



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 25, 2023 – 02:21 PM EDT

PDB ID : 3AZG
Title : Crystal Structure of Human Nucleosome Core Particle Containing H3K115Q mutation
Authors : Iwasaki, W.; Tachiwana, H.; Kawaguchi, K.; Shibata, T.; Kagawa, W.; Kurumizaka, H.
Deposited on : 2011-05-25
Resolution : 2.40 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 ⓘ 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 ⓘ](#)) 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

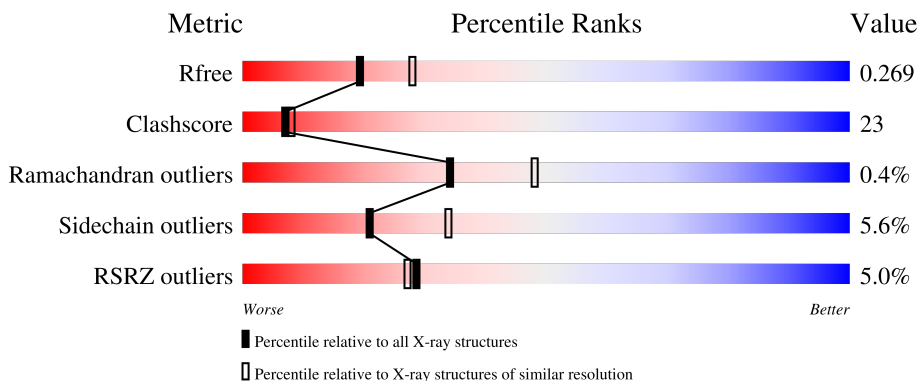
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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 $\leq 5\%$. 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	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	133	<p>62% 14% 2% 22%</p>
4	D	129	<p>2% 50% 21% 27%</p>
4	H	129	<p>58% 10% 10% 28%</p>
5	I	146	<p>14% 14% 85% 1%</p>
5	J	146	<p>14% 11% 88% 1%</p>

2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 12065 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	504	155	138	4	0	0	0
1	E	99	816	513	158	141	4	0	0	0

There are 8 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
A	115	GLN	LYS	engineered mutation	UNP P68431
E	-3	GLY	-	expression tag	UNP P68431
E	-2	SER	-	expression tag	UNP P68431
E	-1	HIS	-	expression tag	UNP P68431
E	115	GLN	LYS	engineered mutation	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	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

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
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	94	736	462	134	138	2	0	0	0
4	H	93	725	456	130	137	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
6	A	1	Total 1	Cl 1	0	0
6	D	1	Total 1	Cl 1	0	0
6	E	1	Total 1	Cl 1	0	0
6	G	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	E	1	Total 1	Mn 1	0	0
7	I	5	Total 5	Mn 5	0	0
7	J	3	Total 3	Mn 3	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	15	Total 15	O 15	0	0
8	B	12	Total 12	O 12	0	0
8	C	11	Total 11	O 11	0	0
8	D	6	Total 6	O 6	0	0
8	E	25	Total 25	O 25	0	0
8	F	12	Total 12	O 12	0	0

Continued on next page...

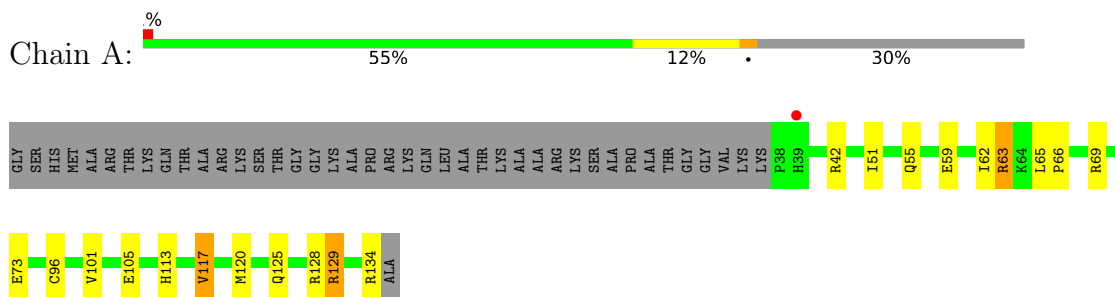
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	G	8	Total O 8 8	0	0
8	H	4	Total O 4 4	0	0
8	I	7	Total O 7 7	0	0
8	J	3	Total O 3 3	0	0

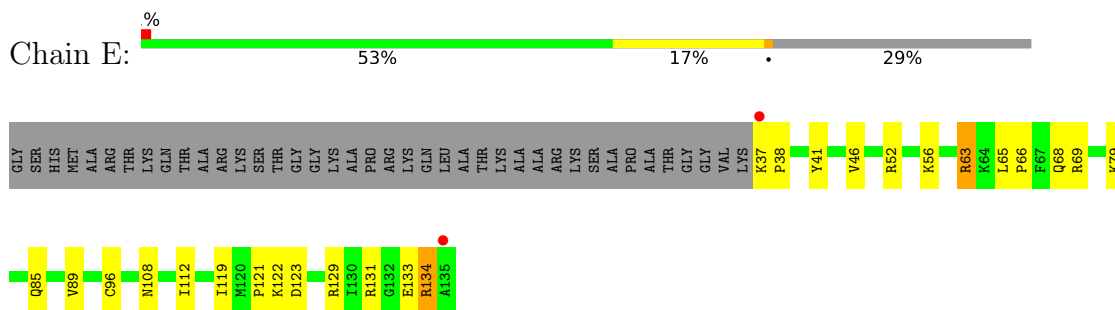
3 Residue-property plots [i](#)

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.

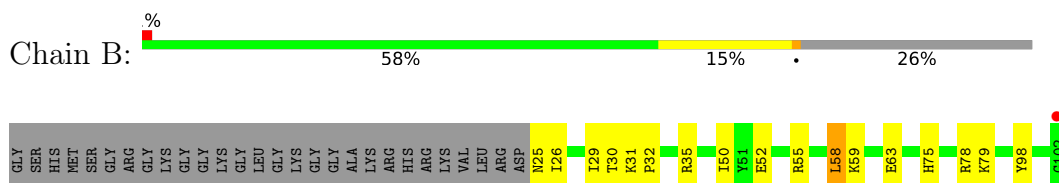
- Molecule 1: Histone H3.1



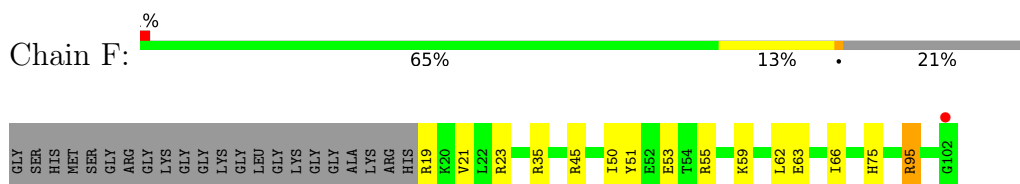
- Molecule 1: Histone H3.1



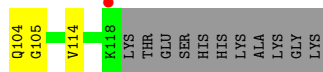
- Molecule 2: Histone H4



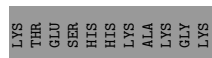
- Molecule 2: Histone H4



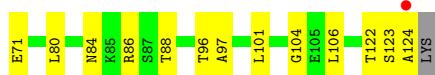
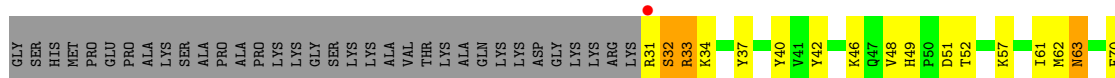
- Molecule 3: Histone H2A type 1-B/E



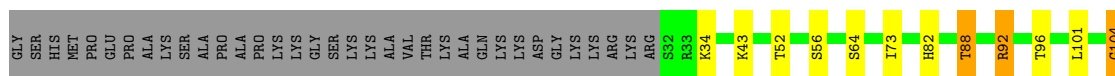
• Molecule 3: Histone H2A type 1-B/E



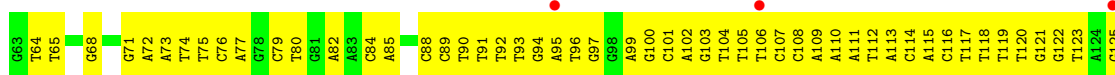
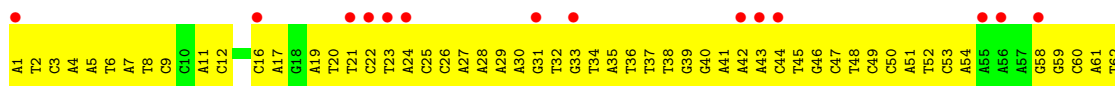
• Molecule 4: Histone H2B type 1-J



• Molecule 4: Histone H2B type 1-J



• Molecule 5: 146-MER DNA



● ● ●
 A126
 A127
 T128
 C129
 T130
 G131
 C132
 A133
 G134
 G135
 T136
 G137
 G138
 A139
 T140
 A141
 T142
 T143
 T144
 G144
 A145
 DT

● Molecule 5: 146-MER DNA

Chain J: 14% 11% 88%

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
 G205
 A207

T208
 G209
 T210
 T211
 C212
 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
 G254
 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, α , β , γ	106.53Å 109.77Å 182.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.16 – 2.40 39.16 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.6 (39.16-2.40) 99.5 (39.16-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.22 (at 2.39Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.226 , 0.269 0.226 , 0.269	Depositor DCC
R_{free} test set	4192 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	44.2	Xtrriage
Anisotropy	0.360	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 50.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.016 for k,h,-l	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	12065	wwPDB-VP
Average B, all atoms (Å ²)	70.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.32% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.48	0/813	0.64	0/1091
1	E	0.53	0/828	0.72	0/1110
2	B	0.48	0/626	0.67	0/837
2	F	0.53	0/680	0.75	1/908 (0.1%)
3	C	0.49	0/845	0.66	0/1139
3	G	0.41	0/815	0.63	0/1100
4	D	0.47	0/747	0.65	0/1004
4	H	0.49	0/736	0.61	0/990
5	I	0.40	0/3332	0.79	0/5141
5	J	0.43	0/3330	0.78	0/5138
All	All	0.45	0/12752	0.74	1/18458 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	95	ARG	NE-CZ-NH2	-5.50	117.55	120.30

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	834	34	0
1	E	816	0	851	22	0
2	B	619	0	659	30	0
2	F	673	0	722	13	0
3	C	835	0	897	40	0
3	G	805	0	861	21	0
4	D	736	0	758	31	0
4	H	725	0	745	19	0
5	I	2970	0	1640	175	0
5	J	2969	0	1641	184	0
6	A	1	0	0	0	0
6	D	1	0	0	0	0
6	E	1	0	0	1	0
6	G	1	0	0	0	0
7	E	1	0	0	0	0
7	I	5	0	0	0	0
7	J	3	0	0	0	0
8	A	15	0	0	1	0
8	B	12	0	0	2	0
8	C	11	0	0	0	0
8	D	6	0	0	0	0
8	E	25	0	0	2	0
8	F	12	0	0	2	0
8	G	8	0	0	0	0
8	H	4	0	0	0	0
8	I	7	0	0	0	0
8	J	3	0	0	0	0
All	All	12065	0	9608	494	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 23.

The worst 5 of 494 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'	1.19	1.17
5:J:286:DT:H2''	5:J:287:DA:H5''	1.17	1.12
5:J:182:DT:H2''	5:J:183:DT:H5'	1.28	1.11
5:I:106:DT:H2''	5:I:107:DC:H5'	1.13	1.10
5:J:197:DA:H2''	5:J:198:DT:H5'	1.33	1.08

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%)	95 (100%)	0	0	100	100
1	E	97/139 (70%)	96 (99%)	1 (1%)	0	100	100
2	B	76/106 (72%)	76 (100%)	0	0	100	100
2	F	82/106 (77%)	81 (99%)	1 (1%)	0	100	100
3	C	106/133 (80%)	102 (96%)	4 (4%)	0	100	100
3	G	102/133 (77%)	99 (97%)	3 (3%)	0	100	100
4	D	92/129 (71%)	88 (96%)	2 (2%)	2 (2%)	6	7
4	H	91/129 (70%)	89 (98%)	1 (1%)	1 (1%)	14	20
All	All	741/1014 (73%)	726 (98%)	12 (2%)	3 (0%)	34	48

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	32	SER
4	D	104	GLY
4	H	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	42
1	E	86/113 (76%)	82 (95%)	4 (5%)	26	42

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	63/81 (78%)	62 (98%)	1 (2%)	62	79
2	F	69/81 (85%)	68 (99%)	1 (1%)	67	82
3	C	85/102 (83%)	77 (91%)	8 (9%)	8	13
3	G	83/102 (81%)	78 (94%)	5 (6%)	19	31
4	D	80/107 (75%)	74 (92%)	6 (8%)	13	21
4	H	79/107 (74%)	73 (92%)	6 (8%)	13	20
All	All	630/806 (78%)	595 (94%)	35 (6%)	21	34

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

Mol	Chain	Res	Type
3	G	114	VAL
4	H	34	LYS
4	H	92	ARG
3	C	101	THR
3	C	95	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	84	GLN
3	G	31	HIS
4	D	49	HIS
4	H	63	ASN
2	F	75	HIS

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 13 ligands modelled in this entry, 13 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%)	0.11	1 (1%) 82 80	28, 41, 67, 104	0
1	E	99/139 (71%)	0.17	2 (2%) 65 63	22, 34, 60, 106	0
2	B	78/106 (73%)	0.06	1 (1%) 77 75	29, 39, 58, 71	0
2	F	84/106 (79%)	0.13	1 (1%) 79 77	24, 34, 56, 90	0
3	C	108/133 (81%)	0.11	4 (3%) 41 41	24, 39, 73, 127	0
3	G	104/133 (78%)	-0.18	0 100 100	30, 43, 70, 97	0
4	D	94/129 (72%)	0.20	2 (2%) 63 61	26, 40, 74, 117	0
4	H	93/129 (72%)	-0.03	0 100 100	29, 44, 82, 100	0
5	I	145/146 (99%)	0.66	20 (13%) 2 2	41, 97, 144, 160	0
5	J	145/146 (99%)	0.67	21 (14%) 2 2	48, 97, 144, 157	0
All	All	1047/1306 (80%)	0.23	52 (4%) 28 27	22, 45, 127, 160	0

The worst 5 of 52 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	11	ARG	8.5
1	E	37	LYS	8.2
3	C	12	ALA	6.6
5	I	24	DA	4.2
5	I	44	DC	4.2

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](#)

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(\AA^2)	Q<0.9
7	MN	I	1004	1/1	0.76	0.16	105,105,105,105	0
7	MN	I	1003	1/1	0.80	0.08	101,101,101,101	0
7	MN	I	1001	1/1	0.89	0.09	112,112,112,112	0
7	MN	J	1001	1/1	0.91	0.07	110,110,110,110	0
7	MN	I	1002	1/1	0.93	0.25	82,82,82,82	0
7	MN	J	1003	1/1	0.94	0.21	69,69,69,69	0
7	MN	I	1005	1/1	0.95	0.15	90,90,90,90	0
6	CL	G	1001	1/1	0.96	0.20	49,49,49,49	0
6	CL	A	1001	1/1	0.97	0.08	58,58,58,58	0
6	CL	D	201	1/1	0.98	0.12	46,46,46,46	0
7	MN	E	1001	1/1	0.98	0.26	36,36,36,36	0
7	MN	J	1002	1/1	0.98	0.23	75,75,75,75	0
6	CL	E	1002	1/1	0.98	0.12	45,45,45,45	0

6.5 Other polymers [i](#)

There are no such residues in this entry.