wwPDB X-ray Structure Validation Summary Report (i)

May 27, 2020 – 03:58 am BST

PDB ID	:	3WTP
Title	:	Crystal Structure of the heterotypic nucleosome containing human CENP-A and H3.3
Authors	:	Arimura, Y.; Shirayama, K.; Horikoshi, N.; Fujita, R.; Kagawa, W.; Fukagawa, T.; Almouzni, G.; Kurumizaka, H.
Deposited on	:	2014-04-14
Resolution	:	2.67 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 following versions of software and data (see references (1)) were used in the production of this report:

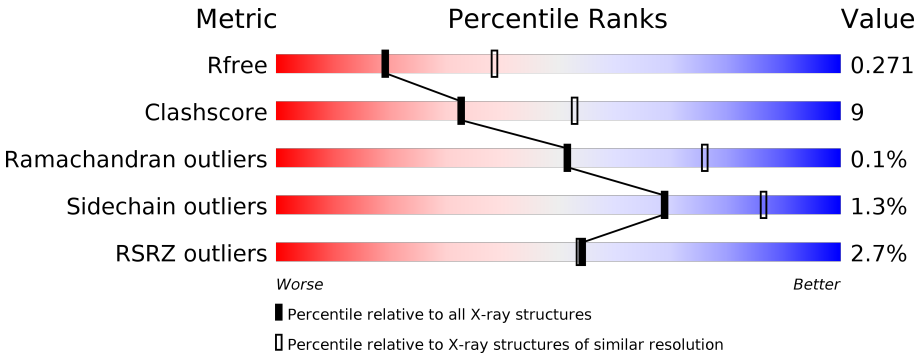
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.67 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.



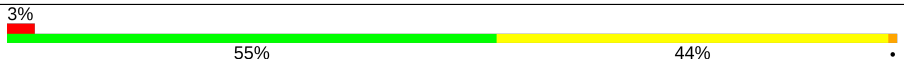
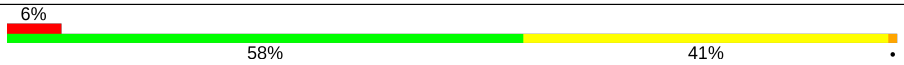
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3863 (2.70-2.66)
Clashscore	141614	4210 (2.70-2.66)
Ramachandran outliers	138981	4141 (2.70-2.66)
Sidechain outliers	138945	4141 (2.70-2.66)
RSRZ outliers	127900	3780 (2.70-2.66)

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 <=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	143	3% 43% 17% • 38%
2	B	106	61% 12% 26%
2	F	106	% 70% 10% 20%
3	C	133	2% 72% 8% • 20%
3	G	133	2% 65% 12% 23%
4	D	129	2% 64% 9% 27%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	H	129	
5	E	140	
6	I	146	
6	J	146	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11916 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-like centromeric protein A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	89	Total	C	N	O	S	0	0	0
			732	477	136	118	1			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP P49450
A	-1	SER	-	EXPRESSION TAG	UNP P49450
A	0	HIS	-	EXPRESSION TAG	UNP P49450

- 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	85	Total	C	N	O	S	0	0	0
			682	430	136	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	107	Total	C	N	O	0	0	0
			824	520	161	143			
3	G	103	Total	C	N	O	0	0	0
			794	500	155	139			

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	94	Total	C	N	O	S	0	0	0
			740	465	135	138	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 protein called Histone H3.3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	100	Total	C	N	O	S	0	0	0
			820	518	160	140	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	-4	GLY	-	EXPRESSION TAG	UNP P84243
E	-3	PRO	-	EXPRESSION TAG	UNP P84243
E	-2	GLY	-	EXPRESSION TAG	UNP P84243
E	-1	HIS	-	EXPRESSION TAG	UNP P84243

- Molecule 6 is a DNA chain called DNA (146-MER).

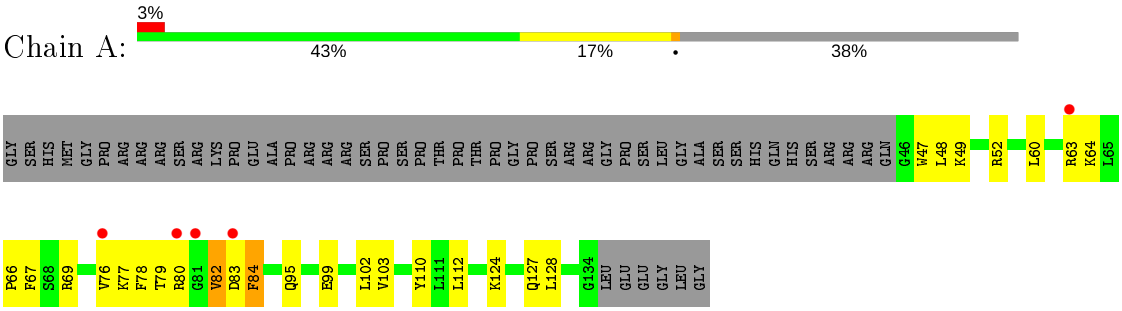
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	I	146	Total	C	N	O	P	0	0	0
			2990	1431	540	874	145			
6	J	146	Total	C	N	O	P	0	0	0
			2990	1431	540	874	145			

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.

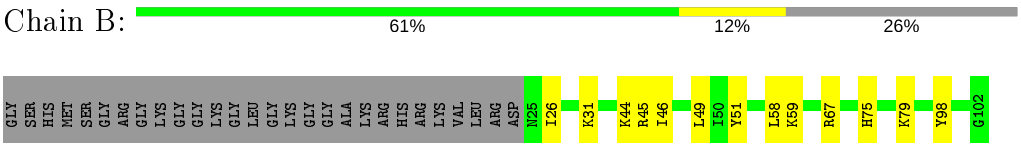
- Molecule 1: Histone H3-like centromeric protein A

Chain A:



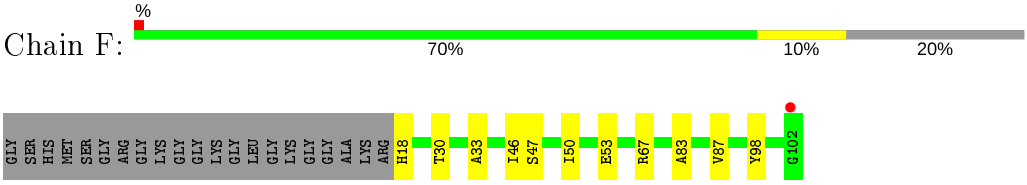
- Molecule 2: Histone H4

Chain B:



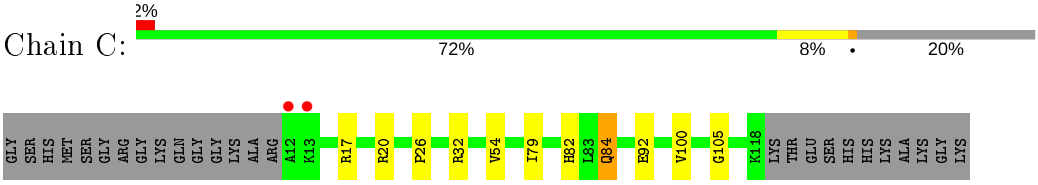
- Molecule 2: Histone H4

Chain F:



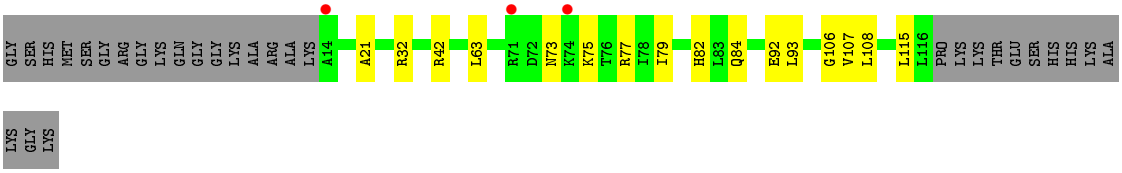
- Molecule 3: Histone H2A type 1-B/E

Chain C:



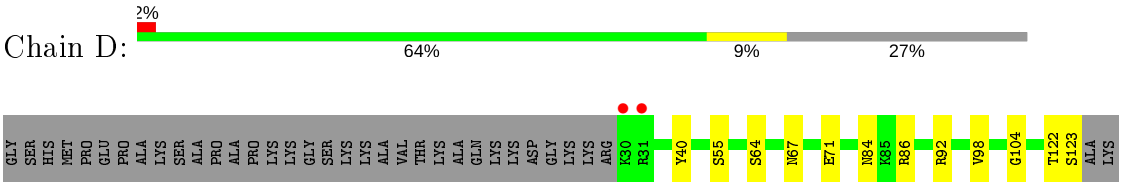
- Molecule 3: Histone H2A type 1-B/E

Chain G:



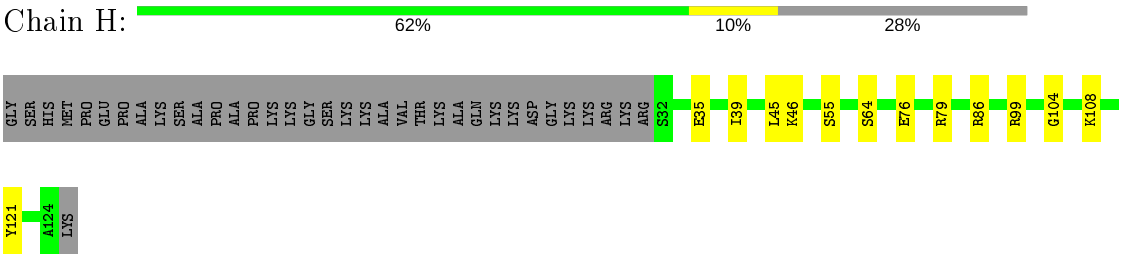
- Molecule 4: Histone H2B type 1-J

Chain D:



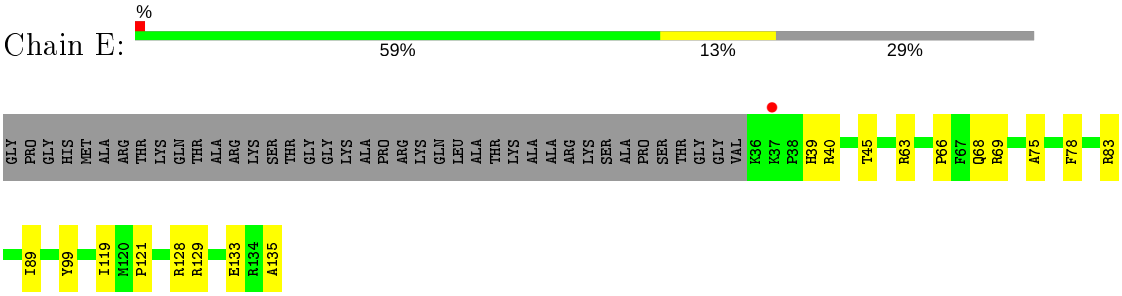
- Molecule 4: Histone H2B type 1-J

Chain H:



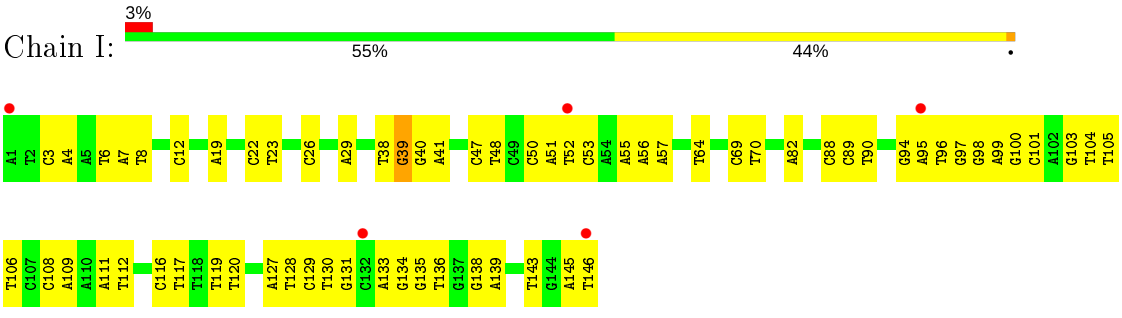
- Molecule 5: Histone H3.3

Chain E:



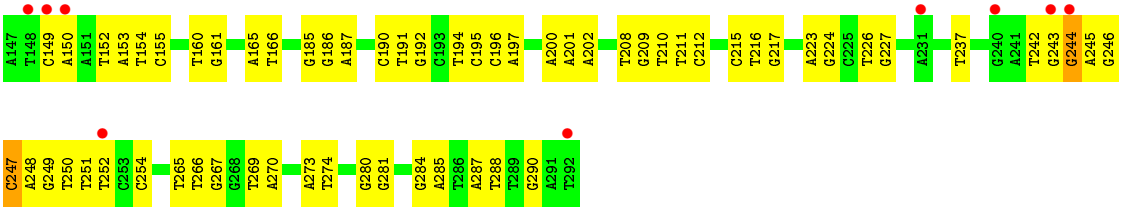
- Molecule 6: DNA (146-MER)

Chain I:



- Molecule 6: DNA (146-MER)

Chain J:


A147 T148 C149 A150 A151 T152 A153 T154 C155 T160 G161 A165 T166 G185 G186 A187 C190 T191 G192 C193 T194 C195 C196 A197 A200 A201 A202 T208 G209 T210 T211 C212 C215 T216 G217 A223 G224 C225 T226 G227 A231 T237 G240 A241 T242 G243 G244 A245 G246
C247 A248 G249 T250 T251 T252 C253 C254 T265 T266 G267 G268 T269 A270 A273 T274 G280 G281 G284 A285 T286 A287 T288 T289 G290 A291 T292

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	98.12Å 107.67Å 168.14Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.67 – 2.67 48.67 – 2.65	Depositor EDS
% Data completeness (in resolution range)	98.3 (48.67-2.67) 97.1 (48.67-2.65)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	2.47 (at 2.65Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.229 , 0.270 0.231 , 0.271	Depositor DCC
R_{free} test set	2577 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	46.8	Xtriage
Anisotropy	0.088	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 46.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	11916	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.65% 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

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	0.35	0/747	0.53	0/1006
2	B	0.33	0/626	0.55	0/837
2	F	0.43	0/690	0.60	0/923
3	C	0.38	0/834	0.55	0/1125
3	G	0.34	0/803	0.55	0/1084
4	D	0.41	0/751	0.59	0/1008
4	H	0.40	0/736	0.58	0/990
5	E	0.43	0/832	0.58	0/1115
6	I	0.79	0/3354	1.04	1/5175 (0.0%)
6	J	0.83	1/3354 (0.0%)	1.06	3/5175 (0.1%)
All	All	0.64	1/12727 (0.0%)	0.87	4/18438 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	J	244	DG	C3'-O3'	-5.74	1.36	1.44

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	J	247	DC	O4'-C1'-N1	7.55	113.29	108.00
6	J	212	DC	O4'-C1'-N1	5.31	111.72	108.00
6	J	191	DT	O4'-C1'-N1	5.30	111.71	108.00
6	I	39	DG	O4'-C1'-N9	5.29	111.70	108.00

There are no chirality outliers.

There are no planarity outliers.

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	732	0	770	26	0
2	B	619	0	659	12	0
2	F	682	0	729	12	0
3	C	824	0	884	8	0
3	G	794	0	846	19	0
4	D	740	0	766	10	0
4	H	725	0	745	11	0
5	E	820	0	865	19	0
6	I	2990	0	1652	72	0
6	J	2990	0	1652	54	0
All	All	11916	0	9568	186	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 9.

The worst 5 of 186 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:I:128:DT:H2"	6:I:129:DC:H5"	1.53	0.90
4:D:86:ARG:HH12	6:I:39:DG:H3'	1.36	0.89
5:E:69:ARG:NH2	6:I:90:DT:OP2	2.07	0.88
1:A:79:THR:HG21	1:A:82:VAL:HG22	1.59	0.84
6:J:194:DT:H2"	6:J:195:DC:H5"	1.60	0.83

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	87/143 (61%)	84 (97%)	2 (2%)	1 (1%)	14 31

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	76/106 (72%)	75 (99%)	1 (1%)	0	100	100
2	F	83/106 (78%)	80 (96%)	3 (4%)	0	100	100
3	C	105/133 (79%)	102 (97%)	3 (3%)	0	100	100
3	G	101/133 (76%)	98 (97%)	3 (3%)	0	100	100
4	D	92/129 (71%)	89 (97%)	3 (3%)	0	100	100
4	H	91/129 (70%)	90 (99%)	1 (1%)	0	100	100
5	E	98/140 (70%)	97 (99%)	1 (1%)	0	100	100
All	All	733/1019 (72%)	715 (98%)	17 (2%)	1 (0%)	51	76

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	82	VAL

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	75/120 (62%)	72 (96%)	3 (4%)	31	57
2	B	63/81 (78%)	62 (98%)	1 (2%)	62	83
2	F	70/81 (86%)	70 (100%)	0	100	100
3	C	84/102 (82%)	83 (99%)	1 (1%)	71	87
3	G	81/102 (79%)	81 (100%)	0	100	100
4	D	81/107 (76%)	79 (98%)	2 (2%)	47	74
4	H	79/107 (74%)	78 (99%)	1 (1%)	69	86
5	E	85/112 (76%)	85 (100%)	0	100	100
All	All	618/812 (76%)	610 (99%)	8 (1%)	69	86

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	49	LEU
4	H	39	ILE
4	D	122	THR
1	A	84	PHE
3	C	84	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	C	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

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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^{th} 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	89/143 (62%)	0.18	5 (5%) 24 22	27, 40, 70, 89	0
2	B	78/106 (73%)	0.08	0 100 100	29, 38, 52, 68	0
2	F	85/106 (80%)	0.04	1 (1%) 79 79	18, 24, 36, 68	0
3	C	107/133 (80%)	0.08	2 (1%) 66 67	19, 29, 45, 66	0
3	G	103/133 (77%)	0.13	3 (2%) 51 51	24, 36, 67, 79	0
4	D	94/129 (72%)	0.15	2 (2%) 63 63	20, 29, 49, 80	0
4	H	93/129 (72%)	0.13	0 100 100	20, 33, 58, 71	0
5	E	100/140 (71%)	0.02	1 (1%) 82 82	17, 26, 51, 80	0
6	I	146/146 (100%)	0.28	5 (3%) 45 44	38, 68, 121, 137	0
6	J	146/146 (100%)	0.34	9 (6%) 20 18	41, 72, 126, 132	0
All	All	1041/1311 (79%)	0.16	28 (2%) 54 54	17, 37, 102, 137	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	G	14	ALA	5.9
4	D	31	ARG	3.9
1	A	80	ARG	3.7
2	F	102	GLY	3.6
4	D	30	LYS	3.5

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.