



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 07:02 am BST

PDB ID : 5GSU  
Title : Crystal structure of nucleosome core particle consisting of human testis-specific histone variants, Th2A and Th2B  
Authors : Kumarevel, T.; Sivaraman, P.  
Deposited on : 2016-08-17  
Resolution : 3.10 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
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

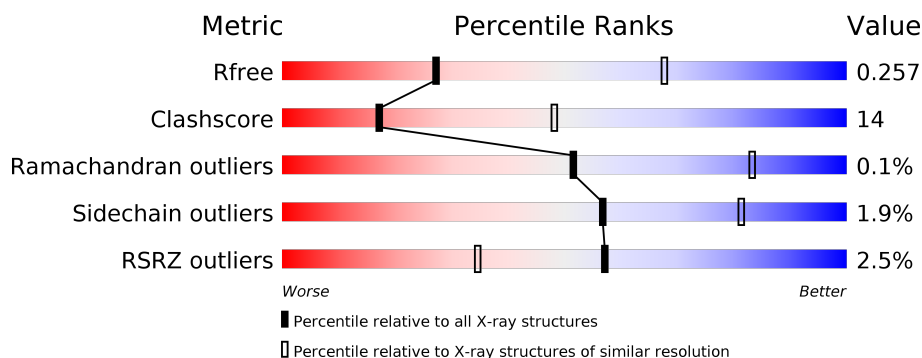
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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 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  $\leq 5\%$ . 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	<div> <div>%</div> <div> <div></div> <div>60%</div> <div>13%</div> <div>27%</div> </div> </div>
1	E	135	<div> <div>%</div> <div> <div></div> <div>57%</div> <div>16%</div> <div>27%</div> </div> </div>
2	B	102	<div> <div></div> <div> <div>60%</div> <div>17%</div> <div>23%</div> </div> </div>
2	F	102	<div> <div>%</div> <div> <div></div> <div>65%</div> <div>19%</div> <div>16%</div> </div> </div>
3	C	130	<div> <div></div> <div> <div>64%</div> <div>18%</div> <div>18%</div> </div> </div>
3	G	130	<div> <div></div> <div> <div>65%</div> <div>15%</div> <div>19%</div> </div> </div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	126	
4	H	126	
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	I	202	-	-	-	X

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 12086 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	99	Total	C	N	O	S	0	0	0
			816	514	158	140	4			
1	E	99	Total	C	N	O	S	0	0	0
			816	514	158	140	4			

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	79	Total	C	N	O	S	0	0	0
			624	394	121	108	1			
2	F	86	Total	C	N	O	S	0	0	0
			694	436	140	117	1			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	106	Total	C	N	O	0	0	0
			814	514	157	143			
3	G	105	Total	C	N	O	0	0	0
			805	508	155	142			

- Molecule 4 is a protein called Histone H2B type 1-A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	97	Total	C	N	O	S	0	0	0
			769	482	140	145	2			
4	H	94	Total	C	N	O	S	0	0	0
			738	464	130	142	2			

- Molecule 5 is a DNA chain called DNA (146-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	146	Total	C	N	O	P	0	0	0
			2990	1431	540	874	145			
5	J	146	Total	C	N	O	P	0	0	0
			2990	1431	540	874	145			

- 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	J	1	Total	Cl	0	0
			1	1		
6	I	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	8	Total	Mn	0	0
			8	8		
7	D	2	Total	Mn	0	0
			2	2		

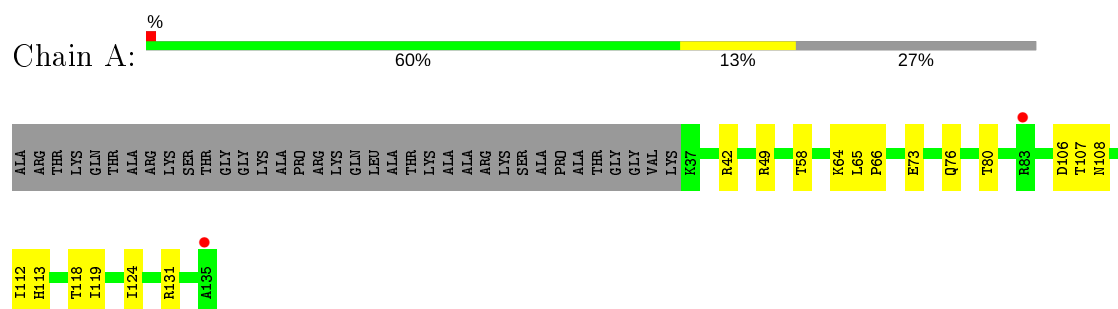
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	B	1	Total	O	0	0
			1	1		
8	F	2	Total	O	0	0
			2	2		
8	C	2	Total	O	0	0
			2	2		
8	G	1	Total	O	0	0
			1	1		
8	D	1	Total	O	0	0
			1	1		
8	I	3	Total	O	0	0
			3	3		
8	J	2	Total	O	0	0
			2	2		

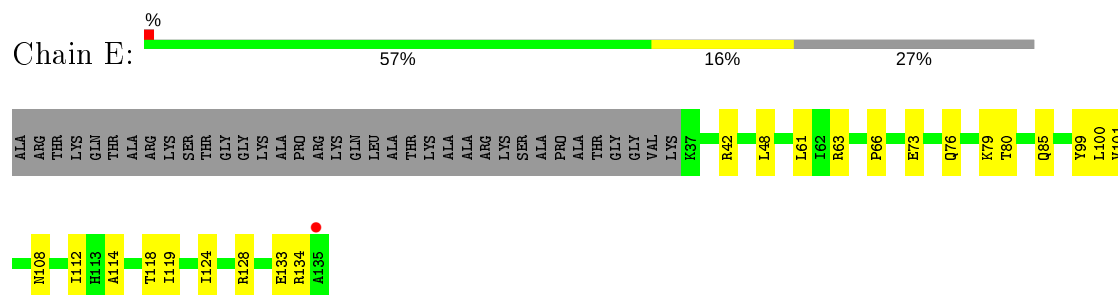
### 3 Residue-property plots [i](#)

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.

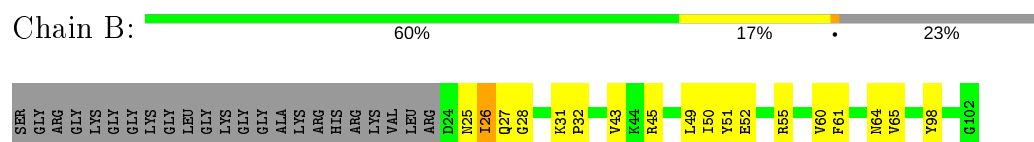
- Molecule 1: Histone H3.1



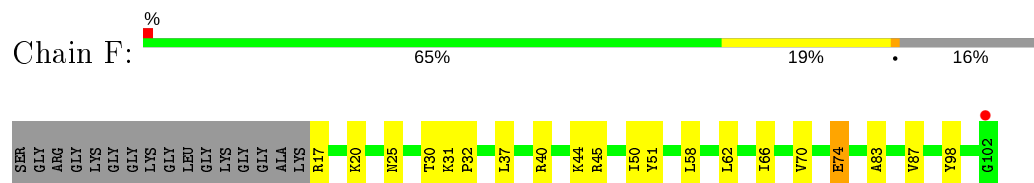
- Molecule 1: Histone H3.1



- Molecule 2: Histone H4

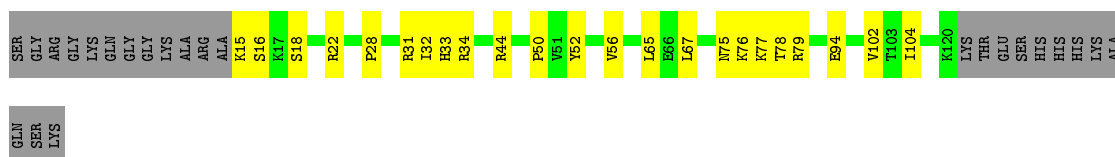


- Molecule 2: Histone H4



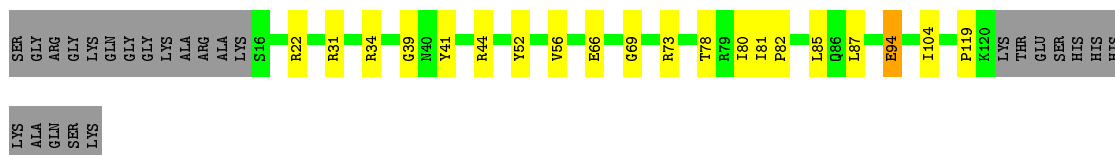
- Molecule 3: Histone H2A type 1-A





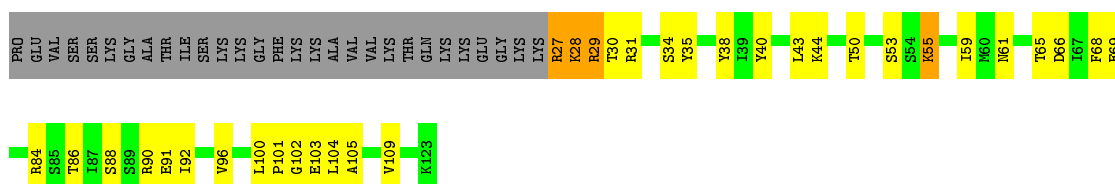
- Molecule 3: Histone H2A type 1-A

Chain G: 65% 15% 19%



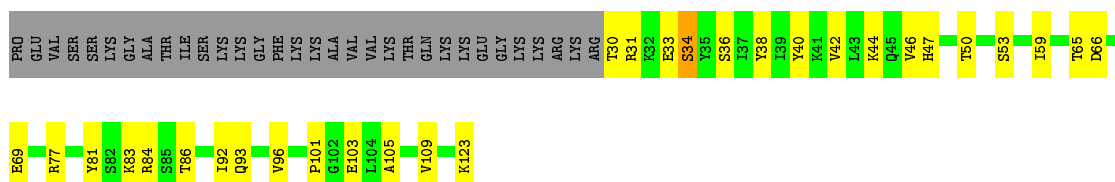
- Molecule 4: Histone H2B type 1-A

Chain D: 50% 24% 23%



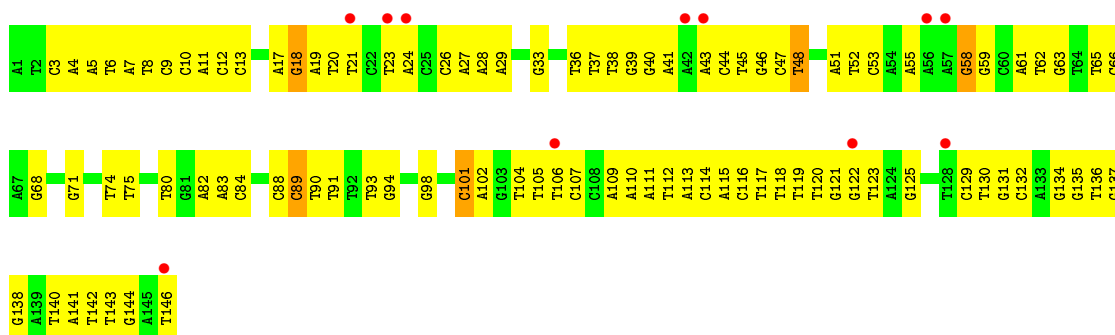
- Molecule 4: Histone H2B type 1-A

Chain H: 51% 23% 25%

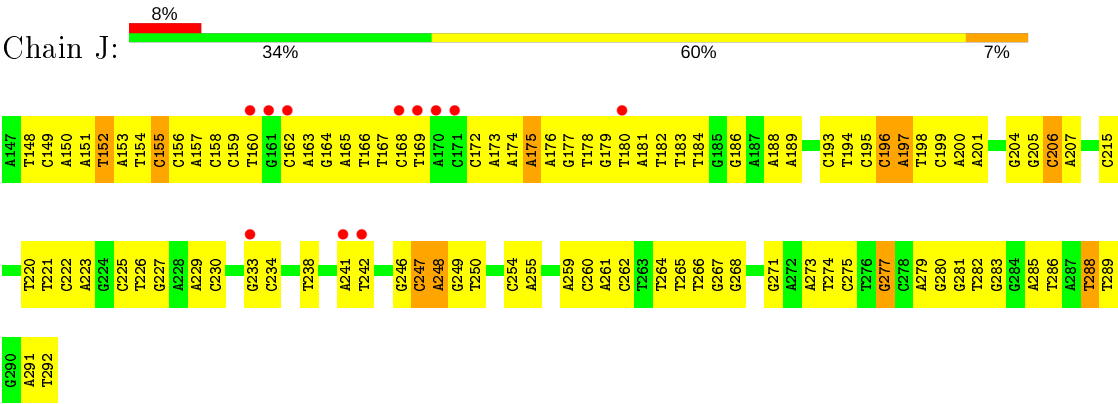


- Molecule 5: DNA (146-MER)

Chain I: 8% 33% 64%



● Molecule 5: DNA (146-MER)





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.09Å 109.74Å 182.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	35.33 – 3.10 37.50 – 3.10	Depositor EDS
% Data completeness (in resolution range)	96.6 (35.33-3.10) 96.7 (37.50-3.10)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.93 (at 3.12Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496, REFMAC	Depositor
R, $R_{free}$	0.197 , 0.257 0.200 , 0.257	Depositor DCC
$R_{free}$ test set	1939 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	79.9	Xtriage
Anisotropy	0.499	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 42.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.024 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	12086	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	94.0	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

### 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	0.53	0/828	0.71	0/1109
1	E	0.67	0/828	0.77	0/1109
2	B	0.62	0/631	0.75	0/844
2	F	0.70	0/702	0.80	0/937
3	C	0.63	0/824	0.78	0/1110
3	G	0.51	0/815	0.74	1/1099 (0.1%)
4	D	0.71	0/780	0.79	1/1044 (0.1%)
4	H	0.61	0/749	0.70	0/1005
5	I	0.90	7/3354 (0.2%)	1.08	9/5175 (0.2%)
5	J	0.90	9/3354 (0.3%)	1.07	7/5175 (0.1%)
All	All	0.78	16/12865 (0.1%)	0.95	18/18607 (0.1%)

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
4	D	0	1

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	J	204	DG	C3'-O3'	-7.87	1.33	1.44
5	J	223	DA	C3'-O3'	-7.01	1.34	1.44
5	J	196	DC	C3'-O3'	-6.80	1.35	1.44
5	I	28	DA	C3'-O3'	-6.75	1.35	1.44
5	I	48	DT	C3'-O3'	-6.51	1.35	1.44

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
5	J	155	DC	O5'-P-OP1	-10.99	95.81	105.70
5	I	58	DG	O4'-C1'-N9	8.16	113.71	108.00
5	J	206	DC	O5'-P-OP1	-7.60	98.86	105.70
5	I	80	DT	OP1-P-OP2	7.01	130.12	119.60
5	I	80	DT	O5'-P-OP2	-6.68	99.69	105.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	D	29	ARG	Sidechain

## 5.2 Too-close contacts

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	816	0	856	14	0
1	E	816	0	856	17	0
2	B	624	0	661	19	0
2	F	694	0	742	16	0
3	C	814	0	871	31	0
3	G	805	0	858	25	0
4	D	769	0	803	39	0
4	H	738	0	764	31	0
5	I	2990	0	1652	83	0
5	J	2990	0	1652	89	0
6	G	1	0	0	0	0
6	I	1	0	0	0	0
6	J	1	0	0	1	0
7	D	2	0	0	0	0
7	I	8	0	0	0	0
7	J	5	0	0	0	0
8	B	1	0	0	0	0
8	C	2	0	0	0	0
8	D	1	0	0	0	0
8	F	2	0	0	0	0
8	G	1	0	0	0	0
8	I	3	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	J	2	0	0	0	0
All	All	12086	0	9715	291	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 14.

The worst 5 of 291 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)
3:C:44:ARG:HB2	4:D:86:THR:HG22	1.50	0.93
5:I:43:DA:H2''	5:I:44:DC:H5''	1.52	0.88
2:B:26:ILE:HG22	2:B:27:GLN:N	1.88	0.88
3:C:79:ARG:HH11	5:I:19:DA:H4'	1.39	0.87
3:C:33:HIS:HD2	3:C:50:PRO:HG3	1.39	0.87

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	97/135 (72%)	97 (100%)	0	0	100	100
1	E	97/135 (72%)	96 (99%)	1 (1%)	0	100	100
2	B	77/102 (76%)	73 (95%)	4 (5%)	0	100	100
2	F	84/102 (82%)	82 (98%)	2 (2%)	0	100	100
3	C	104/130 (80%)	101 (97%)	3 (3%)	0	100	100
3	G	103/130 (79%)	98 (95%)	4 (4%)	1 (1%)	15	49
4	D	95/126 (75%)	89 (94%)	6 (6%)	0	100	100
4	H	92/126 (73%)	86 (94%)	6 (6%)	0	100	100
All	All	749/986 (76%)	722 (96%)	26 (4%)	1 (0%)	51	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	G	119	PRO

### 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	86/110 (78%)	85 (99%)	1 (1%)	71	88
1	E	86/110 (78%)	85 (99%)	1 (1%)	71	88
2	B	63/78 (81%)	62 (98%)	1 (2%)	62	84
2	F	71/78 (91%)	70 (99%)	1 (1%)	67	86
3	C	84/101 (83%)	84 (100%)	0	100	100
3	G	83/101 (82%)	83 (100%)	0	100	100
4	D	86/110 (78%)	81 (94%)	5 (6%)	20	51
4	H	83/110 (76%)	80 (96%)	3 (4%)	35	67
All	All	642/798 (80%)	630 (98%)	12 (2%)	57	81

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	28	LYS
4	D	53	SER
4	H	34	SER
4	D	27	ARG
4	D	104	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	108	ASN
3	C	33	HIS
4	D	61	ASN
4	H	45	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](#)

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 18 ligands modelled in this entry, 18 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 ⓘ

### 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, 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	99/135 (73%)	0.03	2 (2%) 65 44	50, 65, 103, 132	0
1	E	99/135 (73%)	-0.17	1 (1%) 82 67	37, 53, 85, 118	0
2	B	79/102 (77%)	-0.11	0 100 100	49, 63, 82, 107	0
2	F	86/102 (84%)	-0.08	1 (1%) 79 61	39, 51, 74, 121	0
3	C	106/130 (81%)	-0.21	0 100 100	41, 57, 82, 114	0
3	G	105/130 (80%)	-0.20	0 100 100	53, 67, 95, 116	0
4	D	97/126 (76%)	-0.08	0 100 100	43, 59, 101, 139	0
4	H	94/126 (74%)	-0.12	0 100 100	48, 70, 101, 120	0
5	I	146/146 (100%)	0.44	11 (7%) 14 5	74, 126, 162, 171	0
5	J	146/146 (100%)	0.43	11 (7%) 14 5	73, 125, 157, 163	0
All	All	1057/1278 (82%)	0.03	26 (2%) 57 34	37, 68, 145, 171	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	146	DT	12.7
5	I	24	DA	4.4
5	J	169	DT	3.5
5	J	162	DC	3.5
5	J	161	DG	3.4

### 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 ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

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(Å <sup>2</sup> )	Q<0.9
7	MN	I	201	1/1	0.53	0.26	153,153,153,153	0
7	MN	J	302	1/1	0.55	0.32	211,211,211,211	0
7	MN	J	305	1/1	0.56	0.12	155,155,155,155	0
7	MN	I	202	1/1	0.65	0.47	227,227,227,227	0
7	MN	I	207	1/1	0.69	0.15	138,138,138,138	0
7	MN	J	304	1/1	0.86	0.12	98,98,98,98	0
6	CL	J	306	1/1	0.88	0.17	96,96,96,96	0
7	MN	I	205	1/1	0.90	0.13	96,96,96,96	0
7	MN	J	303	1/1	0.91	0.15	100,100,100,100	0
7	MN	I	206	1/1	0.92	0.18	128,128,128,128	0
7	MN	I	208	1/1	0.93	0.46	141,141,141,141	0
7	MN	I	203	1/1	0.93	0.07	159,159,159,159	0
7	MN	I	204	1/1	0.95	0.09	116,116,116,116	0
6	CL	I	209	1/1	0.96	0.09	98,98,98,98	0
6	CL	G	201	1/1	0.96	0.18	79,79,79,79	0
7	MN	J	301	1/1	0.96	0.24	132,132,132,132	0
7	MN	D	202	1/1	0.99	0.15	77,77,77,77	0
7	MN	D	201	1/1	0.99	0.17	69,69,69,69	0

### 6.5 Other polymers ⓘ

There are no such residues in this entry.