



Full wwPDB X-ray Structure Validation Report ⓘ

Feb 28, 2026 – 06:42 PM UTC

PDB ID : 5DUD / pdb_00005dud
Title : Crystal structure of E. coli YbgJK
Authors : Arbing, M.A.; Kaufmann, M.; Shin, A.; Medrano-Soto, A.; Cascio, D.; Eisenberg, D.
Deposited on : 2015-09-18
Resolution : 2.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

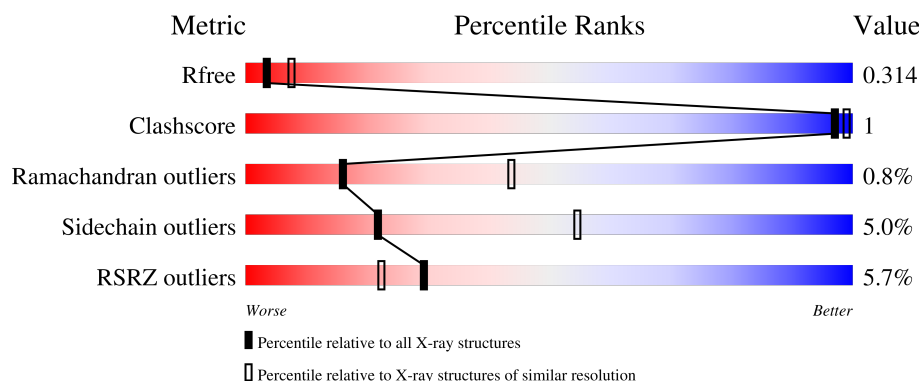
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

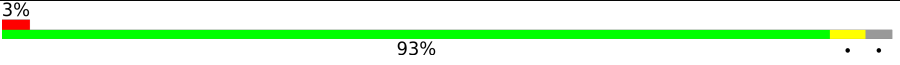
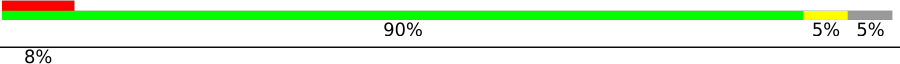
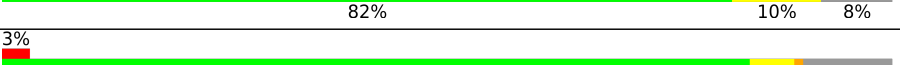

The reported resolution of this entry is 2.80 Å.

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}	180053	3866 (2.80-2.80)
Clashscore	190562	4276 (2.80-2.80)
Ramachandran outliers	187476	4196 (2.80-2.80)
Sidechain outliers	187428	4198 (2.80-2.80)
RSRZ outliers	180081	3869 (2.80-2.80)

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 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	310	
1	C	310	
2	B	218	
2	D	218	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 14331 atoms, of which 6880 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 YbgK.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	300	Total	C	H	N	O	S	0	0	0
			4459	1427	2190	417	411	14			
1	C	295	Total	C	H	N	O	S	0	0	0
			4103	1343	1963	391	392	14			

- Molecule 2 is a protein called YbgJ.

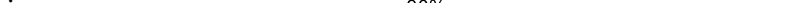
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	200	Total	C	H	N	O	S	0	0	0
			2808	944	1363	241	256	4			
2	D	197	Total	C	H	N	O	S	0	0	0
			2795	921	1364	247	259	4			

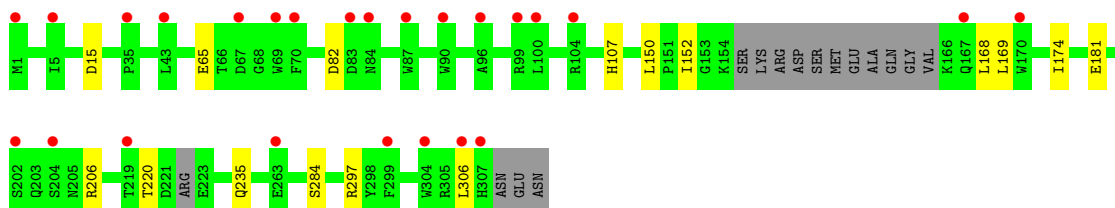
- Molecule 3 is water.

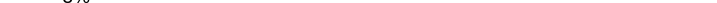
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	67	Total	O	0	0
			67	67		
3	B	29	Total	O	0	0
			29	29		
3	C	39	Total	O	0	0
			39	39		
3	D	31	Total	O	0	0
			31	31		

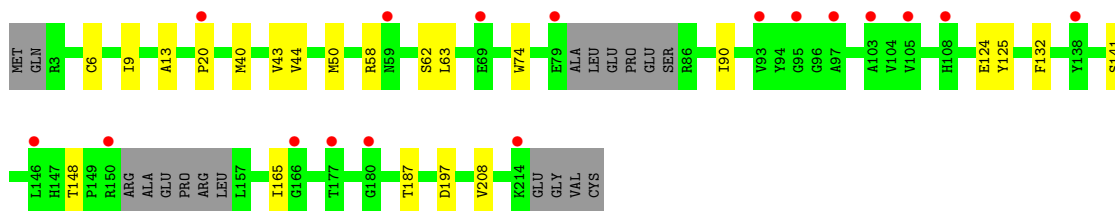
- Molecule 1: YbgK

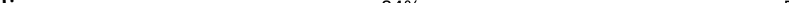


Chain C:  8% 90% 5% 5%



Chain B:  8% 82% 10% 8%



Chain D:  3% 84% 5% 10%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	66.11Å 88.78Å 83.70Å 90.00° 101.91° 90.00°	Depositor
Resolution (Å)	15.00 – 2.80 15.00 – 2.80	Depositor EDS
% Data completeness (in resolution range)	98.6 (15.00-2.80) 98.2 (15.00-2.80)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.79 (at 2.81Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.268 , 0.288 0.296 , 0.314	Depositor DCC
R_{free} test set	1196 reflections (4.59%)	wwPDB-VP
Wilson B-factor (Å ²)	48.3	Xtriage
Anisotropy	0.336	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 46.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.83	EDS
Total number of atoms	14331	wwPDB-VP
Average B, all atoms (Å ²)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.59% 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

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	0.73	0/2319	0.96	0/3146
1	C	0.74	0/2187	0.95	1/2978 (0.0%)
2	B	0.84	0/1485	1.06	1/2042 (0.0%)
2	D	0.77	0/1470	1.01	2/2020 (0.1%)
All	All	0.76	0/7461	0.99	4/10186 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	20	PRO	N-CA-C	5.85	117.84	110.70
2	D	20	PRO	N-CA-C	5.73	117.69	110.70
2	D	9	ILE	N-CA-CB	5.65	118.38	111.60
1	C	15	ASP	CA-CB-CG	5.27	117.87	112.60

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	2269	2190	2200	2	0
1	C	2140	1963	1972	1	0
2	B	1445	1363	1374	7	0
2	D	1431	1364	1364	2	0
3	A	67	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	29	0	0	0	0
3	C	39	0	0	0	0
3	D	31	0	0	0	0
All	All	7451	6880	6910	12	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 1.

All (12) 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)
2:B:9:ILE:HG23	2:B:13:ALA:HB3	1.76	0.67
2:B:90:ILE:HD12	2:B:208:VAL:HG13	1.95	0.49
2:D:65:LEU:O	2:D:68:ILE:HG22	2.14	0.48
2:B:40:MET:HE3	2:B:43:VAL:HG21	1.95	0.47
1:A:43:LEU:HD21	1:A:165:VAL:HG22	2.00	0.44
2:B:44:VAL:HG12	2:B:58:ARG:NH1	2.34	0.43
2:B:125:TYR:O	2:B:208:VAL:N	2.52	0.42
1:C:181:GLU:OE2	1:C:181:GLU:N	2.53	0.42
2:D:44:VAL:HG12	2:D:58:ARG:NH1	2.34	0.42
2:B:40:MET:HE1	2:B:74:TRP:CH2	2.55	0.41
2:B:90:ILE:HD12	2:B:208:VAL:CG1	2.49	0.41
1:A:6:ARG:NH2	1:A:102:LEU:O	2.55	0.40

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	296/310 (96%)	277 (94%)	18 (6%)	1 (0%)	36 66
1	C	289/310 (93%)	272 (94%)	15 (5%)	2 (1%)	18 47

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	194/218 (89%)	181 (93%)	10 (5%)	3 (2%)	8	28
2	D	193/218 (88%)	181 (94%)	10 (5%)	2 (1%)	12	38
All	All	972/1056 (92%)	911 (94%)	53 (6%)	8 (1%)	16	44

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	197	ASP
2	D	141	SER
1	A	82	ASP
2	B	141	SER
1	C	82	ASP
2	D	63	LEU
2	B	132	PHE
1	C	152	ILE

5.3.2 Protein sidechains ⓘ

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	225/251 (90%)	217 (96%)	8 (4%)	31	66
1	C	197/251 (78%)	185 (94%)	12 (6%)	17	46
2	B	139/182 (76%)	131 (94%)	8 (6%)	18	49
2	D	141/182 (78%)	134 (95%)	7 (5%)	22	54
All	All	702/866 (81%)	667 (95%)	35 (5%)	22	54

All (35) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	15	ASP
1	A	32	LEU
1	A	169	LEU
1	A	174	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	201	SER
1	A	217	LYS
1	A	222	ARG
1	A	302	LEU
2	B	6	CYS
2	B	50	MET
2	B	62	SER
2	B	63	LEU
2	B	124	GLU
2	B	148	THR
2	B	165	ILE
2	B	187	THR
1	C	65	GLU
1	C	107	HIS
1	C	150	LEU
1	C	168	LEU
1	C	169	LEU
1	C	174	ILE
1	C	206	ARG
1	C	220	THR
1	C	235	GLN
1	C	284	SER
1	C	297	ARG
1	C	306	LEU
2	D	8	LEU
2	D	9	ILE
2	D	51	ASN
2	D	68	ILE
2	D	187	THR
2	D	198	GLU
2	D	203	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (17) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	14	GLN
1	A	19	HIS
1	A	49	ASN
1	A	235	GLN
1	A	250	GLN
1	A	295	GLN
2	B	52	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	72	GLN
2	B	133	GLN
2	B	213	GLN
1	C	14	GLN
1	C	19	HIS
1	C	23	GLN
1	C	49	ASN
1	C	172	ASN
1	C	295	GLN
2	D	182	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 oligosaccharides 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 ⓘ

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, 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	300/310 (96%)	0.63	8 (2%) 56 46	23, 41, 59, 87	0
1	C	295/310 (95%)	0.93	25 (8%) 16 12	28, 52, 83, 114	0
2	B	200/218 (91%)	0.94	17 (8%) 16 12	32, 52, 74, 101	0
2	D	197/218 (90%)	0.69	7 (3%) 46 37	27, 42, 66, 106	0
All	All	992/1056 (93%)	0.79	57 (5%) 29 22	23, 46, 74, 114	0

All (57) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	83	ASP	5.5
1	C	202	SER	4.6
2	B	150	ARG	3.8
2	B	69	GLU	3.7
2	B	95	GLY	3.5
1	C	104	ARG	3.5
2	B	20	PRO	3.3
2	B	79	GLU	3.1
1	C	67	ASP	3.1
2	B	59	ASN	3.1
1	C	304	TRP	3.0
2	B	214	LYS	3.0
2	D	66	ASP	3.0
2	B	108	HIS	3.0
1	C	306	LEU	3.0
1	C	219	THR	3.0
1	A	305	ARG	2.9
1	C	35	PRO	2.9
1	A	304	TRP	2.9
1	A	196	SER	2.8
2	D	72	GLN	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	177	THR	2.7
2	D	30	ILE	2.7
2	B	180	GLY	2.6
1	C	96	ALA	2.6
1	C	43	LEU	2.5
1	C	170	TRP	2.5
1	C	263	GLU	2.5
1	C	5	ILE	2.4
1	C	90	TRP	2.4
1	C	1	MET	2.4
1	A	82	ASP	2.4
1	A	214	GLN	2.4
1	C	70	PHE	2.4
1	C	167	GLN	2.4
1	A	188	ALA	2.3
1	C	69	TRP	2.3
1	C	87	TRP	2.3
2	B	146	LEU	2.3
1	C	100	LEU	2.3
1	C	307	HIS	2.3
2	B	93	VAL	2.3
1	C	99	ARG	2.2
1	C	204	SER	2.2
2	B	105	VAL	2.2
1	A	179	GLY	2.2
1	C	299	PHE	2.2
2	D	59	ASN	2.2
2	B	103	ALA	2.2
1	C	84	ASN	2.2
1	A	221	ASP	2.2
2	B	166	GLY	2.1
2	D	4	ALA	2.1
2	D	48	PRO	2.1
2	B	97	ALA	2.0
2	B	138	TYR	2.0
2	D	44	VAL	2.0

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 [i](#)

There are no oligosaccharides 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.