



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 02:31 pm GMT

PDB ID : 2VWB
Title : Structure of the archaeal Kae1-Bud32 fusion protein MJ1130: a model for the eukaryotic EKC-KEOPS subcomplex involved in transcription and telomere homeostasis.
Authors : Hecker, A.; Lopreiato, R.; Graille, M.; Collinet, B.; Forterre, P.; Domenico, L.; van Tilbeurgh, H.
Deposited on : 2008-06-20
Resolution : 3.05 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 ⓘ symbol.

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

MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

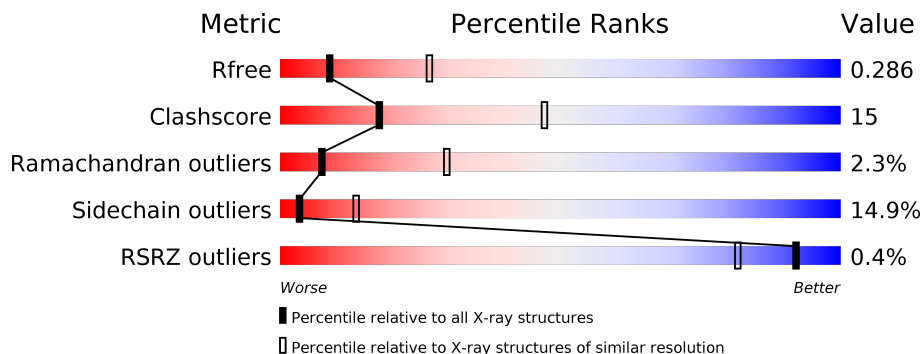
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.05 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}	100719	1348 (3.10-3.02)
Clashscore	112137	1462 (3.10-3.02)
Ramachandran outliers	110173	1410 (3.10-3.02)
Sidechain outliers	110143	1410 (3.10-3.02)
RSRZ outliers	101464	1355 (3.10-3.02)

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 $\leq 5\%$. 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	535	<div> <div></div> <div>62% 30% 5% .</div> </div>
1	B	535	<div> <div></div> <div>56% 33% 6% 5%</div> </div>

2 Entry composition [i](#)

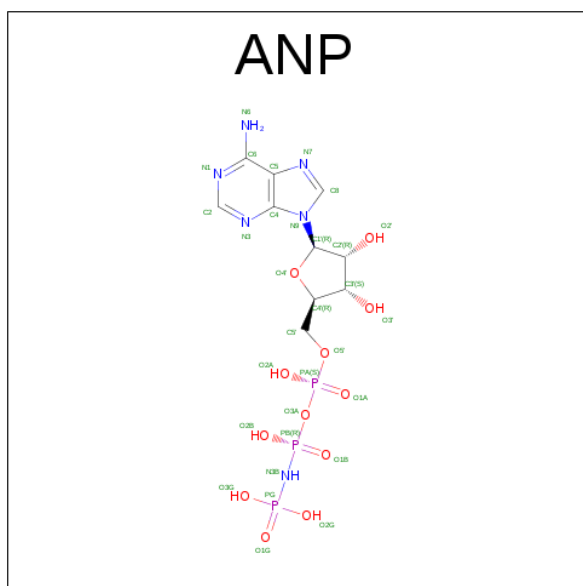
There are 2 unique types of molecules in this entry. The entry contains 8210 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 PUTATIVE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	519	Total	C	N	O	S	0	0	0
			4125	2636	698	770	21			
1	B	507	Total	C	N	O	S	0	0	0
			4023	2568	680	754	21			

- Molecule 2 is PHOSPHOAMINOPHOSPHONIC ACID-ADENYLATE ESTER (three-letter code: ANP) (formula: $C_{10}H_{17}N_6O_{12}P_3$).

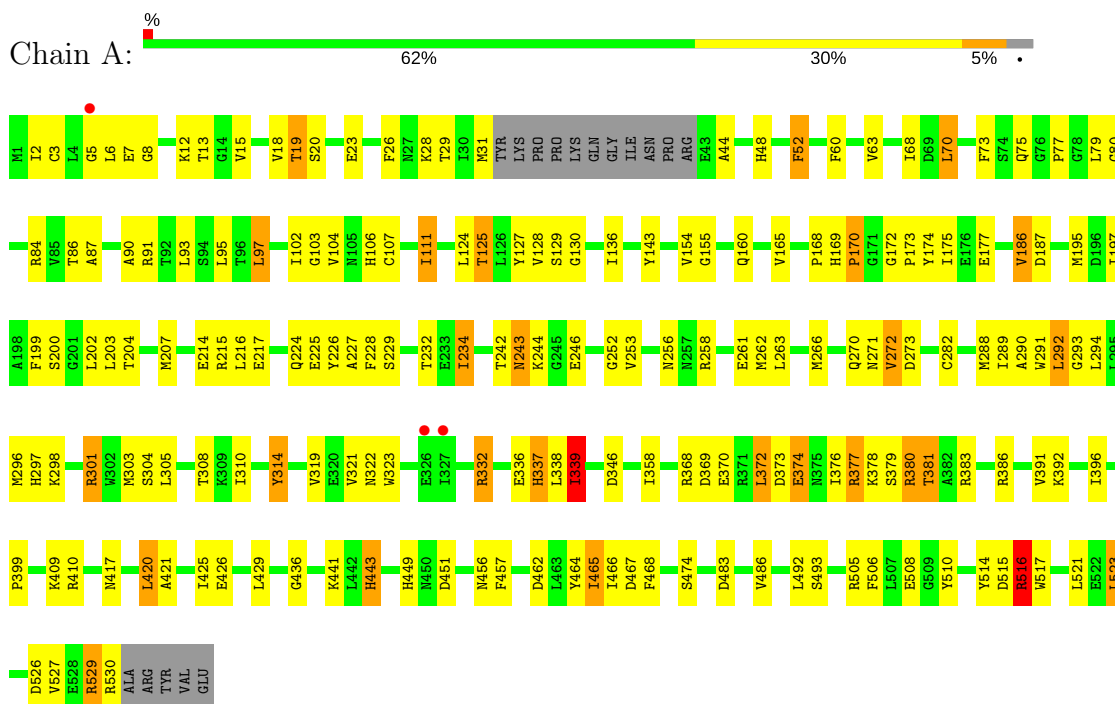


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			31	10	6	12	3		
2	B	1	Total	C	N	O	P	0	0
			31	10	6	12	3		

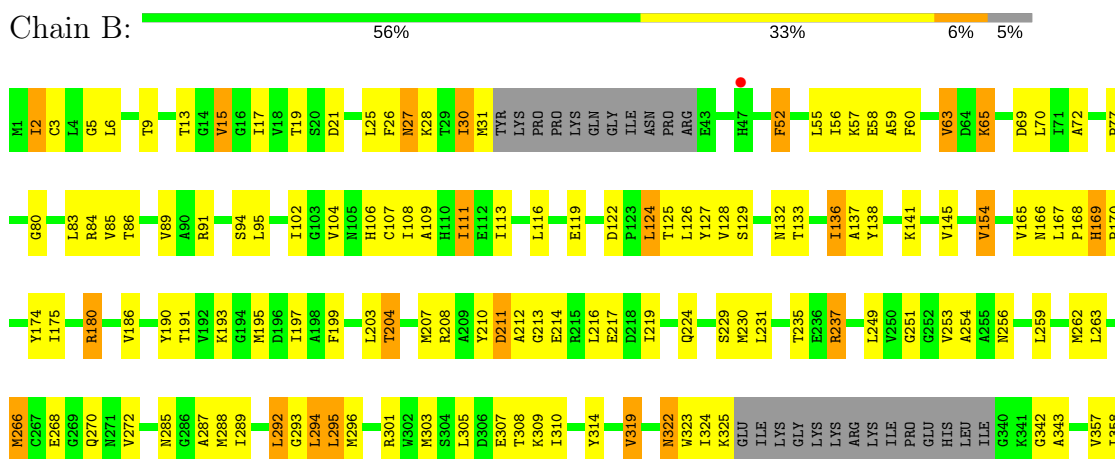
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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: PUTATIVE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE



• Molecule 1: PUTATIVE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE



I359	K360	E361	R362	Y367	E370	R371	L372	D373	E374	I375	I376	R377	R380	T381	A382	R383	E384	A385	R386	Y387	L388	V391	I396	V404	I408	K409	R410	Y415	I416	I417	G418	K419	L420	A421	K422	D423	V424	I425	R426	D427	V428	L429	D430	I431	A432	Y433	I438	K441	L442
H443	K444	I448	H449	M450	D451	F457	L463	Y464	I465	I466	D467	F468	G469	L470	S474	D477	E478	D479	V482	D483	V486	L492	S493	I494	H495	F499	D500	E501	R505	Y510	K511	S512	V513	Y514	D515	R516	W517	I520	L523	A531	R532	TYR	VAL	GLU					

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	147.42Å 148.95Å 65.12Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.95 – 3.05 19.95 – 3.05	Depositor EDS
% Data completeness (in resolution range)	94.4 (19.95-3.05) 94.4 (19.95-3.05)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.28 (at 3.04Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.220 , 0.301 0.210 , 0.286	Depositor DCC
R_{free} test set	1319 reflections (5.26%)	DCC
Wilson B-factor (Å ²)	70.9	Xtriage
Anisotropy	0.301	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 32.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.010 for k,h,-l	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	8210	wwPDB-VP
Average B, all atoms (Å ²)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.85% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ANP

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	0.54	0/4199	0.70	1/5660 (0.0%)
1	B	0.54	0/4094	0.72	1/5520 (0.0%)
All	All	0.54	0/8293	0.71	2/11180 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	124	LEU	CA-CB-CG	5.87	128.80	115.30
1	B	124	LEU	CA-CB-CG	5.04	126.88	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	169	HIS	Peptide

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	4125	0	4181	122	0
1	B	4023	0	4060	131	0
2	A	31	0	13	2	0
2	B	31	0	13	1	0
All	All	8210	0	8267	252	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 15.

The worst 5 of 252 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:A:516:ARG:HG2	1:A:516:ARG:HH11	1.13	1.13
1:B:104:VAL:HG13	1:B:292:LEU:HD12	1.34	1.08
1:A:242:THR:HG21	1:A:244:LYS:HG2	1.37	1.06
1:A:242:THR:CG2	1:A:244:LYS:HG2	1.90	1.01
1:A:107:CYS:HG	1:A:127:TYR:HE2	1.09	0.99

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	515/535 (96%)	454 (88%)	51 (10%)	10 (2%)	9	35
1	B	501/535 (94%)	425 (85%)	63 (13%)	13 (3%)	6	27
All	All	1016/1070 (95%)	879 (86%)	114 (11%)	23 (2%)	7	30

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	170	PRO
1	A	337	HIS
1	A	339	ILE
1	B	166	ASN
1	A	80	GLY

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	443/457 (97%)	372 (84%)	71 (16%)	3	11
1	B	431/457 (94%)	372 (86%)	59 (14%)	4	17
All	All	874/914 (96%)	744 (85%)	130 (15%)	3	13

5 of 130 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	468	PHE
1	B	13	THR
1	B	419	LYS
1	A	492	LEU
1	A	516	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	449	HIS
1	B	224	GLN
1	B	443	HIS
1	A	456	ASN
1	B	160	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](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	ANP	A	1531	-	29,33,33	2.08	5 (17%)	28,52,52	2.00	7 (25%)
2	ANP	B	1533	-	29,33,33	2.19	6 (20%)	28,52,52	2.05	7 (25%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	ANP	A	1531	-	-	0/13/38/38	0/3/3/3
2	ANP	B	1533	-	-	0/13/38/38	0/3/3/3

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1533	ANP	PB-O3A	2.27	1.61	1.59
2	A	1531	ANP	C5-C4	3.08	1.47	1.40
2	B	1533	ANP	C5-C4	3.30	1.48	1.40
2	A	1531	ANP	PG-O1G	4.28	1.51	1.46
2	B	1533	ANP	PG-O1G	4.41	1.51	1.46

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1531	ANP	N3-C2-N1	-6.20	123.46	128.86
2	B	1533	ANP	N3-C2-N1	-6.04	123.60	128.86
2	B	1533	ANP	C4'-O4'-C1'	-3.88	105.64	109.77
2	A	1531	ANP	PA-O3A-PB	-3.00	121.80	132.38
2	B	1533	ANP	C4-C5-N7	-2.98	106.53	109.41

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1531	ANP	2	0
2	B	1533	ANP	1	0

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	519/535 (97%)	-0.32	3 (0%)	89 75	34, 57, 85, 107	0
1	B	507/535 (94%)	-0.34	1 (0%)	94 88	35, 60, 93, 112	0
All	All	1026/1070 (95%)	-0.33	4 (0%)	92 81	34, 58, 90, 112	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	5	GLY	2.8
1	A	326	GLU	2.4
1	B	47	HIS	2.2
1	A	327	ILE	2.2

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 carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q < 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
2	ANP	A	1531	31/31	0.96	0.13	-0.69	43,45,74,74	0
2	ANP	B	1533	31/31	0.96	0.14	-0.73	66,70,87,87	0

6.5 Other polymers [i](#)

There are no such residues in this entry.