wwPDB X-ray Structure Validation Summary Report (i)

Aug 23, 2023 – 06:13 AM EDT

PDB ID	:	3C1C
Title	:	The effect of H3 K79 dimethylation and H4 K20 trimethylation on nucleosome and chromatin structure
Authors	:	Lu, X.; Simon, M.; Chodaparambil, J.; Hansen, J.; Shokat, K.; Luger, K.
Deposited on	:	2008-01-22
Resolution	:	3.15 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 (i)) were used in the production of this report:

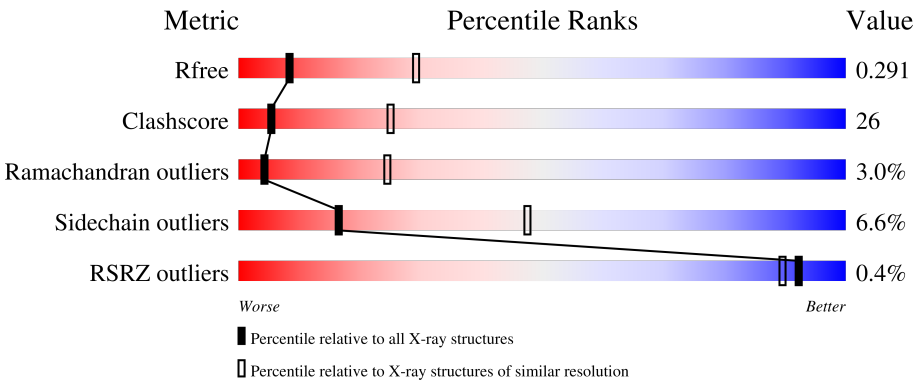
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.15 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	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)
RSRZ outliers	127900	1616 (3.20-3.12)

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	135	36% 30% 5% • 27%
1	E	135	% 42% 28% • 28%
2	B	102	34% 41% • 24%
2	F	102	44% 32% 6% 18%
3	C	129	45% 31% 5% 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	129	40% 36% 5% 19%
4	D	125	42% 27% 5% 26%
4	H	125	35% 34% 5% 26%
5	I	146	2% 38% 62%
5	J	146	32% 68%

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
1	M2L	A	479	-	-	X	-
1	M2L	E	679	-	-	X	-

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 12230 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-like.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	98	809	510	156	140	3	0	0	0
1	E	97	803	507	155	138	3	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	421	ALA	VAL	conflict	UNP P02302
A	426	ARG	LYS	conflict	UNP P02302
A	428	SER	CYS	conflict	UNP P02302
A	486	SER	ARG	conflict	UNP P02302
A	510	ALA	CYS	conflict	UNP P02302
E	621	ALA	VAL	conflict	UNP P02302
E	626	ARG	LYS	conflict	UNP P02302
E	628	SER	CYS	conflict	UNP P02302
E	686	SER	ARG	conflict	UNP P02302
E	710	ALA	CYS	conflict	UNP P02302

- 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

- Molecule 3 is a protein called Histone H2A type 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	105	809	510	158	141	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	G	104	804	507	157	140	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	899	ARG	GLY	conflict	UNP P06897
C	923	SER	ALA	conflict	UNP P06897
C	926	THR	ALA	conflict	UNP P06897
G	1099	ARG	GLY	conflict	UNP P06897
G	1123	SER	ALA	conflict	UNP P06897
G	1126	THR	ALA	conflict	UNP P06897

- Molecule 4 is a protein called Histone 2, H2bf.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	93	728	459	131	136	2	0	0	0
4	H	92	721	454	129	136	2	0	0	0

- Molecule 5 is a DNA chain called Palindromic 146bp Human Alpha satellite DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	I	146	2990	1430	541	874	145	0	0	0
5	J	146	2990	1430	541	874	145	0	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	5	5	5	0	0
6	B	8	8	8	0	0
6	C	23	23	23	0	0
6	D	11	11	11	0	0
6	E	18	18	18	0	0

Continued on next page...

Continued from previous page...

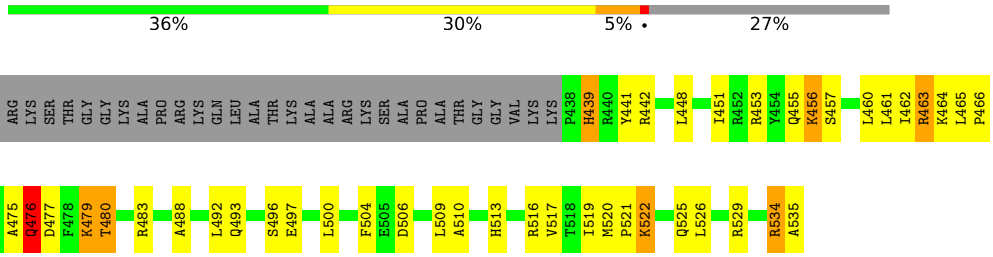
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	F	24	24	24	0	0
			Total	O		
6	G	21	21	21	0	0
			Total	O		
6	H	11	11	11	0	0
			Total	O		
6	I	72	72	72	0	0
			Total	O		
6	J	91	91	91	0	0

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.

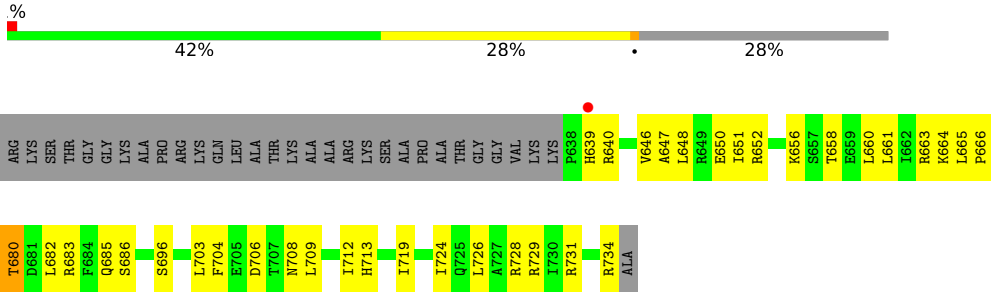
- Molecule 1: Histone H3-like

Chain A:



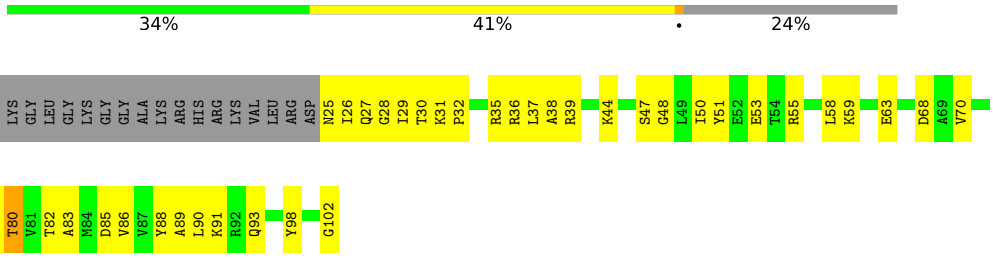
- Molecule 1: Histone H3-like

Chain E:



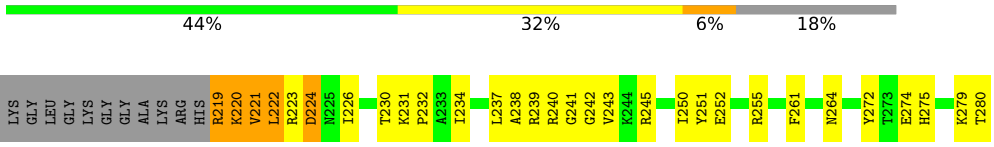
- Molecule 2: Histone H4

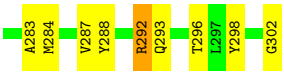
Chain B:



- Molecule 2: Histone H4

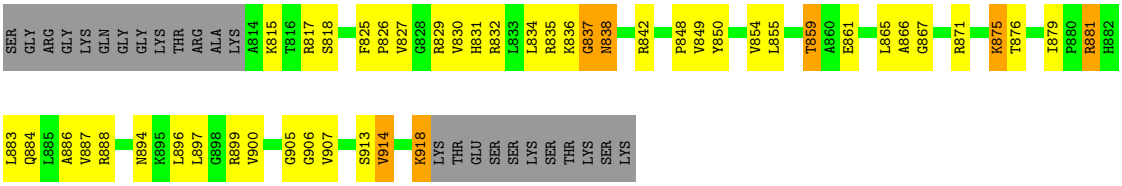
Chain F:





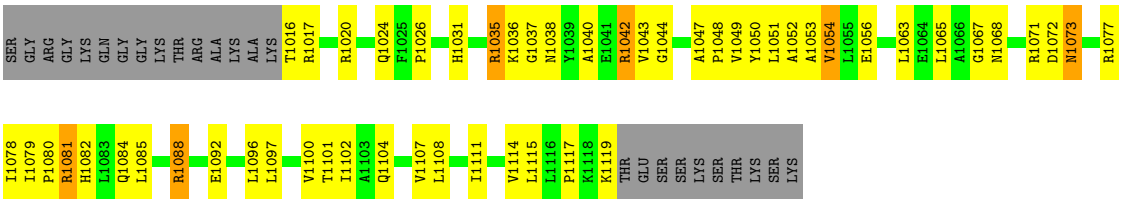
- Molecule 3: Histone H2A type 1

Chain C:



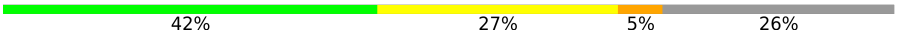
- Molecule 3: Histone H2A type 1

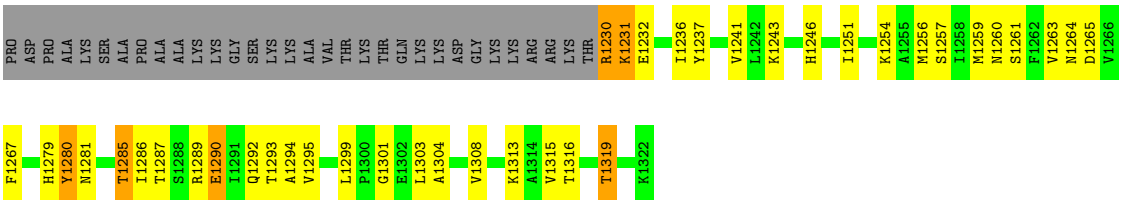
Chain G:



- Molecule 4: Histone 2, H2bf

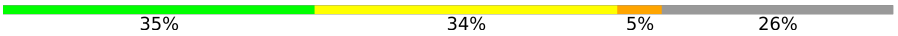
Chain D:

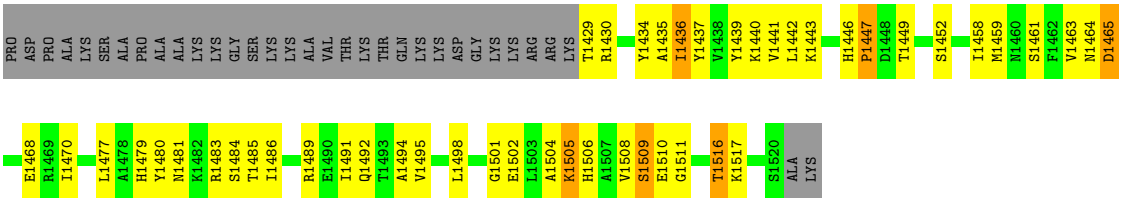




- Molecule 4: Histone 2, H2bf

Chain H:





- Molecule 5: Palindromic 146bp Human Alpha satellite DNA

Chain I:

A1 T2 C3 A4 A5 T6 C10 A11 C12 C13 T14 G15 C16 A17 G18 A19 T20 T21 C22 T23 A24 C25 C26 A27 A28 A29 A30 G31 T34 A35 T36 T37 G46 C49 C50 A51 T52 C53 A54 A55 A56 A57 G58 G59 C60 A61 T64 G71 A72 C77 G78 C79 T80

G81 A82 T86 G87 C88 C89 T90 T93 G94 A95 T96 G97 G98 A99 G100 C101 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 T146

- Molecule 5: Palindromic 146bp Human Alpha satellite DNA

Chain J: 32% 68%

A147 T148 C149 A150 A151 T152 A153 T154 C155 T160 G161 C162 A163 G164 A165 T166 T167 C168 T169 A170 C171 C172 A173 A174 A175 A176 G177 A181 T182 T183 T184 G185 A188 A189 C193 T194 C195 C196 A197 T198 G205 C206 A207 C212 A213 G216 C223 G224 C225 T226 G227 A228

A229 G233 C234 C235 T236 T237 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	107.03Å 110.08Å 182.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	53.00 – 3.15 53.27 – 3.15	Depositor EDS
% Data completeness (in resolution range)	90.8 (53.00-3.15) 88.4 (53.27-3.15)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.87 (at 3.13Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.220 , 0.290 0.225 , 0.291	Depositor DCC
R_{free} test set	1706 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	75.0	Xtriage
Anisotropy	0.556	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 85.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.54$, $< L^2 > = 0.38$	Xtriage
Estimated twinning fraction	0.000 for k,h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	12230	wwPDB-VP
Average B, all atoms (Å^2)	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.64% 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: M2L

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	0.44	0/809	0.71	0/1084
1	E	0.51	0/803	0.69	0/1077
2	B	0.44	0/626	0.68	0/837
2	F	0.55	0/680	0.80	0/908
3	C	0.46	0/819	0.69	0/1106
3	G	0.40	0/814	0.65	0/1099
4	D	0.46	0/739	0.70	0/994
4	H	0.44	0/732	0.66	0/986
5	I	0.37	0/3354	0.71	0/5175
5	J	0.37	0/3354	0.71	0/5175
All	All	0.42	0/12730	0.71	0/18441

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	809	0	846	67	0
1	E	803	0	841	43	0
2	B	619	0	659	54	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	F	673	0	722	49	0
3	C	809	0	864	50	0
3	G	804	0	859	53	0
4	D	728	0	753	37	0
4	H	721	0	742	46	0
5	I	2990	0	1651	124	0
5	J	2990	0	1651	151	0
6	A	5	0	0	0	0
6	B	8	0	0	0	0
6	C	23	0	0	1	0
6	D	11	0	0	1	0
6	E	18	0	0	0	0
6	F	24	0	0	0	0
6	G	21	0	0	2	0
6	H	11	0	0	0	0
6	I	72	0	0	20	0
6	J	91	0	0	18	0
All	All	12230	0	9588	559	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 26.

The worst 5 of 559 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:479:M2L:SG	1:A:479:M2L:HM1A	1.84	1.17
1:A:522:LYS:H	1:A:522:LYS:CD	1.66	1.08
1:E:683:ARG:HB2	2:F:280:THR:HG23	1.32	1.04
1:A:522:LYS:HD2	1:A:522:LYS:N	1.70	1.03
3:G:1102:ILE:HG23	4:H:1458:ILE:HD13	1.42	1.00

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	95/135 (70%)	82 (86%)	12 (13%)	1 (1%)	14	48
1	E	94/135 (70%)	87 (93%)	7 (7%)	0	100	100
2	B	76/102 (74%)	70 (92%)	6 (8%)	0	100	100
2	F	82/102 (80%)	73 (89%)	4 (5%)	5 (6%)	1	10
3	C	103/129 (80%)	92 (89%)	6 (6%)	5 (5%)	2	14
3	G	102/129 (79%)	93 (91%)	7 (7%)	2 (2%)	7	34
4	D	91/125 (73%)	72 (79%)	15 (16%)	4 (4%)	2	16
4	H	90/125 (72%)	70 (78%)	15 (17%)	5 (6%)	2	12
All	All	733/982 (75%)	639 (87%)	72 (10%)	22 (3%)	4	24

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	836	LYS
2	F	220	LYS
4	H	1501	GLY
1	A	476	GLN
3	C	838	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	83/108 (77%)	74 (89%)	9 (11%)	6	25
1	E	83/108 (77%)	78 (94%)	5 (6%)	19	51
2	B	63/78 (81%)	61 (97%)	2 (3%)	39	70
2	F	69/78 (88%)	66 (96%)	3 (4%)	29	62
3	C	83/102 (81%)	76 (92%)	7 (8%)	11	37
3	G	83/102 (81%)	77 (93%)	6 (7%)	14	43

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	D	79/104 (76%)	74 (94%)	5 (6%)	18	49
4	H	79/104 (76%)	75 (95%)	4 (5%)	24	56
All	All	622/784 (79%)	581 (93%)	41 (7%)	16	47

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	219	ARG
3	G	1081	ARG
2	F	224	ASP
3	G	1042	ARG
4	H	1436	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	293	GLN
3	G	1031	HIS
4	H	1464	ASN
3	G	1038	ASN
1	E	668	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)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
1	M2L	E	679	1	9,10,11	0.77	0	6,11,13	1.20	1 (16%)
1	M2L	A	479	1	9,10,11	1.16	1 (11%)	6,11,13	1.39	1 (16%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	M2L	E	679	1	-	1/7/9/11	-
1	M2L	A	479	1	-	4/7/9/11	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	479	M2L	CD-CE	-2.63	1.43	1.51

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	479	M2L	CE-CD-SG	-3.13	103.76	112.73
1	E	679	M2L	CE-CD-SG	-2.59	105.30	112.73

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	479	M2L	N-CA-CB-SG
1	A	479	M2L	C-CA-CB-SG
1	A	479	M2L	CD-CE-NZ-CM1
1	A	479	M2L	CD-CE-NZ-CM2
1	E	679	M2L	CE-CD-SG-CB

There are no ring outliers.

2 monomers are involved in 20 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	E	679	M2L	12	0
1	A	479	M2L	8	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	97/135 (71%)	-0.44	0	100	100	26, 55, 102, 155	0
1	E	96/135 (71%)	-0.38	1 (1%)	82	73	18, 43, 92, 174	0
2	B	78/102 (76%)	-0.47	0	100	100	24, 49, 84, 137	0
2	F	84/102 (82%)	-0.38	0	100	100	22, 41, 80, 110	0
3	C	105/129 (81%)	-0.44	0	100	100	21, 43, 84, 167	0
3	G	104/129 (80%)	-0.45	0	100	100	23, 52, 92, 148	0
4	D	93/125 (74%)	-0.43	0	100	100	26, 50, 89, 130	0
4	H	92/125 (73%)	-0.35	0	100	100	24, 54, 91, 116	0
5	I	146/146 (100%)	-0.17	3 (2%)	63	49	54, 109, 171, 202	0
5	J	146/146 (100%)	-0.29	0	100	100	55, 104, 170, 202	0
All	All	1041/1274 (81%)	-0.37	4 (0%)	92	89	18, 59, 145, 202	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	56	DA	3.1
5	I	104	DT	2.2
5	I	1	DA	2.2
1	E	639	HIS	2.1

6.2 Non-standard residues in protein, DNA, RNA chains (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
1	M2L	A	479	11/12	0.92	0.16	67,120,120,120	0
1	M2L	E	679	11/12	0.92	0.25	70,135,135,135	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.