



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 11:14 pm BST

PDB ID : 5B24  
Title : The crystal structure of the nucleosome containing cyclobutane pyrimidine dimer  
Authors : Horikoshi, N.; Tachiwana, H.; Kagawa, W.; Osakabe, A.; Kurumizaka, H.  
Deposited on : 2015-12-31  
Resolution : 3.60 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:

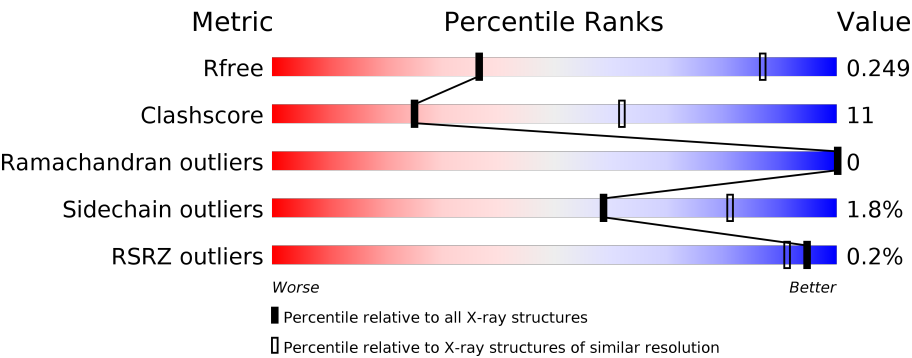
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

# 1 Overall quality at a glance i

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

The reported resolution of this entry is 3.60 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	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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	139	<div><div></div><div><div></div><div>58%</div><div></div><div>12%</div><div></div><div>30%</div></div></div>
1	E	139	<div><div>%</div><div><div></div><div>51%</div><div></div><div>18%</div><div></div><div>29%</div></div></div>
2	B	106	<div><div></div><div><div></div><div>50%</div><div></div><div>23%</div><div></div><div>26%</div></div></div>
2	F	106	<div><div></div><div><div></div><div>58%</div><div></div><div>22%</div><div></div><div>21%</div></div></div>
3	C	133	<div><div></div><div><div></div><div>63%</div><div></div><div>18%</div><div></div><div>19%</div></div></div>
3	G	133	<div><div></div><div><div></div><div>64%</div><div></div><div>14%</div><div></div><div>22%</div></div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	129	<div><div></div><div>54%18%28%</div></div>
4	H	129	<div><div></div><div>52%19%29%</div></div>
5	I	145	<div><div></div><div>49%50%</div></div>
5	J	145	<div><div></div><div>50%50%</div></div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 11973 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	99	Total	C	N	O	S	0	0	0
			816	514	158	140	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	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			

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
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	108	Total	C	N	O	0	0	0
			835	526	165	144			
3	G	104	Total	C	N	O	0	0	0
			805	508	157	140			

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	93	Total	C	N	O	S	0	0	0
			725	456	130	137	2			
4	H	92	Total	C	N	O	S	0	0	0
			719	453	129	135	2			

There are 6 discrepancies between the modelled and reference sequences:

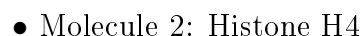
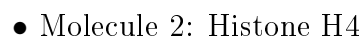
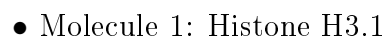
Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	GLY	-	expression tag	UNP P06899
D	-2	SER	-	expression tag	UNP P06899
D	-1	HIS	-	expression tag	UNP P06899
H	-3	GLY	-	expression tag	UNP P06899
H	-2	SER	-	expression tag	UNP P06899
H	-1	HIS	-	expression tag	UNP P06899

- Molecule 5 is a DNA chain called DNA (145-MER).

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



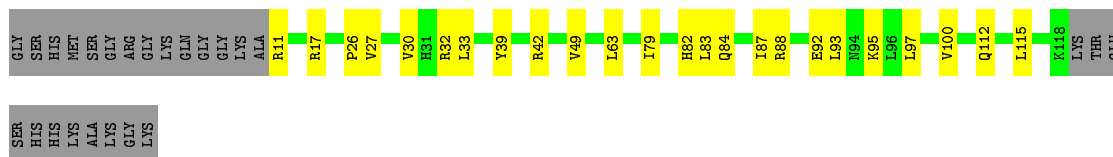
- Molecule 1: Histone H3.1





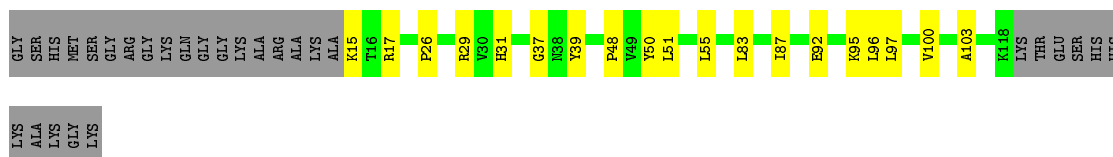
- Molecule 3: Histone H2A type 1-B/E

Chain C: 63% 18% 19%



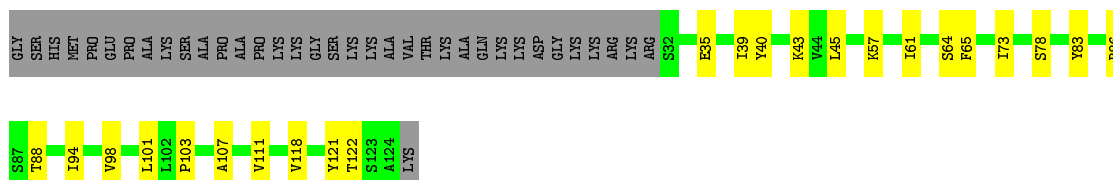
- Molecule 3: Histone H2A type 1-B/E

Chain G: 64% 14% 22%



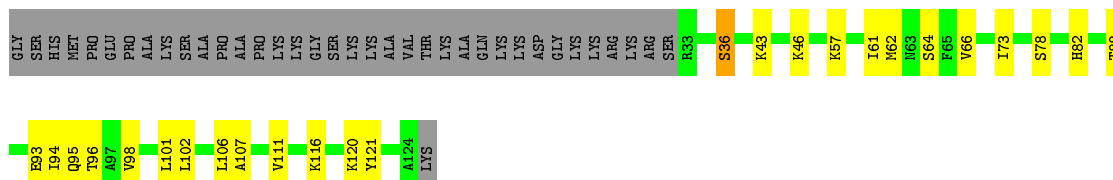
- Molecule 4: Histone H2B type 1-J

Chain D: 54% 18% 28%



- Molecule 4: Histone H2B type 1-J

Chain H: 52% 19% 29%



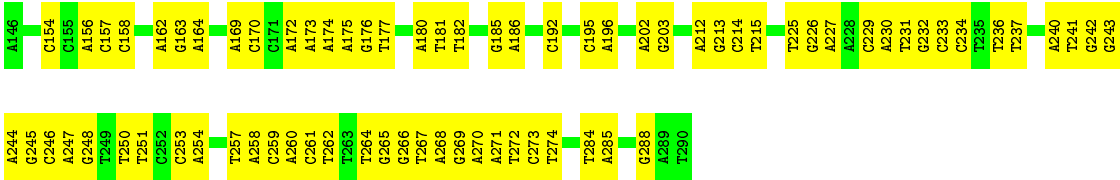
- Molecule 5: DNA (145-MER)

Chain I: 49% 50% 1%





● Molecule 5: DNA (145-MER)





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	104.85Å 108.87Å 174.05Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.64 – 3.60 45.58 – 3.60	Depositor EDS
% Data completeness (in resolution range)	95.9 (34.64-3.60) 88.6 (45.58-3.60)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.31 (at 3.57Å)	Xtriage
Refinement program	PHENIX 1.10.1-2155-000	Depositor
R, $R_{free}$	0.205 , 0.249 0.205 , 0.249	Depositor DCC
$R_{free}$ test set	1168 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	84.9	Xtriage
Anisotropy	0.930	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 44.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.027 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11973	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% 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: TTD

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	0.28	0/813	0.44	0/1090
1	E	0.35	0/828	0.58	0/1109
2	B	0.29	0/626	0.50	0/837
2	F	0.33	0/680	0.48	0/908
3	C	0.27	0/845	0.45	0/1139
3	G	0.30	0/815	0.46	0/1100
4	D	0.29	0/736	0.47	0/990
4	H	0.27	0/730	0.46	0/982
5	I	0.50	0/3309	0.91	1/5103 (0.0%)
5	J	0.52	0/3309	0.93	0/5103
All	All	0.42	0/12691	0.76	1/18361 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	I	101	DC	O4'-C4'-C3'	-5.16	102.44	104.50

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	801	0	839	17	0
1	E	816	0	856	27	0
2	B	619	0	659	25	0
2	F	673	0	722	27	0
3	C	835	0	897	22	0
3	G	805	0	861	19	0
4	D	725	0	745	25	0
4	H	719	0	740	22	0
5	I	2990	0	1652	61	0
5	J	2990	0	1652	59	0
All	All	11973	0	9623	229	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 229 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)
5:I:12:DC:H2"	5:I:13:DC:H5"	1.61	0.81
2:F:79:LYS:NZ	5:I:100:DG:OP1	2.14	0.80
5:J:258:DA:H2"	5:J:259:DC:H5"	1.63	0.79
4:D:64:SER:HB2	2:F:98:TYR:HD1	1.47	0.77
2:B:92:ARG:HH21	4:D:101:LEU:HD12	1.51	0.76

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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%)	93 (98%)	2 (2%)	0	100	100
1	E	97/139 (70%)	92 (95%)	5 (5%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	76/106 (72%)	72 (95%)	4 (5%)	0	100	100
2	F	82/106 (77%)	77 (94%)	5 (6%)	0	100	100
3	C	106/133 (80%)	104 (98%)	2 (2%)	0	100	100
3	G	102/133 (77%)	100 (98%)	2 (2%)	0	100	100
4	D	91/129 (70%)	89 (98%)	2 (2%)	0	100	100
4	H	90/129 (70%)	88 (98%)	2 (2%)	0	100	100
All	All	739/1014 (73%)	715 (97%)	24 (3%)	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	85/113 (75%)	85 (100%)	0	100	100
1	E	86/113 (76%)	82 (95%)	4 (5%)	26	61
2	B	63/81 (78%)	60 (95%)	3 (5%)	25	60
2	F	69/81 (85%)	69 (100%)	0	100	100
3	C	85/102 (83%)	84 (99%)	1 (1%)	71	87
3	G	83/102 (81%)	83 (100%)	0	100	100
4	D	79/107 (74%)	79 (100%)	0	100	100
4	H	78/107 (73%)	75 (96%)	3 (4%)	33	66
All	All	628/806 (78%)	617 (98%)	11 (2%)	59	81

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	37	LYS
1	E	42	ARG
4	H	36	SER
3	C	95	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	134	ARG

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

Mol	Chain	Res	Type
2	B	25	ASN
2	B	64	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	$\# Z  > 2$
5	TTD	I	117	5	42,45,46	1.96	14 (33%)	62,74,77	2.81	24 (38%)
5	TTD	J	262	5	42,45,46	2.00	14 (33%)	62,74,77	2.53	25 (40%)

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
5	TTD	I	117	5	-	13/22/109/110	0/5/6/6
5	TTD	J	262	5	-	13/22/109/110	0/5/6/6

The worst 5 of 28 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	I	117	TTD	C6T-N1T	4.72	1.54	1.46
5	J	262	TTD	C6T-N1T	4.49	1.53	1.46
5	J	262	TTD	C2'-C3R	-4.16	1.43	1.52
5	J	262	TTD	C2-N3	-4.03	1.30	1.38
5	I	117	TTD	C2-N3	-3.92	1.31	1.38

The worst 5 of 49 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	I	117	TTD	O4R-C1R-N1T	10.74	121.38	108.65
5	J	262	TTD	C5T-C4T-N3T	7.29	122.41	116.06
5	I	117	TTD	N3T-C2T-N1T	7.05	124.01	116.69
5	I	117	TTD	C5T-C4T-N3T	6.72	121.91	116.06
5	I	117	TTD	N3-C2-N1	6.31	123.24	116.69

There are no chirality outliers.

5 of 26 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	I	117	TTD	C2'-C3R-O3R-PB
5	I	117	TTD	O4'-C1'-N1-C2
5	I	117	TTD	O4R-C4'-C5R-O5R
5	I	117	TTD	C3'-C4'-C5R-O5R
5	J	262	TTD	O4'-C1'-N1-C2

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	I	117	TTD	4	0
5	J	262	TTD	2	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

### 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	97/139 (69%)	-0.29	0 100 100	70, 94, 133, 150	0
1	E	99/139 (71%)	-0.27	1 (1%) 82 70	57, 71, 107, 148	0
2	B	78/106 (73%)	-0.29	0 100 100	69, 88, 107, 110	0
2	F	84/106 (79%)	-0.31	0 100 100	55, 70, 89, 105	0
3	C	108/133 (81%)	-0.32	0 100 100	55, 75, 107, 162	0
3	G	104/133 (78%)	-0.28	0 100 100	69, 89, 119, 136	0
4	D	93/129 (72%)	-0.33	0 100 100	62, 77, 103, 127	0
4	H	92/129 (71%)	-0.26	0 100 100	66, 87, 111, 136	0
5	I	144/145 (99%)	-0.42	1 (0%) 87 78	94, 153, 191, 202	0
5	J	144/145 (99%)	-0.48	0 100 100	94, 153, 186, 200	0
All	All	1043/1304 (79%)	-0.34	2 (0%) 95 91	55, 90, 173, 202	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	145	DT	6.9
1	E	38	PRO	2.0

### 6.2 Non-standard residues in protein, DNA, RNA chains

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
5	TTD	I	117	40/41	0.80	0.20	159,171,183,189	0
5	TTD	J	262	40/41	0.86	0.19	160,170,178,193	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.