15:LYS:HB2	3:C:20:ARG:HH12	1.10	1.12
5:I:135:DG:H2"	5:I:136:DT:H5'	1.30	1.12
5:J:253:DC:H2"	5:J:254:DC:H5'	1.33	1.11
5:I:105:DT:H2"	5:I:106:DT:H5"	1.31	1.10

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	96/139 (69%)	84 (88%)	10 (10%)	2 (2%)	7	26
1	E	97/139 (70%)	84 (87%)	11 (11%)	2 (2%)	7	26
2	B	75/106 (71%)	64 (85%)	9 (12%)	2 (3%)	5	19
2	F	82/106 (77%)	61 (74%)	18 (22%)	3 (4%)	3	13
3	C	104/133 (78%)	90 (86%)	12 (12%)	2 (2%)	8	28
3	G	102/133 (77%)	85 (83%)	15 (15%)	2 (2%)	7	27
4	D	93/129 (72%)	78 (84%)	13 (14%)	2 (2%)	6	24
4	H	90/129 (70%)	72 (80%)	15 (17%)	3 (3%)	4	15
All	All	739/1014 (73%)	618 (84%)	103 (14%)	18 (2%)	6	22

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	71	VAL
2	F	30	THR
2	F	101	GLY
3	C	38	ASN
4	D	104	GLY

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%)	81 (95%)	4 (5%)	26	59
1	E	86/113 (76%)	83 (96%)	3 (4%)	36	70
2	B	63/81 (78%)	60 (95%)	3 (5%)	25	58
2	F	69/81 (85%)	61 (88%)	8 (12%)	5	16
3	C	84/102 (82%)	75 (89%)	9 (11%)	6	20
3	G	83/102 (81%)	79 (95%)	4 (5%)	25	58
4	D	81/107 (76%)	72 (89%)	9 (11%)	6	19
4	H	78/107 (73%)	73 (94%)	5 (6%)	17	45

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	629/806 (78%)	584 (93%)	45 (7%)	14 39

5 of 45 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	19	ARG
2	F	95	ARG
2	F	45	ARG
2	F	79	GLN
3	G	73	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	64	ASN
3	G	73	ASN
3	G	89	ASN
2	F	93	GLN
3	C	84	GLN

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 10 ligands modelled in this entry, 10 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	98/139 (70%)	-0.01	2 (2%)	65	63	40, 79, 109, 137	0
1	E	99/139 (71%)	-0.19	0	100	100	38, 63, 88, 131	0
2	B	77/106 (72%)	0.01	2 (2%)	56	52	50, 73, 92, 113	0
2	F	84/106 (79%)	-0.13	0	100	100	39, 58, 73, 113	0
3	C	106/133 (79%)	-0.10	1 (0%)	84	84	46, 64, 98, 128	0
3	G	104/133 (78%)	-0.07	1 (0%)	82	82	50, 75, 108, 121	0
4	D	95/129 (73%)	-0.07	1 (1%)	80	80	46, 67, 103, 157	0
4	H	92/129 (71%)	-0.01	2 (2%)	62	59	46, 75, 101, 126	0
5	I	145/146 (99%)	-0.28	0	100	100	73, 134, 174, 191	0
5	J	145/146 (99%)	-0.27	0	100	100	80, 136, 182, 201	0
All	All	1045/1306 (80%)	-0.13	9 (0%)	84	84	38, 77, 162, 201	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	H	52	THR	3.5
4	D	31	ARG	3.4
2	B	26	ILE	2.8
3	C	14	ALA	2.8
1	A	38	PRO	2.7

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	I	1003	1/1	0.78	0.22	119,119,119,119	0
7	MN	I	1001	1/1	0.83	0.09	137,137,137,137	0
6	CL	A	1001	1/1	0.87	0.10	85,85,85,85	0
7	MN	I	1002	1/1	0.92	0.06	123,123,123,123	0
7	MN	J	1002	1/1	0.92	0.22	122,122,122,122	0
7	MN	J	1003	1/1	0.92	0.07	118,118,118,118	0
6	CL	G	1001	1/1	0.94	0.09	74,74,74,74	0
7	MN	J	1001	1/1	0.96	0.05	137,137,137,137	0
7	MN	E	1001	1/1	0.97	0.29	61,61,61,61	0
6	CL	C	1001	1/1	0.98	0.07	87,87,87,87	0

6.5 Other polymers (i)

There are no such residues in this entry.