wwPDB X-ray Structure Validation Summary Report (i)

Nov 6, 2023 – 01:58 PM EST

PDB ID	:	5VE8
Title	:	Crystal structure of full-length Kluyveromyces lactis Kap123 with histone H3 1-28
Authors	:	An, S.; Yoon, J.; Song, J.-J.; Cho, U.-S.
Deposited on	:	2017-04-04
Resolution	:	2.70 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 (i) 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 (i)) were used in the production of this report:

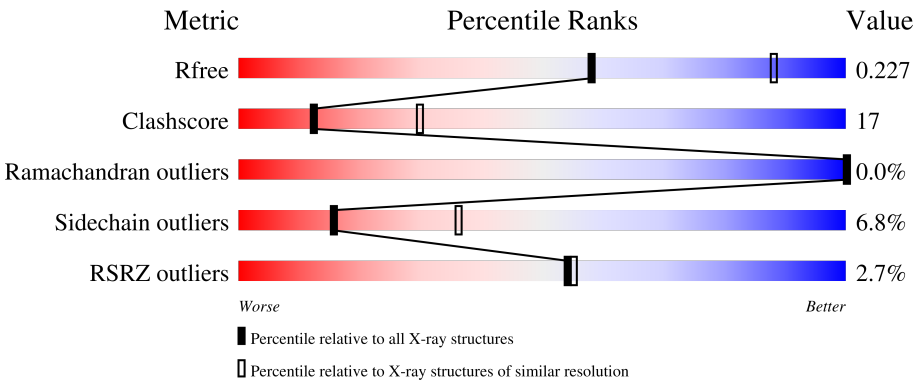
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 <=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	3% 65% 25% • 8%
1	B	1116	2% 64% 25% • 8%
2	C	28	7% 14% 18% 7% 61%
2	E	28	4% 18% 18% 7% 57%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 16167 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
			Total	C	H	N	O	S	Se			
1	A	1025	7957	5045	38	1296	1557	10	11	0	0	0
1	B	1030	7977	5082		1303	1571	10	11	0	0	0

There are 8 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
A	1	MSE	-	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
B	1	MSE	-	expression tag	UNP Q6CMF0

- Molecule 2 is a protein called Histone H3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	11	70	42	16	12	0	0	0
2	E	12	81	48	20	13	0	0	0

- Molecule 3 is water.

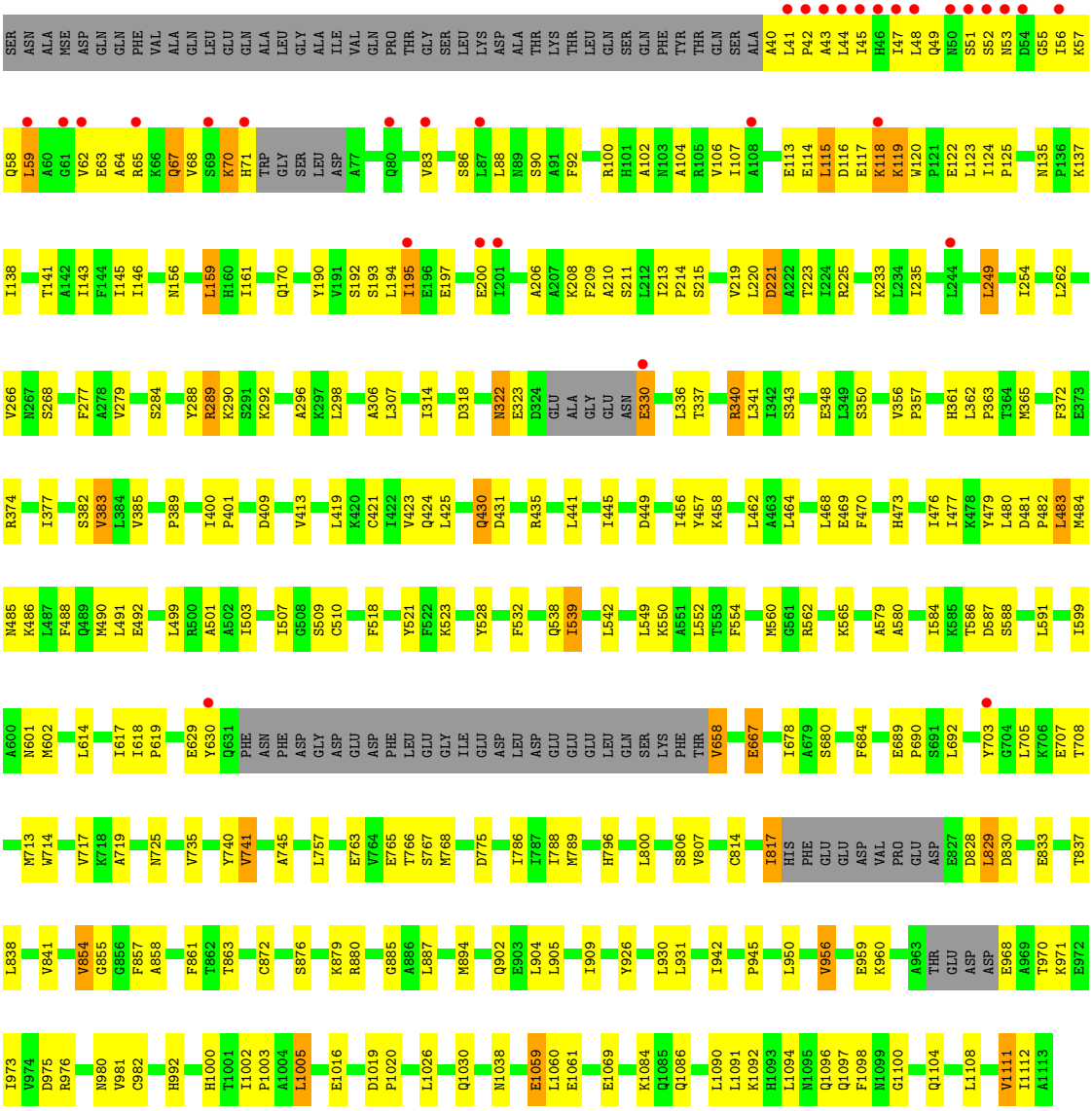
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	35	35	35	0	0
3	B	47	47	47	0	0

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: Kap123

Chain A: 3% 65% 25% 8%



- Molecule 1: Kap123

Chain B: 2% 64% 25% • 8%

SER ASN ALA MSE ASP GLN GLN PHE VAL ALA GLN LEU GLU GLN ALA LEU GLY ALA ILE VAL GLN PRO THR GLY SER LEU LYS ASP ALA THR LYS THR LEU GLN SER GLN PHE TYR THR GLN SER ALA A40 L41 P42 A43 L44 I45 H46 I47 L48 Q49 N50 S51 S52 N53 D54 G55 I56 K57

Q58 L59 A60 E63 A64 R65 K66 Q67 V68 S69 K70 H71 W72 G73 S74 L75 D76 A77 A78 T79 Q80 T81 S82 V83 K84 Q85 S86 L87 A91 F92 N93 E94 G95 K96 D97 R100 H101 A102 N103 A104 R105 V106 S109 E113 E114 L115 D116 E117 K118 K119 W120 P121 E122 L128

N135 P136 K137 I138 T141 I145 I146 L147 F153 N156 L159 L165 N166 L167 S175 L178 S182 Q186 A187 L188 S192 S193 L194 I195 E196 E197 N202 Y205 F209 A210 S211 L212 I213 P214 S215 V219 T223 E226 N237 L249

T250 T253 I254 A255 D256 L257 V258 K259 L260 A261 L262 V266 R275 S284 K290 S291 K292 A296 K297 L298 I302 T303 L304 A305 A306 L307 K308 E312 E313 V316 E317 T321 N322 E323 D324 GLU ALA GLY GLU ASN E330 E331 N332 T333 L336 R340 S343

V356 P357 H361 F372 L378 S382 V383 P389 D397 I400 P401 V413 L416 Q424 E432 R435 Y436 H437 E438 L441 I456 Y457 K458 L464 L467 L468 E469 F470 I471 A472 H473 I476 Y479 P482 L483 N484 F488 E492

T493 Q494 K498 L499 R500 A501 F512 A513 Y528 F532 I539 K550 A551 L552 N560 G561 R562 A563 V564 K565 Y581 E582 A583 I584 K585 T586 D587 S588 L591 Y596 A597 F598 I599 A600 N601 M602 V605 T616 F622 E626 Y630 Q631 F632 ASN PHE

ASP GLY ASP GLU ASP PHE LEU GLU GLY ILE GLU ASP LEU ASP GLU GLU GLU LEU GLN SER LYS PHE T657 Y664 E675 S680 H683 F684 L685 E686 Y687 V688 E689 P690 K693 Y703 G704 L705 K706 E707 L710 W714 V717 L722 L726 Y731 V735

P736 S737 Y740 V752 T766 S767 M768 V772 D775 L776 S777 E778 M779 L780 R781 L782 F783 G784 P785 I786 M789 D790 N791 G792 D793 S794 T795 H796 L797 C801 R802 C814 I817 HIS PHE GLU GLU ASP VAL PRO GLU ASP GLU D828 L829 D830 E833 T834 E835

L849 N852 L853 V854 F857 V860 F861 T862 T863 C872 Q873 S874 K875 R880 S881 F882 G885 L892 E897 I901 L905 S911 S917 R921 C922 N923 A924 S925 Y926 G927 V928 G929 L930 V939 I942 Y943 S944 P945 L950 V956 N961 LEU

ALA THR GLU ASP D967 E968 A969 T970 D975 R976 V981 I1002 P1003 A1004 L1005 T1013 E1016 E1017 Y1018 D1019 P1020 I1021 F1022 L1026 E1031 Q1032 I1036 I1037 P1041 I1044 E1059 T1065 R1068 N1071 L1072 R1075 E1082 L1090 L1091 K1092 H1093 L1094 N1095

Q1096 Q1097 F1098 N1099 G1100 A1101 V1102 A1103 Q1104 L1108 V1111 I1112 A1113

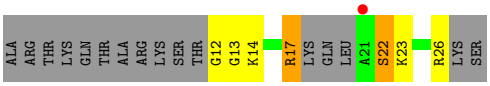
• Molecule 2: Histone H3

Chain C: 7% 14% 18% 7% 61%

ALA ARG THR LYS GLN THR ALA ARG LYS SER THR G12 G13 K14 A15 P16 R17 LYS GLN LEU A21 S22 K23 A24 A25 ARG LYS SER

• Molecule 2: Histone H3

Chain E: 4% 18% 18% 7% 57%


ALA
ARG
THR
LYS
GLN
THR
ALA
ARG
LYS
SER
THR
G12
G13
K14
R17
LYS
GLN
LEU
A21
S22
K23
R26
LYS
SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	79.10Å 88.11Å 102.28Å 79.59° 80.77° 70.74°	Depositor
Resolution (Å)	44.61 – 2.70 44.62 – 2.70	Depositor EDS
% Data completeness (in resolution range)	98.8 (44.61-2.70) 98.8 (44.62-2.70)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.81 (at 2.69Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.185 , 0.230 0.186 , 0.227	Depositor DCC
R_{free} test set	2000 reflections (2.89%)	wwPDB-VP
Wilson B-factor (Å^2)	58.5	Xtriage
Anisotropy	0.383	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 51.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	16167	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.86% 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 (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	0.49	0/8025	0.61	1/10873 (0.0%)
1	B	0.74	5/8087 (0.1%)	0.65	2/10959 (0.0%)
2	C	1.67	0/69	0.94	0/88
2	E	0.93	0/80	1.41	1/102 (1.0%)
All	All	0.64	5/16261 (0.0%)	0.64	4/22022 (0.0%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	675	GLU	CD-OE1	-6.20	1.18	1.25
1	B	436	TYR	CE1-CZ	-6.12	1.30	1.38
1	B	432	GLU	CD-OE2	-5.73	1.19	1.25
1	B	801	CYS	CB-SG	-5.21	1.73	1.81
1	B	436	TYR	CB-CG	-5.19	1.43	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	22	SER	N-CA-C	9.57	136.85	111.00
1	B	801	CYS	CA-CB-SG	-5.77	103.62	114.00
1	A	430	GLN	N-CA-C	5.27	125.22	111.00
1	B	1091	LEU	CB-CG-CD1	-5.22	102.13	111.00

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	7919	38	8002	235	0
1	B	7977	0	8055	303	0
2	C	70	0	75	13	0
2	E	81	0	88	13	0
3	A	35	0	0	5	0
3	B	47	0	0	4	0
All	All	16129	38	16220	546	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 17.

The worst 5 of 546 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:735:VAL:HG13	1:B:789:MSE:HE1	1.16	1.15
1:B:630:TYR:HE2	1:B:703:TYR:CE1	1.67	1.13
1:B:195:ILE:HD13	1:B:205:TYR:HB3	1.24	1.11
1:B:92:PHE:HA	1:B:100:ARG:HH12	1.15	1.08
1:B:435:ARG:NH1	1:B:968:GLU:OE2	1.87	1.08

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	1013/1116 (91%)	994 (98%)	18 (2%)	1 (0%)	51 78
1	B	1020/1116 (91%)	1004 (98%)	16 (2%)	0	100 100
2	C	7/28 (25%)	6 (86%)	1 (14%)	0	100 100
2	E	8/28 (29%)	6 (75%)	2 (25%)	0	100 100
All	All	2048/2288 (90%)	2010 (98%)	37 (2%)	1 (0%)	100 100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	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	876/946 (93%)	818 (93%)	58 (7%)	16	38
1	B	885/946 (94%)	826 (93%)	59 (7%)	16	37
2	C	5/20 (25%)	3 (60%)	2 (40%)	0	0
2	E	6/20 (30%)	5 (83%)	1 (17%)	2	5
All	All	1772/1932 (92%)	1652 (93%)	120 (7%)	16	36

5 of 120 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1111	VAL
1	B	1072	LEU
1	B	147	LEU
1	B	1059	GLU
2	C	17	ARG

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

Mol	Chain	Res	Type
1	A	156	ASN
1	B	322	ASN
1	B	332	ASN
1	B	430	GLN
1	B	1063	ASN

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	1014/1116 (90%)	-0.02	32 (3%)	47	48	37, 63, 110, 140	0
1	B	1019/1116 (91%)	-0.03	21 (2%)	63	65	32, 63, 108, 130	0
2	C	11/28 (39%)	0.46	2 (18%)	1	1	47, 64, 88, 88	0
2	E	12/28 (42%)	0.57	1 (8%)	11	9	49, 64, 80, 85	0
All	All	2056/2288 (89%)	-0.02	56 (2%)	54	55	32, 63, 108, 140	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	41	LEU	5.2
1	B	45	ILE	5.0
1	A	52	SER	4.8
1	A	83	VAL	4.6
1	A	45	ILE	4.2

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 (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.