wwPDB EM Validation Summary Report

Nov 20, 2022 – 07:26 AM EST

PDB ID	:	7MBM
EMDB ID	:	EMD-23738
Title	:	Cryo-EM structure of MLL1-NCP (H3K4M) complex, mode01
Authors	:	Park, S.H.; Ayoub, A.; Lee, Y.T.; Dou, Y.; Cho, U.
Deposited on	:	2021-04-01
Resolution	:	Not provided

This is a wwPDB EM 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/EMValidationReportHelp
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 (i)) were used in the production of this report:

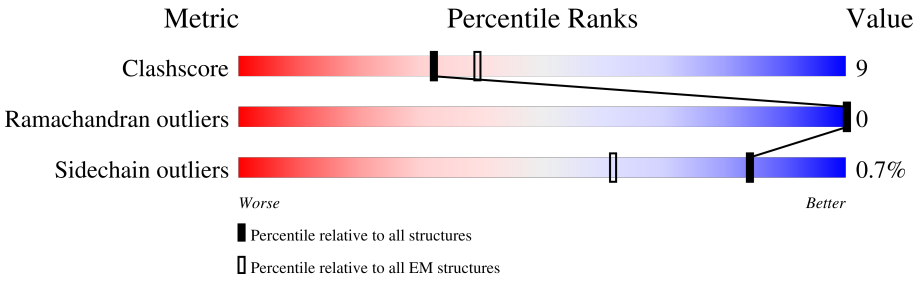
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is unknown.

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)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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%

Mol	Chain	Length	Quality of chain
1	A	538	54% 9% 37%
2	B	313	75% 20% .
3	C	209	70% 16% 14%
4	D	534	27% 6% 67%
5	G	136	54% 17% . 29%
5	K	136	57% 13% . 29%
6	H	103	65% 13% 22%
6	L	103	64% 12% 24%
7	I	129	73% 10% 17%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	M	129	
8	J	123	
8	N	123	
9	O	147	
10	P	147	

2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 19757 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Retinoblastoma-binding protein 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	339	Total	C	N	O	S	0	0
			2665	1681	466	503	15		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP Q15291

- Molecule 2 is a protein called WD repeat-containing protein 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	300	Total	C	N	O	S	0	0
			2326	1485	388	444	9		

- Molecule 3 is a protein called Histone-lysine N-methyltransferase 2A.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	179	Total	C	N	O	S	0	0
			1435	904	266	250	15		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	3761	SER	-	expression tag	UNP Q03164
C	3861	ILE	ASN	conflict	UNP Q03164
C	3867	LEU	GLN	conflict	UNP Q03164

- Molecule 4 is a protein called Set1/Ash2 histone methyltransferase complex subunit ASH2.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	176	Total	C	N	O	S	0	0
			1399	906	234	253	6		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	SER	-	expression tag	UNP Q9UBL3

- Molecule 5 is a protein called Histone H3.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	G	97	Total	C	N	O	S	0	0
			802	506	155	138	3		
5	K	97	Total	C	N	O	S	0	0
			802	506	155	138	3		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	4	MET	LYS	conflict	UNP A0A310TTQ1
K	4	MET	LYS	conflict	UNP A0A310TTQ1

- Molecule 6 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	H	80	Total	C	N	O	S	0	0
			641	405	125	110	1		
6	L	78	Total	C	N	O	S	0	0
			622	393	120	108	1		

- Molecule 7 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				AltConf	Trace
7	I	107	Total	C	N	O	0	0
			825	519	163	143		
7	M	105	Total	C	N	O	0	0
			809	510	158	141		

- Molecule 8 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	J	94	Total	C	N	O	S	0	0
			741	466	135	138	2		
8	N	95	Total	C	N	O	S	0	0
			745	469	134	140	2		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	0	MET	-	initiating methionine	UNP P02281
J	29	THR	SER	engineered mutation	UNP P02281
N	0	MET	-	initiating methionine	UNP P02281
N	29	THR	SER	engineered mutation	UNP P02281

- Molecule 9 is a DNA chain called DNA (145-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
9	O	145	Total	C	N	O	P	0	0
			2990	1415	559	871	145		

- Molecule 10 is a DNA chain called DNA (145-MER).

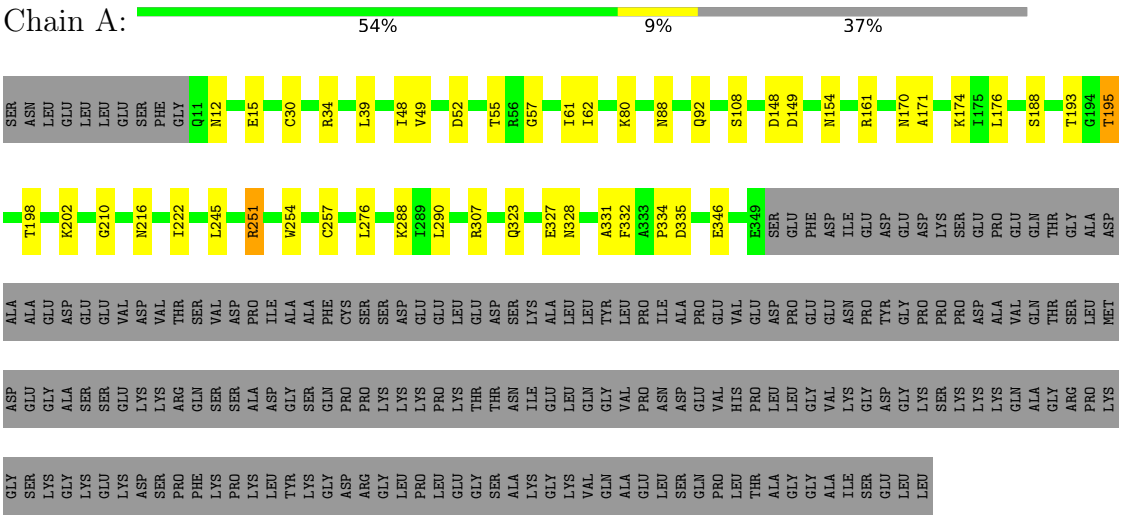
Mol	Chain	Residues	Atoms					AltConf	Trace
10	P	145	Total	C	N	O	P	0	0
			2955	1403	538	869	145		

3 Residue-property plots

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. 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. 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.

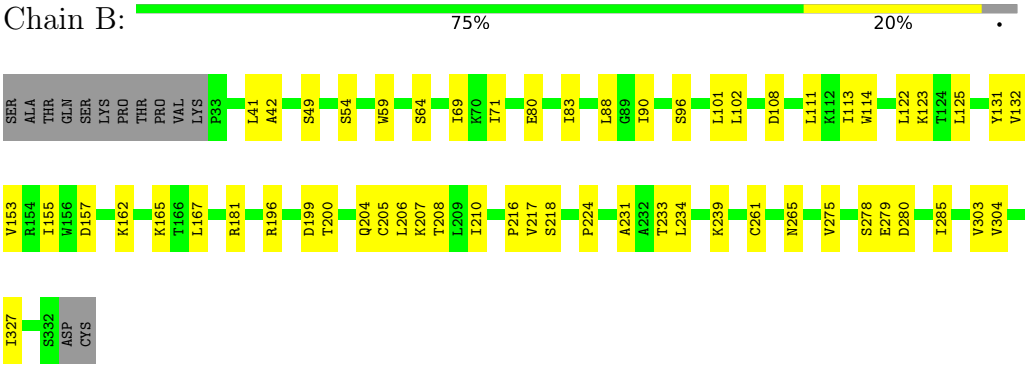
- Molecule 1: Retinoblastoma-binding protein 5

Chain A:



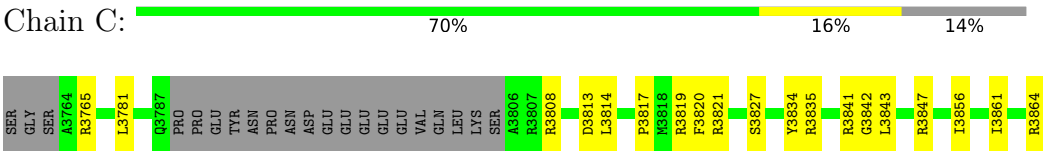
- Molecule 2: WD repeat-containing protein 5

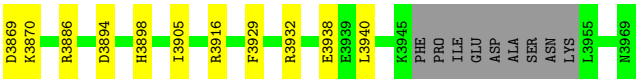
Chain B:



- Molecule 3: Histone-lysine N-methyltransferase 2A

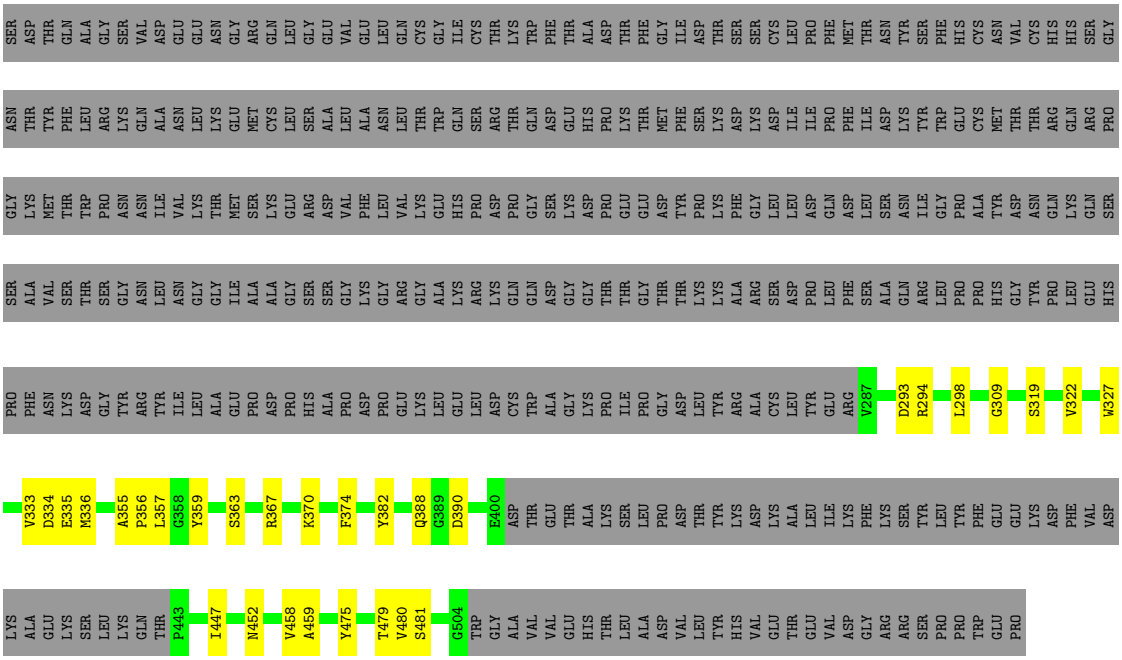
Chain C:





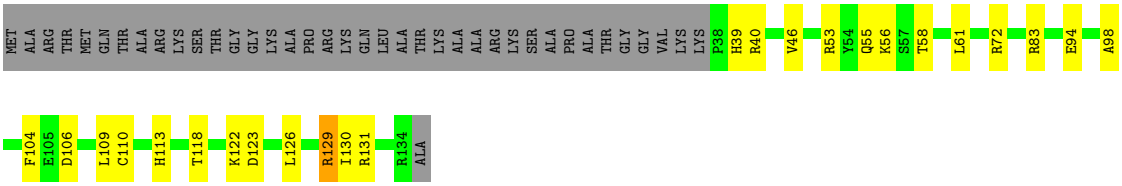
• Molecule 4: Set1/Ash2 histone methyltransferase complex subunit ASH2

Chain D:



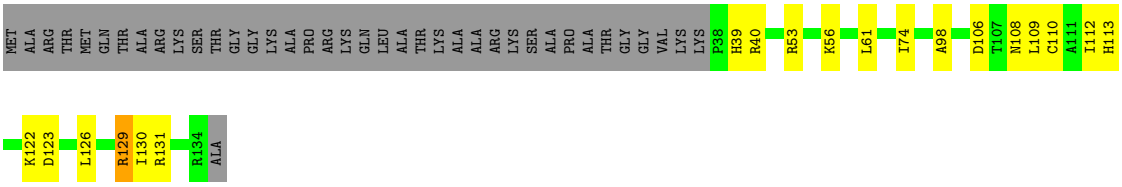
• Molecule 5: Histone H3

Chain G:



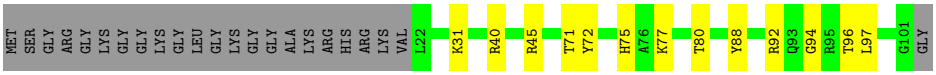
• Molecule 5: Histone H3

Chain K:



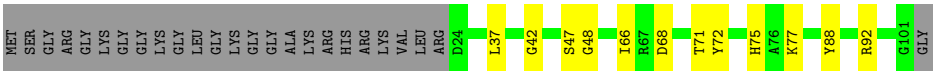
• Molecule 6: Histone H4

Chain H:



• Molecule 6: Histone H4

Chain L:

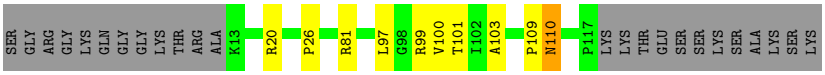


• Molecule 7: Histone H2A

Chain I:

• Molecule 7: Histone H2A

Chain M:



• Molecule 8: Histone H2B 1.1

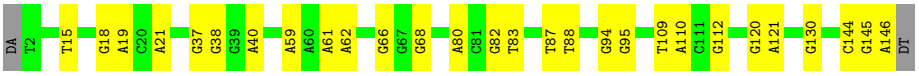
Chain J:

• Molecule 8: Histone H2B 1.1

Chain N:

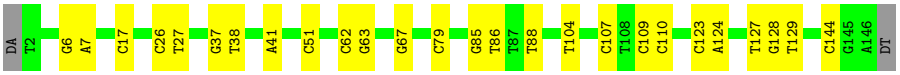
• Molecule 9: DNA (145-MER)

Chain O:



• Molecule 10: DNA (145-MER)

Chain P:


DA
T2
G6
A7
C17
C26
T27
G37
T38
A41
C51
C62
G63
G67
C79
G85
T86
T87
T88
T104
C107
T108
C109
C110
C123
A124
T127
G128
T129
C144
G145
A146
DT

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	30322	Depositor
Resolution determination method	Not provided	
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	53.4	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

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	0.28	0/2723	0.56	1/3699 (0.0%)
2	B	0.28	0/2382	0.52	0/3231
3	C	0.33	0/1461	0.56	1/1951 (0.1%)
4	D	0.27	0/1442	0.51	0/1950
5	G	0.37	0/814	0.52	0/1092
5	K	0.37	0/814	0.52	0/1092
6	H	0.39	0/648	0.60	0/868
6	L	0.41	0/629	0.57	0/843
7	I	0.34	0/835	0.54	0/1127
7	M	0.34	0/819	0.53	0/1106
8	J	0.37	0/752	0.52	0/1011
8	N	0.38	0/756	0.52	0/1015
9	O	0.70	0/3357	0.99	0/5184
10	P	0.71	0/3311	0.98	0/5103
All	All	0.48	0/20743	0.73	2/29272 (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
1	A	0	1
3	C	0	1
6	H	0	1
All	All	0	3

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	195	THR	N-CA-CB	5.43	120.61	110.30
3	C	3865	SER	CA-C-N	5.25	128.74	117.20

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	251	ARG	Peptide
3	C	3864	ARG	Peptide
6	H	31	LYS	Peptide

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	2665	0	2631	83	0
2	B	2326	0	2308	94	0
3	C	1435	0	1437	25	0
4	D	1399	0	1363	23	0
5	G	802	0	841	45	0
5	K	802	0	841	36	0
6	H	641	0	684	35	0
6	L	622	0	660	45	0
7	I	825	0	882	17	0
7	M	809	0	864	14	0
8	J	741	0	768	36	0
8	N	745	0	773	45	0
9	O	2990	0	1628	40	0
10	P	2955	0	1627	40	0
All	All	19757	0	17307	336	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 9.

The worst 5 of 336 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:A:334:PRO:HD3	2:B:208:THR:CB	1.39	1.47
1:A:334:PRO:CG	2:B:208:THR:O	1.67	1.41
1:A:334:PRO:CD	2:B:208:THR:HB	1.54	1.38
1:A:334:PRO:HG3	2:B:208:THR:N	1.37	1.35

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:334:PRO:CD	2:B:208:THR:O	1.78	1.32

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 PDB entries followed by that with respect to all EM entries.

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	337/538 (63%)	291 (86%)	46 (14%)	0	100	100
2	B	298/313 (95%)	270 (91%)	28 (9%)	0	100	100
3	C	173/209 (83%)	153 (88%)	20 (12%)	0	100	100
4	D	172/534 (32%)	158 (92%)	14 (8%)	0	100	100
5	G	95/136 (70%)	91 (96%)	4 (4%)	0	100	100
5	K	95/136 (70%)	90 (95%)	5 (5%)	0	100	100
6	H	78/103 (76%)	69 (88%)	9 (12%)	0	100	100
6	L	76/103 (74%)	72 (95%)	4 (5%)	0	100	100
7	I	105/129 (81%)	98 (93%)	7 (7%)	0	100	100
7	M	103/129 (80%)	98 (95%)	5 (5%)	0	100	100
8	J	92/123 (75%)	88 (96%)	4 (4%)	0	100	100
8	N	93/123 (76%)	89 (96%)	4 (4%)	0	100	100
All	All	1717/2576 (67%)	1567 (91%)	150 (9%)	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 PDB entries followed by that with respect to all EM

entries.

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	296/462 (64%)	292 (99%)	4 (1%)	67 67
2	B	262/274 (96%)	262 (100%)	0	100 100
3	C	151/182 (83%)	150 (99%)	1 (1%)	84 84
4	D	149/460 (32%)	148 (99%)	1 (1%)	84 84
5	G	85/111 (77%)	83 (98%)	2 (2%)	49 49
5	K	85/111 (77%)	83 (98%)	2 (2%)	49 49
6	H	66/79 (84%)	66 (100%)	0	100 100
6	L	64/79 (81%)	64 (100%)	0	100 100
7	I	84/101 (83%)	84 (100%)	0	100 100
7	M	83/101 (82%)	82 (99%)	1 (1%)	71 71
8	J	81/103 (79%)	81 (100%)	0	100 100
8	N	81/103 (79%)	81 (100%)	0	100 100
All	All	1487/2166 (69%)	1476 (99%)	11 (1%)	84 84

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	G	129	ARG
5	K	53	ARG
7	M	110	ASN
5	K	129	ARG
3	C	3886	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	328	ASN
4	D	388	GLN
7	M	110	ASN
1	A	92	GLN
1	A	88	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

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 Map visualisation

This section contains visualisations of the EMDB entry EMD-23738. These allow visual inspection of the internal detail of the map and identification of artifacts.

No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

This section was not generated.

6.2 Central slices

This section was not generated.

6.3 Largest variance slices

This section was not generated.

6.4 Orthogonal surface views

This section was not generated.

6.5 Mask visualisation

This section was not generated. No masks/segmentation were deposited.

7 Map analysis

This section contains the results of statistical analysis of the map.

7.1 Map-value distribution

This section was not generated.

7.2 Volume estimate versus contour level

This section was not generated.

7.3 Rotationally averaged power spectrum

This section was not generated. The rotationally averaged power spectrum had issues being displayed.

8 Fourier-Shell correlation

This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit

This section was not generated.