wwPDB X-ray Structure Validation Summary Report (i)

Apr 19, 2016 – 08:15 PM EDT

PDB ID	:	5HQ2
Title	:	Structural model of Set8 histone H4 Lys20 methyltransferase bound to nucleosome core particle
Authors	:	Tavarekere, G.; McGinty, R.K.; Tan, S.
Deposited on	:	2016-01-21
Resolution	:	4.50 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:

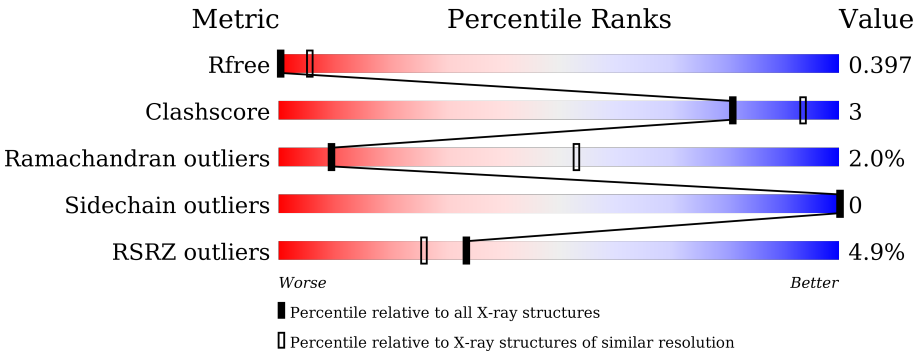
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027257
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027257

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 4.50 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.



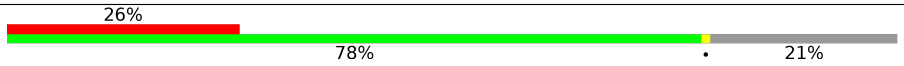
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1071 (5.40-3.60)
Clashscore	102246	1003 (5.30-3.62)
Ramachandran outliers	100387	1117 (5.40-3.60)
Sidechain outliers	100360	1099 (5.40-3.60)
RSRZ outliers	91569	1075 (5.40-3.60)

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	135	64% 10% 27%
2	B	102	75% · 24%
3	G	129	75% 6% 19%
4	H	122	70% 7% 22%
5	I	149	% 48% · 51%
6	J	149	46% •• 52%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	K	483	
8	M	202	

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 7738 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	99	Total	C	N	O	0	0	0
			495	297	99	99			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	ALA	GLY	conflict	UNP P84233

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	78	Total	C	N	O	0	0	0
			384	228	78	78			

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	G	105	Total	C	N	O	0	0	0
			516	306	105	105			

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	H	95	Total	C	N	O	0	0	0
			473	283	95	95			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	29	THR	SER	conflict	UNP P02281

- Molecule 5 is a DNA chain called DNA (149-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	73	Total	C	N	O	P	0	0	0
			1487	707	271	436	73			

- Molecule 6 is a DNA chain called DNA (149-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	J	72	Total	C	N	O	P	0	0	0
			1484	704	274	434	72			

- Molecule 7 is a protein called Guanine nucleotide exchange factor SRM1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
7	K	422	Total	C	N	O	0	0	0
			2112	1268	422	422			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	0	GLY	-	expression tag	UNP P21827
K	1	SER	-	expression tag	UNP P21827

- Molecule 8 is a protein called N-lysine methyltransferase SETD8.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
8	M	160	Total	C	N	O	0	0	0
			787	467	160	160			

There are 3 discrepancies between the modelled and reference sequences:

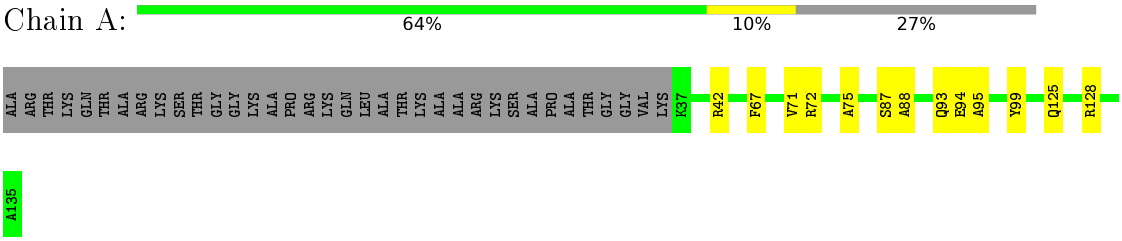
Chain	Residue	Modelled	Actual	Comment	Reference
M	151	GLY	-	expression tag	UNP Q9NQR1
M	152	SER	-	expression tag	UNP Q9NQR1
M	347	PHE	HIS	conflict	UNP Q9NQR1

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.

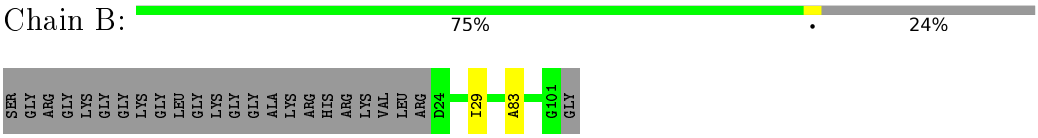
- Molecule 1: Histone H3.2

Chain A:



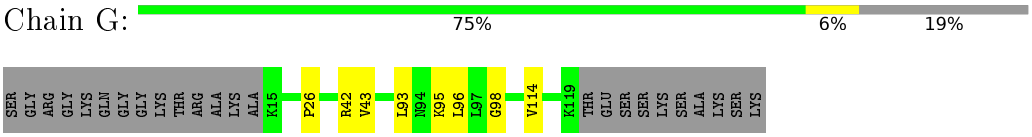
- Molecule 2: Histone H4

Chain B:



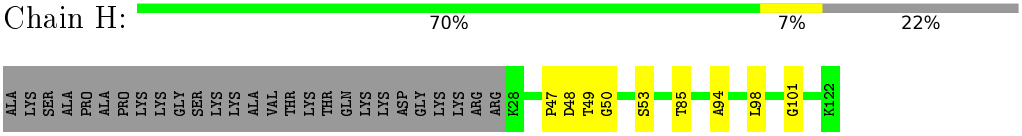
- Molecule 3: Histone H2A

Chain G:



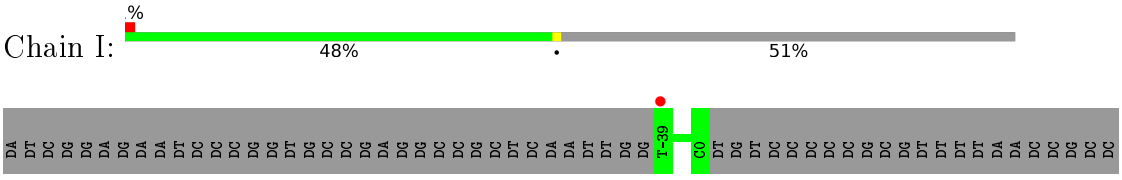
- Molecule 4: Histone H2B 1.1

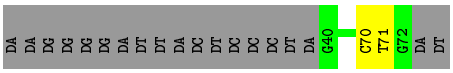
Chain H:



- Molecule 5: DNA (149-MER)

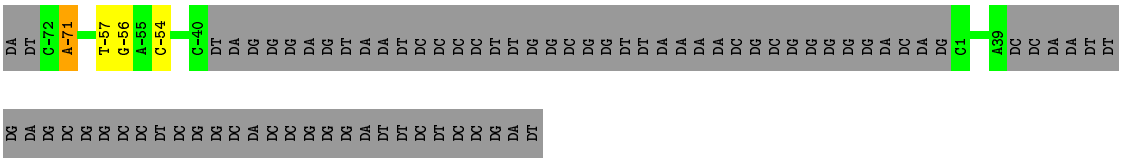
Chain I:





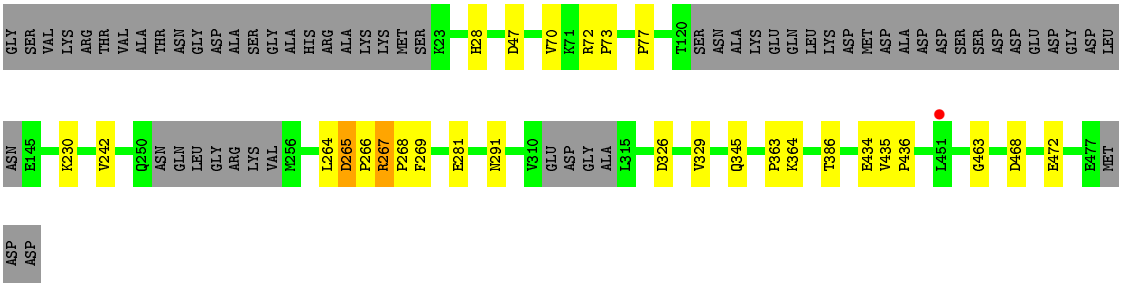
- Molecule 6: DNA (149-MER)

Chain J:



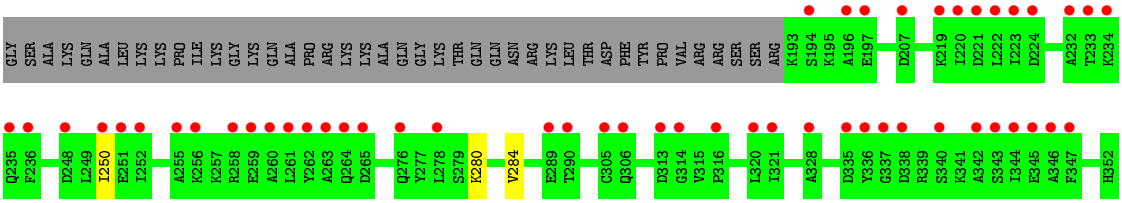
- Molecule 7: Guanine nucleotide exchange factor SRM1

Chain K:



- Molecule 8: N-lysine methyltransferase SETD8

Chain M:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	100.73Å 300.75Å 182.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.00 – 4.50 44.07 – 4.50	Depositor EDS
% Data completeness (in resolution range)	99.3 (46.00-4.50) 99.5 (44.07-4.50)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.25 (at 4.45Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.339 , 0.397 0.345 , 0.397	Depositor DCC
R_{free} test set	845 reflections (5.31%)	DCC
Wilson B-factor (Å^2)	213.1	Xtriage
Anisotropy	0.174	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.15 , 306.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	7738	wwPDB-VP
Average B, all atoms (Å^2)	221.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.71% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.46	0/495	0.60	0/691
2	B	0.47	0/384	0.71	0/532
3	G	0.44	0/515	0.58	0/714
4	H	0.45	0/473	0.68	1/659 (0.2%)
5	I	0.30	0/1665	0.73	0/2562
6	J	0.33	1/1664 (0.1%)	0.71	0/2566
7	K	0.37	0/2130	0.56	0/2968
8	M	0.35	0/786	0.50	0/1091
All	All	0.37	1/8112 (0.0%)	0.64	1/11783 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	J	-71	DA	C1'-N9	-5.27	1.39	1.47

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	H	101	GLY	N-CA-C	6.40	129.09	113.10

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	495	0	234	9	0
2	B	384	0	186	3	0
3	G	516	0	249	4	0
4	H	473	0	227	6	0
5	I	1487	0	821	4	0
6	J	1484	0	812	3	0
7	K	2112	0	1078	17	0
8	M	787	0	362	0	0
All	All	7738	0	3969	39	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 3.

The worst 5 of 39 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:70:DC:H2"	5:I:71:DT:H71	1.52	0.89
3:G:42:ARG:CB	4:H:85:THR:CB	2.59	0.80
5:I:71:DT:O4	6:J:-71:DA:N1	2.25	0.70
5:I:70:DC:H2"	5:I:71:DT:C7	2.22	0.68
7:K:266:PRO:HA	7:K:267:ARG:CB	2.27	0.64

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	97/135 (72%)	77 (79%)	19 (20%)	1 (1%)	19 65
2	B	76/102 (74%)	59 (78%)	16 (21%)	1 (1%)	15 60
3	G	103/129 (80%)	82 (80%)	18 (18%)	3 (3%)	6 45
4	H	93/122 (76%)	77 (83%)	15 (16%)	1 (1%)	17 64

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
7	K	414/483 (86%)	316 (76%)	88 (21%)	10 (2%)	7 49
8	M	158/202 (78%)	137 (87%)	18 (11%)	3 (2%)	10 54
All	All	941/1173 (80%)	748 (80%)	174 (18%)	19 (2%)	9 53

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
7	K	265	ASP
7	K	267	ARG
8	M	280	LYS
7	K	70	VAL
7	K	386	THR

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	1/110 (1%)	1 (100%)	0	100 100
2	B	1/78 (1%)	1 (100%)	0	100 100
4	H	1/102 (1%)	1 (100%)	0	100 100
7	K	22/400 (6%)	22 (100%)	0	100 100
All	All	25/690 (4%)	25 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

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.

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	99/135 (73%)	-0.88	0 100 100	42, 88, 196, 253	0
2	B	78/102 (76%)	-0.76	0 100 100	38, 86, 140, 180	0
3	G	105/129 (81%)	-0.84	0 100 100	39, 103, 172, 195	0
4	H	95/122 (77%)	-0.87	0 100 100	54, 100, 211, 276	0
5	I	73/149 (48%)	-0.12	1 (1%) 78 69	136, 222, 304, 352	0
6	J	72/149 (48%)	-0.31	0 100 100	141, 216, 296, 318	0
7	K	422/483 (87%)	-0.49	1 (0%) 95 94	120, 236, 335, 381	0
8	M	160/202 (79%)	1.43	52 (32%) 1 2	201, 444, 500, 500	0
All	All	1104/1471 (75%)	-0.29	54 (4%) 33 26	38, 205, 488, 500	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
8	M	222	LEU	10.9
8	M	220	ILE	7.4
8	M	221	ASP	6.6
8	M	259	GLU	6.5
8	M	343	SER	6.4

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.