



wwPDB X-ray Structure Validation Summary Report

Feb 14, 2017 – 10:46 pm GMT

PDB ID : 3AYW
Title : Crystal Structure of Human Nucleosome Core Particle Containing H3K56Q mutation
Authors : Iwasaki, W.; Tachiwana, H.; Kawaguchi, K.; Shibata, T.; Kagawa, W.; Kurumizaka, H.
Deposited on : 2011-05-19
Resolution : 2.90 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

<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

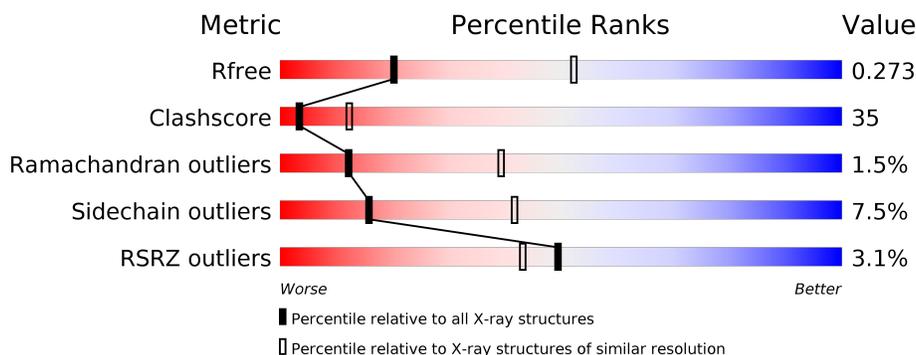
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 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	<p>%</p> <p>36% 35% 26%</p>
4	H	129	<p>34% 35% 29%</p>
5	I	146	<p>10% 7% 92%</p>
5	J	146	<p>10% 12% 88%</p>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	CL	A	1001	-	-	X	-
7	MN	D	201	-	-	-	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 11972 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	97	799	504	153	138	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	56	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	56	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	103	796	502	155	139	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	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	146	2990	1431	540	874	145	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	D	1	Total 1	Cl 1	0	0
6	E	1	Total 1	Cl 1	0	0

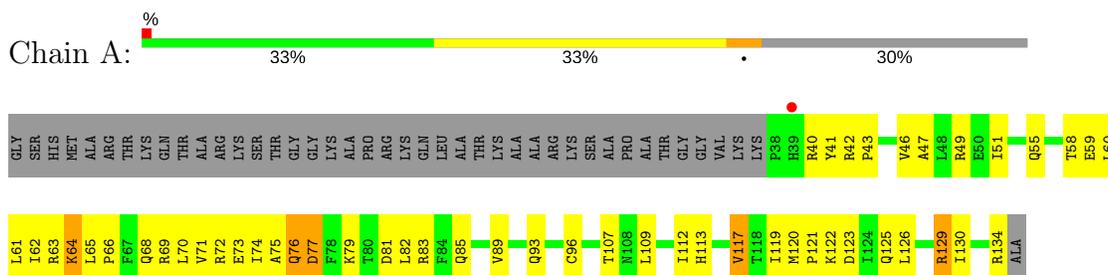
- Molecule 7 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	5	Total 5	Mn 5	0	0
7	I	5	Total 5	Mn 5	0	0
7	D	1	Total 1	Mn 1	0	0

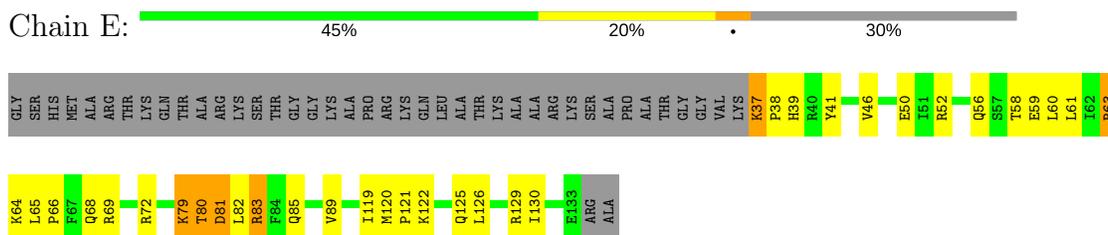
3 Residue-property plots [i](#)

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.

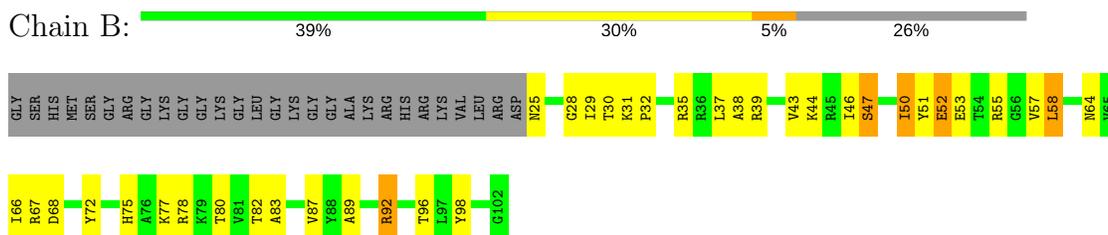
- Molecule 1: Histone H3.1



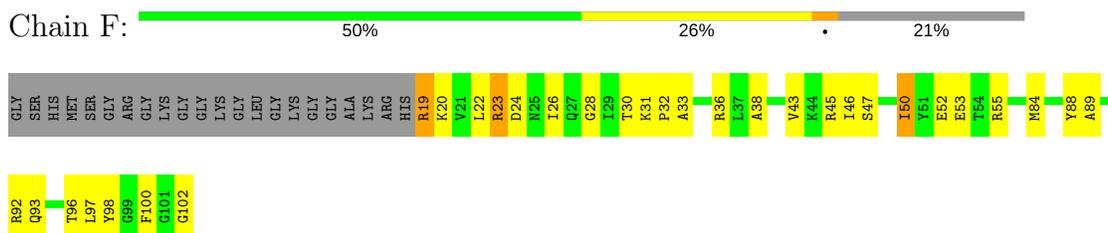
- Molecule 1: Histone H3.1



- Molecule 2: Histone H4



- Molecule 2: Histone H4



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: 10%
12% 88%

A147
T148
C149
A150
A151
T152
A153
T154
C155
C156
A157
C158
C159
T160
G161
C162
A163
G164
A165
T166
T169
A170
C171
C172
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
A202
A203
G204
G205
C206
A207
T208
G209

T210
T211
C212
C215
T216
G217
A218
A219
T220
T221
C222
A223
T226
G227
A228
A229
C230
A231
T232
G233
C234
A235
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	105.95Å 109.48Å 181.56Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.99 – 2.90 38.99 – 2.90	Depositor EDS
% Data completeness (in resolution range)	86.1 (38.99-2.90) 86.1 (38.99-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.93 (at 2.90Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.218 , 0.273 0.219 , 0.273	Depositor DCC
R_{free} test set	2057 reflections (5.02%)	DCC
Wilson B-factor (Å ²)	60.1	Xtrriage
Anisotropy	0.375	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 59.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.25$	Xtrriage
Estimated twinning fraction	0.054 for k,h,-l	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	11972	wwPDB-VP
Average B, all atoms (Å ²)	77.0	wwPDB-VP

Xtrriage'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.*

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

5.1 Standard geometry

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	0.40	0/813	0.63	0/1091
1	E	0.44	0/811	0.65	0/1089
2	B	0.41	0/626	0.69	0/837
2	F	0.45	0/680	0.73	0/908
3	C	0.39	0/845	0.64	0/1139
3	G	0.38	0/806	0.62	0/1089
4	D	0.41	0/766	0.66	0/1026
4	H	0.39	0/730	0.62	0/982
5	I	0.39	0/3332	0.78	1/5141 (0.0%)
5	J	0.37	0/3354	0.79	0/5175
All	All	0.39	0/12763	0.73	1/18477 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
5	I	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	I	117	DT	N1-C1'-C2'	5.08	122.25	112.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	I	117	DT	Sidechain

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	72	0
1	E	799	0	833	46	0
2	B	619	0	659	44	0
2	F	673	0	722	40	0
3	C	835	0	897	81	0
3	G	796	0	848	57	0
4	D	755	0	784	70	0
4	H	719	0	740	55	0
5	I	2970	0	1640	237	0
5	J	2990	0	1652	215	0
6	A	1	0	0	2	0
6	D	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	5	0	0	0	0
7	J	5	0	0	0	0
All	All	11972	0	9609	761	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 35.

The worst 5 of 761 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:248:DA:H2''	5:J:249:DG:H5''	1.24	1.16
5:J:152:DT:H2''	5:J:153:DA:H5''	1.26	1.14
5:J:277:DG:H2''	5:J:278:DC:H5'	1.28	1.12
5:I:43:DA:H2''	5:I:44:DC:H5''	1.15	1.10
5:I:5:DA:H2''	5:I:6:DT:H5''	1.34	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%)	93 (98%)	1 (1%)	1 (1%)	17	48
1	E	95/139 (68%)	89 (94%)	6 (6%)	0	100	100
2	B	76/106 (72%)	71 (93%)	5 (7%)	0	100	100
2	F	82/106 (77%)	79 (96%)	3 (4%)	0	100	100
3	C	106/133 (80%)	96 (91%)	8 (8%)	2 (2%)	9	33
3	G	101/133 (76%)	96 (95%)	2 (2%)	3 (3%)	5	20
4	D	94/129 (73%)	83 (88%)	8 (8%)	3 (3%)	5	19
4	H	90/129 (70%)	77 (86%)	11 (12%)	2 (2%)	8	29
All	All	739/1014 (73%)	684 (93%)	44 (6%)	11 (2%)	12	39

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	72	ASP
4	D	31	ARG
4	H	123	SER
4	D	104	GLY
3	G	38	ASN

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%)	79 (93%)	6 (7%)	17	44
1	E	85/113 (75%)	79 (93%)	6 (7%)	17	44
2	B	63/81 (78%)	58 (92%)	5 (8%)	14	39
2	F	69/81 (85%)	64 (93%)	5 (7%)	17	43
3	C	85/102 (83%)	77 (91%)	8 (9%)	10	30
3	G	82/102 (80%)	75 (92%)	7 (8%)	12	35
4	D	82/107 (77%)	77 (94%)	5 (6%)	22	53
4	H	78/107 (73%)	73 (94%)	5 (6%)	20	50
All	All	629/806 (78%)	582 (92%)	47 (8%)	16	42

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	63	ASN
1	E	79	LYS
4	H	86	ARG
4	D	101	LEU
1	E	80	THR

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

Mol	Chain	Res	Type
4	D	49	HIS
4	D	63	ASN
3	G	73	ASN
3	C	84	GLN
3	G	38	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, 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%)	0.01	1 (1%) 82 81	31, 49, 71, 104	0
1	E	97/139 (69%)	-0.09	0 100 100	24, 42, 64, 95	0
2	B	78/106 (73%)	-0.07	0 100 100	30, 47, 67, 87	0
2	F	84/106 (79%)	-0.04	0 100 100	23, 40, 65, 83	0
3	C	108/133 (81%)	-0.12	1 (0%) 84 83	27, 45, 76, 123	0
3	G	103/133 (77%)	-0.16	0 100 100	33, 52, 74, 85	0
4	D	96/129 (74%)	-0.02	1 (1%) 82 81	32, 48, 92, 118	0
4	H	92/129 (71%)	-0.08	0 100 100	30, 52, 78, 99	0
5	I	145/146 (99%)	0.52	15 (10%) 7 5	48, 100, 152, 173	0
5	J	146/146 (100%)	0.54	14 (9%) 9 6	49, 103, 147, 185	0
All	All	1046/1306 (80%)	0.09	32 (3%) 49 43	23, 53, 134, 185	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	56	DA	6.7
5	J	147	DA	6.7
4	D	125	LYS	5.8
5	I	55	DA	5.6
5	I	106	DT	4.7

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 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, 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	LLDF	B-factors(\AA^2)	Q<0.9
7	MN	D	201	1/1	0.98	0.28	3.55	46,46,46,46	0
6	CL	D	202	1/1	0.91	0.15	-1.45	53,53,53,53	0
6	CL	G	1001	1/1	0.99	0.14	-1.69	44,44,44,44	0
7	MN	I	1001	1/1	0.68	0.14	-	121,121,121,121	0
7	MN	J	1001	1/1	0.80	0.14	-	119,119,119,119	0
7	MN	I	1005	1/1	0.86	0.07	-	95,95,95,95	0
7	MN	J	1004	1/1	0.97	0.13	-	65,65,65,65	0
7	MN	I	1002	1/1	0.84	0.09	-	126,126,126,126	0
6	CL	A	1001	1/1	0.95	0.07	-	69,69,69,69	0
7	MN	J	1002	1/1	0.96	0.10	-	101,101,101,101	0
7	MN	J	1003	1/1	0.90	0.14	-	103,103,103,103	0
7	MN	I	1003	1/1	0.94	0.17	-	86,86,86,86	0
7	MN	J	1005	1/1	0.97	0.04	-	108,108,108,108	0
7	MN	I	1004	1/1	0.97	0.07	-	102,102,102,102	0
6	CL	E	1001	1/1	0.99	0.08	-	53,53,53,53	0

6.5 Other polymers [i](#)

There are no such residues in this entry.