



# Full wwPDB/EMDatabank EM Map/Model Validation Report ⓘ

Jan 22, 2018 – 07:20 PM EST

PDB ID : 6BUZ  
EMDB ID: : EMD-7293  
Title : Cryo-EM structure of CENP-A nucleosome in complex with kinetochore protein CENP-N  
Authors : Chittori, S.; Hong, J.; Kelly, A.E.; Bai, Y.; Subramaniam, S.  
Deposited on : 2017-12-11  
Resolution : 3.92 Å(reported)

This is a Full wwPDB/EMDatabank EM Map/Model Validation Report  
for a publicly released PDB/EMDB 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/EMValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030736

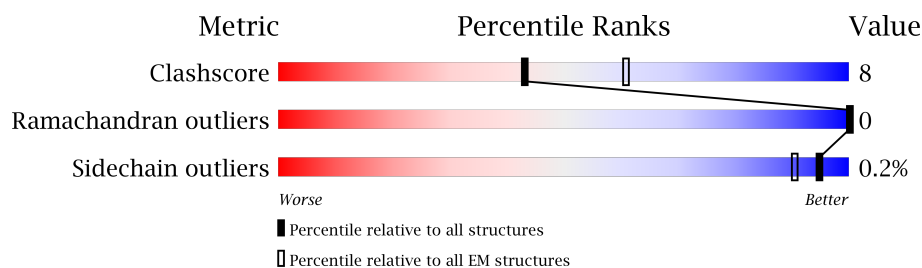
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.92 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)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	160	53% 43%
1	E	160	53% 43%
2	B	103	67% 11% 22%
2	F	103	69% 9% 22%
3	C	130	75% 8% 18%
3	G	130	75% 7% 18%
4	D	126	67% 8% 25%
4	H	126	67% 8% 25%
5	I	147	75% 24%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	J	147	<div><div></div><div>71%</div><div>27%</div><div></div></div>
7	N	668	<div><div></div><div>20%</div><div>80%</div><div></div></div>

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 12509 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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					AltConf	Trace
1	A	91	Total	C	N	O	S	0	0
			749	488	139	121	1		
1	E	91	Total	C	N	O	S	0	0
			749	488	139	121	1		

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	expression tag	UNP P49450
A	-18	GLY	-	expression tag	UNP P49450
A	-17	SER	-	expression tag	UNP P49450
A	-16	SER	-	expression tag	UNP P49450
A	-15	HIS	-	expression tag	UNP P49450
A	-14	HIS	-	expression tag	UNP P49450
A	-13	HIS	-	expression tag	UNP P49450
A	-12	HIS	-	expression tag	UNP P49450
A	-11	HIS	-	expression tag	UNP P49450
A	-10	HIS	-	expression tag	UNP P49450
A	-9	SER	-	expression tag	UNP P49450
A	-8	SER	-	expression tag	UNP P49450
A	-7	GLY	-	expression tag	UNP P49450
A	-6	LEU	-	expression tag	UNP P49450
A	-5	VAL	-	expression tag	UNP P49450
A	-4	PRO	-	expression tag	UNP P49450
A	-3	ARG	-	expression tag	UNP P49450
A	-2	GLY	-	expression tag	UNP P49450
A	-1	SER	-	expression tag	UNP P49450
A	0	HIS	-	expression tag	UNP P49450
E	-19	MET	-	expression tag	UNP P49450
E	-18	GLY	-	expression tag	UNP P49450
E	-17	SER	-	expression tag	UNP P49450
E	-16	SER	-	expression tag	UNP P49450
E	-15	HIS	-	expression tag	UNP P49450
E	-14	HIS	-	expression tag	UNP P49450

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	-13	HIS	-	expression tag	UNP P49450
E	-12	HIS	-	expression tag	UNP P49450
E	-11	HIS	-	expression tag	UNP P49450
E	-10	HIS	-	expression tag	UNP P49450
E	-9	SER	-	expression tag	UNP P49450
E	-8	SER	-	expression tag	UNP P49450
E	-7	GLY	-	expression tag	UNP P49450
E	-6	LEU	-	expression tag	UNP P49450
E	-5	VAL	-	expression tag	UNP P49450
E	-4	PRO	-	expression tag	UNP P49450
E	-3	ARG	-	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					AltConf	Trace
2	B	80	Total	C	N	O	S	0	0
			638	401	125	111	1		
2	F	80	Total	C	N	O	S	0	0
			638	401	125	111	1		

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	107	Total	C	N	O		0	0
			826	520	163	143			
3	G	107	Total	C	N	O		0	0
			826	520	163	143			

- Molecule 4 is a protein called Histone H2B.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	94	Total	C	N	O	S	0	0
			736	462	134	138	2		
4	H	94	Total	C	N	O	S	0	0
			736	462	134	138	2		

- Molecule 5 is a DNA chain called DNA (147-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
5	I	145	Total	C	N	O	P	0	0
			2952	1403	538	867	144		

- Molecule 6 is a DNA chain called DNA (147-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	J	145	Total	C	N	O	P	0	0
			2987	1415	559	869	144		

- Molecule 7 is a protein called Maltose-binding periplasmic protein, Centromere protein N chimera.

Mol	Chain	Residues	Atoms				AltConf	Trace
7	N	134	Total	C	N	O	0	0
			672	402	135	135		

There are 18 discrepancies between the modelled and reference sequences:


Chain	Residue	Modelled	Actual	Comment	Reference
N	-381	MET	-	expression tag	UNP P0AEY0
N	-380	LYS	-	expression tag	UNP P0AEY0
N	-379	GLU	-	expression tag	UNP P0AEY0
N	-378	HIS	-	expression tag	UNP P0AEY0
N	-377	HIS	-	expression tag	UNP P0AEY0
N	-376	HIS	-	expression tag	UNP P0AEY0
N	-375	HIS	-	expression tag	UNP P0AEY0
N	-374	HIS	-	expression tag	UNP P0AEY0
N	-373	HIS	-	expression tag	UNP P0AEY0
N	-372	HIS	-	expression tag	UNP P0AEY0
N	-371	HIS	-	expression tag	UNP P0AEY0
N	-11	ALA	GLU	engineered mutation	UNP P0AEY0
N	-8	ALA	LYS	engineered mutation	UNP P0AEY0
N	-7	ALA	ASP	engineered mutation	UNP P0AEY0
N	-3	ASN	ARG	engineered mutation	UNP P0AEY0
N	-2	ALA	ILE	engineered mutation	UNP P0AEY0
N	-1	ALA	-	linker	UNP P0AEY0
N	0	ALA	-	linker	UNP P0AEY0

### 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. 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. 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

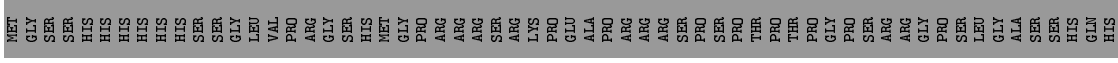
Chain A: 





- Molecule 1: Histone H3-like centromeric protein A


Chain E: 






- Molecule 2: Histone H4

Chain B: 



- Molecule 2: Histone H4

Chain F: 



- Molecule 3: Histone H2A

Chain C: 



- Molecule 3: Histone H2A

WORLDWIDE  
**PDB**  
PROTEIN DATA BANK

 **EMDataBank**  
Unified Data Resource for 3DEM



[illegible]

## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	61749	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	40	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

## 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  >2	RMSZ	# Z  >2
1	A	0.41	0/764	0.56	0/1029
1	E	0.42	0/764	0.56	0/1029
2	B	0.42	0/645	0.54	0/862
2	F	0.42	0/645	0.54	0/862
3	C	0.38	0/836	0.55	0/1128
3	G	0.38	0/836	0.54	0/1128
4	D	0.39	0/747	0.49	0/1004
4	H	0.41	0/747	0.50	0/1004
5	I	0.92	0/3308	0.99	0/5099
6	J	0.93	0/3354	1.00	0/5180
7	N	0.26	0/664	0.45	0/915
All	All	0.71	0/13310	0.81	0/19240

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

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	749	0	789	5	0
1	E	749	0	789	4	0
2	B	638	0	676	10	0
2	F	638	0	676	7	0
3	C	826	0	884	9	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	826	0	884	8	0
4	D	736	0	758	8	0
4	H	736	0	758	8	0
5	I	2952	0	1628	32	0
6	J	2987	0	1629	38	0
7	N	672	0	291	2	0
All	All	12509	0	9762	84	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 8.

All (84) 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:27:DC:N3	6:J:26:DA:N6	2.48	0.61
3:G:78:ARG:NH1	6:J:-53:DA:OP1	2.34	0.60
5:I:61:DG:H1	6:J:61:DC:H42	1.48	0.60
5:I:8:DC:H42	6:J:-8:DG:H1	1.49	0.60
5:I:7:DC:H42	6:J:-7:DG:H1	1.50	0.58
5:I:52:DG:O6	6:J:-53:DA:N6	2.35	0.58
1:A:69:ARG:NH1	6:J:17:DA:OP1	2.37	0.58
5:I:-49:DG:H1	6:J:49:DC:H42	1.51	0.57
1:E:62:ILE:O	1:E:95:GLN:NE2	2.38	0.57
5:I:5:DC:H42	6:J:-5:DG:H1	1.51	0.57
3:C:30:ARG:NH1	4:D:37:SER:O	2.38	0.56
1:A:62:ILE:O	1:A:95:GLN:NE2	2.37	0.56
4:H:37:SER:OG	4:H:38:TYR:N	2.39	0.56
3:C:45:GLY:HA2	6:J:38:DG:H5"	1.87	0.55
1:A:108:ASP:OD2	1:A:133:ARG:NH2	2.38	0.55
1:E:108:ASP:OD2	1:E:133:ARG:NH2	2.38	0.55
5:I:27:DG:O6	6:J:-28:DC:N4	2.40	0.55
3:C:78:ARG:HH21	6:J:57:DG:H4'	1.73	0.54
3:G:30:ARG:NH1	4:H:37:SER:O	2.40	0.54
5:I:17:DA:H61	6:J:-17:DT:H3	1.55	0.53
5:I:45:DC:N4	6:J:-46:DT:O4	2.41	0.53
3:C:30:ARG:NH2	6:J:49:DC:OP1	2.42	0.53
2:B:92:LYS:NZ	4:H:69:ASP:OD1	2.42	0.52
2:B:52:TYR:O	2:B:56:ARG:NH1	2.43	0.52
2:F:52:TYR:O	2:F:56:ARG:NH1	2.43	0.51
3:G:86:LEU:O	3:G:90:ASN:ND2	2.45	0.50
2:B:76:HIS:HE2	4:D:94:GLU:HG3	1.76	0.50

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:86:LEU:O	3:C:90:ASN:ND2	2.45	0.50
1:A:61:LEU:HD12	2:B:38:LEU:HD23	1.93	0.50
5:I:69:DC:O2	5:I:70:DC:N4	2.45	0.50
6:J:-15:DA:H1'	6:J:-14:DA:H5'	1.93	0.50
1:A:120:THR:OG1	2:B:46:ARG:NH1	2.44	0.49
2:F:93:ARG:HH21	4:H:102:LEU:HG	1.78	0.49
4:D:80:ARG:O	4:D:84:TYR:N	2.45	0.49
7:N:1:MET:O	7:N:5:VAL:N	2.44	0.48
4:D:105:GLY:O	4:D:109:LYS:NZ	2.39	0.48
3:G:43:ARG:HH21	5:I:38:DT:H4'	1.79	0.47
5:I:-17:DT:O4	6:J:16:DA:N6	2.47	0.47
5:I:-60:DG:O6	6:J:59:DA:N6	2.48	0.47
1:E:120:THR:OG1	2:F:46:ARG:NH1	2.39	0.46
5:I:-33:DA:N6	6:J:32:DG:O6	2.47	0.46
6:J:5:DT:H2''	6:J:6:DA:C8	2.50	0.46
2:B:93:ARG:HH21	4:D:102:LEU:HG	1.80	0.46
5:I:-15:DA:H1'	5:I:-14:DA:H5'	1.97	0.46
5:I:-16:DT:H4'	5:I:-15:DA:H5'	1.97	0.46
1:E:49:LYS:NZ	5:I:-65:DT:OP1	2.49	0.46
6:J:-20:DC:H2''	6:J:-19:DG:H5'	1.98	0.45
2:B:97:THR:HB	3:G:101:VAL:HG22	1.98	0.45
5:I:-62:DC:H2''	5:I:-61:DG:C8	2.52	0.45
3:C:40:TYR:O	4:D:79:SER:OG	2.31	0.45
5:I:12:DT:O4	6:J:-13:DA:N6	2.50	0.45
3:C:17:THR:HG22	3:C:19:SER:H	1.82	0.44
2:B:40:ARG:NH1	2:B:45:LYS:O	2.50	0.44
5:I:6:DC:O2	6:J:-5:DG:N2	2.50	0.44
3:G:17:THR:HG22	3:G:19:SER:H	1.82	0.44
4:D:37:SER:OG	4:D:38:TYR:N	2.49	0.44
2:F:40:ARG:NH1	2:F:45:LYS:O	2.50	0.44
5:I:-45:DA:N6	6:J:44:DT:O4	2.51	0.44
5:I:36:DC:H42	6:J:-36:DG:H1	1.67	0.43
5:I:-64:DC:H42	6:J:64:DG:H1	1.67	0.43
2:B:76:HIS:HD2	4:D:97:THR:HG21	1.84	0.42
3:G:65:GLU:HA	4:H:50:HIS:CE1	2.54	0.42
2:F:36:ARG:HH21	2:F:40:ARG:HH21	1.66	0.42
5:I:31:DT:H2''	5:I:32:DA:C8	2.53	0.42
3:G:65:GLU:HA	4:H:50:HIS:HE1	1.84	0.42
5:I:19:DC:H42	6:J:-19:DG:H1	1.65	0.42
6:J:-18:DG:H4'	6:J:-17:DT:H5'	2.02	0.42
2:F:37:ARG:NH2	6:J:-13:DA:OP2	2.42	0.42

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:J:-57:DT:H2''	6:J:-56:DG:C8	2.55	0.42
5:I:-68:DG:H22	6:J:68:DC:H2'	1.85	0.42
6:J:-63:DT:H1'	6:J:-62:DA:H5'	2.02	0.41
2:B:36:ARG:HH21	2:B:40:ARG:HH21	1.66	0.41
4:H:80:ARG:O	4:H:84:TYR:N	2.46	0.41
5:I:-21:DC:OP1	7:N:170:ARG:N	2.53	0.41
6:J:-2:DC:H2''	6:J:-1:DA:C8	2.55	0.41
5:I:-53:DG:N2	6:J:54:DT:O2	2.54	0.41
4:H:62:ILE:O	4:H:66:PHE:N	2.53	0.41
5:I:-8:DC:H2'	5:I:-7:DG:C4	2.56	0.41
3:C:21:ARG:NH2	5:I:-42:DT:OP1	2.39	0.41
6:J:38:DG:H2''	6:J:39:DA:H8	1.86	0.40
3:C:101:VAL:HG22	2:F:97:THR:HB	2.03	0.40
6:J:-32:DT:H2''	6:J:-31:DA:H8	1.85	0.40
5:I:36:DC:N4	6:J:-37:DG:O6	2.54	0.40
6:J:43:DA:H2'	6:J:44:DT:H71	2.03	0.40

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 PDB entries followed by that with respect to all EM entries.

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	89/160 (56%)	85 (96%)	4 (4%)	0	100	100
1	E	89/160 (56%)	85 (96%)	4 (4%)	0	100	100
2	B	78/103 (76%)	75 (96%)	3 (4%)	0	100	100
2	F	78/103 (76%)	75 (96%)	3 (4%)	0	100	100
3	C	105/130 (81%)	98 (93%)	7 (7%)	0	100	100
3	G	105/130 (81%)	98 (93%)	7 (7%)	0	100	100
4	D	92/126 (73%)	90 (98%)	2 (2%)	0	100	100

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	H	92/126 (73%)	88 (96%)	4 (4%)	0	100	100
7	N	118/668 (18%)	105 (89%)	13 (11%)	0	100	100
All	All	846/1706 (50%)	799 (94%)	47 (6%)	0	100	100

There are no Ramachandran outliers to report.

### 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 PDB entries followed by that with respect to all EM entries.

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	77/135 (57%)	77 (100%)	0	100	100
1	E	77/135 (57%)	77 (100%)	0	100	100
2	B	65/79 (82%)	65 (100%)	0	100	100
2	F	65/79 (82%)	65 (100%)	0	100	100
3	C	84/100 (84%)	84 (100%)	0	100	100
3	G	84/100 (84%)	84 (100%)	0	100	100
4	D	80/105 (76%)	80 (100%)	0	100	100
4	H	80/105 (76%)	79 (99%)	1 (1%)	73	88
7	N	1/566 (0%)	1 (100%)	0	100	100
All	All	613/1404 (44%)	612 (100%)	1 (0%)	95	97

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	H	32	ARG

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

Mol	Chain	Res	Type
4	D	85	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.