



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 11, 2024 – 12:39 PM JST

PDB ID : 8YJF
Title : Structure of human SPT16 MD-CTD and MCM2 HBD chaperoning a histone H3-H4 tetramer and an H2A-H2B dimer
Authors : Gan, S.L.; Yang, W.S.; Xu, R.M.
Deposited on : 2024-03-01
Resolution : 4.40 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

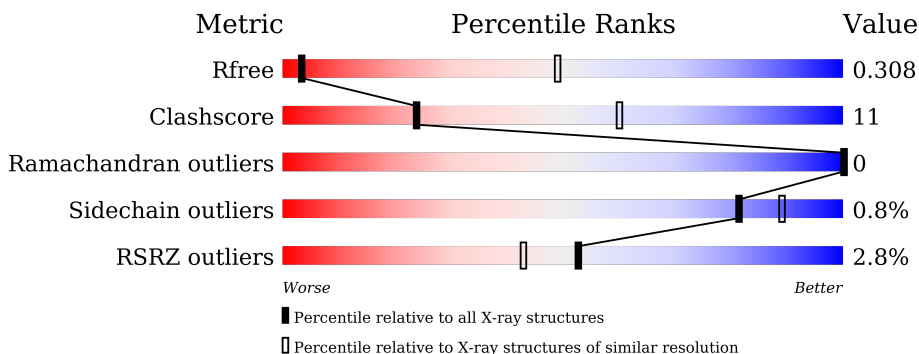
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.40 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	1043 (5.00-3.80)
Clashscore	141614	1111 (5.00-3.80)
Ramachandran outliers	138981	1059 (5.00-3.80)
Sidechain outliers	138945	1041 (5.00-3.80)
RSRZ outliers	127900	1095 (5.08-3.70)

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 $\leq 5\%$. 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	350	2% 64% 24% 12%
2	B	93	44% 20% 35%
3	C	81	2% 64% 32% .
3	E	81	2% 73% 23% .
4	D	103	2% 44% 26% 30%
4	F	103	47% 22% 31%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	G	126	 3% 51% 21% 29%
6	H	136	 7% 51% 15% 34%

2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 6797 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 FACT complex subunit SPT16.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	307	2525	1606	419	488	12	0	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	639	GLY	-	expression tag	UNP Q9Y5B9
A	640	PRO	-	expression tag	UNP Q9Y5B9
A	641	LEU	-	expression tag	UNP Q9Y5B9
A	642	GLY	-	expression tag	UNP Q9Y5B9
A	643	SER	-	expression tag	UNP Q9Y5B9

- Molecule 2 is a protein called DNA replication licensing factor MCM2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	60	470	276	86	106	2	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	62	SER	-	expression tag	UNP P49736

- Molecule 3 is a protein called Histone H3.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	78	630	399	115	112	4	0	0	0
3	E	78	631	399	117	111	4	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	55	MET	-	initiating methionine	UNP P68431
E	55	MET	-	initiating methionine	UNP P68431

- Molecule 4 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	72	Total	C	N	O	S	0	0	0
			576	361	114	100	1			
4	F	71	Total	C	N	O	S	0	0	0
			567	357	112	97	1			

- Molecule 5 is a protein called Histone H2B type 2-E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	G	90	Total	C	N	O	S	0	0	0
			699	441	123	133	2			

- Molecule 6 is a protein called Histone H2A type 1-D.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
6	H	90	Total	C	N	O	0	0	0
			699	438	137	124			

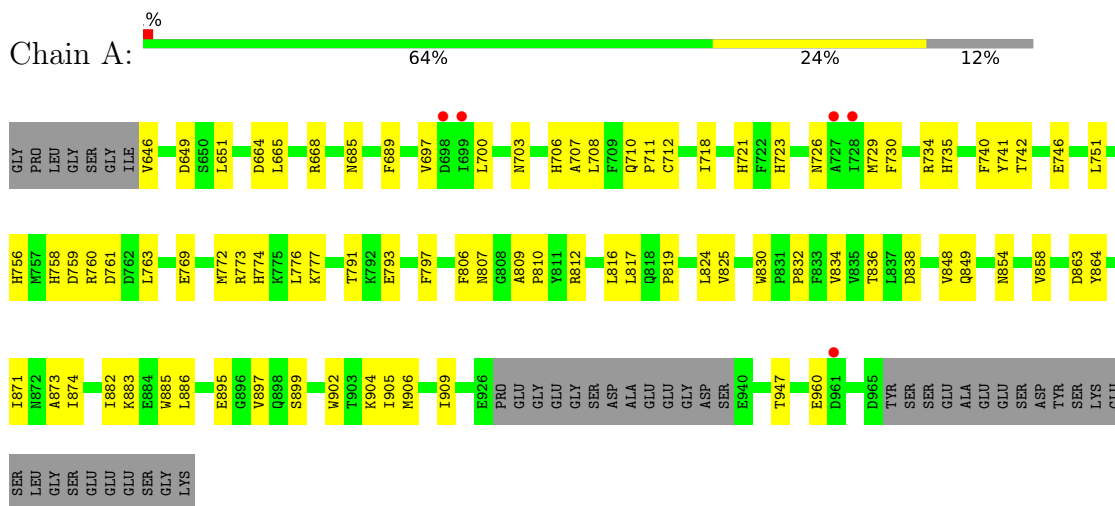
There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	-6	MET	-	initiating methionine	UNP P20671
H	-5	VAL	-	expression tag	UNP P20671
H	-4	MET	-	expression tag	UNP P20671
H	-3	LYS	-	expression tag	UNP P20671
H	-2	ASP	-	expression tag	UNP P20671
H	-1	LEU	-	expression tag	UNP P20671
H	0	LEU	-	expression tag	UNP P20671

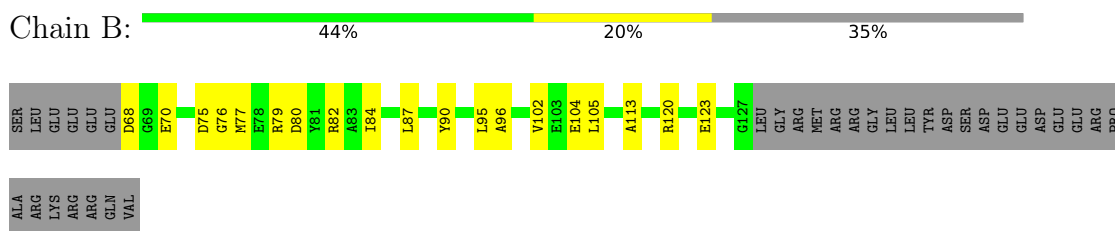
3 Residue-property plots [i](#)

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.

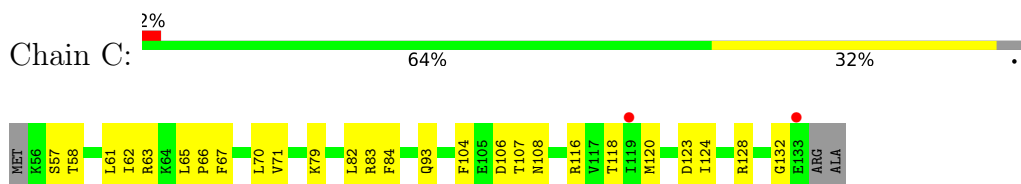
- Molecule 1: FACT complex subunit SPT16



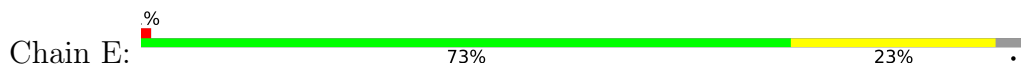
- Molecule 2: DNA replication licensing factor MCM2



- Molecule 3: Histone H3.1



- Molecule 3: Histone H3.1



4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α , β , γ	190.41Å 190.41Å 79.83Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.13 – 4.40 49.13 – 4.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.13-4.40) 100.0 (49.13-4.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	877.02 (at 4.45Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.293 , 0.312 0.290 , 0.308	Depositor DCC
R_{free} test set	1071 reflections (10.03%)	wwPDB-VP
Wilson B-factor (Å ²)	202.7	Xtrriage
Anisotropy	0.219	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 166.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.37$, $\langle L^2 \rangle = 0.20$	Xtrriage
Estimated twinning fraction	0.137 for h,-h-k,-l	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	6797	wwPDB-VP
Average B, all atoms (Å ²)	231.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.19% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [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 >5	RMSZ	# Z >5
1	A	0.27	0/2582	0.50	0/3487
2	B	0.26	0/472	0.59	0/631
3	C	0.28	0/637	0.56	0/854
3	E	0.27	0/638	0.56	0/856
4	D	0.27	0/581	0.62	0/778
4	F	0.26	0/572	0.57	0/766
5	G	0.25	0/710	0.47	0/957
6	H	0.26	0/707	0.56	0/952
All	All	0.27	0/6899	0.54	0/9281

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	2525	0	2439	61	0
2	B	470	0	427	16	0
3	C	630	0	661	24	0
3	E	631	0	661	19	0
4	D	576	0	618	30	0
4	F	567	0	615	20	0
5	G	699	0	711	17	0
6	H	699	0	737	16	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	6797	0	6869	151	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 151 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)
3:C:58:THR:HA	4:D:40:ARG:HH22	1.38	0.87
1:A:774:HIS:HA	1:A:777:LYS:HE2	1.59	0.84
4:F:78:ARG:NH1	4:F:80:THR:O	2.23	0.72
1:A:960:GLU:HB3	6:H:28:GLY:HA3	1.72	0.71
4:D:78:ARG:NH1	4:D:80:THR:O	2.23	0.70

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	303/350 (87%)	282 (93%)	21 (7%)	0	100	100
2	B	58/93 (62%)	56 (97%)	2 (3%)	0	100	100
3	C	76/81 (94%)	69 (91%)	7 (9%)	0	100	100
3	E	76/81 (94%)	72 (95%)	4 (5%)	0	100	100
4	D	70/103 (68%)	67 (96%)	3 (4%)	0	100	100
4	F	69/103 (67%)	69 (100%)	0	0	100	100
5	G	88/126 (70%)	86 (98%)	2 (2%)	0	100	100
6	H	88/136 (65%)	82 (93%)	6 (7%)	0	100	100
All	All	828/1073 (77%)	783 (95%)	45 (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	282/315 (90%)	281 (100%)	1 (0%)	91	94
2	B	45/75 (60%)	45 (100%)	0	100	100
3	C	68/70 (97%)	66 (97%)	2 (3%)	42	64
3	E	67/70 (96%)	67 (100%)	0	100	100
4	D	60/79 (76%)	60 (100%)	0	100	100
4	F	59/79 (75%)	59 (100%)	0	100	100
5	G	76/106 (72%)	74 (97%)	2 (3%)	46	67
6	H	71/106 (67%)	70 (99%)	1 (1%)	67	81
All	All	728/900 (81%)	722 (99%)	6 (1%)	81	89

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	G	37	TYR
5	G	65	PHE
6	H	17	ARG
3	C	84	PHE
1	A	838	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
4	F	27	GLN
5	G	49	HIS
1	A	723	HIS
1	A	735	HIS
1	A	755	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 monosaccharides 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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	307/350 (87%)	-0.16	5 (1%) 72 62	162, 212, 269, 282	0
2	B	60/93 (64%)	-0.60	0 100 100	173, 220, 258, 258	0
3	C	78/81 (96%)	0.00	2 (2%) 56 46	192, 225, 259, 267	0
3	E	78/81 (96%)	-0.04	1 (1%) 77 68	176, 220, 237, 241	0
4	D	72/103 (69%)	0.11	2 (2%) 53 43	209, 242, 266, 271	0
4	F	71/103 (68%)	-0.01	0 100 100	192, 223, 265, 267	0
5	G	90/126 (71%)	0.06	4 (4%) 34 28	232, 273, 283, 288	0
6	H	90/136 (66%)	0.46	10 (11%) 5 5	235, 267, 279, 284	0
All	All	846/1073 (78%)	-0.04	24 (2%) 53 43	162, 232, 277, 288	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	G	53	GLY	5.0
6	H	27	VAL	3.8
1	A	961	ASP	3.5
3	C	133	GLU	3.1
6	H	98	GLY	3.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

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.