wwPDB X-ray Structure Validation Summary Report (i)

Oct 31, 2023 – 07:28 PM JST

PDB ID	:	5GXQ
Title	:	The crystal structure of the nucleosome containing H3.6
Authors	:	Taguchi, H.; Xie, Y.; Horikoshi, N.; Kurumizaka, H.
Deposited on	:	2016-09-19
Resolution	:	2.85 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 (i)) were used in the production of this report:

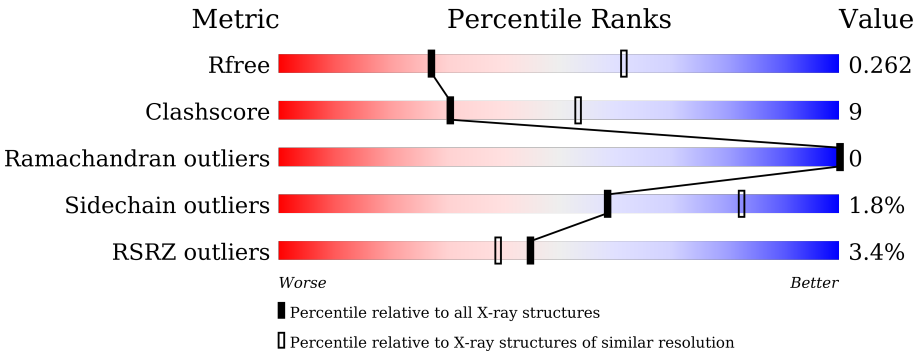
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.85 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.



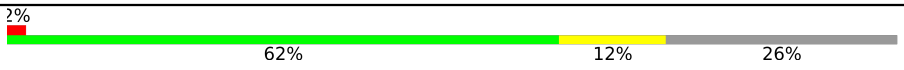
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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	58% 12% 30%
1	E	139	3% 58% 12% 29%
2	B	106	61% 11% 26%
2	F	106	2% 73% 7% 21%
3	C	133	2% 71% 9% 19%
3	G	133	61% 17% 22%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	129	
4	H	129	
5	I	146	
5	J	146	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 11981 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 H3.6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	97	Total	C	N	O	S	0	0	0
			790	499	151	138	2			
1	E	99	Total	C	N	O	S	0	0	0
			805	508	154	141	2			

- 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	96	Total	C	N	O	S	0	0	0
			755	474	138	141	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

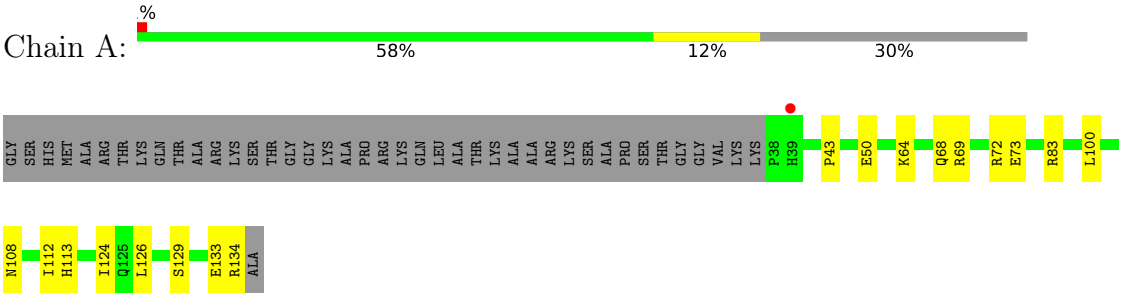
- Molecule 5 is a DNA chain called DNA (146-MER).

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

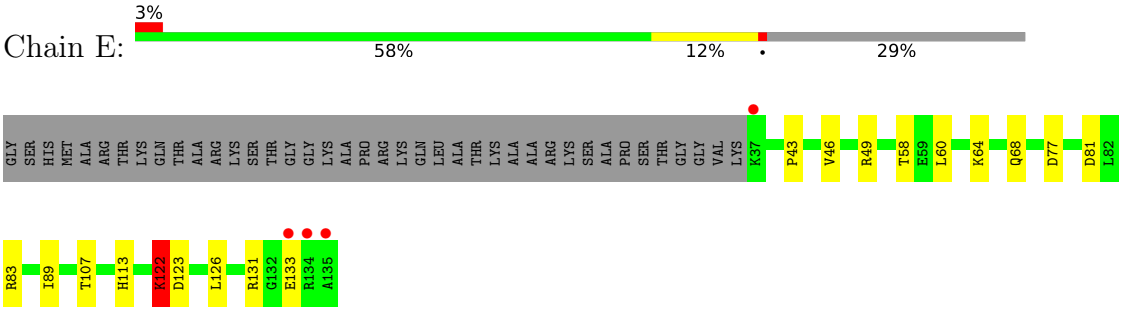
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, 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.

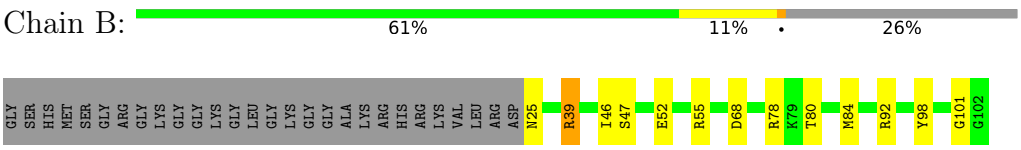
• Molecule 1: Histone H3.6



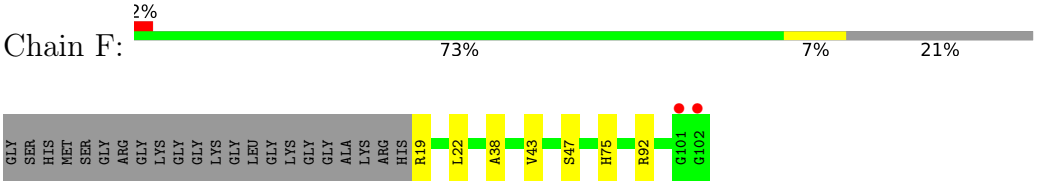
• Molecule 1: Histone H3.6



• Molecule 2: Histone H4



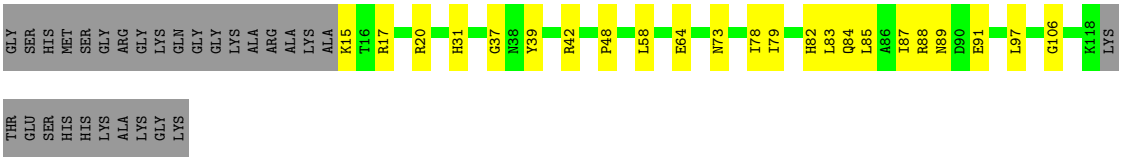
• Molecule 2: Histone H4



• Molecule 3: Histone H2A type 1-B/E

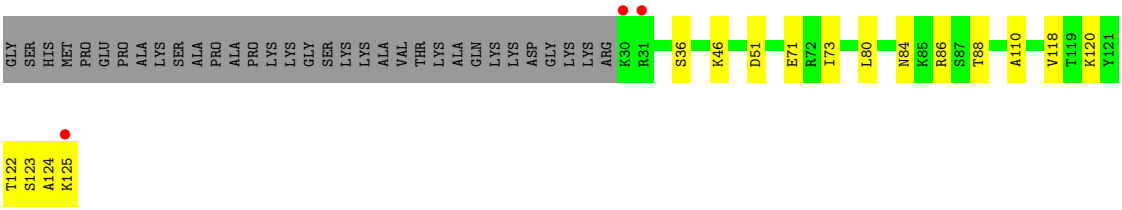
• Molecule 3: Histone H2A type 1-B/E

Chain G:



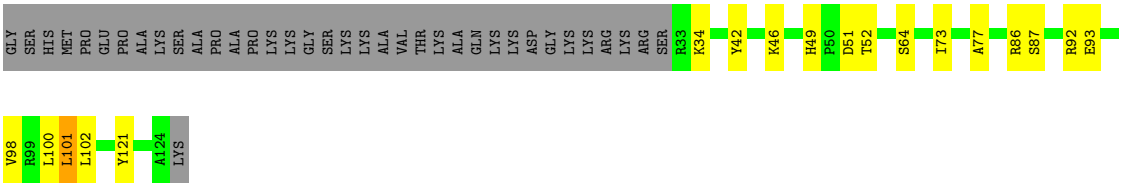
• Molecule 4: Histone H2B type 1-J

Chain D:



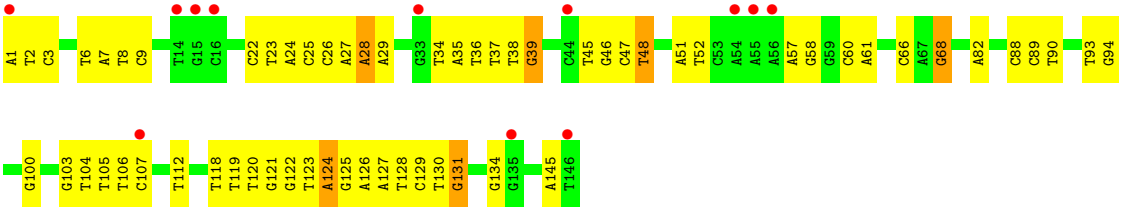
• Molecule 4: Histone H2B type 1-J

Chain H:



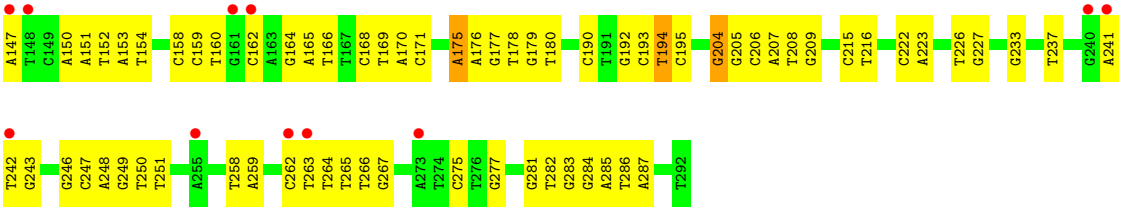
• Molecule 5: DNA (146-MER)

Chain I:



• Molecule 5: DNA (146-MER)

Chain J:


A147 T148 C149 A150 A151 T152 A153 T154 C158 C159 T160 G161 C162 A163 G164 A165 T166 T167 C168 T169 A170 C171 A175 A176 G177 T178 G179 T180 C190 T191 G192 C193 T194 C195 G204 G205 C206 A207 T208 G209 C215 T216 C222 A223 T226 G227 G233 T237 G240 A241
T242 G243 G246 C247 A248 G249 T250 T251 A255 T258 A259 C262 T263 T264 T265 T266 G267 A273 T274 C275 T276 G277 G281 T282 G283 G284 A285 T286 A287 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	106.03Å 109.76Å 181.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.13 – 2.85 40.64 – 2.84	Depositor EDS
% Data completeness (in resolution range)	99.6 (38.13-2.85) 96.8 (40.64-2.84)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.33 (at 2.86Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.219 , 0.262 0.219 , 0.262	Depositor DCC
R_{free} test set	2561 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	58.3	Xtriage
Anisotropy	0.289	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 55.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.026 for k,h,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	11981	wwPDB-VP
Average B, all atoms (Å^2)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.91% 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)

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	0.45	0/803	0.63	0/1079
1	E	0.67	2/818 (0.2%)	1.03	3/1098 (0.3%)
2	B	0.45	0/626	0.71	1/837 (0.1%)
2	F	0.57	0/680	0.70	0/908
3	C	0.51	0/845	0.65	0/1139
3	G	0.44	0/815	0.65	0/1100
4	D	0.51	0/766	0.62	0/1026
4	H	0.51	0/730	0.71	1/982 (0.1%)
5	I	0.82	2/3354 (0.1%)	1.09	7/5175 (0.1%)
5	J	0.84	3/3354 (0.1%)	1.08	2/5175 (0.0%)
All	All	0.70	7/12791 (0.1%)	0.94	14/18519 (0.1%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	I	68	DG	C3'-O3'	-6.86	1.35	1.44
5	J	277	DG	C3'-O3'	-6.36	1.35	1.44
1	E	77	ASP	CB-CG	5.83	1.64	1.51
5	J	175	DA	C3'-O3'	-5.68	1.36	1.44
5	J	204	DG	C3'-O3'	-5.21	1.37	1.44

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	77	ASP	CB-CG-OD1	17.53	134.08	118.30
1	E	77	ASP	CB-CG-OD2	-14.70	105.07	118.30
5	I	131	DG	O4'-C4'-C3'	-7.80	101.32	106.00
1	E	122	LYS	CD-CE-NZ	-7.58	94.27	111.70
2	B	39	ARG	NE-CZ-NH1	-6.89	116.85	120.30

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	790	0	819	17	0
1	E	805	0	836	12	0
2	B	619	0	659	10	0
2	F	673	0	722	5	0
3	C	835	0	897	11	0
3	G	805	0	861	16	0
4	D	755	0	784	14	0
4	H	719	0	740	10	0
5	I	2990	0	1652	61	0
5	J	2990	0	1652	62	0
All	All	11981	0	9622	176	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 176 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:36:DT:H2”	5:I:37:DT:H5”	1.61	0.81
5:I:47:DC:H2”	5:I:48:DT:H5”	1.63	0.78
5:I:131:DG:H1	5:J:162:DC:H42	1.29	0.77
3:C:39:TYR:OH	4:D:71:GLU:OE1	2.04	0.76
5:I:26:DC:H2”	5:I:27:DA:C8	2.21	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%)	94 (99%)	1 (1%)	0	100	100
1	E	97/139 (70%)	93 (96%)	4 (4%)	0	100	100
2	B	76/106 (72%)	74 (97%)	2 (3%)	0	100	100
2	F	82/106 (77%)	81 (99%)	1 (1%)	0	100	100
3	C	106/133 (80%)	105 (99%)	1 (1%)	0	100	100
3	G	102/133 (77%)	99 (97%)	3 (3%)	0	100	100
4	D	94/129 (73%)	89 (95%)	5 (5%)	0	100	100
4	H	90/129 (70%)	86 (96%)	4 (4%)	0	100	100
All	All	742/1014 (73%)	721 (97%)	21 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	83/112 (74%)	83 (100%)	0	100	100
1	E	84/112 (75%)	82 (98%)	2 (2%)	49	77
2	B	63/81 (78%)	62 (98%)	1 (2%)	62	84
2	F	69/81 (85%)	68 (99%)	1 (1%)	67	86
3	C	85/102 (83%)	84 (99%)	1 (1%)	71	89
3	G	83/102 (81%)	81 (98%)	2 (2%)	49	77
4	D	82/107 (77%)	82 (100%)	0	100	100
4	H	78/107 (73%)	74 (95%)	4 (5%)	24	52
All	All	627/804 (78%)	616 (98%)	11 (2%)	59	82

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

Mol	Chain	Res	Type
4	H	34	LYS
4	H	87	SER
4	H	101	LEU
4	H	100	LEU
2	F	47	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	76	GLN
3	G	31	HIS
3	G	73	ASN
3	G	82	HIS
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)

There are no non-standard protein/DNA/RNA residues in this entry.

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

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	97/139 (69%)	0.01	1 (1%) 82 81	31, 44, 67, 93	0
1	E	99/139 (71%)	-0.01	4 (4%) 38 32	26, 36, 61, 92	0
2	B	78/106 (73%)	-0.11	0 100 100	33, 42, 55, 58	0
2	F	84/106 (79%)	-0.10	2 (2%) 59 56	26, 35, 52, 77	0
3	C	108/133 (81%)	-0.09	3 (2%) 53 48	26, 39, 66, 126	0
3	G	104/133 (78%)	-0.18	0 100 100	32, 45, 67, 86	0
4	D	96/129 (74%)	0.05	3 (3%) 49 44	29, 42, 73, 118	0
4	H	92/129 (71%)	-0.14	0 100 100	31, 45, 65, 89	0
5	I	146/146 (100%)	0.44	12 (8%) 11 8	45, 97, 129, 143	0
5	J	146/146 (100%)	0.35	11 (7%) 14 10	50, 96, 131, 144	0
All	All	1050/1306 (80%)	0.06	36 (3%) 45 39	26, 47, 119, 144	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	146	DT	8.0
5	J	148	DT	7.6
4	D	125	LYS	6.6
1	E	37	LYS	5.4
3	C	13	LYS	5.0

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 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.