



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 05:38 pm GMT

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](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

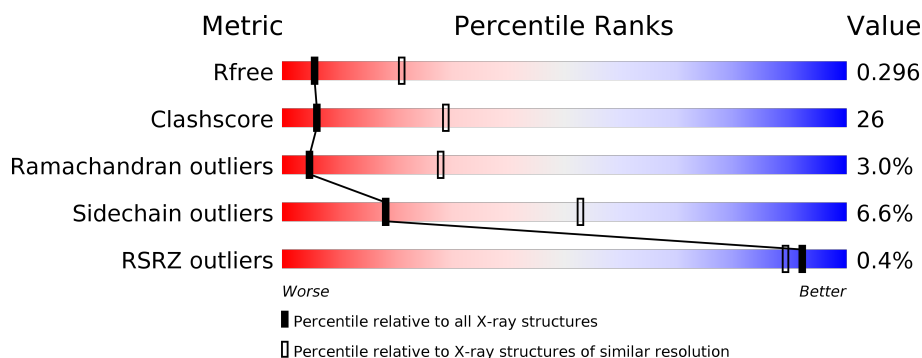
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.15 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))
$R_{free}$	111664	1389 (3.20-3.12)
Clashscore	122126	1522 (3.20-3.12)
Ramachandran outliers	120053	1493 (3.20-3.12)
Sidechain outliers	120020	1492 (3.20-3.12)
RSRZ outliers	108989	1344 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	135	
1	E	135	
2	B	102	
2	F	102	
3	C	129	
3	G	129	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	125	
4	H	125	
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
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
1	A	98	Total	C	N	O	S	0	0	0
			809	510	156	140	3			
1	E	97	Total	C	N	O	S	0	0	0
			803	507	155	138	3			

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
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
			673	424	133	115	1			

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

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

*Continued on next page...*

*Continued from previous page...*

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

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
4	D	93	Total	C	N	O	S	0	0	0
			728	459	131	136	2			
4	H	92	Total	C	N	O	S	0	0	0
			721	454	129	136	2			

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

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

- Molecule 6 is water.

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

*Continued on next page...*

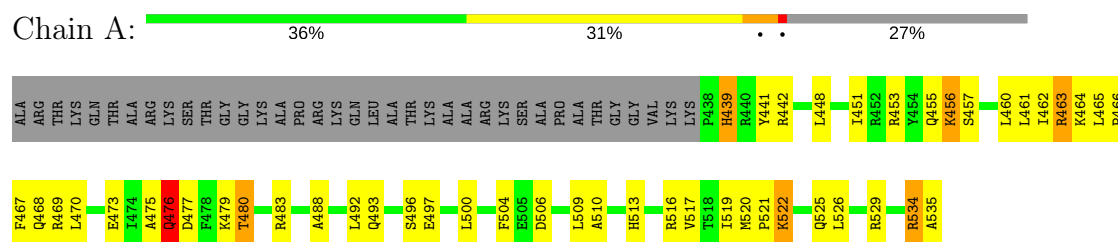
*Continued from previous page...*

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

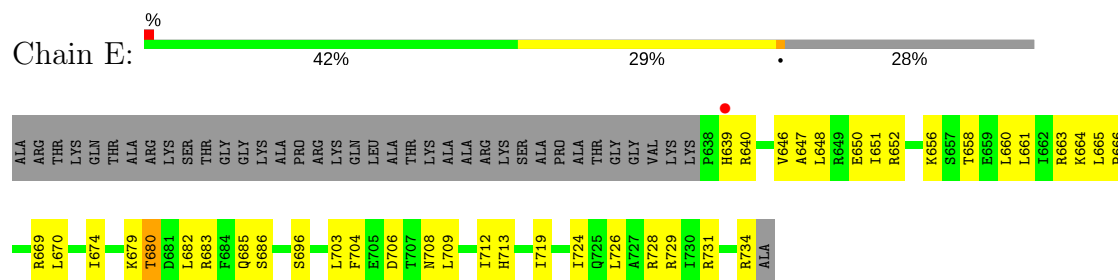
### 3 Residue-property plots [i](#)

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.

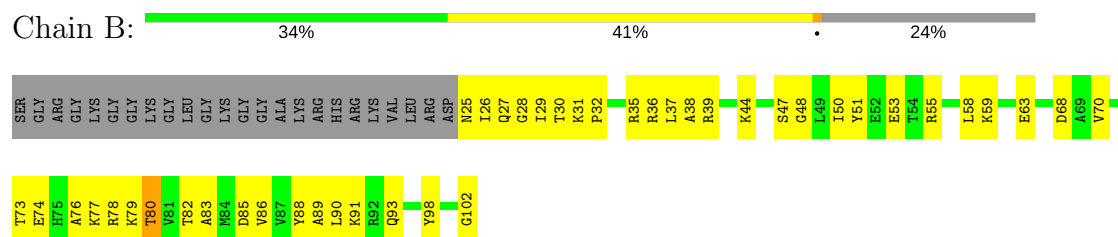
#### • Molecule 1: Histone H3-like



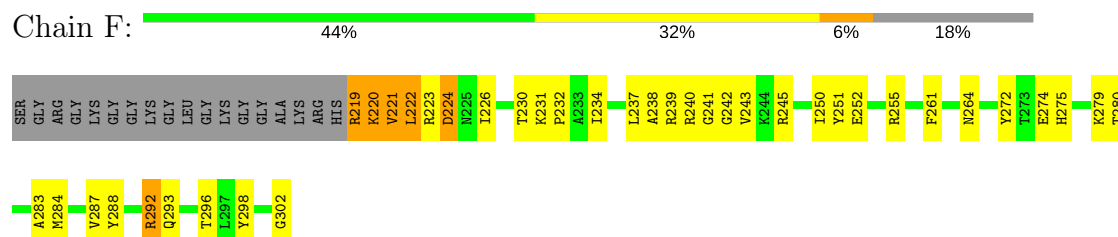
#### • Molecule 1: Histone H3-like



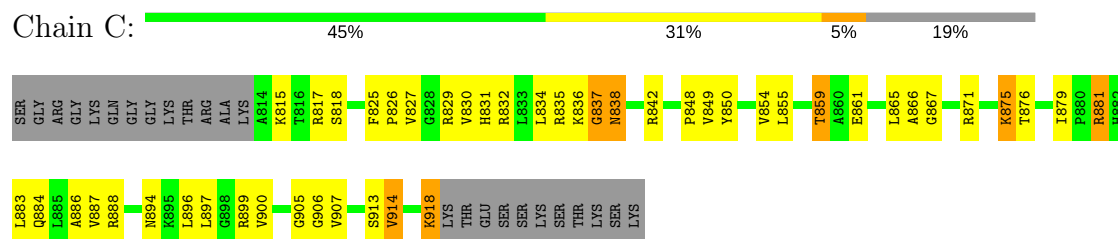
#### • Molecule 2: Histone H4



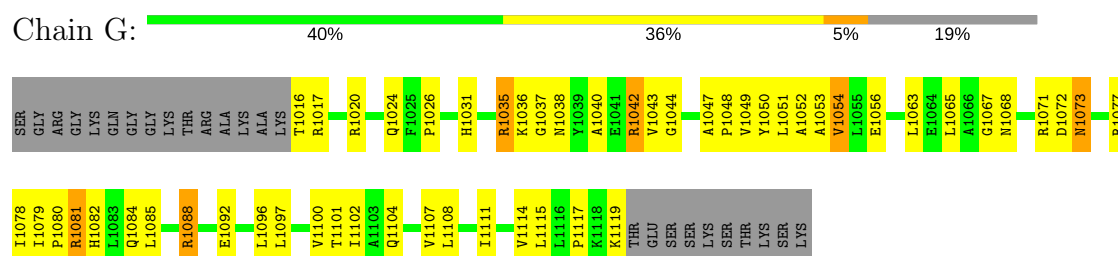
#### • Molecule 2: Histone H4



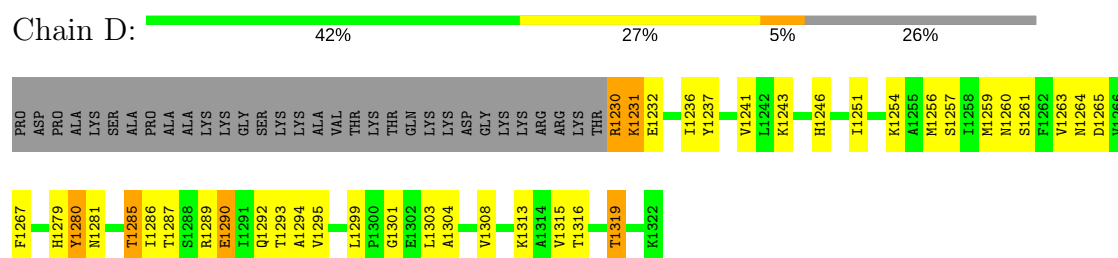
- Molecule 3: Histone H2A type 1



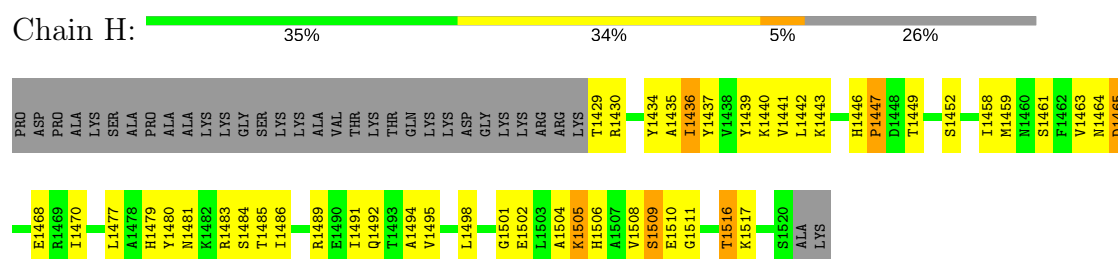
- Molecule 3: Histone H2A type 1



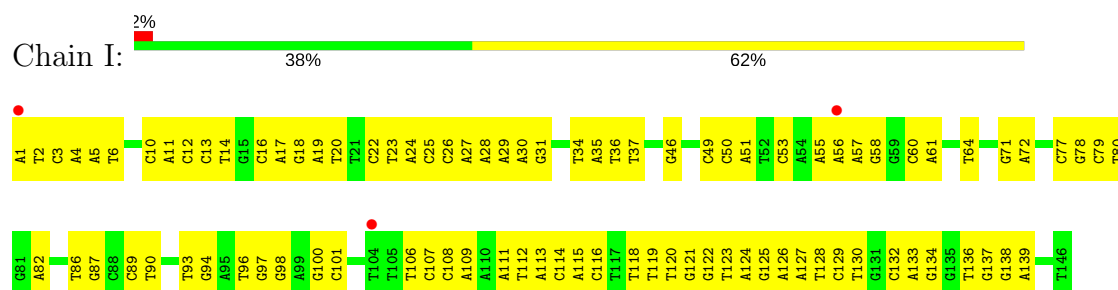
- Molecule 4: Histone 2, H2bf



- Molecule 4: Histone 2, H2bf

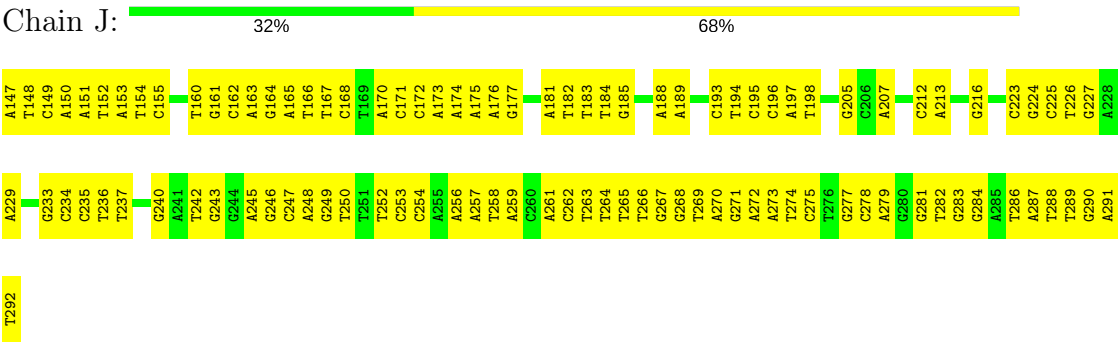


- Molecule 5: Palindromic 146bp Human Alpha satellite DNA





● Molecule 5: Palindromic 146bp Human Alpha satellite DNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	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
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.87 (at 3.13Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.220 , 0.290 0.228 , 0.296	Depositor DCC
$R_{free}$ test set	1706 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	75.0	Xtriage
Anisotropy	0.556	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 85.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.54$ , $\langle L^2 \rangle = 0.38$	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 <sup>2</sup> )	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.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: 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:HM1A	1:A:479:M2L:SG	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%)	16	54
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	11
3	C	103/129 (80%)	92 (89%)	6 (6%)	5 (5%)	2	17
3	G	102/129 (79%)	93 (91%)	7 (7%)	2 (2%)	8	39
4	D	91/125 (73%)	72 (79%)	15 (16%)	4 (4%)	3	18
4	H	90/125 (72%)	70 (78%)	15 (17%)	5 (6%)	2	14
All	All	733/982 (75%)	639 (87%)	72 (10%)	22 (3%)	5	28

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 ⓘ

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%)	7	28
1	E	83/108 (77%)	78 (94%)	5 (6%)	21	56
2	B	63/78 (81%)	61 (97%)	2 (3%)	42	74
2	F	69/78 (88%)	66 (96%)	3 (4%)	32	67
3	C	83/102 (81%)	76 (92%)	7 (8%)	12	41
3	G	83/102 (81%)	77 (93%)	6 (7%)	16	48

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	D	79/104 (76%)	74 (94%)	5 (6%)	20	55
4	H	79/104 (76%)	75 (95%)	4 (5%)	26	62
All	All	622/784 (79%)	581 (93%)	41 (7%)	18	53

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

Mol	Chain	Res	Type
4	D	1230	ARG
4	D	1319	THR
4	H	1436	ILE
4	D	1285	THR
4	D	1289	ARG

Some 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
4	H	1464	ASN
3	G	1031	HIS
2	B	93	GLN
3	G	1038	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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	A	479	1	10,10,11	1.30	2 (20%)	7,11,13	1.37	1 (14%)
1	M2L	E	679	1	10,10,11	0.93	0	7,11,13	1.45	2 (28%)

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	A	479	1	-	0/7/9/11	0/0/0/0
1	M2L	E	679	1	-	0/7/9/11	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	479	M2L	CB-SG	-2.87	1.73	1.81
1	A	479	M2L	CD-CE	-2.57	1.43	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	479	M2L	CE-CD-SG	-3.12	103.76	112.71
1	E	679	M2L	CE-CD-SG	-2.58	105.30	112.71
1	E	679	M2L	O-C-CA	-2.31	118.61	124.96

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 20 short contacts:

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates 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, 95<sup>th</sup> 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 <sup>2</sup> )	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 72	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 47	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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
1	M2L	E	679	11/12	0.92	0.25	70,135,135,135	0
1	M2L	A	479	11/12	0.92	0.16	67,120,120,120	0

### 6.3 Carbohydrates [i](#)

There are no carbohydrates 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.