



wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 01:25 pm BST

PDB ID : 5Z23
Title : Crystal structure of the nucleosome containing a chimeric histone H3/CENP-A CATD
Authors : Arimura, Y.; Tachiwana, H.; Takagi, H.
Deposited on : 2017-12-28
Resolution : 2.73 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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	:	2.11
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

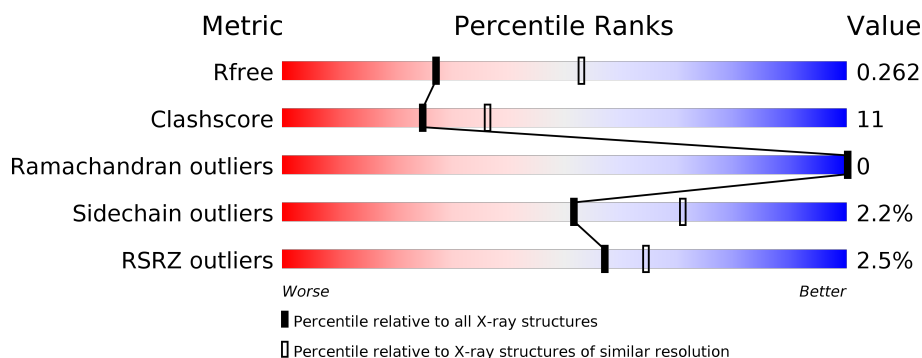
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.73 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	1271 (2.76-2.72)
Clashscore	141614	1322 (2.76-2.72)
Ramachandran outliers	138981	1297 (2.76-2.72)
Sidechain outliers	138945	1298 (2.76-2.72)
RSRZ outliers	127900	1243 (2.76-2.72)

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 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	141	<div> <div> <div></div> <div>59%</div> <div>11%</div> <div>•</div> <div>30%</div> </div> </div>
1	E	141	<div> <div> <div></div> <div>54%</div> <div>15%</div> <div>••</div> <div>29%</div> </div> </div>
2	B	106	<div> <div> <div></div> <div>58%</div> <div>14%</div> <div>•</div> <div>26%</div> </div> </div>
2	F	106	<div> <div> <div></div> <div>69%</div> <div>10%</div> <div></div> <div>21%</div> </div> </div>
3	C	167	<div> <div> <div>4%</div> <div>51%</div> <div>11%</div> <div>•</div> <div>37%</div> </div> </div>
3	G	167	<div> <div> <div>2%</div> <div>49%</div> <div>13%</div> <div>•</div> <div>38%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	129	<div><div><div>%</div><div><div></div><div></div><div></div></div><div>56%16%29%</div></div></div>
4	H	129	<div><div><div></div><div></div><div></div></div><div>57%13%29%</div></div>
5	I	146	<div><div><div>4%</div><div><div></div><div></div><div></div></div><div>48%51%</div></div></div>
5	J	146	<div><div><div>3%</div><div><div></div><div></div><div></div></div><div>52%47%</div></div></div>

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 12007 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,Histone H3-like centromeric protein A,Histone H3.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	99	Total	C	N	O	S	0	0	0
			830	533	160	135	2			
1	E	100	Total	C	N	O	S	0	0	0
			836	536	161	137	2			

There are 6 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
E	-3	GLY	-	expression tag	UNP P68431
E	-2	SER	-	expression tag	UNP P68431
E	-1	HIS	-	expression tag	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
			629	397	123	108	1			
2	F	84	Total	C	N	O	S	0	0	0
			673	424	133	115	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
F	-2	SER	-	expression tag	UNP P62805

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
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	Se	0	0	0
			815	511	159	142	3			
3	G	104	Total	C	N	O	Se	0	0	0
			805	505	157	140	3			

There are 80 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
C	51	MSE	LEU	engineered mutation	UNP P04908
C	58	MSE	LEU	engineered mutation	UNP P04908
C	93	MSE	LEU	engineered mutation	UNP P04908
C	130	GLN	-	expression tag	UNP P04908
C	131	LEU	-	expression tag	UNP P04908
C	132	ALA	-	expression tag	UNP P04908
C	133	ILE	-	expression tag	UNP P04908
C	134	ARG	-	expression tag	UNP P04908
C	135	ASN	-	expression tag	UNP P04908
C	136	ASP	-	expression tag	UNP P04908
C	137	GLU	-	expression tag	UNP P04908
C	138	GLU	-	expression tag	UNP P04908
C	139	MSE	-	expression tag	UNP P04908
C	140	ASN	-	expression tag	UNP P04908
C	141	LYS	-	expression tag	UNP P04908
C	142	LEU	-	expression tag	UNP P04908
C	143	LEU	-	expression tag	UNP P04908
C	144	GLY	-	expression tag	UNP P04908
C	145	ARG	-	expression tag	UNP P04908
C	146	VAL	-	expression tag	UNP P04908
C	147	THR	-	expression tag	UNP P04908
C	148	ILE	-	expression tag	UNP P04908
C	149	ALA	-	expression tag	UNP P04908
C	150	GLN	-	expression tag	UNP P04908
C	151	GLY	-	expression tag	UNP P04908
C	152	GLY	-	expression tag	UNP P04908

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	153	VAL	-	expression tag	UNP P04908
C	154	LEU	-	expression tag	UNP P04908
C	155	PRO	-	expression tag	UNP P04908
C	156	ASN	-	expression tag	UNP P04908
C	157	ILE	-	expression tag	UNP P04908
C	158	GLN	-	expression tag	UNP P04908
C	159	ALA	-	expression tag	UNP P04908
C	160	VAL	-	expression tag	UNP P04908
C	161	LEU	-	expression tag	UNP P04908
C	162	LEU	-	expression tag	UNP P04908
C	163	PRO	-	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
G	51	MSE	LEU	engineered mutation	UNP P04908
G	58	MSE	LEU	engineered mutation	UNP P04908
G	93	MSE	LEU	engineered mutation	UNP P04908
G	130	GLN	-	expression tag	UNP P04908
G	131	LEU	-	expression tag	UNP P04908
G	132	ALA	-	expression tag	UNP P04908
G	133	ILE	-	expression tag	UNP P04908
G	134	ARG	-	expression tag	UNP P04908
G	135	ASN	-	expression tag	UNP P04908
G	136	ASP	-	expression tag	UNP P04908
G	137	GLU	-	expression tag	UNP P04908
G	138	GLU	-	expression tag	UNP P04908
G	139	MSE	-	expression tag	UNP P04908
G	140	ASN	-	expression tag	UNP P04908
G	141	LYS	-	expression tag	UNP P04908
G	142	LEU	-	expression tag	UNP P04908
G	143	LEU	-	expression tag	UNP P04908
G	144	GLY	-	expression tag	UNP P04908
G	145	ARG	-	expression tag	UNP P04908
G	146	VAL	-	expression tag	UNP P04908
G	147	THR	-	expression tag	UNP P04908
G	148	ILE	-	expression tag	UNP P04908
G	149	ALA	-	expression tag	UNP P04908
G	150	GLN	-	expression tag	UNP P04908
G	151	GLY	-	expression tag	UNP P04908
G	152	GLY	-	expression tag	UNP P04908
G	153	VAL	-	expression tag	UNP P04908
G	154	LEU	-	expression tag	UNP P04908

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	155	PRO	-	expression tag	UNP P04908
G	156	ASN	-	expression tag	UNP P04908
G	157	ILE	-	expression tag	UNP P04908
G	158	GLN	-	expression tag	UNP P04908
G	159	ALA	-	expression tag	UNP P04908
G	160	VAL	-	expression tag	UNP P04908
G	161	LEU	-	expression tag	UNP P04908
G	162	LEU	-	expression tag	UNP P04908
G	163	PRO	-	expression tag	UNP P04908

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	92	Total	C	N	O	Se	0	0	0
			720	453	129	136	2			
4	H	92	Total	C	N	O	Se	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

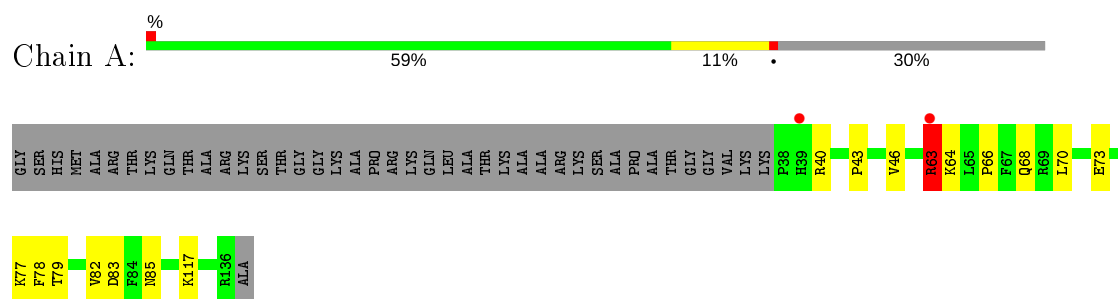
- Molecule 5 is a DNA chain called DNA (146-MER).

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			

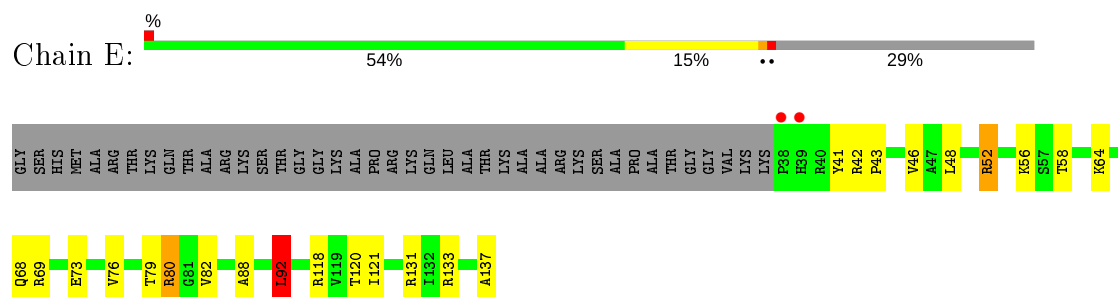
3 Residue-property plots [i](#)

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.

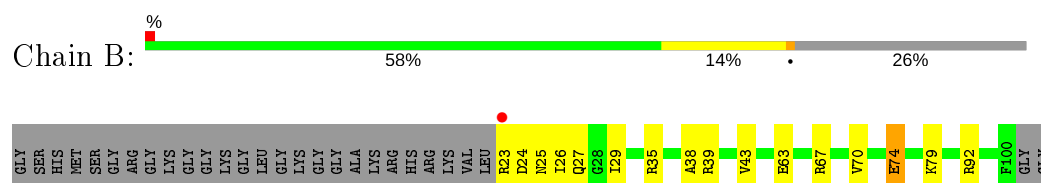
- Molecule 1: Histone H3.1,Histone H3-like centromeric protein A,Histone H3.1



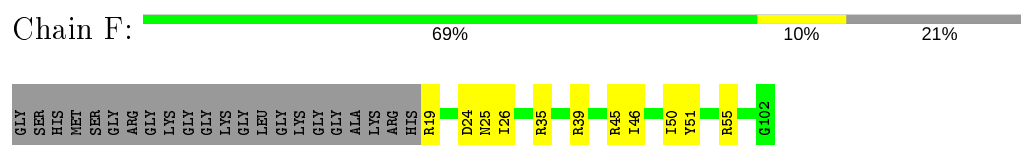
- Molecule 1: Histone H3.1,Histone H3-like centromeric protein A,Histone H3.1



- Molecule 2: Histone H4



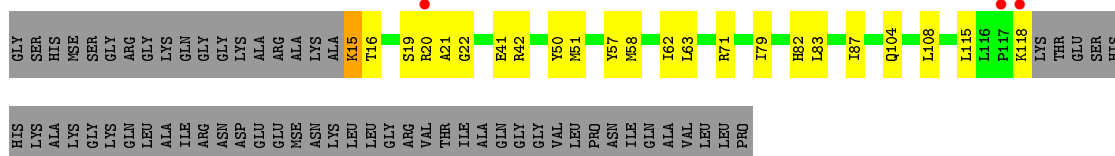
- Molecule 2: Histone H4



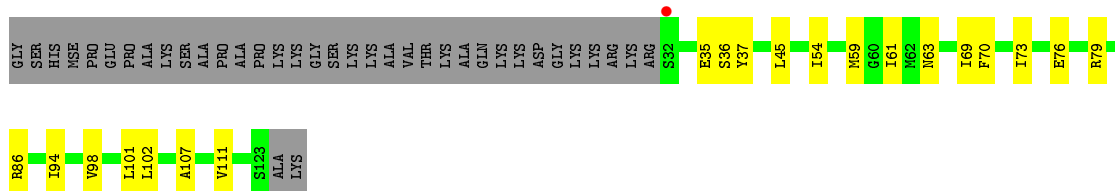
- Molecule 3: Histone H2A type 1-B/E



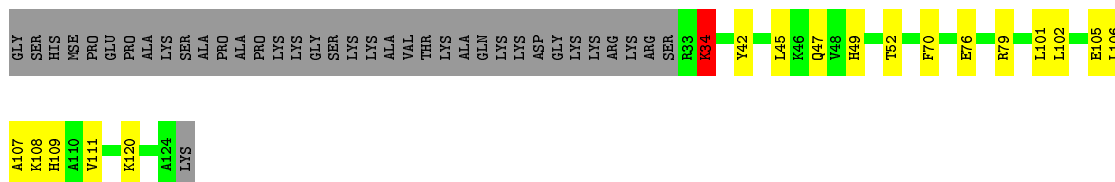
- Molecule 3: Histone H2A type 1-B/E



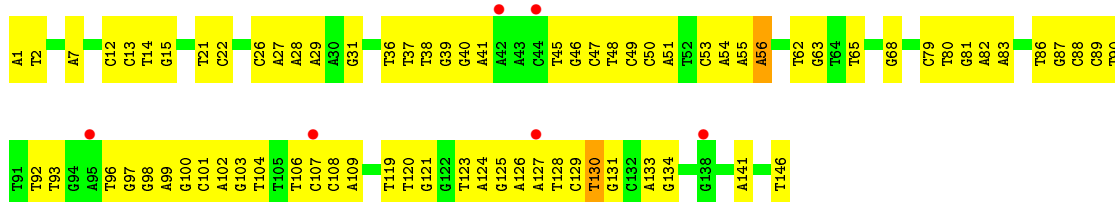
- Molecule 4: Histone H2B type 1-J



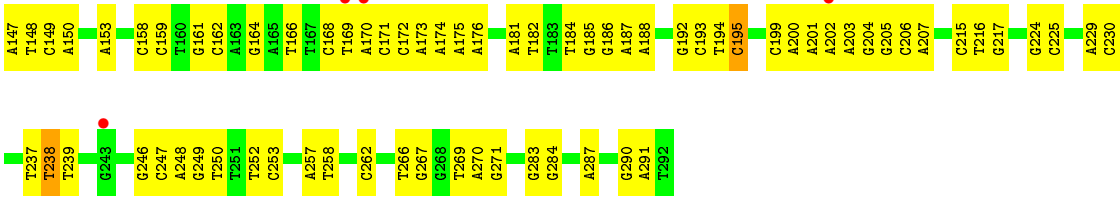
- Molecule 4: Histone H2B type 1-J



- Molecule 5: DNA (146-MER)



- 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, α , β , γ	99.64Å 100.75Å 173.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.88 – 2.73 47.88 – 2.73	Depositor EDS
% Data completeness (in resolution range)	96.1 (47.88-2.73) 96.1 (47.88-2.73)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.47 (at 2.73Å)	Xtriage
Refinement program	PHENIX 1.12 _2829	Depositor
R, R_{free}	0.203 , 0.261 0.203 , 0.262	Depositor DCC
R_{free} test set	1985 reflections (4.38%)	wwPDB-VP
Wilson B-factor (Å ²)	65.3	Xtriage
Anisotropy	0.329	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 53.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.022 for k,h,-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	12007	wwPDB-VP
Average B, all atoms (Å ²)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.07% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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	0.54	0/846	0.75	1/1138 (0.1%)
1	E	0.53	0/852	0.85	4/1145 (0.3%)
2	B	0.59	0/636	0.83	1/852 (0.1%)
2	F	0.55	0/680	0.72	0/908
3	C	0.52	0/822	0.69	1/1102 (0.1%)
3	G	0.54	0/812	0.76	2/1088 (0.2%)
4	D	0.51	0/729	0.67	0/977
4	H	0.48	0/728	0.69	1/976 (0.1%)
5	I	0.86	2/3354 (0.1%)	1.04	4/5175 (0.1%)
5	J	0.88	3/3354 (0.1%)	1.03	2/5175 (0.0%)
All	All	0.73	5/12813 (0.0%)	0.92	16/18536 (0.1%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	J	225	DC	C3'-O3'	-6.00	1.36	1.44
5	J	195	DC	C3'-O3'	-5.69	1.36	1.44
5	J	224	DG	C3'-O3'	-5.67	1.36	1.44
5	I	65	DT	C1'-N1	5.62	1.56	1.49
5	I	130	DT	C3'-O3'	-5.35	1.36	1.44

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	80	ARG	NE-CZ-NH1	-8.74	115.93	120.30
1	E	80	ARG	CB-CG-CD	8.67	134.15	111.60
1	E	92	LEU	CA-CB-CG	-8.10	96.68	115.30
3	G	41	GLU	CA-CB-CG	-7.48	96.95	113.40
1	A	63	ARG	NE-CZ-NH2	-7.08	116.76	120.30

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	830	0	873	23	0
1	E	836	0	878	25	0
2	B	629	0	670	13	0
2	F	673	0	722	11	0
3	C	815	0	865	25	0
3	G	805	0	855	29	0
4	D	720	0	740	18	0
4	H	719	0	740	17	0
5	I	2990	0	1652	65	0
5	J	2990	0	1652	73	0
All	All	12007	0	9647	231	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 11.

The worst 5 of 231 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:83:LEU:HD13	4:D:61:ILE:HD11	1.34	1.05
1:A:63:ARG:HH12	5:J:238:DT:C5'	1.75	0.99
1:A:63:ARG:NH1	5:J:238:DT:OP1	1.97	0.97
5:I:36:DT:H2''	5:I:37:DT:H5''	1.48	0.96
1:A:63:ARG:HH12	5:J:238:DT:H5''	1.25	0.96

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	97/141 (69%)	97 (100%)	0	0	100	100
1	E	98/141 (70%)	97 (99%)	1 (1%)	0	100	100
2	B	76/106 (72%)	76 (100%)	0	0	100	100
2	F	82/106 (77%)	81 (99%)	1 (1%)	0	100	100
3	C	104/167 (62%)	101 (97%)	3 (3%)	0	100	100
3	G	102/167 (61%)	99 (97%)	3 (3%)	0	100	100
4	D	90/129 (70%)	90 (100%)	0	0	100	100
4	H	90/129 (70%)	88 (98%)	2 (2%)	0	100	100
All	All	739/1086 (68%)	729 (99%)	10 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	86/114 (75%)	84 (98%)	2 (2%)	50	70
1	E	86/114 (75%)	83 (96%)	3 (4%)	36	57
2	B	65/81 (80%)	63 (97%)	2 (3%)	40	60
2	F	69/81 (85%)	68 (99%)	1 (1%)	67	80
3	C	83/125 (66%)	82 (99%)	1 (1%)	71	83
3	G	83/125 (66%)	81 (98%)	2 (2%)	49	68
4	D	79/104 (76%)	79 (100%)	0	100	100
4	H	78/104 (75%)	75 (96%)	3 (4%)	33	54
All	All	629/848 (74%)	615 (98%)	14 (2%)	52	71

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	92	LEU
1	E	133	ARG
4	H	34	LYS
1	E	56	LYS
3	G	20	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
4	D	63	ASN
4	H	109	HIS

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](#)

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 ⓘ

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	99/141 (70%)	0.10	2 (2%) 65 72	35, 50, 93, 105	0
1	E	100/141 (70%)	0.11	2 (2%) 65 72	33, 52, 99, 124	0
2	B	78/106 (73%)	0.25	1 (1%) 77 82	38, 50, 79, 105	0
2	F	84/106 (79%)	0.05	0 100 100	36, 50, 85, 107	0
3	C	103/167 (61%)	0.12	7 (6%) 17 19	37, 53, 101, 116	0
3	G	101/167 (60%)	0.24	3 (2%) 50 57	37, 61, 89, 114	0
4	D	90/129 (69%)	-0.03	1 (1%) 80 85	37, 52, 78, 115	0
4	H	90/129 (69%)	0.17	0 100 100	39, 58, 87, 124	0
5	I	146/146 (100%)	0.06	6 (4%) 37 41	58, 109, 145, 157	0
5	J	146/146 (100%)	0.09	4 (2%) 54 61	52, 107, 150, 158	0
All	All	1037/1378 (75%)	0.11	26 (2%) 57 64	33, 61, 134, 158	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	116	LEU	6.1
1	E	38	PRO	5.5
3	C	117	PRO	4.4
3	C	12	ALA	4.3
1	A	39	HIS	3.3

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.