



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 08:57 am BST

PDB ID : 2F8N  
Title : 2.9 Angstrom X-ray structure of hybrid macroH2A nucleosomes  
Authors : Chakravarthy, S.; Luger, K.  
Deposited on : 2005-12-02  
Resolution : 2.90 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:

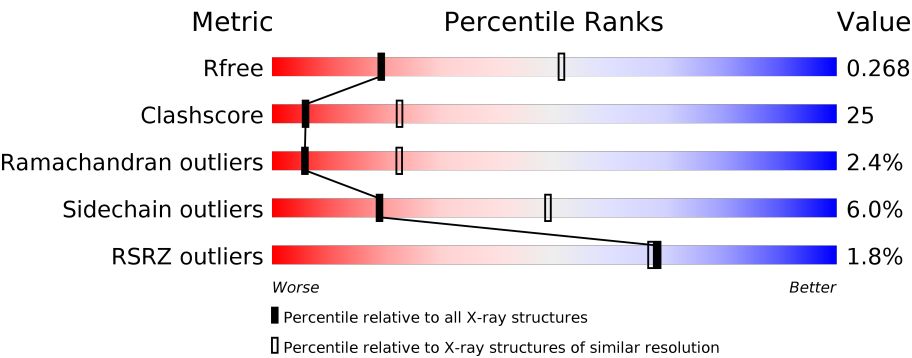
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:  
*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (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  $\geq 3$ , 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	I	146	<div><div>2%</div><div><div></div><div></div><div></div><div></div></div><div>36%64%</div><div>.</div></div>
1	J	146	<div><div>%</div><div><div></div><div></div><div></div><div></div></div><div>36%64%</div><div>.</div></div>
2	A	136	<div><div>%</div><div><div></div><div></div><div></div><div></div></div><div>39%29%28%</div><div>.</div></div>
2	E	136	<div><div>%</div><div><div></div><div></div><div></div><div></div></div><div>42%27%28%</div><div>.</div></div>
3	B	103	<div><div>%</div><div><div></div><div></div><div></div><div></div></div><div>37%37%23%</div><div>..</div></div>
3	F	103	<div><div>2%</div><div><div></div><div></div><div></div><div></div></div><div>48%31%18%</div><div>.</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	126	<div><div><div></div><div></div><div></div><div></div></div><div><div>2%</div><div>40%</div><div>31%</div><div>•</div><div>26%</div></div></div>
5	H	123	<div><div><div></div><div></div><div></div><div></div></div><div><div>2%</div><div>39%</div><div>30%</div><div>7%</div><div>24%</div></div></div>
6	G	120	<div><div><div></div><div></div><div></div><div></div></div><div><div>2%</div><div>43%</div><div>43%</div><div>5%</div><div>10%</div></div></div>
7	K	149	<div><div><div></div><div></div><div></div><div></div></div><div><div>%</div><div>42%</div><div>25%</div><div>•</div><div>30%</div></div></div>

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 12066 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 DNA chain called alpha-satellite DNA (146 bp).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	I	145	Total	C	N	O	P	0	0	0
			2970	1420	539	867	144			
1	J	145	Total	C	N	O	P	0	0	0
			2969	1420	536	869	144			

- Molecule 2 is a protein called Histone H3.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	98	Total	C	N	O	S	0	0	0
			807	508	156	140	3			
2	E	98	Total	C	N	O	S	0	0	0
			807	508	156	140	3			

- Molecule 3 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	79	Total	C	N	O	S	0	0	0
			627	395	121	110	1			
3	F	84	Total	C	N	O	S	0	0	0
			673	424	133	115	1			

- Molecule 4 is a protein called Histone 3, H2ba.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	93	Total	C	N	O	S	0	0	0
			731	459	131	139	2			

- Molecule 5 is a protein called Histone H2B.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	H	93	Total	C	N	O	S	0	0	0
			729	459	131	137	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	1400	MET	-	INITIATING METHIONINE	UNP P02281
H	1429	THR	SER	CONFLICT	UNP P02281

- Molecule 6 is a protein called Core histone macro-H2A.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	108	Total	C	N	O	S	0	0	0
			823	527	151	143	2			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	1067	VAL	GLY	CONFLICT	UNP O75367

- Molecule 7 is a protein called Histone H2A type 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	K	105	Total	C	N	O		0	0	0
			810	511	158	141				

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	-19	MET	-	CLONING ARTIFACT	UNP Q8CGP6
K	-18	GLY	-	CLONING ARTIFACT	UNP Q8CGP6
K	-17	SER	-	CLONING ARTIFACT	UNP Q8CGP6
K	-16	SER	-	CLONING ARTIFACT	UNP Q8CGP6
K	-15	HIS	-	EXPRESSION TAG	UNP Q8CGP6
K	-14	HIS	-	EXPRESSION TAG	UNP Q8CGP6
K	-13	HIS	-	EXPRESSION TAG	UNP Q8CGP6
K	-12	HIS	-	EXPRESSION TAG	UNP Q8CGP6
K	-11	HIS	-	EXPRESSION TAG	UNP Q8CGP6
K	-10	HIS	-	EXPRESSION TAG	UNP Q8CGP6
K	-9	SER	-	CLONING ARTIFACT	UNP Q8CGP6
K	-8	SER	-	CLONING ARTIFACT	UNP Q8CGP6
K	-7	GLY	-	CLONING ARTIFACT	UNP Q8CGP6
K	-6	LEU	-	CLONING ARTIFACT	UNP Q8CGP6
K	-5	VAL	-	CLONING ARTIFACT	UNP Q8CGP6
K	-4	PRO	-	CLONING ARTIFACT	UNP Q8CGP6
K	-3	ARG	-	CLONING ARTIFACT	UNP Q8CGP6
K	-2	GLY	-	CLONING ARTIFACT	UNP Q8CGP6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
K	-1	SER	-	CLONING ARTIFACT	UNP Q8CGP6

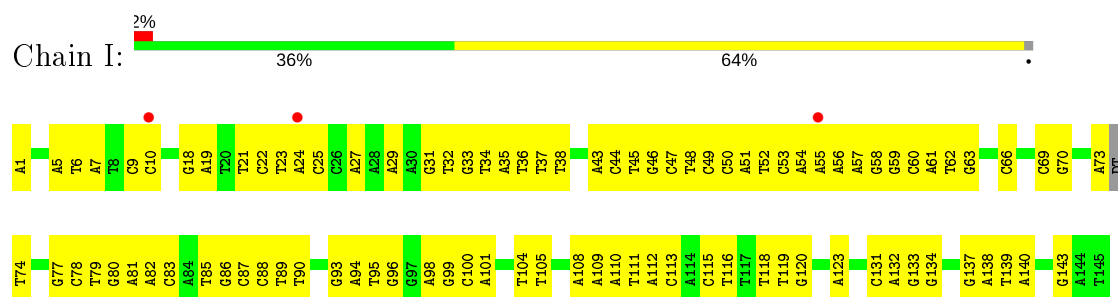
- Molecule 8 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	I	23	Total O 23 23	0	0
8	J	23	Total O 23 23	0	0
8	A	6	Total O 6 6	0	0
8	B	4	Total O 4 4	0	0
8	D	11	Total O 11 11	0	0
8	E	17	Total O 17 17	0	0
8	F	13	Total O 13 13	0	0
8	H	1	Total O 1 1	0	0
8	G	7	Total O 7 7	0	0
8	K	15	Total O 15 15	0	0

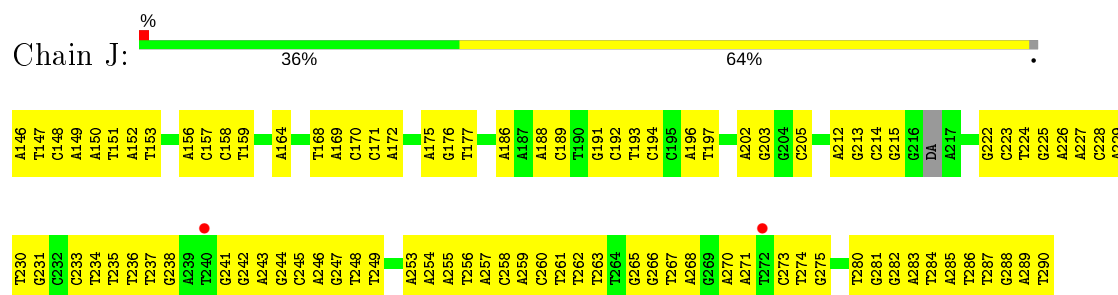
### 3 Residue-property plots

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.

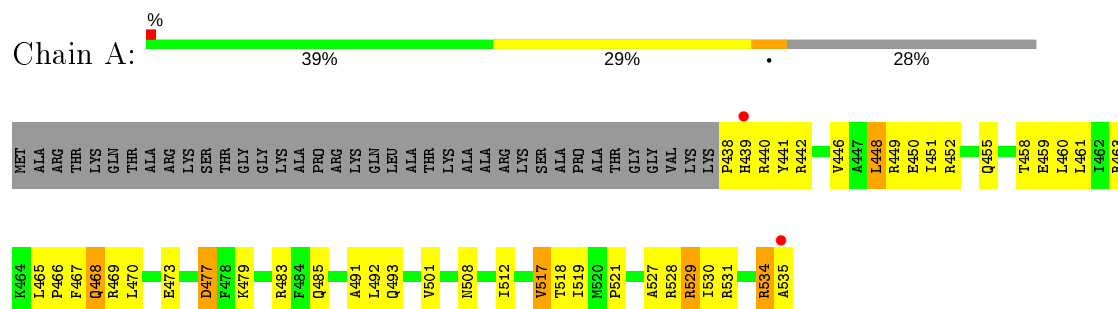
#### • Molecule 1: alpha-satellite DNA (146 bp)



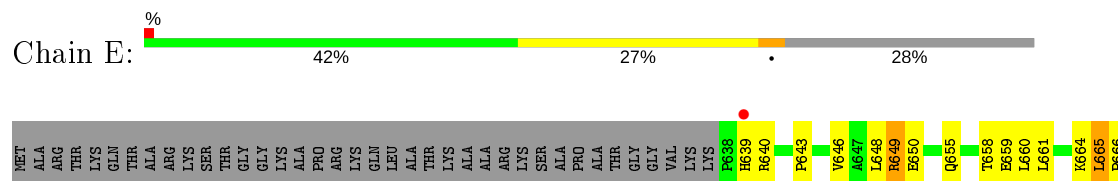
#### • Molecule 1: alpha-satellite DNA (146 bp)



#### • Molecule 2: Histone H3.1

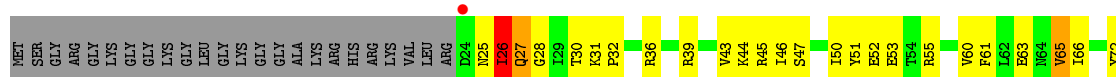


#### • Molecule 2: Histone H3.1

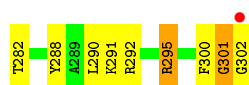




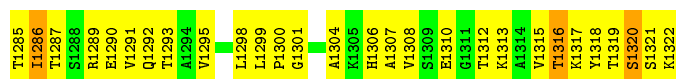
- Molecule 3: Histone H4



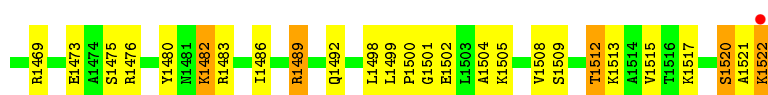
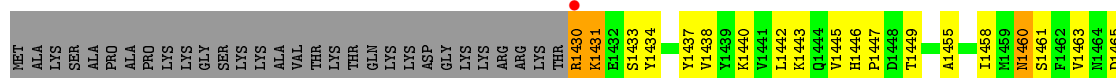
- Molecule 3: Histone H4



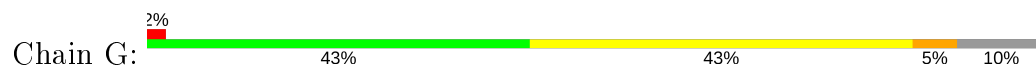
- Molecule 4: Histone 3, H2ba



- Molecule 5: Histone H2B.1



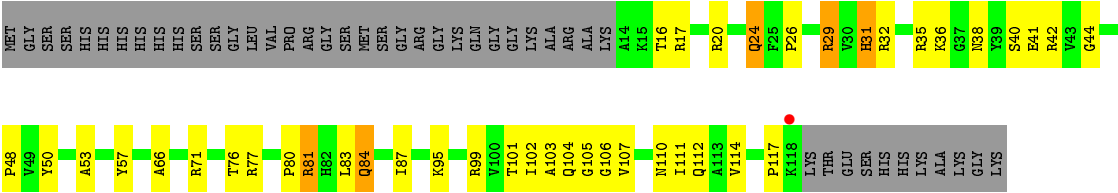
- Molecule 6: Core histone macro-H2A.1







● Molecule 7: Histone H2A type 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.14Å 109.27Å 176.27Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	31.40 – 2.90 31.45 – 2.90	Depositor EDS
% Data completeness (in resolution range)	(Not available) (31.40-2.90) 94.0 (31.45-2.90)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.18 (at 2.90Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.217 , 0.269 0.216 , 0.268	Depositor DCC
$R_{free}$ test set	2243 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	73.7	Xtriage
Anisotropy	0.485	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 90.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.022 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	12066	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	99.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.39% 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](#)

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	I	0.33	0/3332	0.67	0/5141
1	J	0.31	0/3330	0.67	0/5138
2	A	0.43	0/819	0.62	0/1097
2	E	0.49	0/819	0.70	0/1097
3	B	0.45	0/634	0.75	0/848
3	F	0.52	0/680	0.75	0/908
4	D	0.47	0/742	0.67	0/996
5	H	0.42	0/740	0.62	0/994
6	G	0.36	0/836	0.65	0/1129
7	K	0.42	0/820	0.65	0/1107
All	All	0.38	0/12752	0.67	0/18455

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
1	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
1	J	212	DA	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	I	2970	0	1639	120	0
1	J	2969	0	1640	125	0
2	A	807	0	844	70	0
2	E	807	0	844	43	0
3	B	627	0	663	61	0
3	F	673	0	722	41	0
4	D	731	0	753	49	1
5	H	729	0	753	54	0
6	G	823	0	885	78	0
7	K	810	0	866	58	0
8	A	6	0	0	1	0
8	B	4	0	0	5	0
8	D	11	0	0	2	0
8	E	17	0	0	3	1
8	F	13	0	0	6	0
8	G	7	0	0	4	0
8	H	1	0	0	0	0
8	I	23	0	0	18	0
8	J	23	0	0	7	0
8	K	15	0	0	3	0
All	All	12066	0	9609	539	1

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 539 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)
6:G:1039:PRO:HD3	8:G:425:HOH:O	1.46	1.14
2:E:677:ASP:OD2	8:E:300:HOH:O	1.65	1.13
1:I:87:DC:H4'	8:I:426:HOH:O	1.57	1.04
1:I:69:DC:H4'	3:B:45:ARG:HH12	1.20	1.01
1:I:21:DT:H2''	1:I:22:DC:H5'	1.40	1.00

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:D:1245:VAL:O	8:E:300:HOH:O[3_445]	2.04	0.16

## 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	
2	A	96/136 (71%)	86 (90%)	7 (7%)	3 (3%)	4	16
2	E	96/136 (71%)	92 (96%)	3 (3%)	1 (1%)	15	45
3	B	77/103 (75%)	67 (87%)	7 (9%)	3 (4%)	3	12
3	F	82/103 (80%)	77 (94%)	4 (5%)	1 (1%)	13	40
4	D	91/126 (72%)	84 (92%)	5 (6%)	2 (2%)	6	24
5	H	91/123 (74%)	78 (86%)	10 (11%)	3 (3%)	4	15
6	G	106/120 (88%)	97 (92%)	5 (5%)	4 (4%)	3	13
7	K	103/149 (69%)	94 (91%)	8 (8%)	1 (1%)	15	45
All	All	742/996 (74%)	675 (91%)	49 (7%)	18 (2%)	6	22

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	B	26	ILE
4	D	1320	SER
6	G	1039	PRO
6	G	1118	LYS
7	K	117	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	
2	A	85/111 (77%)	79 (93%)	6 (7%)	14	40
2	E	85/111 (77%)	81 (95%)	4 (5%)	26	59
3	B	64/79 (81%)	62 (97%)	2 (3%)	40	74
3	F	69/79 (87%)	65 (94%)	4 (6%)	20	50
4	D	81/108 (75%)	76 (94%)	5 (6%)	18	47
5	H	79/103 (77%)	73 (92%)	6 (8%)	13	36
6	G	86/95 (90%)	82 (95%)	4 (5%)	26	59
7	K	83/116 (72%)	76 (92%)	7 (8%)	11	31
All	All	632/802 (79%)	594 (94%)	38 (6%)	19	49

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	219	ARG
5	H	1430	ARG
7	K	84	GLN
3	F	223	ARG
5	H	1460	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
5	H	1446	HIS
5	H	1464	ASN
7	K	31	HIS
3	F	293	GLN
7	K	24	GLN

### 5.3.3 RNA ⓘ

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, 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	I	145/146 (99%)	0.25	3 (2%) 63 61	61, 130, 172, 189	0
1	J	145/146 (99%)	0.22	2 (1%) 75 75	62, 127, 170, 199	0
2	A	98/136 (72%)	0.10	2 (2%) 65 63	41, 70, 118, 165	0
2	E	98/136 (72%)	0.10	2 (2%) 65 63	29, 51, 87, 163	0
3	B	79/103 (76%)	-0.02	1 (1%) 77 77	48, 69, 109, 182	0
3	F	84/103 (81%)	-0.04	2 (2%) 59 56	32, 51, 74, 176	0
4	D	93/126 (73%)	-0.10	2 (2%) 62 59	38, 66, 104, 132	0
5	H	93/123 (75%)	0.01	2 (2%) 62 59	44, 70, 114, 176	0
6	G	108/120 (90%)	-0.01	2 (1%) 66 65	43, 73, 136, 187	0
7	K	105/149 (70%)	-0.07	1 (0%) 82 82	37, 60, 102, 177	0
All	All	1048/1288 (81%)	0.06	19 (1%) 68 67	29, 73, 161, 199	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	H	1430	ARG	5.1
2	E	735	ALA	4.8
7	K	118	LYS	4.4
6	G	1014	LYS	4.4
3	F	302	GLY	3.2

### 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.