wwPDB X-ray Structure Validation Summary Report (i)

Oct 19, 2016 – 10:58 AM EDT

PDB ID	:	5BT1
Title	:	histone chaperone Hif1 playing with histone H2A-H2B dimer
Authors	:	Liu, H.; Zhang, M.; Gao, Y.; Teng, M.; Niu, L.
Deposited on	:	2015-06-02
Resolution	:	2.62 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
http://wwpdb.org/validation/2016/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:

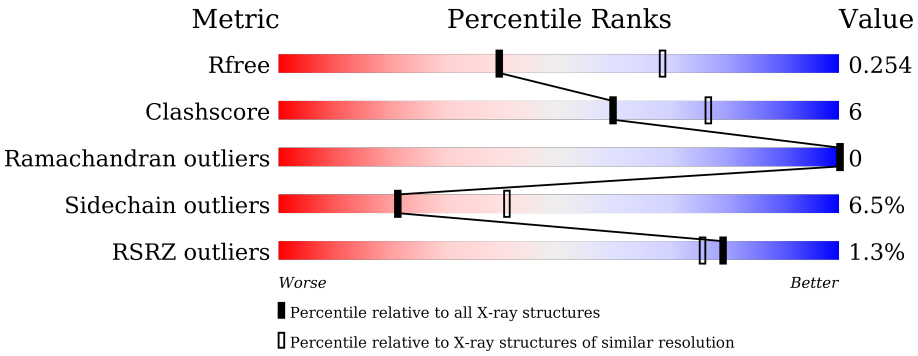
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 2.62 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}	91344	2700 (2.64-2.60)
Clashscore	102246	3065 (2.64-2.60)
Ramachandran outliers	100387	3015 (2.64-2.60)
Sidechain outliers	100360	3015 (2.64-2.60)
RSRZ outliers	91569	2706 (2.64-2.60)

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. 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	393	% 51% 10% • 38%
1	B	393	% 52% 11% • 37%
2	C	143	% 52% 7% • 41%
3	D	142	53% 10% • 37%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5317 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 HAT1-interacting factor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	244	Total	C	N	O	S	0	0	0
			1932	1219	334	371	8			
1	B	249	Total	C	N	O	S	0	0	0
			1990	1256	346	380	8			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	386	LEU	-	expression tag	UNP Q12373
A	387	GLU	-	expression tag	UNP Q12373
A	388	HIS	-	expression tag	UNP Q12373
A	389	HIS	-	expression tag	UNP Q12373
A	390	HIS	-	expression tag	UNP Q12373
A	391	HIS	-	expression tag	UNP Q12373
A	392	HIS	-	expression tag	UNP Q12373
A	393	HIS	-	expression tag	UNP Q12373
B	386	LEU	-	expression tag	UNP Q12373
B	387	GLU	-	expression tag	UNP Q12373
B	388	HIS	-	expression tag	UNP Q12373
B	389	HIS	-	expression tag	UNP Q12373
B	390	HIS	-	expression tag	UNP Q12373
B	391	HIS	-	expression tag	UNP Q12373
B	392	HIS	-	expression tag	UNP Q12373
B	393	HIS	-	expression tag	UNP Q12373

- Molecule 2 is a protein called Histone H2A.1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	85	Total	C	N	O	0	0	0
			650	407	128	115			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-11	MET	-	initiating methionine	UNP P04911
C	-10	GLY	-	expression tag	UNP P04911
C	-9	HIS	-	expression tag	UNP P04911
C	-8	HIS	-	expression tag	UNP P04911
C	-7	HIS	-	expression tag	UNP P04911
C	-6	HIS	-	expression tag	UNP P04911
C	-5	HIS	-	expression tag	UNP P04911
C	-4	HIS	-	expression tag	UNP P04911
C	-3	GLY	-	expression tag	UNP P04911
C	-2	SER	-	expression tag	UNP P04911
C	-1	HIS	-	expression tag	UNP P04911

- Molecule 3 is a protein called Histone H2B.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	90	Total	C	N	O	S	0	0	0
			698	440	120	137	1			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-11	MET	-	initiating methionine	UNP P02293
D	-10	GLY	-	expression tag	UNP P02293
D	-9	HIS	-	expression tag	UNP P02293
D	-8	HIS	-	expression tag	UNP P02293
D	-7	HIS	-	expression tag	UNP P02293
D	-6	HIS	-	expression tag	UNP P02293
D	-5	HIS	-	expression tag	UNP P02293
D	-4	HIS	-	expression tag	UNP P02293
D	-3	GLY	-	expression tag	UNP P02293
D	-2	SER	-	expression tag	UNP P02293
D	-1	HIS	-	expression tag	UNP P02293

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	20	Total	O	0	0
			20	20		
4	B	15	Total	O	0	0
			15	15		
4	C	4	Total	O	0	0
			4	4		

Continued on next page...

Continued from previous page...

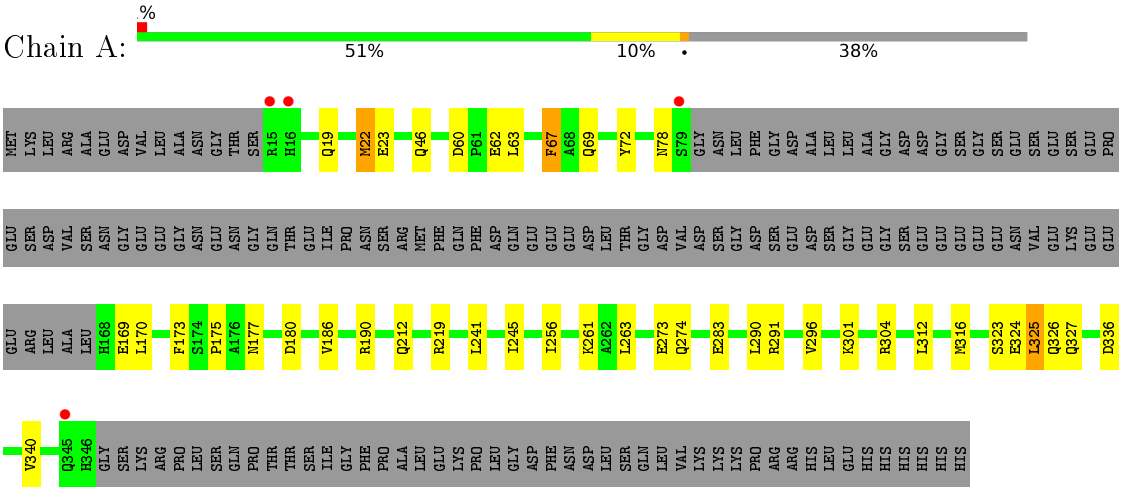
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	D	8	Total O 8 8	0	0

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 errors 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.

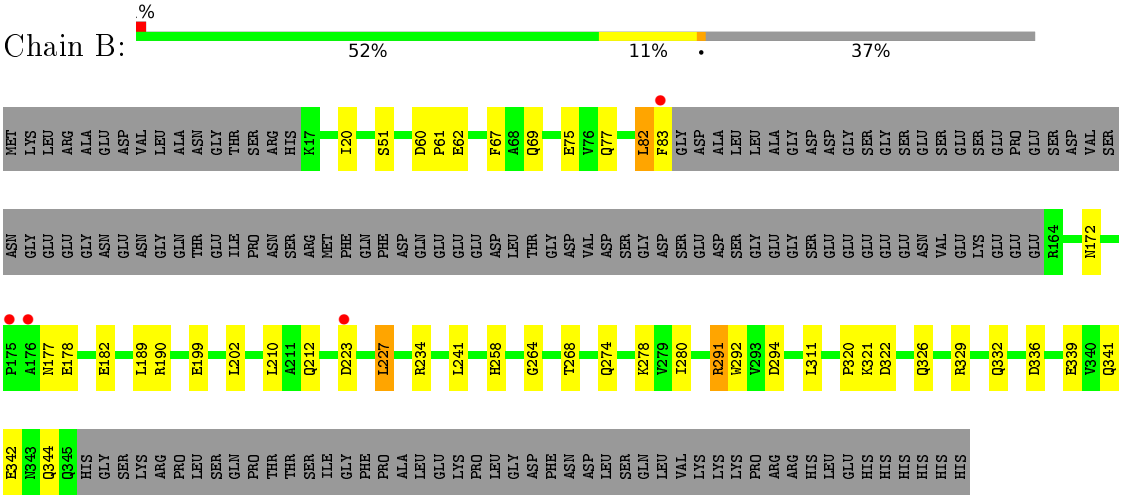
- Molecule 1: HAT1-interacting factor 1

Chain A:



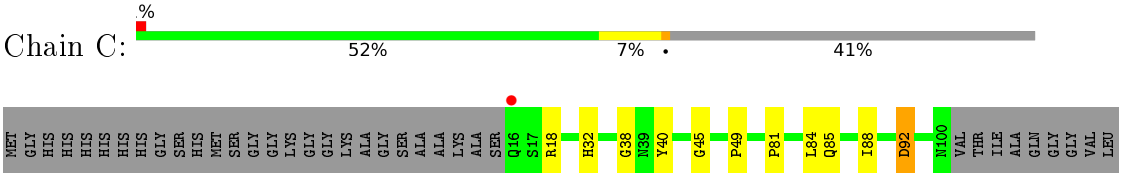
- Molecule 1: HAT1-interacting factor 1

Chain B:



- Molecule 2: Histone H2A.1

Chain C:



- Molecule 3: Histone H2B.1

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	102.67Å 48.36Å 160.71Å 90.00° 107.05° 90.00°	Depositor
Resolution (Å)	43.38 – 2.62 43.38 – 2.62	Depositor EDS
% Data completeness (in resolution range)	98.1 (43.38-2.62) 98.0 (43.38-2.62)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.48 (at 2.61Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.207 , 0.262 0.203 , 0.254	Depositor DCC
R_{free} test set	1162 reflections (5.40%)	DCC
Wilson B-factor (Å^2)	44.5	Xtriage
Anisotropy	0.146	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 37.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.022 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5317	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.10% 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	0.44	0/1962	0.58	1/2649 (0.0%)
1	B	0.41	0/2022	0.56	0/2729
2	C	0.36	0/658	0.51	0/889
3	D	0.41	0/708	0.52	0/955
All	All	0.42	0/5350	0.55	1/7222 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	325	LEU	CA-CB-CG	5.41	127.75	115.30

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	1932	0	1892	20	0
1	B	1990	0	1954	29	0
2	C	650	0	670	8	0
3	D	698	0	712	8	0
4	A	20	0	0	0	0
4	B	15	0	0	0	0
4	C	4	0	0	0	0
4	D	8	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	5317	0	5228	60	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 6.

The worst 5 of 60 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)
3:D:102:ARG:HH11	3:D:102:ARG:HG2	1.49	0.76
1:B:227:LEU:H	1:B:227:LEU:HD12	1.53	0.73
1:A:296:VAL:HG13	1:A:301:LYS:HE3	1.73	0.70
1:B:291:ARG:HG2	1:B:292:TRP:CD1	2.26	0.70
1:B:326:GLN:HE22	1:B:329:ARG:HH11	1.39	0.69

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	240/393 (61%)	229 (95%)	11 (5%)	0	100	100
1	B	245/393 (62%)	242 (99%)	3 (1%)	0	100	100
2	C	83/143 (58%)	82 (99%)	1 (1%)	0	100	100
3	D	88/142 (62%)	85 (97%)	3 (3%)	0	100	100
All	All	656/1071 (61%)	638 (97%)	18 (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	202/341 (59%)	187 (93%)	15 (7%)	17	33
1	B	209/341 (61%)	196 (94%)	13 (6%)	23	44
2	C	64/108 (59%)	63 (98%)	1 (2%)	70	88
3	D	77/118 (65%)	70 (91%)	7 (9%)	12	22
All	All	552/908 (61%)	516 (94%)	36 (6%)	21	41

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	77	GLN
1	B	199	GLU
3	D	102	ARG
1	B	172	ASN
1	B	223	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	78	ASN
1	B	177	ASN
2	C	85	GLN
1	B	47	GLN
1	B	326	GLN

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

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	244/393 (62%)	0.12	4 (1%) 74 69	26, 42, 73, 89	0
1	B	249/393 (63%)	0.18	4 (1%) 74 69	25, 42, 69, 92	0
2	C	85/143 (59%)	0.06	1 (1%) 81 77	28, 53, 75, 83	0
3	D	90/142 (63%)	0.08	0 100 100	31, 48, 61, 69	0
All	All	668/1071 (62%)	0.13	9 (1%) 79 75	25, 44, 71, 92	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	83	PHE	4.8
1	A	15	ARG	4.3
1	A	16	HIS	3.7
1	B	175	PRO	3.6
2	C	16	GLN	2.8

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

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.