wwPDB X-ray Structure Validation Summary Report (i)

Oct 30, 2023 – 07:32 PM JST

PDB ID	:	4YM6
Title	:	Crystal structure of the human nucleosome containing 6-4PP (outside)
Authors	:	Osakabe, A.; Tachiwana, H.; Kagawa, W.; Horikoshi, N.; Matsumoto, S.; Hasegawa, M.; Matsumoto, N.; Toga, T.; Yamamoto, J.; Hanaoka, F.; Thoma, N.H.; Sugasawa, K.; Iwai, S.; Kurumizaka, H.
Deposited on	:	2015-03-06
Resolution	:	3.51 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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

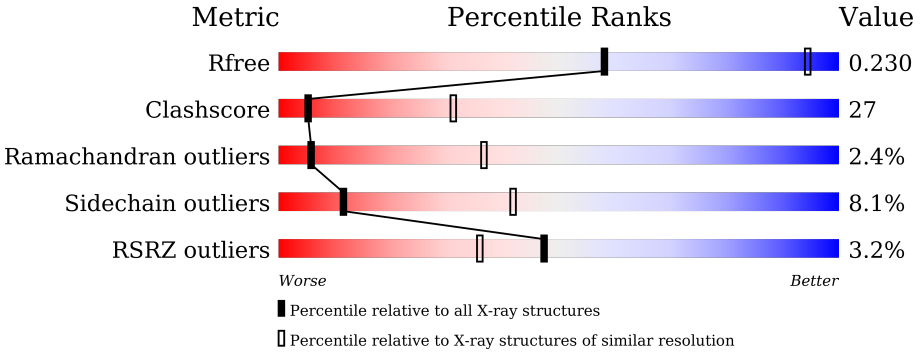
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.51 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	1161 (3.60-3.44)
Clashscore	141614	1244 (3.60-3.44)
Ramachandran outliers	138981	1206 (3.60-3.44)
Sidechain outliers	138945	1207 (3.60-3.44)
RSRZ outliers	127900	1080 (3.60-3.44)

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 of 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 <=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	% 44% 24% • 30%
1	E	139	4% 50% 20% • 29%
2	B	106	2% 45% 25% • 26%
2	F	106	47% 29% • 21%
3	C	133	% 54% 22% 5% 19%
3	G	133	% 52% 21% 5% 22%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	129	2% 37% 31% 6% 26%
4	H	129	% 37% 29% 5% 29%
5	I	145	8% 17% 66% 18%
5	J	145	5% 18% 69% 13%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 12003 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
			Total	C	N	O	S			
1	A	97	801	505	155	137	4	0	0	0
1	E	99	816	514	158	140	4	0	0	0

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
			Total	C	N	O	S			
2	B	78	619	391	120	107	1	0	0	0
2	F	84	673	424	133	115	1	0	0	0

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

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
			Total	C	N	O	S			
4	D	96	755	474	138	141	2	0	0	0
4	H	92	719	453	129	135	2	0	0	0

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 145-MER DNA.

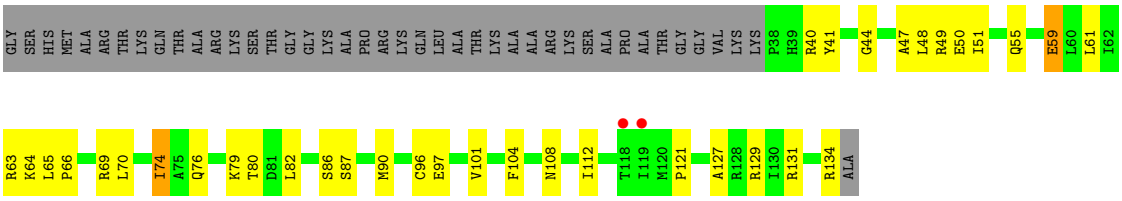
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	I	145	2990	1431	540	874	145	0	0	0
5	J	145	2990	1431	540	874	145	0	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

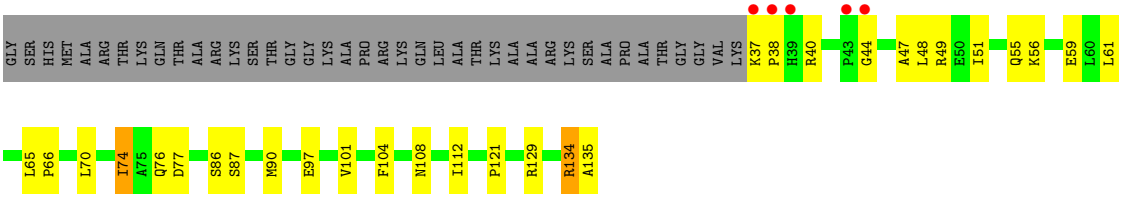
• Molecule 1: Histone H3.1

Chain A:



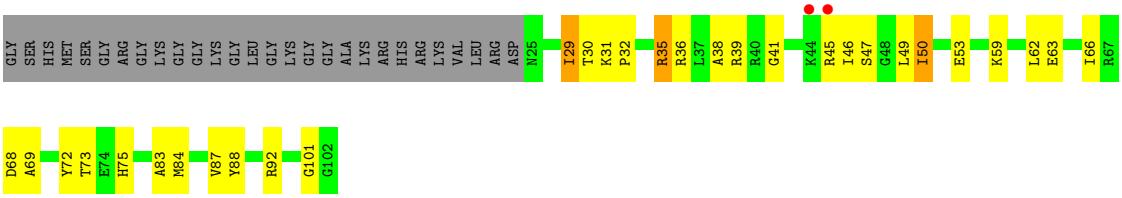
• Molecule 1: Histone H3.1

Chain E:



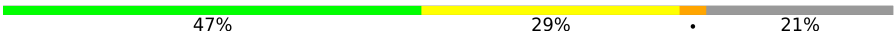
• Molecule 2: Histone H4

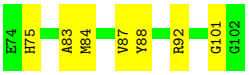
Chain B:



• Molecule 2: Histone H4

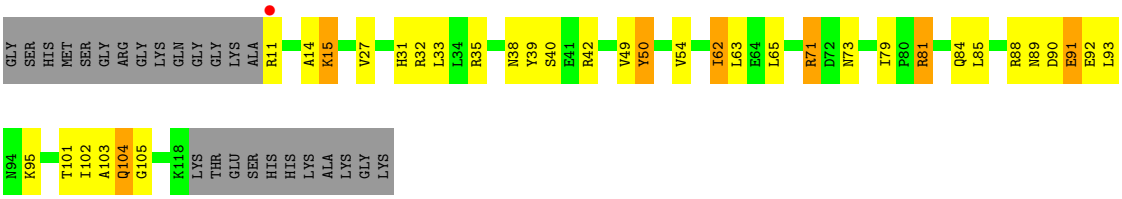
Chain F:





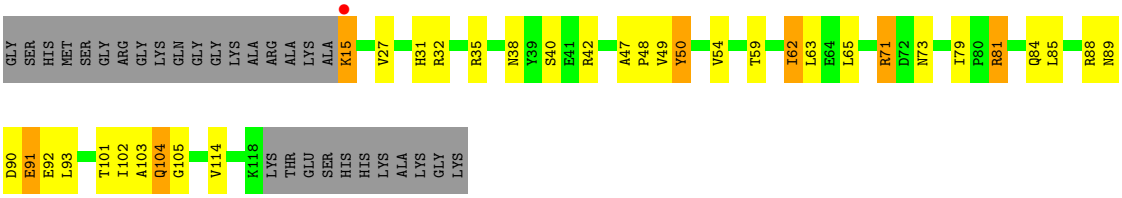
• Molecule 3: Histone H2A type 1-B/E

Chain C:



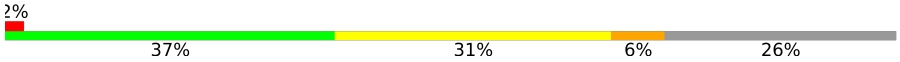
• Molecule 3: Histone H2A type 1-B/E

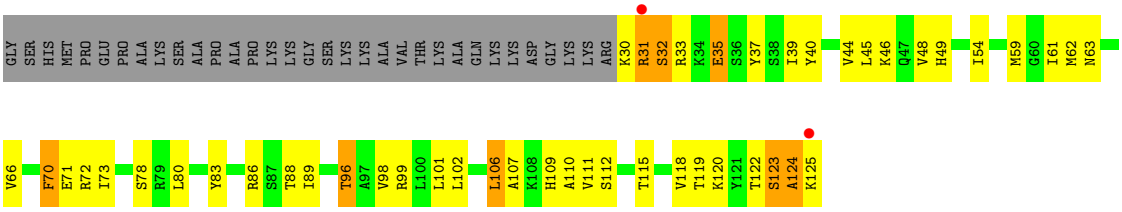
Chain G:



• Molecule 4: Histone H2B type 1-J

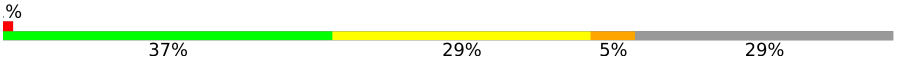
Chain D:

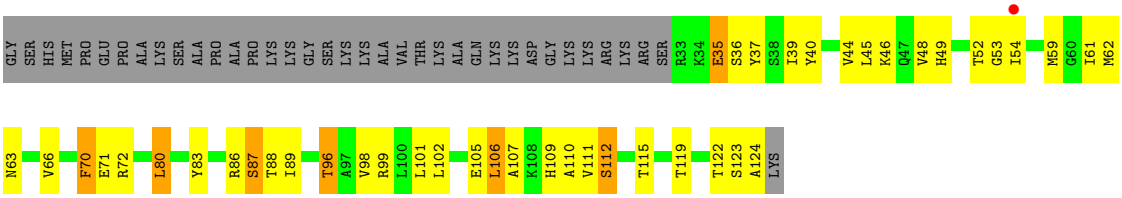




• Molecule 4: Histone H2B type 1-J

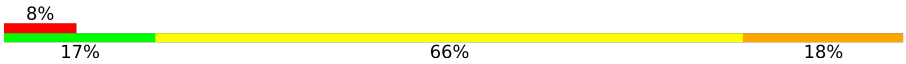
Chain H:

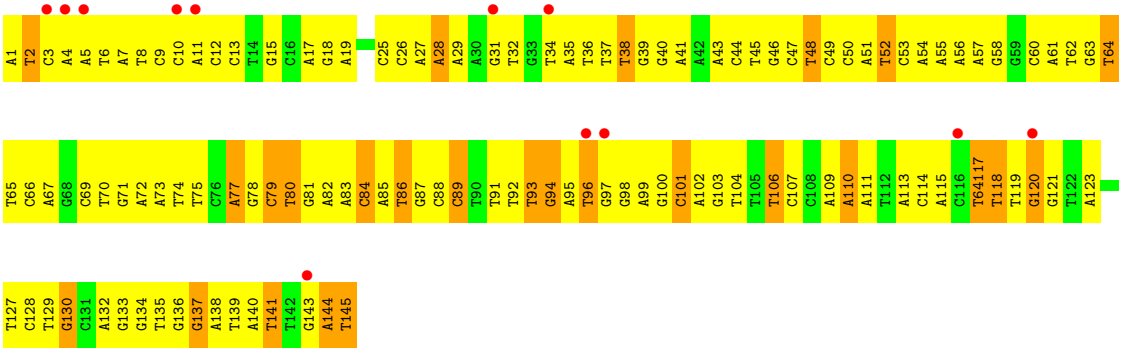




• Molecule 5: 145-MER DNA

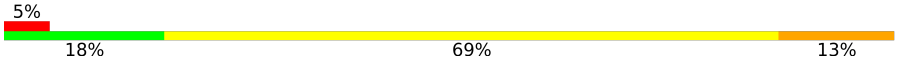
Chain I:

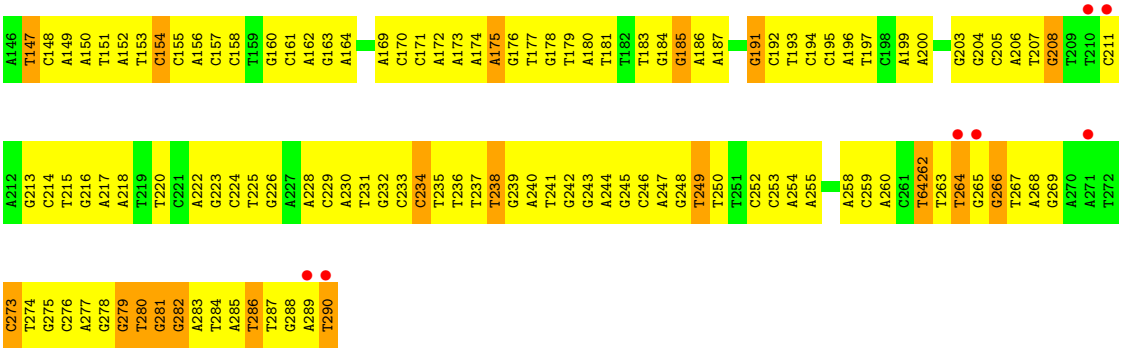




- Molecule 5: 145-MER DNA

Chain J:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	105.71Å 109.48Å 178.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.83 – 3.51 48.61 – 3.51	Depositor EDS
% Data completeness (in resolution range)	95.8 (29.83-3.51) 98.6 (48.61-3.51)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	10.70 (at 3.48Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, R_{free}	0.244 , 0.287 0.232 , 0.230	Depositor DCC
R_{free} test set	1291 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	107.0	Xtriage
Anisotropy	0.590	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 62.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.037 for k,h,-l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	12003	wwPDB-VP
Average B, all atoms (Å^2)	151.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.49% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: T64

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	0.55	1/813 (0.1%)	0.68	0/1090
1	E	0.66	0/828	0.72	1/1109 (0.1%)
2	B	0.53	0/626	0.72	0/837
2	F	0.73	0/680	0.76	0/908
3	C	0.60	0/845	0.69	0/1139
3	G	0.49	0/815	0.65	0/1100
4	D	0.64	0/766	0.66	0/1026
4	H	0.54	0/730	0.64	0/982
5	I	0.75	1/3309 (0.0%)	1.42	36/5103 (0.7%)
5	J	0.68	0/3309	1.40	30/5103 (0.6%)
All	All	0.66	2/12721 (0.0%)	1.15	67/18397 (0.4%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	96	CYS	CB-SG	-5.30	1.73	1.81
5	I	118	DT	C1'-N1	5.05	1.55	1.49

The worst 5 of 67 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	J	282	DG	O4'-C1'-N9	8.52	113.96	108.00
5	I	89	DC	O4'-C1'-N1	8.03	113.62	108.00
5	J	280	DT	C5-C4-O4	-7.55	119.61	124.90
5	J	273	DC	O4'-C4'-C3'	-7.47	101.51	104.50
5	J	211	DC	O4'-C1'-N1	7.27	113.09	108.00

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	26	0
1	E	816	0	856	25	0
2	B	619	0	659	32	0
2	F	673	0	722	37	0
3	C	835	0	897	39	0
3	G	805	0	861	40	0
4	D	755	0	784	63	0
4	H	719	0	740	47	0
5	I	2990	0	1652	198	0
5	J	2990	0	1652	183	0
All	All	12003	0	9662	577	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 27.

The worst 5 of 577 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:F:75:HIS:HD2	4:H:96:THR:HG21	1.01	1.13
2:F:75:HIS:CD2	4:H:96:THR:HG21	1.86	1.11
2:B:75:HIS:CD2	4:D:96:THR:HG21	1.88	1.08
2:B:75:HIS:HD2	4:D:96:THR:HG21	1.10	1.08
5:J:157:DC:H2"	5:J:158:DC:H5"	1.30	1.07

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%)	87 (92%)	7 (7%)	1 (1%)	14	53
1	E	97/139 (70%)	91 (94%)	5 (5%)	1 (1%)	15	54
2	B	76/106 (72%)	66 (87%)	8 (10%)	2 (3%)	5	34
2	F	82/106 (77%)	70 (85%)	9 (11%)	3 (4%)	3	28
3	C	106/133 (80%)	91 (86%)	13 (12%)	2 (2%)	8	41
3	G	102/133 (77%)	89 (87%)	11 (11%)	2 (2%)	7	41
4	D	94/129 (73%)	82 (87%)	8 (8%)	4 (4%)	2	23
4	H	90/129 (70%)	81 (90%)	6 (7%)	3 (3%)	4	30
All	All	742/1014 (73%)	657 (88%)	67 (9%)	18 (2%)	6	36

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	91	GLU
4	D	124	ALA
3	G	91	GLU
2	B	50	ILE
3	C	104	GLN

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%)	11	41
1	E	86/113 (76%)	80 (93%)	6 (7%)	15	47
2	B	63/81 (78%)	60 (95%)	3 (5%)	25	60
2	F	69/81 (85%)	66 (96%)	3 (4%)	29	62
3	C	85/102 (83%)	75 (88%)	10 (12%)	5	26
3	G	83/102 (81%)	75 (90%)	8 (10%)	8	35
4	D	82/107 (77%)	75 (92%)	7 (8%)	10	40

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	H	78/107 (73%)	71 (91%)	7 (9%)	9 38
All	All	631/806 (78%)	580 (92%)	51 (8%)	11 42

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	74	ILE
2	F	73	THR
4	H	96	THR
1	E	76	GLN
1	E	134	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	108	ASN
2	F	64	ASN
4	H	63	ASN
3	G	84	GLN
4	H	49	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)

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	T64	J	262	5	41,44,45	1.67	7 (17%)	53,69,72	2.07	10 (18%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
5	T64	I	117	5	41,44,45	1.65	7 (17%)	53,69,72	2.08	11 (20%)

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	T64	J	262	5	-	8/22/76/77	0/3/5/5
5	T64	I	117	5	-	9/22/76/77	0/3/5/5

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	J	262	T64	C1'-N1	4.88	1.52	1.45
5	I	117	T64	C5T-C4T	-4.25	1.40	1.47
5	J	262	T64	C5T-C4T	-4.20	1.40	1.47
5	I	117	T64	C1'-N1	4.13	1.51	1.45
5	J	262	T64	C2-N1	-4.13	1.28	1.36

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	J	262	T64	C5T-C4T-N3T	-8.08	118.67	123.49
5	I	117	T64	C5T-C4T-N3T	-7.72	118.88	123.49
5	I	117	T64	C4-N3-C2	-6.96	115.97	126.67
5	J	262	T64	O4'-C1'-N1	6.09	115.87	108.65
5	I	117	T64	O4'-C1'-N1	4.89	114.45	108.65

There are no chirality outliers.

5 of 17 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	I	117	T64	C5R-O5R-PB-O5P
5	I	117	T64	C3R-C4R-C5'-O5'
5	I	117	T64	O4'-C4R-C5'-O5'
5	J	262	T64	C5R-O5R-PB-O5P
5	J	262	T64	O4'-C4R-C5'-O5'

There are no ring outliers.

2 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	J	262	T64	6	0
5	I	117	T64	11	0

5.5 Carbohydrates (i)

There are no monosaccharides 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 (i)

6.1 Protein, DNA and RNA chains (i)

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%)	0.08	2 (2%)	63	50	83, 101, 157, 201	0
1	E	99/139 (71%)	0.26	5 (5%)	28	20	79, 98, 160, 217	0
2	B	78/106 (73%)	0.10	2 (2%)	56	42	78, 94, 117, 132	0
2	F	84/106 (79%)	0.20	0	100	100	75, 93, 116, 133	0
3	C	108/133 (81%)	0.16	1 (0%)	84	73	78, 98, 144, 209	0
3	G	104/133 (78%)	0.01	1 (0%)	82	71	80, 101, 131, 176	0
4	D	96/129 (74%)	0.13	2 (2%)	63	50	76, 99, 146, 188	0
4	H	92/129 (71%)	0.08	1 (1%)	80	69	78, 99, 130, 185	0
5	I	144/145 (99%)	0.62	12 (8%)	11	10	131, 202, 236, 252	0
5	J	144/145 (99%)	0.45	7 (4%)	29	22	126, 202, 235, 253	0
All	All	1046/1304 (80%)	0.24	33 (3%)	47	36	75, 106, 224, 253	0

The worst 5 of 33 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	37	LYS	7.3
5	J	290	DT	4.9
5	I	11	DA	3.5
5	J	289	DA	3.3
5	I	97	DG	3.3

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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, 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	B-factors(Å^2)	Q<0.9
5	T64	J	262	40/41	0.47	0.38	227,251,272,276	0
5	T64	I	117	40/41	0.78	0.35	184,215,230,235	0

6.3 Carbohydrates (i)

There are no monosaccharides 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.