



wwPDB X-ray Structure Validation Summary Report ⓘ

May 23, 2020 – 03:22 pm BST

PDB ID : 5W0V
Title : Crystal structure of full-length Kluyveromyces lactis Kap123 with histone H4 1-34
Authors : An, S.; Yoon, J.; Song, J.-J.; Cho, U.-S.
Deposited on : 2017-05-31
Resolution : 2.82 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 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	:	2.11
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

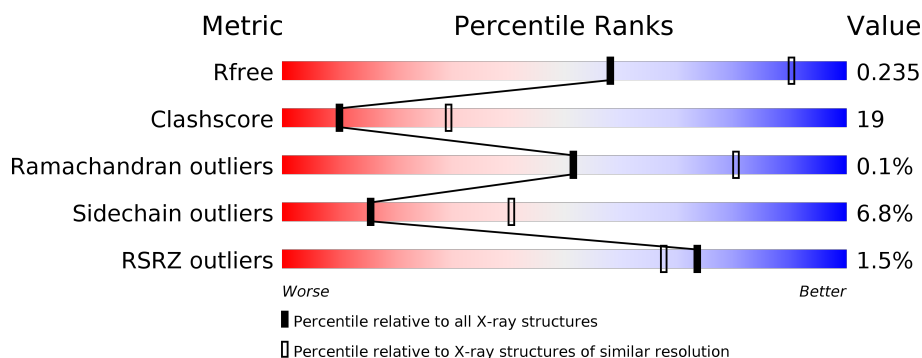
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.82 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	3617 (2.84-2.80)
Clashscore	141614	4060 (2.84-2.80)
Ramachandran outliers	138981	3978 (2.84-2.80)
Sidechain outliers	138945	3980 (2.84-2.80)
RSRZ outliers	127900	3552 (2.84-2.80)

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 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	1116	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, orange 1%, yellow 29%, green 59%, grey 11%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> % 59% 29% • 9% </div> </div>
1	B	1116	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 2%, orange 1%, yellow 27%, green 60%, grey 10%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 2% 60% 27% • 10% </div> </div>
2	C	34	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 15%, yellow 6%, grey 79%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 15% 6% 79% </div> </div>
2	D	34	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 12%, yellow 6%, grey 82%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 12% 6% 82% </div> </div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 15787 atoms, of which 38 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 Kap123.

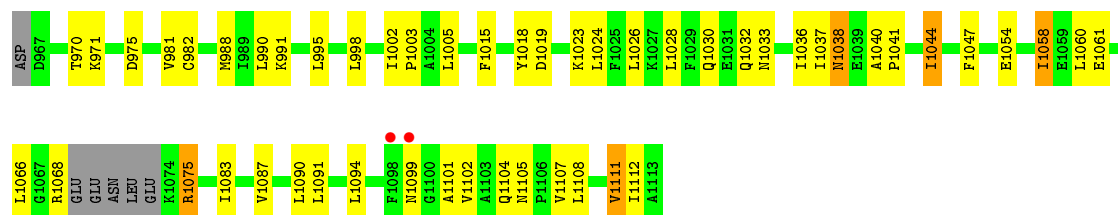
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1017	Total	C	H	N	O	S	0	0	0
			7910	5020	38	1288	1543	21			
1	B	1003	Total	C	N	O	S		0	0	0
			7773	4957	1266	1529	21				

There are 6 discrepancies between the modelled and reference sequences:

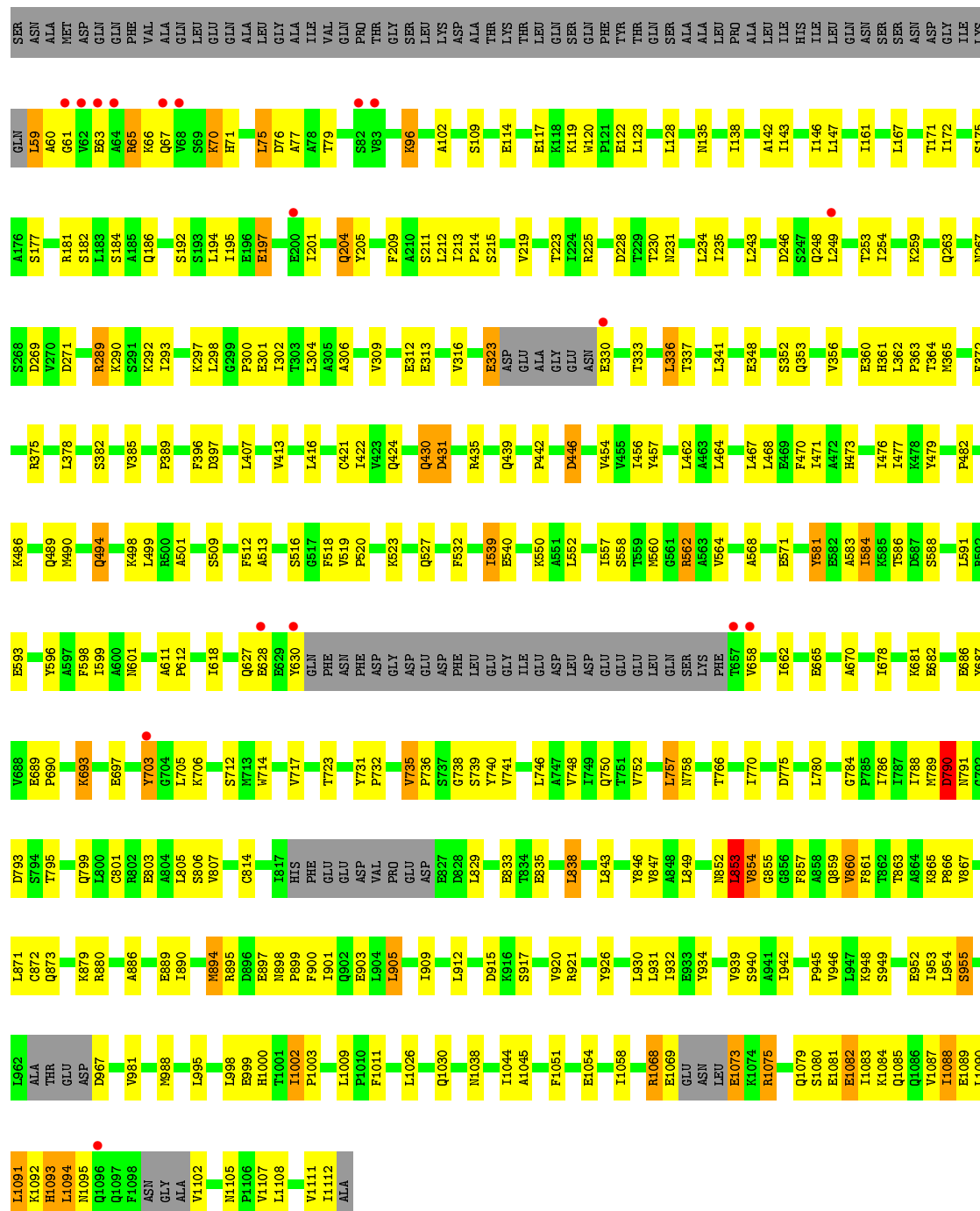
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP Q6CMF0
A	-1	ASN	-	expression tag	UNP Q6CMF0
A	0	ALA	-	expression tag	UNP Q6CMF0
B	-2	SER	-	expression tag	UNP Q6CMF0
B	-1	ASN	-	expression tag	UNP Q6CMF0
B	0	ALA	-	expression tag	UNP Q6CMF0

- Molecule 2 is a protein called Histone H4 1-34.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	7	Total	C	N	O	0	0	0
			54	31	16	7			
2	D	6	Total	C	N	O	0	0	0
			50	29	15	6			



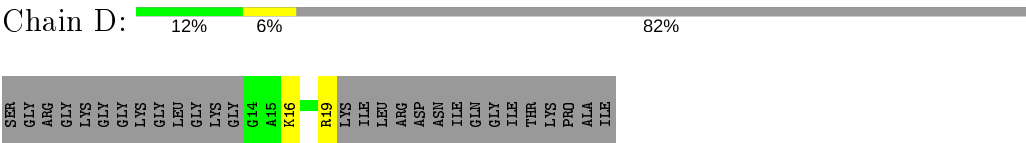
• Molecule 1: Kap123



● Molecule 2: Histone H4 1-34



● Molecule 2: Histone H4 1-34



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	78.79 Å 87.70 Å 101.62 Å 79.59° 81.45° 71.70°	Depositor
Resolution (Å)	33.79 – 2.82 33.79 – 2.82	Depositor EDS
% Data completeness (in resolution range)	98.3 (33.79-2.82) 98.3 (33.79-2.82)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.79 (at 2.81 Å)	Xtriage
Refinement program	PHENIX 1.11.1 _2575	Depositor
R, R_{free}	0.170 , 0.234 0.172 , 0.235	Depositor DCC
R_{free} test set	1998 reflections (3.34%)	wwPDB-VP
Wilson B-factor (Å ²)	75.8	Xtriage
Anisotropy	0.512	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 53.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	15787	wwPDB-VP
Average B, all atoms (Å ²)	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.38% 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	0.46	0/7990	0.62	1/10843 (0.0%)
1	B	0.49	0/7889	0.63	6/10705 (0.1%)
2	C	0.34	0/54	0.78	0/68
2	D	0.40	0/50	0.86	0/63
All	All	0.47	0/15983	0.63	7/21679 (0.0%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	581	TYR	CA-CB-CG	8.67	129.87	113.40
1	B	854	VAL	N-CA-CB	6.53	125.86	111.50
1	A	852	ASN	CB-CA-C	-6.50	97.41	110.40
1	B	562	ARG	NE-CZ-NH1	-6.02	117.29	120.30
1	B	853	LEU	CB-CA-C	5.73	121.09	110.20

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	7872	38	7970	280	0
1	B	7773	0	7854	320	0
2	C	54	0	56	9	0
2	D	50	0	53	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	15749	38	15933	599	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 19.

The worst 5 of 599 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)
1:B:736:PRO:HD2	1:B:789:MET:CE	1.56	1.33
1:B:739:SER:N	1:B:791:ASN:OD1	1.61	1.32
1:B:738:GLY:HA2	1:B:791:ASN:OD1	1.29	1.24
1:B:738:GLY:CA	1:B:791:ASN:OD1	1.86	1.21
1:B:738:GLY:HA2	1:B:791:ASN:CG	1.64	1.18

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	1003/1116 (90%)	983 (98%)	19 (2%)	1 (0%)	51	80
1	B	989/1116 (89%)	968 (98%)	20 (2%)	1 (0%)	51	80
2	C	5/34 (15%)	5 (100%)	0	0	100	100
2	D	4/34 (12%)	4 (100%)	0	0	100	100
All	All	2001/2300 (87%)	1960 (98%)	39 (2%)	2 (0%)	51	80

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	854	VAL
1	B	854	VAL

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	873/958 (91%)	816 (94%)	57 (6%)	17	43
1	B	863/958 (90%)	801 (93%)	62 (7%)	14	37
2	C	4/23 (17%)	4 (100%)	0	100	100
2	D	4/23 (17%)	4 (100%)	0	100	100
All	All	1744/1962 (89%)	1625 (93%)	119 (7%)	16	40

5 of 119 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1061	GLU
1	B	177	SER
1	B	1068	ARG
1	A	1099	ASN
1	B	71	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	67	GLN
1	A	71	HIS
1	A	156	ASN
1	B	204	GLN
1	B	799	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 [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	1017/1116 (91%)	-0.41	13 (1%) 77 72	49, 77, 116, 142	0
1	B	1003/1116 (89%)	-0.31	17 (1%) 70 63	45, 82, 126, 152	0
2	C	7/34 (20%)	-0.46	0 100 100	73, 98, 103, 104	0
2	D	6/34 (17%)	-0.46	0 100 100	64, 85, 88, 91	0
All	All	2033/2300 (88%)	-0.36	30 (1%) 73 67	45, 79, 122, 152	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	703	TYR	6.9
1	B	67	GLN	3.6
1	B	630	TYR	3.4
1	B	63	GLU	3.2
1	B	703	TYR	3.1

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 carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.