



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 27, 2020 – 01:46 am BST

PDB ID : 3AN2  
Title : The structure of the centromeric nucleosome containing CENP-A  
Authors : Tachiwana, H.; Kagawa, W.; Shiga, T.; Saito, K.; Osakabe, A.; Hayashi-Takanaka, Y.; Park, S.-Y.; Kimura, H.; Kurumizaka, H.  
Deposited on : 2010-08-27  
Resolution : 3.60 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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)  
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

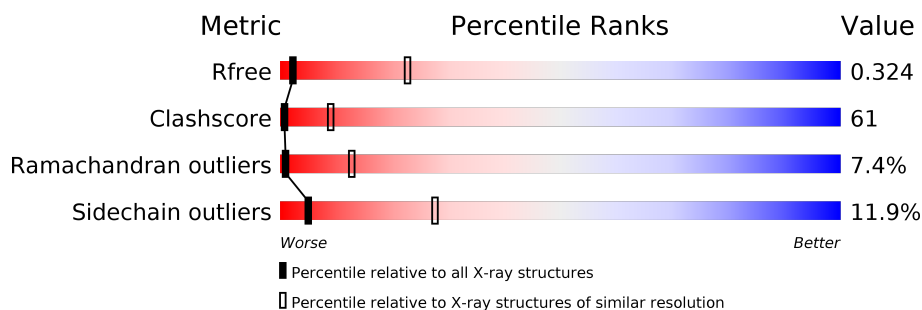
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.60 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.




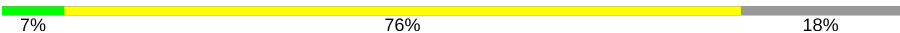
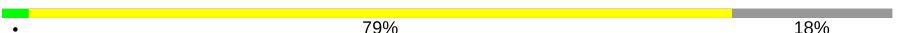
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)

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 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	143	
1	E	143	
2	B	106	
2	F	106	
3	C	133	
3	G	133	
4	D	129	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	H	129	 25% 39% 6% 30%
5	I	147	 7% 76% 18%
5	J	147	 • 79% 18%

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10542 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			
1	E	83	Total	C	N	O	S	0	0	0
			685	449	126	109	1			

There are 6 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
E	-2	GLY	-	EXPRESSION TAG	UNP P49450
E	-1	SER	-	EXPRESSION TAG	UNP P49450
E	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	79	Total	C	N	O	S	0	0	0
			627	395	121	110	1			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	EXPRESSION TAG	UNP B2R4R0
B	-2	SER	-	EXPRESSION TAG	UNP B2R4R0
B	-1	HIS	-	EXPRESSION TAG	UNP B2R4R0
F	-3	GLY	-	EXPRESSION TAG	UNP B2R4R0
F	-2	SER	-	EXPRESSION TAG	UNP B2R4R0
F	-1	HIS	-	EXPRESSION TAG	UNP B2R4R0

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	97	Total	C	N	O	0	0	0
			752	471	148	133			
3	G	100	Total	C	N	O	0	0	0
			773	485	152	136			

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	90	Total	C	N	O	Se	0	0	0
			699	441	123	133	2			
4	H	90	Total	C	N	O	Se	0	0	0
			699	441	123	133	2			

There are 6 discrepancies between the modelled and reference sequences:

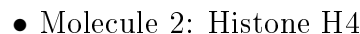
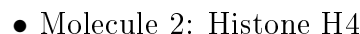
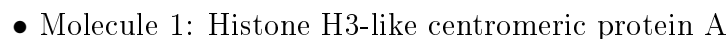
Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	GLY	-	EXPRESSION TAG	UNP P06899
D	-2	SER	-	EXPRESSION TAG	UNP P06899
D	-1	HIS	-	EXPRESSION TAG	UNP P06899
H	-3	GLY	-	EXPRESSION TAG	UNP P06899
H	-2	SER	-	EXPRESSION TAG	UNP P06899
H	-1	HIS	-	EXPRESSION TAG	UNP P06899

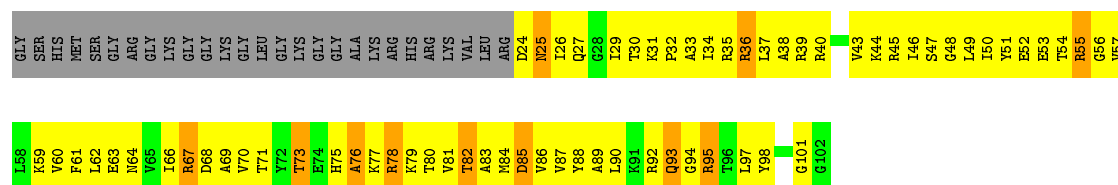
- Molecule 5 is a DNA chain called 147 mer DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	121	Total	C	N	O	P	0	0	0
			2478	1187	448	723	120			
5	J	121	Total	C	N	O	P	0	0	0
			2478	1187	448	723	120			

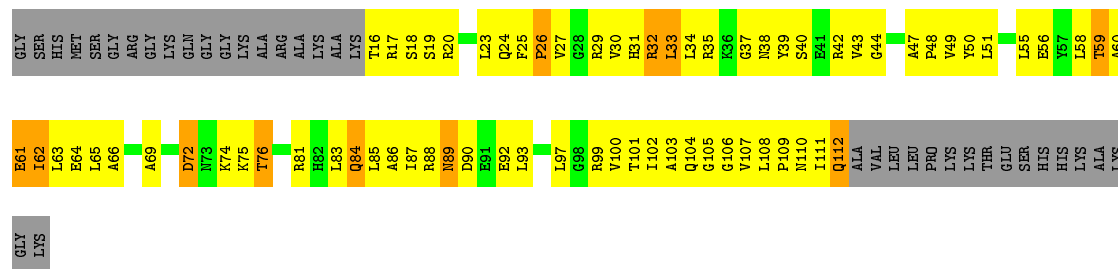
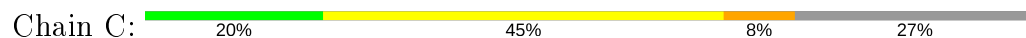


- Molecule 1: Histone H3-like centromeric protein A

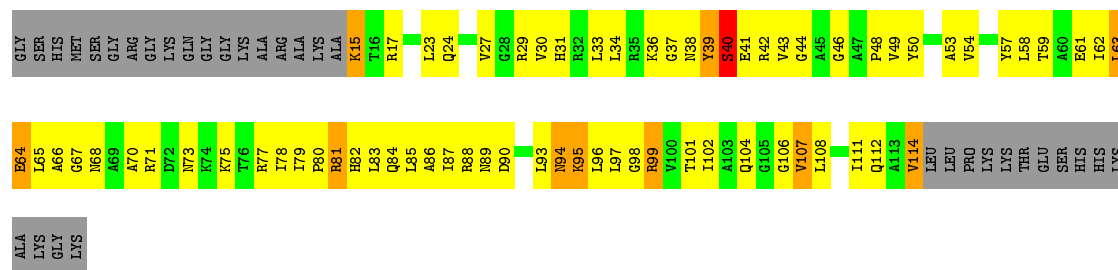
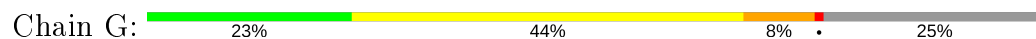




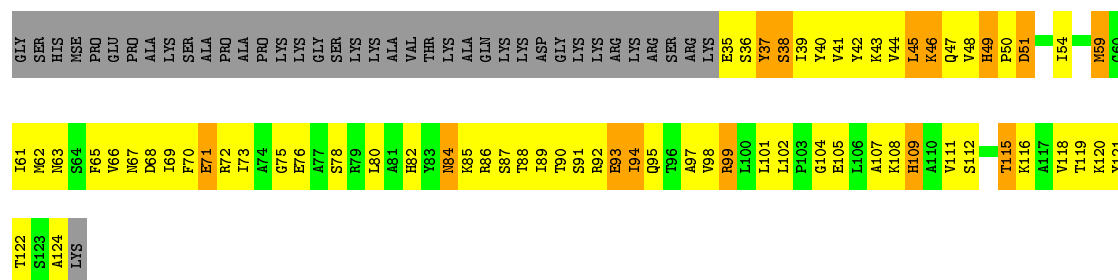
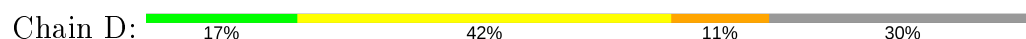
• Molecule 3: Histone H2A type 1-B/E



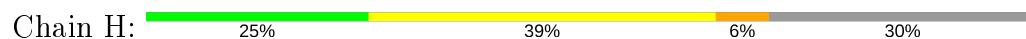
• Molecule 3: Histone H2A type 1-B/E



• Molecule 4: Histone H2B type 1-J



• Molecule 4: Histone H2B type 1-J







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	65.84Å 83.29Å 176.83Å 90.00° 100.70° 90.00°	Depositor
Resolution (Å)	50.00 – 3.60 47.18 – 3.58	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-3.60) 95.0 (47.18-3.58)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.96 (at 3.57Å)	Xtriage
Refinement program	CNS 1.21	Depositor
R, $R_{free}$	0.271 , 0.325 0.271 , 0.324	Depositor DCC
$R_{free}$ test set	1089 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	85.5	Xtriage
Anisotropy	0.441	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 31.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.38$ , $\langle L^2 \rangle = 0.20$	Xtriage
Estimated twinning fraction	0.339 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	10542	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	106.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.91% 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	0.41	0/747	0.67	0/1006
1	E	0.39	0/697	0.70	0/936
2	B	0.44	0/626	0.74	0/837
2	F	0.50	0/634	0.93	2/848 (0.2%)
3	C	0.44	0/761	0.72	0/1027
3	G	0.56	1/782 (0.1%)	0.79	0/1055
4	D	0.51	0/708	0.85	0/951
4	H	0.45	0/708	0.72	0/951
5	I	0.36	0/2780	0.75	0/4289
5	J	0.38	0/2780	0.75	0/4289
All	All	0.42	1/11223 (0.0%)	0.76	2/16189 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	114	VAL	CA-CB	6.13	1.67	1.54

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	76	ALA	N-CA-C	-9.91	84.25	111.00
2	F	76	ALA	C-N-CA	-5.45	108.08	121.70

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	135	0
1	E	685	0	731	135	0
2	B	619	0	659	127	0
2	F	627	0	663	122	0
3	C	752	0	792	114	0
3	G	773	0	819	95	0
4	D	699	0	714	100	0
4	H	699	0	714	76	0
5	I	2478	0	1370	224	0
5	J	2478	0	1370	272	0
All	All	10542	0	8602	1164	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 61.

The worst 5 of 1164 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:-9:DT:H2''	5:I:-8:DG:H5'	1.30	1.14
5:J:-52:DT:H2''	5:J:-51:DT:H5''	1.13	1.11
5:J:52:DA:H2''	5:J:53:DA:H5''	1.24	1.10
5:I:-30:DA:H2''	5:I:-29:DA:H5''	1.32	1.09
5:J:21:DG:H2''	5:J:22:DA:H5'	1.36	1.08

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%)	63 (72%)	15 (17%)	9 (10%)	<b>0</b> <b>7</b>
1	E	79/143 (55%)	59 (75%)	16 (20%)	4 (5%)	<b>2</b> <b>20</b>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	76/106 (72%)	44 (58%)	18 (24%)	14 (18%)	0	2
2	F	77/106 (73%)	61 (79%)	13 (17%)	3 (4%)	3	27
3	C	95/133 (71%)	72 (76%)	19 (20%)	4 (4%)	3	25
3	G	98/133 (74%)	72 (74%)	21 (21%)	5 (5%)	2	20
4	D	88/129 (68%)	65 (74%)	16 (18%)	7 (8%)	1	11
4	H	88/129 (68%)	63 (72%)	20 (23%)	5 (6%)	1	18
All	All	688/1022 (67%)	499 (72%)	138 (20%)	51 (7%)	1	13

5 of 51 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	47	SER
2	B	49	LEU
2	B	99	GLY
4	D	49	HIS
1	E	118	ARG

### 5.3.2 Protein sidechains ⓘ

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%)	67 (89%)	8 (11%)	6	32
1	E	71/120 (59%)	64 (90%)	7 (10%)	8	35
2	B	63/81 (78%)	54 (86%)	9 (14%)	3	21
2	F	64/81 (79%)	56 (88%)	8 (12%)	4	25
3	C	77/102 (76%)	68 (88%)	9 (12%)	5	29
3	G	79/102 (78%)	70 (89%)	9 (11%)	5	29
4	D	76/104 (73%)	67 (88%)	9 (12%)	5	28
4	H	76/104 (73%)	66 (87%)	10 (13%)	4	23
All	All	581/814 (71%)	512 (88%)	69 (12%)	5	27

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	93	GLU
1	E	92	LEU
4	H	82	HIS
4	D	99	ARG
1	E	60	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	84	ASN
1	E	55	GLN
3	G	89	ASN
4	D	63	ASN
3	G	104	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 [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 ⓘ

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.