wwPDB X-ray Structure Validation Summary Report (i)

Nov 3, 2023 – 08:35 AM EDT

|		
PDB ID	:	3WA9
Title	:	The nucleosome containing human H2A.Z.1
Authors	:	Horikoshi, N.; Sato, K.; Shimada, K.; Arimura, Y.; Osakabe, A.; Tachiwana, H.; Iwasaki, W.; Kagawa, W.; Harata, M.; Kimura, H.; Kurumizaka, H.
Deposited on	:	2013-04-30
Resolution	:	3.07 Å(reported)

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 3.07 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.
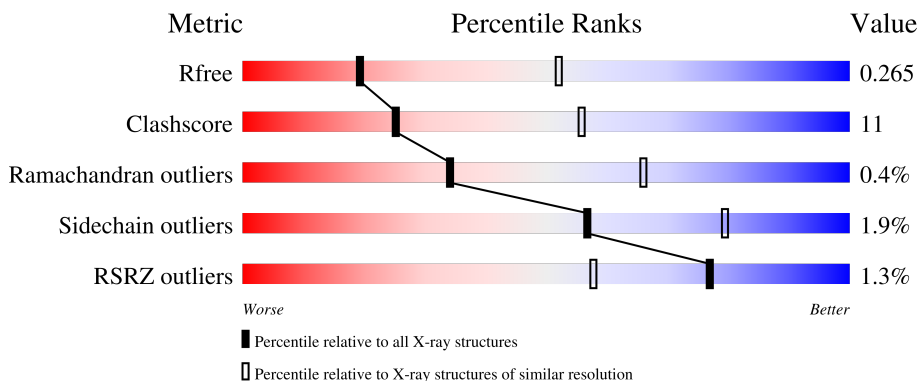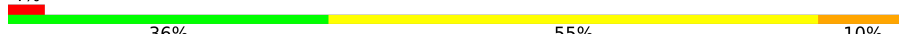
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1447 (3.10-3.06)
Clashscore	141614	1546 (3.10-3.06)
Ramachandran outliers	138981	1487 (3.10-3.06)
Sidechain outliers	138945	1486 (3.10-3.06)
RSRZ outliers	127900	1416 (3.10-3.06)

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	53% 16% • 30%
1	E	139	.% 53% 17% • 29%
2	B	106	63% 10% 26%
2	F	106	68% 11% 21%
3	C	131	2% 63% 16% • 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	131	57% 21% · 21%
4	D	129	% 55% 16% · 28%
4	H	129	59% 12% · 28%
5	I	146	4% 36% 55% 10%
5	J	146	3% 34% 52% 14%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 11925 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 H3.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	97	Total	C	N	O	S	0	0	0
			801	505	155	137	4			
1	E	98	Total	C	N	O	S	0	0	0
			810	511	157	138	4			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP P68431
A	-2	SER	-	expression tag	UNP P68431
A	-1	HIS	-	expression tag	UNP P68431
E	-3	GLY	-	expression tag	UNP P68431
E	-2	SER	-	expression tag	UNP P68431
E	-1	HIS	-	expression tag	UNP P68431

- 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.Z.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	106	Total 803	C 504	N 157	O 142	0	0	0
3	G	104	Total 789	C 495	N 154	O 140	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-3	GLY	-	expression tag	UNP P0C0S5
C	-2	SER	-	expression tag	UNP P0C0S5
C	-1	HIS	-	expression tag	UNP P0C0S5
G	-3	GLY	-	expression tag	UNP P0C0S5
G	-2	SER	-	expression tag	UNP P0C0S5
G	-1	HIS	-	expression tag	UNP P0C0S5

- Molecule 4 is a protein called Histone H2B type 1-J.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	93	Total 725	C 456	N 130	O 137	S 2	0	0	0
4	H	93	Total 725	C 456	N 130	O 137	S 2	0	0	0

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 2990	C 1431	N 540	O 874	P 145	0	0	0
5	J	146	Total 2990	C 1431	N 540	O 874	P 145	0	0	0

3 Residue-property plots ⓘ

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 H3.1

Chain A:
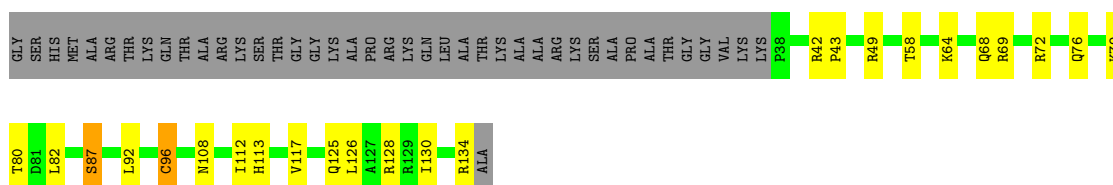
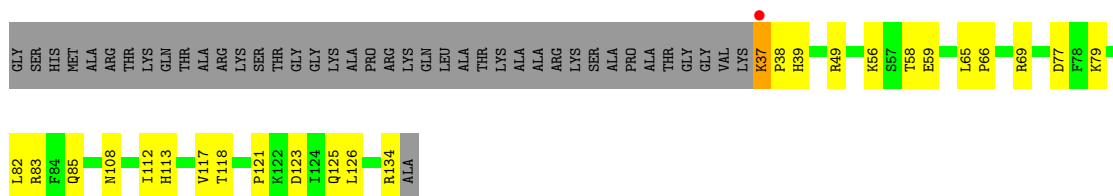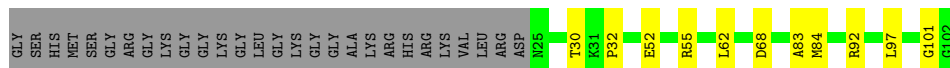• Molecule 1: Histone H3.1

Chain E:

• Molecule 2: Histone H4

Chain B:

• Molecule 2: Histone H4

Chain F:
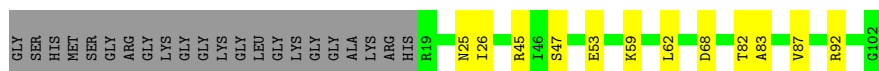
• Molecule 3: Histone H2A.Z

Chain C:

• Molecule 3: Histone H2A.Z

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:

T143 G144 A145 T146

● Molecule 5: DNA (146-MER)

Chain J:

G290 A291 T292
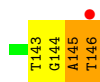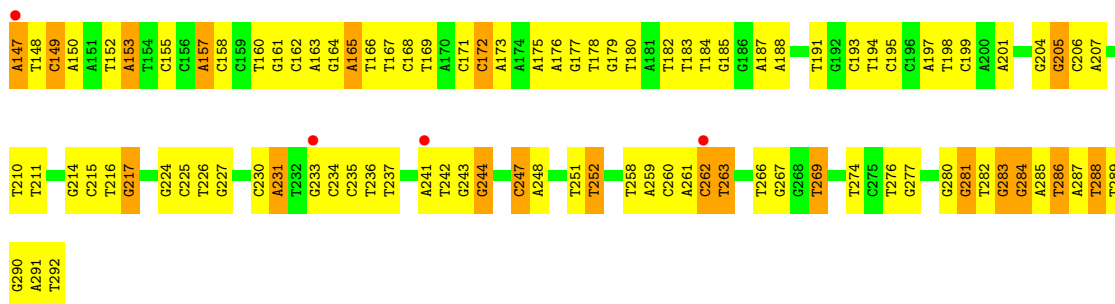
4   Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	104.90Å 109.39Å 181.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.96 – 3.07 38.96 – 3.04	Depositor EDS
% Data completeness (in resolution range)	99.8 (38.96-3.07) 96.4 (38.96-3.04)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	4.38 (at 3.06Å)	Xtriage
Refinement program	PHENIX 1.7.3_928	Depositor
R, R_{free}	0.222 , 0.271 0.218 , 0.265	Depositor DCC
R_{free} test set	2028 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	61.0	Xtriage
Anisotropy	0.216	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 51.8	EDS
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.029 for k,h,-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	11925	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.03% 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	0.50	1/813 (0.1%)	0.65	0/1090
1	E	0.57	1/822 (0.1%)	0.82	2/1102 (0.2%)
2	B	0.48	0/626	0.65	0/837
2	F	0.53	0/680	0.67	0/908
3	C	0.47	0/814	0.81	3/1095 (0.3%)
3	G	0.45	0/800	0.57	0/1077
4	D	0.50	0/736	0.65	0/990
4	H	0.44	0/736	0.58	0/990
5	I	0.69	0/3354	1.46	57/5175 (1.1%)
5	J	0.66	0/3354	1.50	55/5175 (1.1%)
All	All	0.60	2/12735 (0.0%)	1.20	117/18439 (0.6%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	96	CYS	CB-SG	-6.24	1.71	1.82
1	E	77	ASP	CB-CG	5.19	1.62	1.51

The worst 5 of 117 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	77	ASP	CB-CG-OD1	11.49	128.64	118.30
5	J	241	DA	O4'-C1'-N9	10.08	115.05	108.00
3	C	39	ARG	CB-CA-C	9.42	129.23	110.40
3	C	40	THR	N-CA-CB	-9.02	93.15	110.30
5	I	143	DT	O4'-C1'-N1	8.63	114.05	108.00

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	801	0	839	22	0
1	E	810	0	851	23	0
2	B	619	0	659	7	0
2	F	673	0	722	12	0
3	C	803	0	853	27	0
3	G	789	0	835	24	0
4	D	725	0	745	22	0
4	H	725	0	745	12	0
5	I	2990	0	1652	60	0
5	J	2990	0	1652	70	0
All	All	11925	0	9553	213	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 11.

The worst 5 of 213 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:144:DG:H1	5:J:149:DC:H42	1.18	0.92
5:I:43:DA:H2”	5:I:44:DC:H5”	1.64	0.80
5:I:41:DA:N6	5:J:251:DT:O4	2.16	0.79
3:C:39:ARG:O	3:G:41:THR:HG22	1.84	0.76
3:C:34:ARG:NH1	5:I:29:DA:OP1	2.25	0.70

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	96/139 (69%)	94 (98%)	2 (2%)	0	100	100
2	B	76/106 (72%)	72 (95%)	4 (5%)	0	100	100
2	F	82/106 (77%)	79 (96%)	3 (4%)	0	100	100
3	C	104/131 (79%)	97 (93%)	7 (7%)	0	100	100
3	G	102/131 (78%)	94 (92%)	7 (7%)	1 (1%)	15	47
4	D	91/129 (70%)	87 (96%)	3 (3%)	1 (1%)	14	44
4	H	91/129 (70%)	87 (96%)	3 (3%)	1 (1%)	14	44
All	All	737/1010 (73%)	704 (96%)	30 (4%)	3 (0%)	34	66

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	123	SER
4	H	105	GLU
3	G	115	LYS

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	85/113 (75%)	83 (98%)	2 (2%)	49	74
1	E	86/113 (76%)	84 (98%)	2 (2%)	50	75
2	B	63/81 (78%)	63 (100%)	0	100	100
2	F	69/81 (85%)	68 (99%)	1 (1%)	67	84
3	C	83/99 (84%)	80 (96%)	3 (4%)	35	66
3	G	82/99 (83%)	80 (98%)	2 (2%)	49	74
4	D	79/107 (74%)	79 (100%)	0	100	100
4	H	79/107 (74%)	77 (98%)	2 (2%)	47	74

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	626/800 (78%)	614 (98%)	12 (2%)	57 79

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	47	SER
3	G	17	VAL
4	H	115	THR
3	G	38	SER
3	C	73	SER

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

Mol	Chain	Res	Type
3	G	114	HIS

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 monosaccharides 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 (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, 95th 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%)	-0.38	0	100	100	26, 42, 64, 104	0
1	E	98/139 (70%)	-0.42	1 (1%)	82	66	21, 33, 60, 99	0
2	B	78/106 (73%)	-0.46	0	100	100	30, 39, 52, 61	0
2	F	84/106 (79%)	-0.46	0	100	100	23, 33, 52, 78	0
3	C	106/131 (80%)	-0.27	2 (1%)	66	45	25, 42, 87, 112	0
3	G	104/131 (79%)	-0.32	0	100	100	33, 52, 87, 111	0
4	D	93/129 (72%)	-0.29	1 (1%)	80	62	29, 43, 65, 102	0
4	H	93/129 (72%)	-0.38	0	100	100	32, 50, 74, 109	0
5	I	146/146 (100%)	0.20	6 (4%)	37	18	46, 103, 140, 147	0
5	J	146/146 (100%)	0.14	4 (2%)	54	29	51, 104, 140, 154	0
All	All	1045/1302 (80%)	-0.22	14 (1%)	77	58	21, 49, 127, 154	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	146	DT	5.9
1	E	37	LYS	4.5
3	C	15	LYS	3.4
3	C	120	LYS	3.1
5	J	147	DA	3.0

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 ⓘ

There are no monosaccharides 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.