



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:29 AM GMT

PDB ID : 3AZH
Title : Crystal Structure of Human Nucleosome Core Particle Containing H3K122Q mutation
Authors : Iwasaki, W.; Tachiwana, H.; Kawaguchi, K.; Shibata, T.; Kagawa, W.; Kurumizaka, H.
Deposited on : 2011-05-25
Resolution : 3.49 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
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
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) : trunk26865

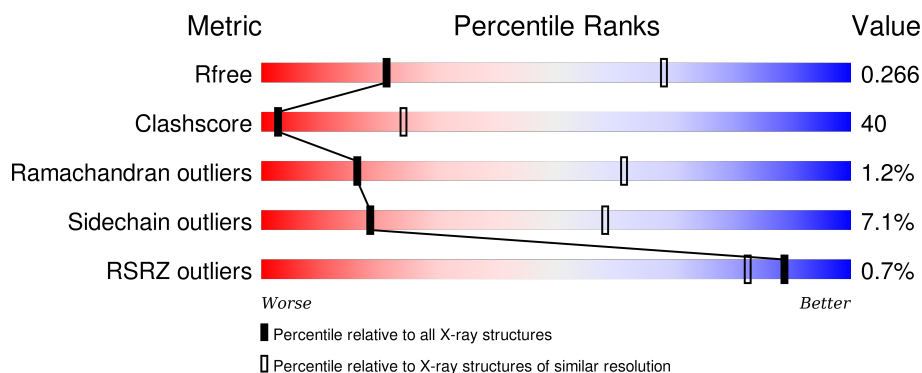
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.49 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}	91344	1051 (3.60-3.40)
Clashscore	102246	1157 (3.60-3.40)
Ramachandran outliers	100387	1120 (3.60-3.40)
Sidechain outliers	100360	1121 (3.60-3.40)
RSRZ outliers	91569	1058 (3.60-3.40)

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 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	<div> <div></div> <div>36% 30% . 30%</div> </div>
1	E	139	<div> <div>%</div> <div>43% 24% . 29%</div> </div>
2	B	106	<div> <div></div> <div>43% 27% . 26%</div> </div>
2	F	106	<div> <div></div> <div>42% 36% . 21%</div> </div>
3	C	133	<div> <div>%</div> <div>39% 37% . 20%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	133	
4	D	129	
4	H	129	
5	I	146	
5	J	146	

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
7	MN	D	201	-	-	-	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 11956 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
1	A	97	Total	C	N	O	S	0	0	0
			801	504	155	138	4			
1	E	99	Total	C	N	O	S	0	0	0
			816	513	158	141	4			

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	122	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	122	GLN	LYS	ENGINEERED MUTATION	UNP P68431

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	78	Total	C	N	O	S	0	0	0
			619	391	120	107	1			
2	F	84	Total	C	N	O	S	0	0	0
			678	428	135	114	1			

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
3	C	106	Total	C	N	O	0	0	0
			819	517	160	142			
3	G	104	Total	C	N	O	0	0	0
			805	508	157	140			

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
4	D	95	Total	C	N	O	S	0	0	0
			745	468	136	139	2			
4	H	92	Total	C	N	O	S	0	0	0
			719	453	129	135	2			

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
5	I	145	Total	C	N	O	P	0	0	0
			2970	1421	538	867	144			
5	J	145	Total	C	N	O	P	0	0	0
			2969	1421	535	869	144			

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	G	1	Total	Cl	0	0
			1	1		
6	C	1	Total	Cl	0	0
			1	1		
6	E	1	Total	Cl	0	0
			1	1		

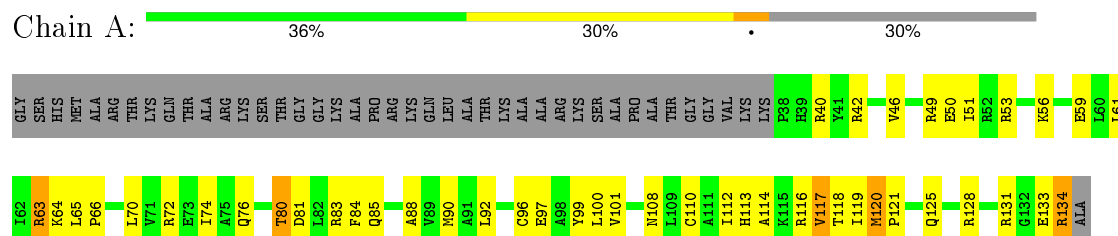
- Molecule 7 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	5	Total	Mn	0	0
			5	5		
7	I	6	Total	Mn	0	0
			6	6		
7	D	1	Total	Mn	0	0
			1	1		

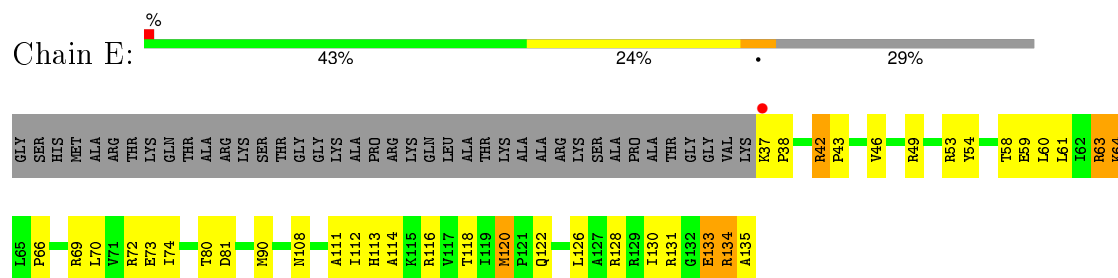
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

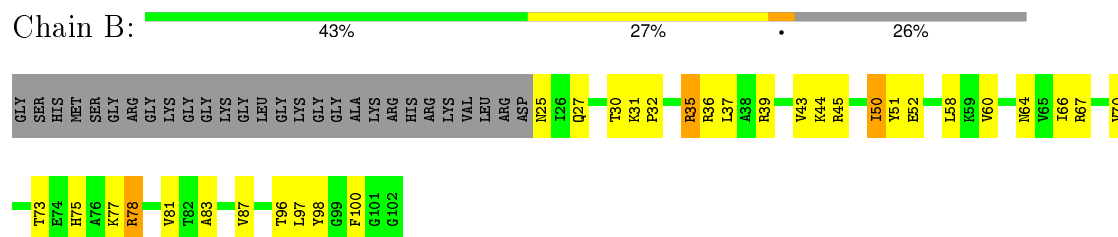
• Molecule 1: Histone H3.1



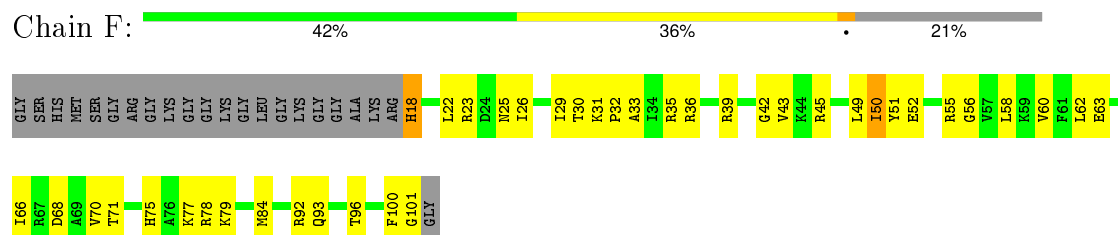
• Molecule 1: Histone H3.1



• Molecule 2: Histone H4



• Molecule 2: Histone H4



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



DA	T148	C149	A150	A151	T152	A153	T154	C155	C156	A157	C158	C159	T160	G161	G164	A165	T166	T167	C168	T169	A170	C171	C172	A173	A174	A175	T176	G177	T178	G179	T180	A181	T182	T183	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	C158	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	G262	T263	T264	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	106.08Å 109.66Å 181.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.81 – 3.49 45.81 – 3.49	Depositor EDS
% Data completeness (in resolution range)	98.8 (45.81-3.49) 98.8 (45.81-3.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.31 (at 3.48Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.190 , 0.266 0.190 , 0.266	Depositor DCC
R_{free} test set	1368 reflections (5.02%)	DCC
Wilson B-factor (Å ²)	55.7	Xtriage
Anisotropy	0.359	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 50.4	EDS
Estimated twinning fraction	0.024 for k,h,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 27271 reflections	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11956	wwPDB-VP
Average B, all atoms (Å ²)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.51% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.42	0/813	0.64	0/1091
1	E	0.44	0/828	0.69	0/1110
2	B	0.45	0/626	0.70	0/837
2	F	0.44	0/686	0.71	0/918
3	C	0.41	0/829	0.64	0/1118
3	G	0.41	0/815	0.63	0/1100
4	D	0.40	0/756	0.67	0/1015
4	H	0.41	0/730	0.64	0/982
5	I	0.45	0/3332	0.80	0/5141
5	J	0.41	0/3330	0.78	0/5138
All	All	0.42	0/12745	0.73	0/18450

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	801	0	834	62	0
1	E	816	0	851	45	0
2	B	619	0	659	41	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	F	678	0	726	35	0
3	C	819	0	879	66	0
3	G	805	0	861	79	0
4	D	745	0	771	75	0
4	H	719	0	740	65	0
5	I	2970	0	1640	246	0
5	J	2969	0	1641	275	0
6	C	1	0	0	0	0
6	E	1	0	0	1	0
6	G	1	0	0	0	0
7	D	1	0	0	0	0
7	I	6	0	0	0	0
7	J	5	0	0	0	0
All	All	11956	0	9602	855	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 40.

The worst 5 of 855 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:149:DC:H2''	5:J:150:DA:H5'	1.22	1.16
5:J:211:DT:H2''	5:J:212:DC:H5'	1.28	1.16
5:J:189:DA:H2''	5:J:190:DC:H5'	1.20	1.14
5:J:199:DC:H2''	5:J:200:DA:H5'	1.25	1.13
5:I:91:DT:H2''	5:I:92: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	95/139 (68%)	88 (93%)	7 (7%)	0	100	100
1	E	97/139 (70%)	93 (96%)	3 (3%)	1 (1%)	19	66
2	B	76/106 (72%)	68 (90%)	8 (10%)	0	100	100
2	F	82/106 (77%)	68 (83%)	13 (16%)	1 (1%)	16	61
3	C	104/133 (78%)	90 (86%)	13 (12%)	1 (1%)	19	66
3	G	102/133 (77%)	86 (84%)	13 (13%)	3 (3%)	6	42
4	D	93/129 (72%)	72 (77%)	19 (20%)	2 (2%)	8	49
4	H	90/129 (70%)	77 (86%)	12 (13%)	1 (1%)	17	63
All	All	739/1014 (73%)	642 (87%)	88 (12%)	9 (1%)	16	61

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	H	104	GLY
4	D	55	SER
4	D	104	GLY
3	C	104	GLN
1	E	64	LYS

5.3.2 Protein sidechains ⓘ

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%)	77 (91%)	8 (9%)	11	44
1	E	86/113 (76%)	78 (91%)	8 (9%)	11	45
2	B	63/81 (78%)	60 (95%)	3 (5%)	31	71
2	F	70/81 (86%)	67 (96%)	3 (4%)	35	74
3	C	84/102 (82%)	76 (90%)	8 (10%)	11	43
3	G	83/102 (81%)	77 (93%)	6 (7%)	18	57
4	D	81/107 (76%)	76 (94%)	5 (6%)	23	63
4	H	78/107 (73%)	74 (95%)	4 (5%)	29	69

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	630/806 (78%)	585 (93%)	45 (7%)	18	58

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

Mol	Chain	Res	Type
4	D	34	LYS
1	E	59	GLU
4	H	88	THR
4	D	88	THR
1	E	63	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	39	HIS
1	E	122	GLN
3	G	84	GLN
1	E	108	ASN
1	E	113	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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%)	-0.48	0 100 100	12, 36, 69, 103	0
1	E	99/139 (71%)	-0.48	1 (1%) 84 76	11, 30, 65, 115	0
2	B	78/106 (73%)	-0.61	0 100 100	12, 33, 57, 96	0
2	F	84/106 (79%)	-0.61	0 100 100	12, 29, 55, 84	0
3	C	106/133 (79%)	-0.53	1 (0%) 85 78	13, 32, 72, 91	0
3	G	104/133 (78%)	-0.47	1 (0%) 84 76	14, 41, 76, 109	0
4	D	95/129 (73%)	-0.47	1 (1%) 82 73	18, 36, 85, 136	0
4	H	92/129 (71%)	-0.47	0 100 100	16, 41, 67, 85	0
5	I	145/146 (99%)	0.21	3 (2%) 67 58	41, 93, 140, 196	0
5	J	145/146 (99%)	0.06	0 100 100	40, 96, 138, 155	0
All	All	1045/1306 (80%)	-0.33	7 (0%) 89 82	11, 43, 125, 196	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	37	LYS	3.5
4	D	31	ARG	2.9
5	I	105	DT	2.6
5	I	104	DT	2.4
3	C	13	LYS	2.3

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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	0.99	0.25	3.18	46,46,46,46	0
6	CL	G	1001	1/1	0.97	0.18	1.53	46,46,46,46	0
6	CL	C	1001	1/1	0.97	0.11	-2.17	48,48,48,48	0
7	MN	I	1004	1/1	0.94	0.09	-7.70	59,59,59,59	0
7	MN	I	1002	1/1	0.86	0.11	-	99,99,99,99	0
7	MN	J	1003	1/1	0.90	0.15	-	67,67,67,67	0
6	CL	E	1001	1/1	0.93	0.26	-	74,74,74,74	0
7	MN	J	1002	1/1	0.97	0.16	-	94,94,94,94	0
7	MN	J	1005	1/1	0.94	0.12	-	108,108,108,108	0
7	MN	I	1005	1/1	0.70	0.18	-	108,108,108,108	0
7	MN	I	1003	1/1	0.96	0.22	-	70,70,70,70	0
7	MN	J	1001	1/1	0.90	0.09	-	94,94,94,94	0
7	MN	I	1001	1/1	0.50	0.28	-	118,118,118,118	0
7	MN	I	1006	1/1	0.80	0.11	-	105,105,105,105	0
7	MN	J	1004	1/1	0.95	0.18	-	48,48,48,48	0

6.5 Other polymers [i](#)

There are no such residues in this entry.