wwPDB X-ray Structure Validation Summary Report (i)

Sep 22, 2016 – 03:57 PM EDT

PDB ID	:	5DUD
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 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

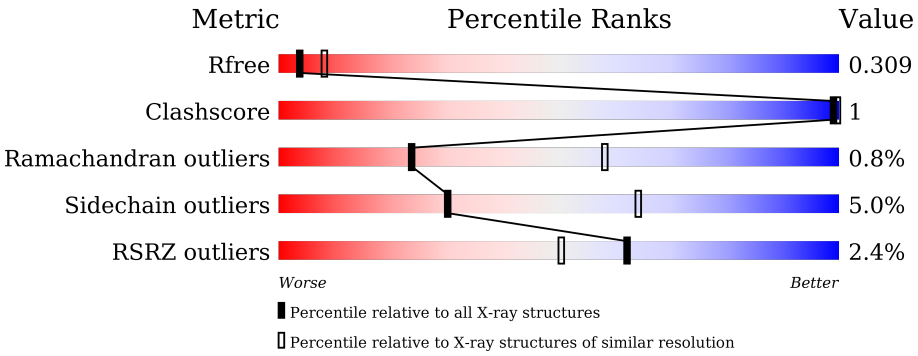
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 2.80 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}	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (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 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 <=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	2% 93% . .
1	C	310	3% 90% 5% 5%
2	B	218	2% 83% 8% 8%
2	D	218	% 85% 5% 10%

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.

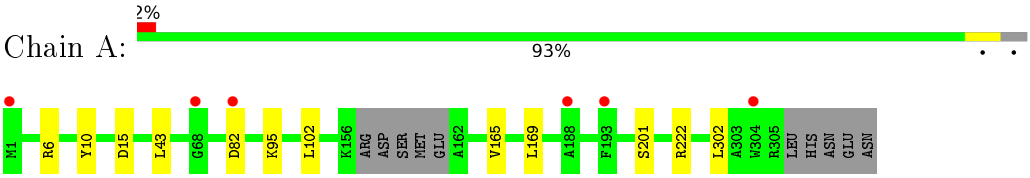
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	67	Total	O	0	0
			67	67		
3	B	29	Total	O	0	0
			29	29		
3	C	39	Total	O	0	0
			39	39		
3	D	31	Total	O	0	0
			31	31		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

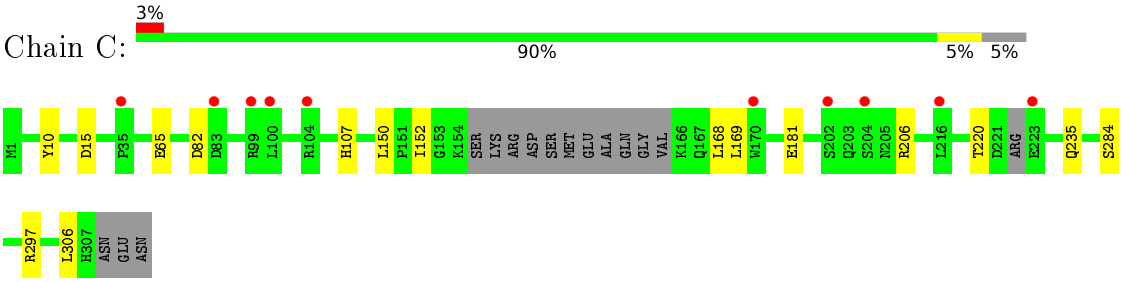
- Molecule 1: YbgK

Chain A:



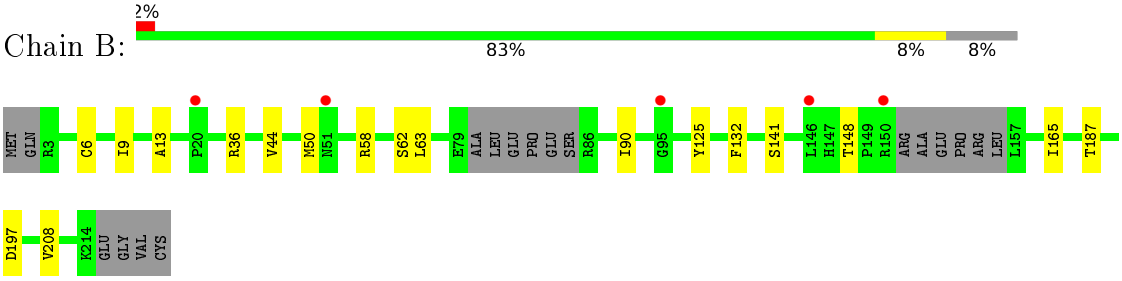
- Molecule 1: YbgK

Chain C:



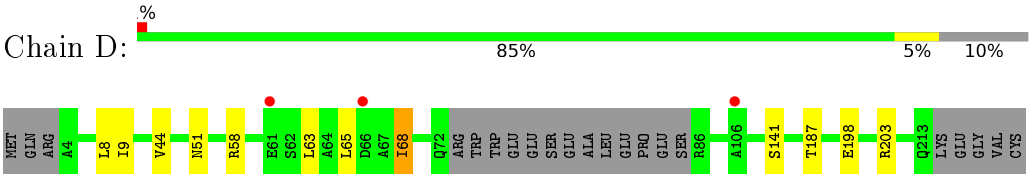
- Molecule 2: YbgJ

Chain B:



- Molecule 2: YbgJ

Chain D:



4 Data and refinement statistics (i)

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 60.20 – 2.80	Depositor EDS
% Data completeness (in resolution range)	98.6 (15.00-2.80) 99.0 (60.20-2.80)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.79 (at 2.81Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.268 , 0.288 0.288 , 0.309	Depositor DCC
R_{free} test set	1194 reflections (5.16%)	DCC
Wilson B-factor (Å^2)	48.3	Xtriage
Anisotropy	0.336	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 56.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.82	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.*

[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

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	0.38	0/2319	0.56	0/3146
1	C	0.37	0/2187	0.55	0/2978
2	B	0.43	0/1485	0.63	0/2042
2	D	0.39	0/1470	0.57	0/2020
All	All	0.39	0/7461	0.57	0/10186

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

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	5	0
2	D	1431	1364	1364	2	0
3	A	67	0	0	0	0
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	10	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.

The worst 5 of 10 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
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

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%)	46	79
1	C	289/310 (93%)	272 (94%)	15 (5%)	2 (1%)	26	62
2	B	194/218 (89%)	181 (93%)	10 (5%)	3 (2%)	13	40
2	D	193/218 (88%)	181 (94%)	10 (5%)	2 (1%)	19	52
All	All	972/1056 (92%)	911 (94%)	53 (6%)	8 (1%)	24	58

5 of 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

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%)	218 (97%)	7 (3%)	47	81
1	C	197/251 (78%)	184 (93%)	13 (7%)	21	51
2	B	139/182 (76%)	131 (94%)	8 (6%)	25	57
2	D	141/182 (78%)	134 (95%)	7 (5%)	30	64
All	All	702/866 (81%)	667 (95%)	35 (5%)	30	64

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	10	TYR
1	C	150	LEU
2	D	187	THR
1	C	15	ASP
1	C	65	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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, 95^{th} 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%)	0.31	6 (2%) 68 58	24, 40, 58, 91	0
1	C	295/310 (95%)	0.41	10 (3%) 49 36	29, 53, 84, 116	0
2	B	200/218 (91%)	0.49	5 (2%) 61 48	33, 51, 74, 104	0
2	D	197/218 (90%)	0.35	3 (1%) 76 68	28, 42, 67, 107	0
All	All	992/1056 (93%)	0.39	24 (2%) 62 50	24, 47, 75, 116	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	202	SER	4.0
2	B	20	PRO	3.6
1	A	188	ALA	3.3
1	C	35	PRO	3.2
1	A	304	TRP	3.1

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

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.