



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:30 AM GMT

PDB ID : 3AZJ
Title : Crystal Structure of Human Nucleosome Core Particle Containing H4K44Q mutation
Authors : Iwasaki, W.; Tachiwana, H.; Kawaguchi, K.; Shibata, T.; Kagawa, W.; Kurumizaka, H.
Deposited on : 2011-05-25
Resolution : 2.89 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
<http://wwpdb.org/validation/2016/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 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

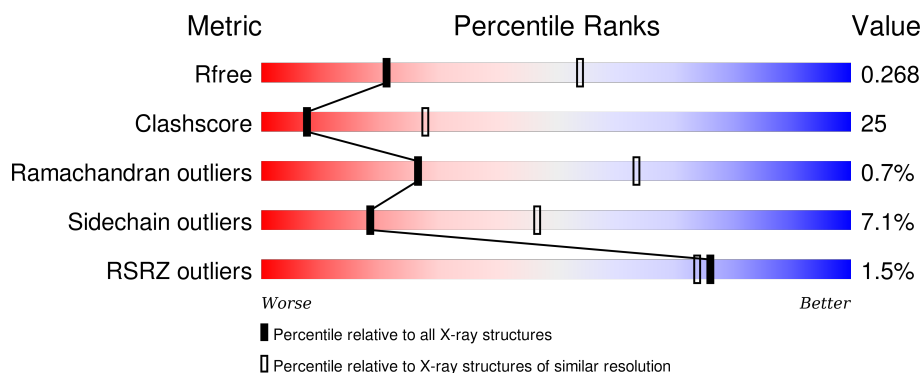
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.89 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}	91344	1451 (2.90-2.90)
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

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. 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	139	<div> <div></div> <div>46% 22% • 30%</div> </div>
1	E	139	<div> <div>2%</div> <div>46% 23% • 29%</div> </div>
2	B	106	<div> <div></div> <div>51% 24% 25%</div> </div>
2	F	106	<div> <div></div> <div>50% 27% • 22%</div> </div>
3	C	133	<div> <div>2%</div> <div>47% 23% 5% 25%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	133	
4	D	129	
4	H	129	
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	D	202	-	-	-	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 11860 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	97	Total	C	N	O	S	0	0	0
			801	505	155	137	4			
1	E	98	Total	C	N	O	S	0	0	0
			810	511	157	138	4			

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	79	Total	C	N	O	S	0	0	0
			627	394	121	111	1			
2	F	83	Total	C	N	O	S	0	0	0
			668	421	132	114	1			

There are 8 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
B	44	GLN	LYS	ENGINEERED MUTATION	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
F	44	GLN	LYS	ENGINEERED MUTATION	UNP P62805

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	100	Total	C	N	O	0	0	0
			774	484	155	135			
3	G	96	Total	C	N	O	0	0	0
			744	466	147	131			

There are 6 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
G	-3	GLY	-	EXPRESSION TAG	UNP P04908
G	-2	SER	-	EXPRESSION TAG	UNP P04908
G	-1	HIS	-	EXPRESSION TAG	UNP P04908

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	96	Total	C	N	O	S	0	0	0
			756	474	140	140	2			
4	H	93	Total	C	N	O	S	0	0	0
			725	456	130	137	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 146-MER DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	145	Total	C	N	O	P	0	0	0
			2970	1421	538	867	144			
5	J	145	Total	C	N	O	P	0	0	0
			2969	1421	535	869	144			

- 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	A	1	Total	Cl	0	0
			1	1		
6	D	1	Total	Cl	0	0
			1	1		
6	E	1	Total	Cl	0	0
			1	1		

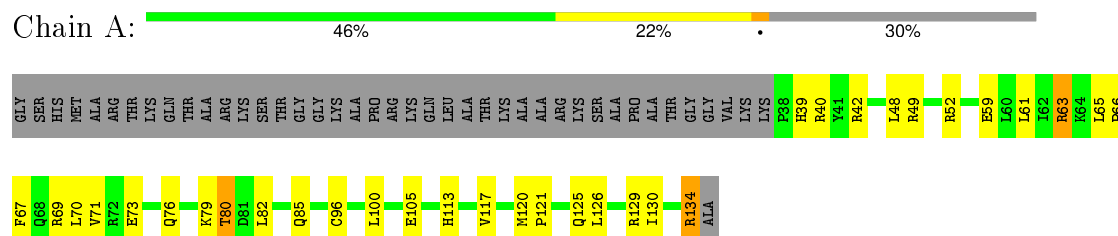
- Molecule 7 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	5	Total	Mn	0	0
			5	5		
7	I	6	Total	Mn	0	0
			6	6		
7	D	1	Total	Mn	0	0
			1	1		

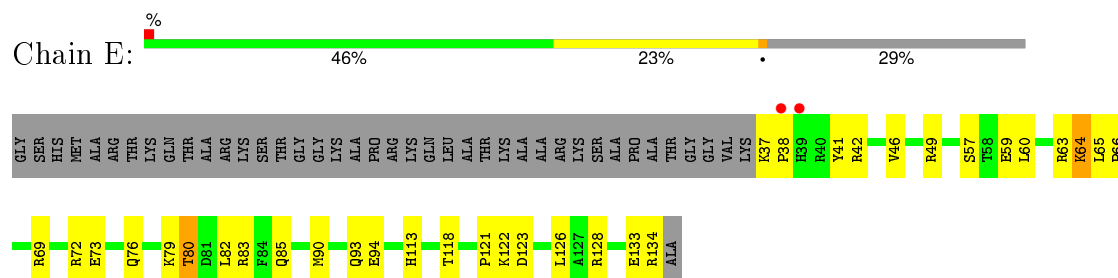
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

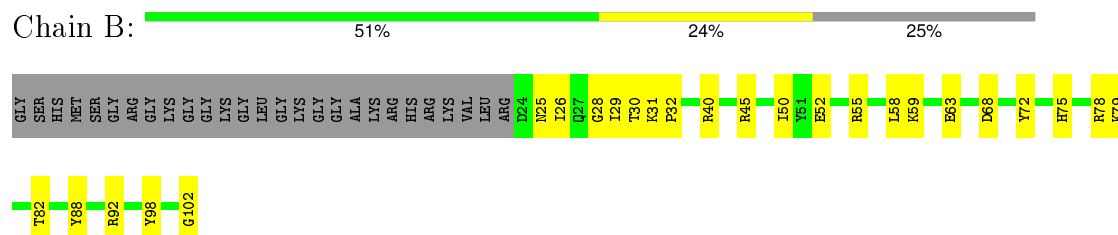
• Molecule 1: Histone H3.1



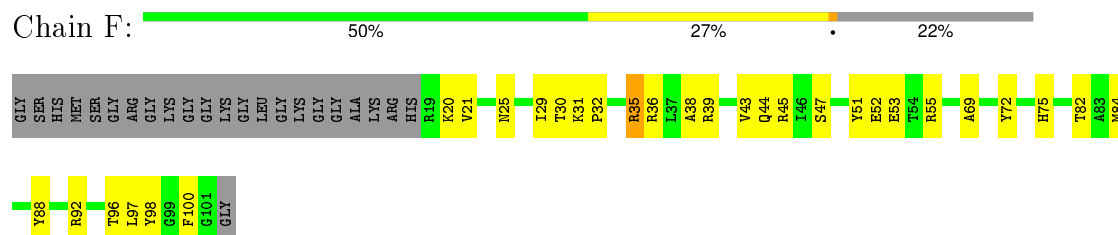
• Molecule 1: Histone H3.1



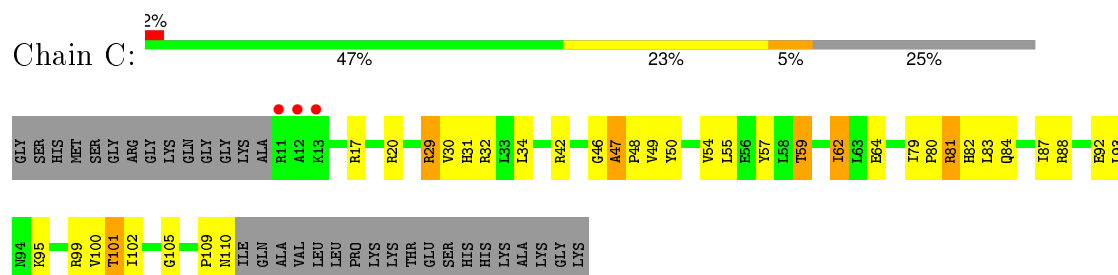
• Molecule 2: Histone H4



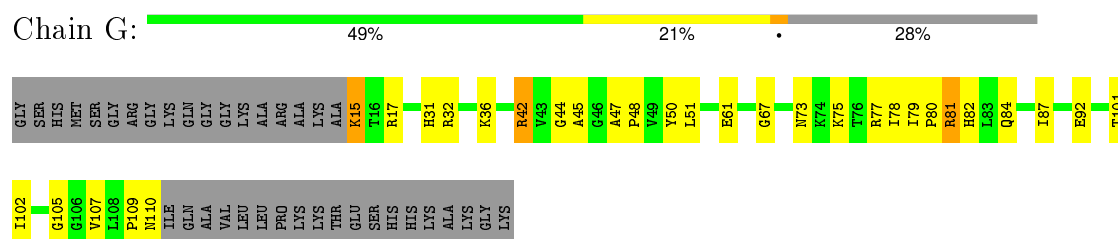
• Molecule 2: Histone H4



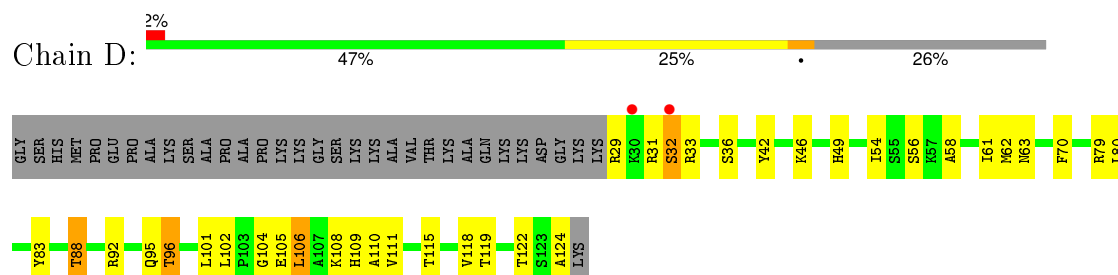
- Molecule 3: Histone H2A type 1-B/E



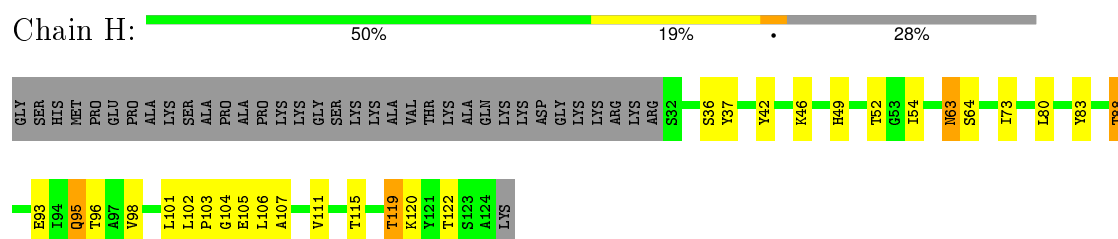
- Molecule 3: Histone H2A type 1-B/E



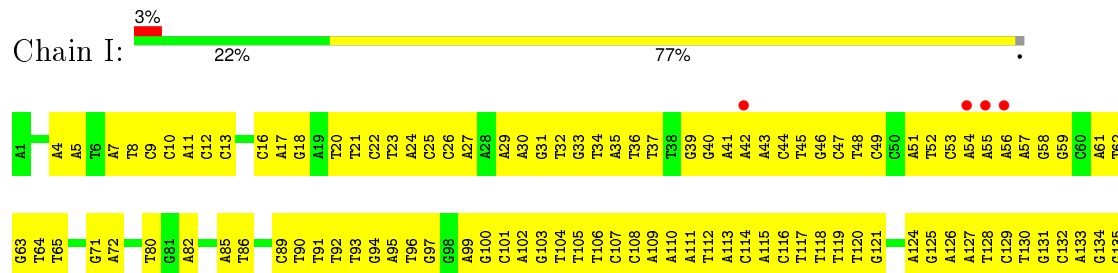
- Molecule 4: Histone H2B type 1-J



- Molecule 4: Histone H2B type 1-J



- Molecule 5: 146-MER DNA



T136
G137
G138
A139
T140
A141
T142
T143
G144
A145
DT

● Molecule 5: 146-MER DNA



DA	T148	C149	A150	A151	T152	A153	T154	C155	C158	C159	T160	G161	C162	A163	G164	A165	T166	T167	C168	T169	A170	C171	C172	A173	A174	A175	A176	G177	T178	G179	T180	A181	T182	T183	T184	G185	G186	A187	A188	A189	C190	T191	G192	C193	T194	C195	C196	A197	T198	C199	A200	A201	A202	A203	G204	G205	C206	A207
----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

T208	G209	T210	T211	C212	A213	G214	C215	T216	A219	T220	T221	C222	C225	T226	G227	A228	A229	C230	A231	T232	G233	C234	C235	T236	T237	G240	A241	T242	G243	G244	A245	G246	C247	A248	G249	T250	T251	T252	C253	G254	A255	A256	A257	T258	A259	C260	A261	C262	T263	T264	T265	T266	G267	G268	T269	A270
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

G271	A272	A273	T274	C275	T276	G280	G281	T282	G283	G284	A285	T286	A287	T288	T289	T292
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	105.95Å 109.43Å 180.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.61 – 2.89 48.61 – 2.89	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.61-2.89) 99.7 (48.61-2.89)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.31 (at 2.91Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.220 , 0.268 0.220 , 0.268	Depositor DCC
R_{free} test set	2412 reflections (5.06%)	DCC
Wilson B-factor (Å ²)	48.6	Xtriage
Anisotropy	0.318	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 46.4	EDS
Estimated twinning fraction	0.016 for k,h,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 47693 reflections	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11860	wwPDB-VP
Average B, all atoms (Å ²)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.16% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0/813	0.71	0/1090
1	E	0.56	0/822	0.69	0/1102
2	B	0.49	0/634	0.75	0/849
2	F	0.59	0/675	0.83	0/904
3	C	0.53	0/783	0.71	0/1054
3	G	0.48	0/753	0.67	0/1015
4	D	0.53	0/767	0.70	0/1029
4	H	0.50	0/736	0.70	0/990
5	I	0.48	0/3332	0.84	0/5141
5	J	0.47	0/3330	0.83	0/5138
All	All	0.50	0/12645	0.79	0/18312

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
5	J	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	J	280	DG	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	801	0	839	34	0
1	E	810	0	851	50	0
2	B	627	0	658	31	0
2	F	668	0	714	35	0
3	C	774	0	822	40	0
3	G	744	0	786	25	0
4	D	756	0	784	33	0
4	H	725	0	745	25	0
5	I	2970	0	1640	161	0
5	J	2969	0	1641	180	0
6	A	1	0	0	1	0
6	D	1	0	0	0	0
6	E	1	0	0	1	0
6	G	1	0	0	1	0
7	D	1	0	0	0	0
7	I	6	0	0	0	0
7	J	5	0	0	0	0
All	All	11860	0	9480	536	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 25.

The worst 5 of 536 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)
2:B:75:HIS:CD2	4:D:96:THR:HG21	1.82	1.13
5:J:219:DA:H2''	5:J:220:DT:H5'	1.32	1.11
5:I:128:DT:H2''	5:I:129:DC:H5'	1.32	1.11
2:B:75:HIS:HD2	4:D:96:THR:HG21	1.07	1.10
5:I:115:DA:H2''	5:I:116:DC:H5''	1.25	1.10

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	95/139 (68%)	92 (97%)	3 (3%)	0	100	100
1	E	96/139 (69%)	93 (97%)	3 (3%)	0	100	100
2	B	77/106 (73%)	73 (95%)	4 (5%)	0	100	100
2	F	81/106 (76%)	79 (98%)	2 (2%)	0	100	100
3	C	98/133 (74%)	94 (96%)	2 (2%)	2 (2%)	9	33
3	G	94/133 (71%)	93 (99%)	1 (1%)	0	100	100
4	D	94/129 (73%)	87 (93%)	5 (5%)	2 (2%)	9	32
4	H	91/129 (70%)	86 (94%)	4 (4%)	1 (1%)	17	51
All	All	726/1014 (72%)	697 (96%)	24 (3%)	5 (1%)	26	63

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	32	SER
4	D	104	GLY
4	H	104	GLY
3	C	47	ALA
3	C	109	PRO

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	85/113 (75%)	78 (92%)	7 (8%)	14	39

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	86/113 (76%)	83 (96%)	3 (4%)	43	78
2	B	64/81 (79%)	64 (100%)	0	100	100
2	F	69/81 (85%)	66 (96%)	3 (4%)	35	71
3	C	78/102 (76%)	71 (91%)	7 (9%)	12	34
3	G	76/102 (74%)	68 (90%)	8 (10%)	8	25
4	D	82/107 (77%)	73 (89%)	9 (11%)	8	23
4	H	79/107 (74%)	72 (91%)	7 (9%)	12	35
All	All	619/806 (77%)	575 (93%)	44 (7%)	18	47

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	96	THR
1	E	80	THR
4	H	95	GLN
4	D	101	LEU
4	D	108	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	110	ASN
4	D	95	GLN
3	G	31	HIS
3	C	84	GLN
3	G	73	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 [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 16 ligands modelled in this entry, 16 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, 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	97/139 (69%)	-0.29	0 100 100	17, 34, 62, 84	0
1	E	98/139 (70%)	-0.28	2 (2%) 68 64	13, 26, 53, 85	0
2	B	79/106 (74%)	-0.42	0 100 100	20, 32, 50, 67	0
2	F	83/106 (78%)	-0.32	0 100 100	15, 26, 41, 63	0
3	C	100/133 (75%)	-0.17	3 (3%) 54 47	16, 30, 64, 113	0
3	G	96/133 (72%)	-0.31	0 100 100	21, 36, 65, 72	0
4	D	96/129 (74%)	-0.19	2 (2%) 67 62	17, 31, 66, 90	0
4	H	93/129 (72%)	-0.21	0 100 100	21, 36, 66, 80	0
5	I	145/146 (99%)	0.20	4 (2%) 56 50	38, 87, 134, 158	0
5	J	145/146 (99%)	0.11	4 (2%) 56 50	41, 88, 137, 154	0
All	All	1032/1306 (79%)	-0.15	15 (1%) 76 74	13, 38, 115, 158	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	12	ALA	7.0
3	C	13	LYS	4.2
3	C	11	ARG	3.6
1	E	39	HIS	3.0
5	I	54	DA	2.9

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å ²)	Q<0.9
7	MN	D	202	1/1	0.98	0.26	4.15	31,31,31,31	0
6	CL	D	201	1/1	0.97	0.14	-0.86	40,40,40,40	0
6	CL	G	1001	1/1	0.98	0.14	-0.90	34,34,34,34	0
7	MN	I	1004	1/1	0.86	0.12	-2.14	74,74,74,74	0
7	MN	I	1005	1/1	0.67	0.12	-	96,96,96,96	0
6	CL	E	1001	1/1	0.99	0.07	-	54,54,54,54	0
7	MN	I	1006	1/1	0.58	0.14	-	97,97,97,97	0
7	MN	J	1004	1/1	0.98	0.17	-	60,60,60,60	0
7	MN	I	1002	1/1	0.88	0.11	-	94,94,94,94	0
6	CL	A	1001	1/1	0.95	0.12	-	51,51,51,51	0
7	MN	J	1002	1/1	0.95	0.09	-	96,96,96,96	0
7	MN	J	1003	1/1	0.94	0.09	-	77,77,77,77	0
7	MN	I	1003	1/1	0.90	0.19	-	68,68,68,68	0
7	MN	J	1005	1/1	0.97	0.05	-	86,86,86,86	0
7	MN	J	1001	1/1	0.90	0.07	-	94,94,94,94	0
7	MN	I	1001	1/1	0.69	0.10	-	104,104,104,104	0

6.5 Other polymers ⓘ

There are no such residues in this entry.