



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 13, 2016 – 11:10 AM EDT

PDB ID : 5B40  
Title : The nucleosome structure containing H2B-K120 and H4-K31 monoubiquitinations  
Authors : Machida, S.; Sekine, S.; Nishiyama, Y.; Horikoshi, N.; Kurumizaka, H.  
Deposited on : 2016-03-22  
Resolution : 3.33 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  
<http://wwpdb.org/validation/2016/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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027790
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-20027790

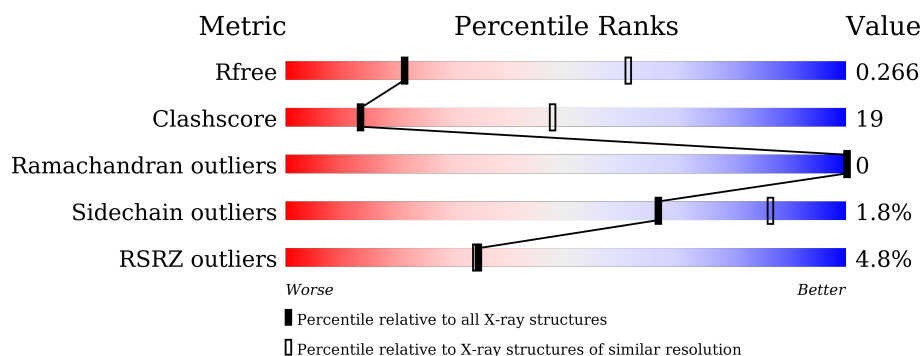
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.33 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	1004 (3.40-3.28)
Clashscore	102246	1072 (3.40-3.28)
Ramachandran outliers	100387	1055 (3.40-3.28)
Sidechain outliers	100360	1054 (3.40-3.28)
RSRZ outliers	91569	1009 (3.40-3.28)

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. 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>6%</div> <div> <div>40%</div> <div>26%</div> <div>•</div> <div>32%</div> </div> </div>
1	E	139	<div> <div>8%</div> <div> <div>37%</div> <div>29%</div> <div>•</div> <div>32%</div> </div> </div>
2	B	106	<div> <div>6%</div> <div> <div>46%</div> <div>27%</div> <div>26%</div> </div> </div>
2	F	106	<div> <div>5%</div> <div> <div>45%</div> <div>30%</div> <div>25%</div> </div> </div>
3	C	133	<div> <div>2%</div> <div> <div>55%</div> <div>23%</div> <div>•</div> <div>22%</div> </div> </div>
3	G	133	<div> <div>3%</div> <div> <div>47%</div> <div>30%</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><div></div><div></div></div><div><div>2%</div><div>49%</div><div>23%</div><div>28%</div></div></div>
4	H	129	<div><div><div></div><div></div><div></div></div><div><div>4%</div><div>49%</div><div>22%</div><div>29%</div></div></div>
5	I	146	<div><div><div></div><div></div><div></div></div><div><div>%</div><div>26%</div><div>69%</div><div>5%</div></div></div>
5	J	146	<div><div><div></div><div></div><div></div></div><div><div>%</div><div>29%</div><div>68%</div><div>.</div></div></div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11817 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 H3.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	94	Total	C	N	O	S	0	0	0
			772	488	147	135	2			
1	E	95	Total	C	N	O	S	0	0	0
			778	491	148	137	2			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP Q71DI3
A	-2	SER	-	expression tag	UNP Q71DI3
A	-1	HIS	-	expression tag	UNP Q71DI3
A	110	ALA	CYS	engineered mutation	UNP Q71DI3
E	-3	GLY	-	expression tag	UNP Q71DI3
E	-2	SER	-	expression tag	UNP Q71DI3
E	-1	HIS	-	expression tag	UNP Q71DI3
E	110	ALA	CYS	engineered mutation	UNP Q71DI3

- 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
			616	388	119	107	2			
2	F	80	Total	C	N	O	S	0	0	0
			638	402	124	110	2			

There are 8 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
B	31	CYS	LYS	engineered mutation	UNP P62805

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	-3	GLY	-	expression tag	UNP P62805
F	-2	SER	-	expression tag	UNP P62805
F	-1	HIS	-	expression tag	UNP P62805
F	31	CYS	LYS	engineered mutation	UNP P62805

- Molecule 3 is a protein called Histone H2A type 1-B/E.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	104	Total	C	N	O	0	0	0
			801	505	156	140			
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
			722	453	129	137	3			
4	H	91	Total	C	N	O	S	0	0	0
			705	444	124	134	3			

There are 8 discrepancies between the modelled and reference sequences:

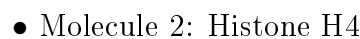
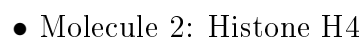
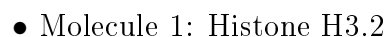
Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	GLY	-	expression tag	UNP P06899
D	-2	SER	-	expression tag	UNP P06899
D	-1	HIS	-	expression tag	UNP P06899
D	120	CYS	LYS	engineered mutation	UNP P06899
H	-3	GLY	-	expression tag	UNP P06899
H	-2	SER	-	expression tag	UNP P06899
H	-1	HIS	-	expression tag	UNP P06899
H	120	CYS	LYS	engineered mutation	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			

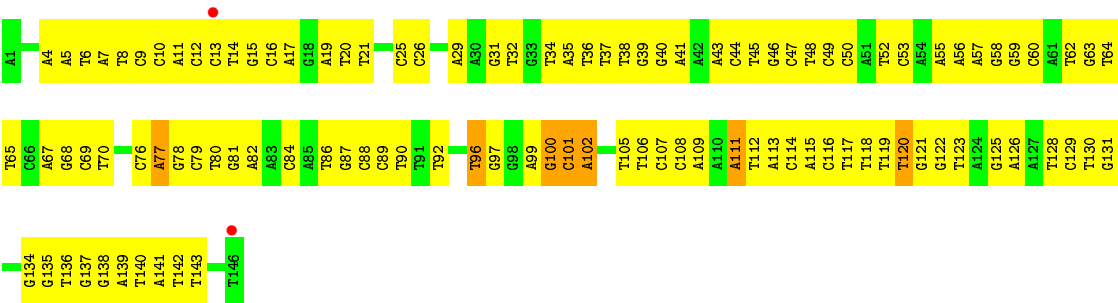


- Molecule 1: Histone H3.2

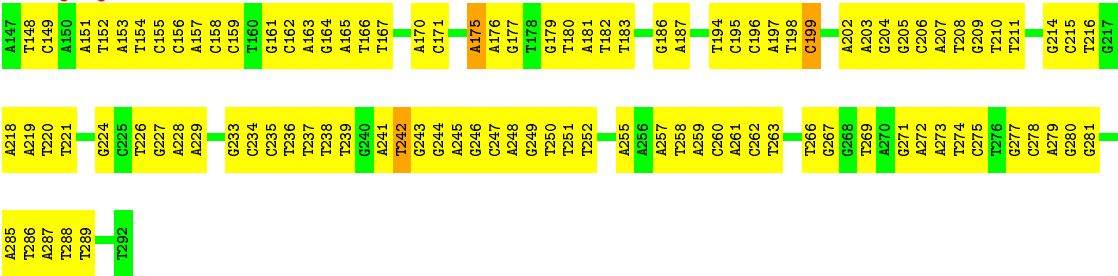








● Molecule 5: DNA (146-MER)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	100.42Å 100.42Å 186.03Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.39 – 3.33 44.18 – 3.33	Depositor EDS
% Data completeness (in resolution range)	97.6 (39.39-3.33) 97.7 (44.18-3.33)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.79 (at 3.32Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.201 , 0.263 0.201 , 0.266	Depositor DCC
$R_{free}$ test set	1995 reflections (6.66%)	DCC
Wilson B-factor (Å <sup>2</sup> )	101.4	Xtriage
Anisotropy	0.013	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 74.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	0.068 for -h,-k,l 0.458 for h,-h-k,-l 0.068 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11817	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	122.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.55% 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

### 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	0.66	0/782	0.85	0/1049
1	E	0.69	0/788	0.92	0/1056
2	B	0.72	0/623	0.93	0/834
2	F	0.71	0/645	0.89	1/865 (0.1%)
3	C	0.67	0/811	0.90	0/1096
3	G	0.67	0/815	1.00	1/1100 (0.1%)
4	D	0.70	0/733	0.81	0/987
4	H	0.67	0/716	0.86	1/965 (0.1%)
5	I	0.91	3/3354 (0.1%)	1.10	10/5175 (0.2%)
5	J	0.89	2/3354 (0.1%)	1.07	2/5175 (0.0%)
All	All	0.81	5/12621 (0.0%)	1.01	15/18302 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	E	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	J	175	DA	C3'-O3'	-6.71	1.35	1.44
5	I	81	DG	C3'-O3'	-6.45	1.35	1.44
5	J	224	DG	C3'-O3'	-6.12	1.35	1.44
5	I	77	DA	C3'-O3'	-6.04	1.36	1.44
5	I	49	DC	C1'-N1	5.26	1.56	1.49

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	I	120	DT	O4'-C1'-N1	6.84	112.79	108.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	J	199	DC	O4'-C1'-N1	5.82	112.08	108.00
3	G	108	LEU	CA-CB-CG	5.56	128.09	115.30
5	I	50	DC	O5'-P-OP1	-5.46	100.78	105.70
5	I	101	DC	P-O5'-C5'	-5.45	112.17	120.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	115	LYS	Peptide

## 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	772	0	811	39	0
1	E	778	0	816	53	0
2	B	616	0	651	27	0
2	F	638	0	676	35	0
3	C	801	0	853	37	0
3	G	805	0	861	45	0
4	D	722	0	737	34	0
4	H	705	0	719	32	0
5	I	2990	0	1652	108	0
5	J	2990	0	1652	107	0
All	All	11817	0	9428	397	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 19.

The worst 5 of 397 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)
1:E:49:ARG:HD3	1:E:52:ARG:HE	1.14	1.07
5:J:182:DT:H2''	5:J:183:DT:H5''	1.52	0.91

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:62:ILE:HD13	2:F:29:ILE:HD12	1.54	0.89
3:G:104:GLN:HE21	4:H:57:LYS:HB3	1.36	0.89
3:G:15:LYS:O	3:G:20:ARG:NH2	2.08	0.87

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	92/139 (66%)	87 (95%)	5 (5%)	0	100	100
1	E	93/139 (67%)	86 (92%)	7 (8%)	0	100	100
2	B	76/106 (72%)	68 (90%)	8 (10%)	0	100	100
2	F	78/106 (74%)	72 (92%)	6 (8%)	0	100	100
3	C	102/133 (77%)	97 (95%)	5 (5%)	0	100	100
3	G	102/133 (77%)	98 (96%)	4 (4%)	0	100	100
4	D	91/129 (70%)	89 (98%)	2 (2%)	0	100	100
4	H	89/129 (69%)	89 (100%)	0	0	100	100
All	All	723/1014 (71%)	686 (95%)	37 (5%)	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	81/112 (72%)	77 (95%)	4 (5%)	31	69
1	E	81/112 (72%)	77 (95%)	4 (5%)	31	69
2	B	63/81 (78%)	63 (100%)	0	100	100
2	F	66/81 (82%)	66 (100%)	0	100	100
3	C	82/102 (80%)	81 (99%)	1 (1%)	78	90
3	G	83/102 (81%)	82 (99%)	1 (1%)	78	90
4	D	79/107 (74%)	79 (100%)	0	100	100
4	H	77/107 (72%)	76 (99%)	1 (1%)	76	90
All	All	612/804 (76%)	601 (98%)	11 (2%)	66	86

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

Mol	Chain	Res	Type
3	C	74	LYS
1	E	52	ARG
1	E	120	MET
1	A	117	VAL
1	E	117	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	76	GLN
4	D	63	ASN
1	E	93	GLN
3	G	104	GLN

### 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 [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	94/139 (67%)	0.71	8 (8%) 13 13	58, 80, 118, 144	0
1	E	95/139 (68%)	0.79	11 (11%) 6 6	60, 83, 115, 142	0
2	B	78/106 (73%)	0.72	6 (7%) 16 16	52, 75, 97, 102	0
2	F	80/106 (75%)	0.71	5 (6%) 23 24	51, 77, 103, 144	0
3	C	104/133 (78%)	0.51	3 (2%) 55 55	51, 75, 106, 114	0
3	G	104/133 (78%)	0.41	4 (3%) 44 43	54, 78, 109, 128	0
4	D	93/129 (72%)	0.51	3 (3%) 51 51	53, 76, 109, 136	0
4	H	91/129 (70%)	0.62	5 (5%) 29 28	55, 75, 100, 122	0
5	I	146/146 (100%)	-0.51	2 (1%) 78 79	119, 163, 204, 221	0
5	J	146/146 (100%)	-0.52	2 (1%) 78 79	115, 160, 200, 223	0
All	All	1031/1306 (78%)	0.29	49 (4%) 34 34	51, 86, 183, 223	0

The worst 5 of 49 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	22	LEU	5.5
1	E	119	ILE	5.0
2	F	44	LYS	4.2
1	E	44	GLY	4.0
1	E	118	THR	3.8

### 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

There are no such residues in this entry.