wwPDB X-ray Structure Validation Summary Report (i)

Nov 20, 2023 – 11:30 AM JST

PDB ID	:	7CFY
Title	:	Crystal Structure of YbeA CP74 W48F
Authors	:	Liu, C.Y.; Lai, C.H.; Hsu, S.T.D.; Lyu, P.C.
Deposited on	:	2020-06-29
Resolution	:	2.41 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 (1)) were used in the production of this report:

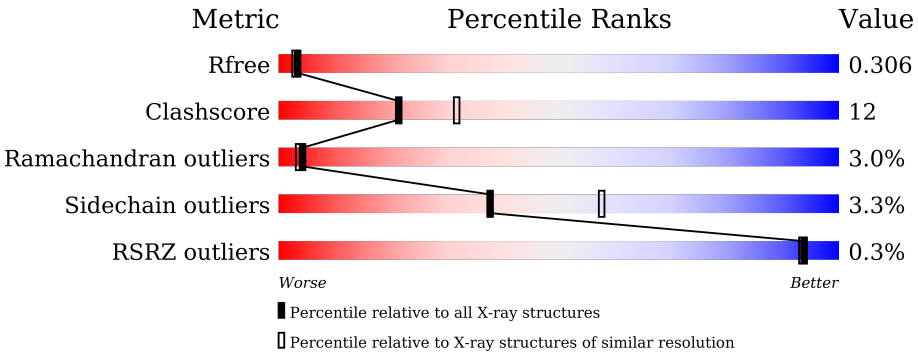
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.41 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	158	61% 29% • 8%
1	B	158	68% 23% • 9%
1	C	158	65% 24% • 8%
1	D	158	% 70% 25% • •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4465 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 Ribosomal RNA large subunit methyltransferase H,Ribosomal RNA large subunit methyltransferase H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	146	Total	C	N	O	S	0	0	0
			1094	699	185	205	5			
1	B	144	Total	C	N	O	S	0	0	0
			1094	700	187	202	5			
1	C	145	Total	C	N	O	S	0	0	0
			1091	699	181	206	5			
1	D	152	Total	C	N	O	S	0	0	0
			1146	730	196	215	5			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P0A8I8
A	48	PHE	TRP	engineered mutation	UNP P0A8I8
A	84	GLY	-	linker	UNP P0A8I8
A	85	SER	-	linker	UNP P0A8I8
B	1	MET	-	initiating methionine	UNP P0A8I8
B	48	PHE	TRP	engineered mutation	UNP P0A8I8
B	84	GLY	-	linker	UNP P0A8I8
B	85	SER	-	linker	UNP P0A8I8
C	1	MET	-	initiating methionine	UNP P0A8I8
C	48	PHE	TRP	engineered mutation	UNP P0A8I8
C	84	GLY	-	linker	UNP P0A8I8
C	85	SER	-	linker	UNP P0A8I8
D	1	MET	-	initiating methionine	UNP P0A8I8
D	48	PHE	TRP	engineered mutation	UNP P0A8I8
D	84	GLY	-	linker	UNP P0A8I8
D	85	SER	-	linker	UNP P0A8I8

- Molecule 2 is water.

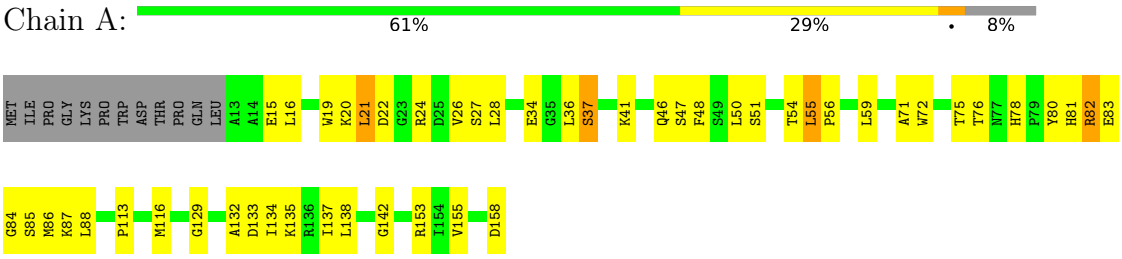
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	7	Total O 7 7	0	0
2	B	12	Total O 12 12	0	0
2	C	11	Total O 11 11	0	0
2	D	10	Total O 10 10	0	0

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.

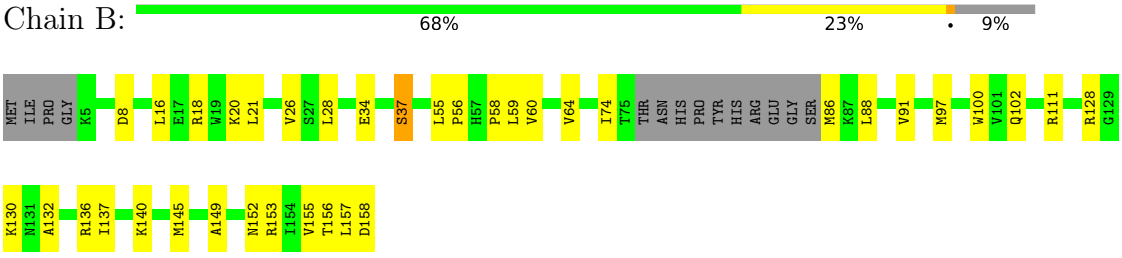
• Molecule 1: Ribosomal RNA large subunit methyltransferase H,Ribosomal RNA large subunit methyltransferase H

Chain A:



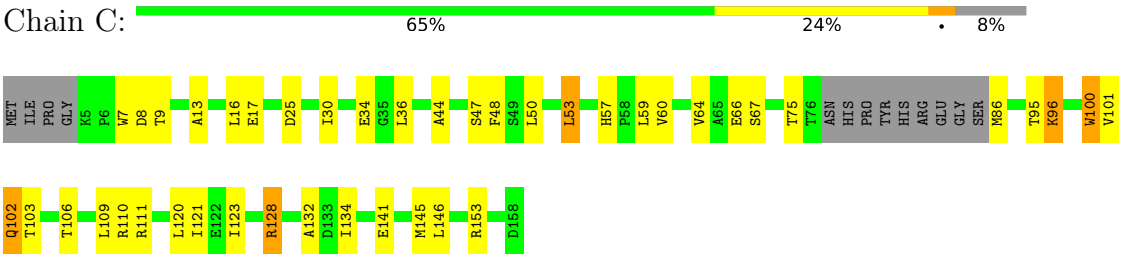
• Molecule 1: Ribosomal RNA large subunit methyltransferase H,Ribosomal RNA large subunit methyltransferase H

Chain B:



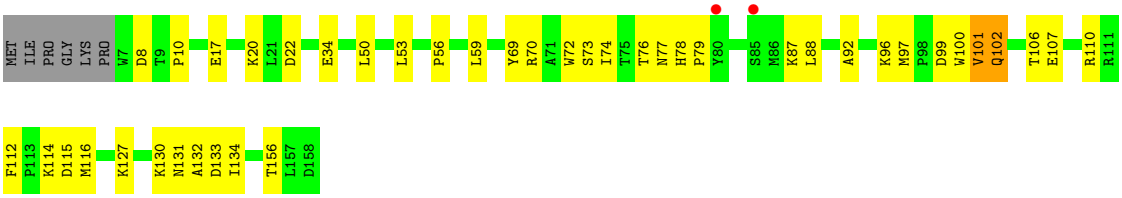
• Molecule 1: Ribosomal RNA large subunit methyltransferase H,Ribosomal RNA large subunit methyltransferase H

Chain C:



• Molecule 1: Ribosomal RNA large subunit methyltransferase H,Ribosomal RNA large subunit methyltransferase H

Chain D:


MET
ILE
PRO
GLY
LYS
PRO
W7
D8
T9
P10
E17
K20
L21
D22
E34
L50
L53
P56
L59
Y69
R70
A71
W72
S73
I74
T75
T76
N77
H78
P79
Y80
S85
M86
K87
L88
A92
K96
M97
P98
D99
W100
V101
Q102
T106
E107
R110
R111
F112
P113
K114
D115
M116
K127
K130
N131
A132
D133
I134
T156
L157
D158

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	39.14Å 57.83Å 69.63Å 67.17° 90.05° 90.15°	Depositor
Resolution (Å)	23.72 – 2.41 23.72 – 2.41	Depositor EDS
% Data completeness (in resolution range)	94.9 (23.72-2.41) 90.1 (23.72-2.41)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.84 (at 2.41Å)	Xtriage
Refinement program	PHENIX 1.16-3549	Depositor
R, R_{free}	0.217 , 0.307 0.217 , 0.306	Depositor DCC
R_{free} test set	2004 reflections (9.67%)	wwPDB-VP
Wilson B-factor (Å^2)	33.3	Xtriage
Anisotropy	0.510	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 17.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.437 for h,-k,-l 0.000 for -h,k,k-l 0.000 for -h,-k,-k+l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4465	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 28.64 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.7795e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	0.46	0/1118	0.63	1/1521 (0.1%)
1	B	0.44	0/1117	0.61	0/1518
1	C	0.46	0/1114	0.66	1/1518 (0.1%)
1	D	0.45	0/1172	0.60	0/1597
All	All	0.45	0/4521	0.63	2/6154 (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(^{o})	Ideal(^{o})
1	C	53	LEU	CA-CB-CG	5.28	127.43	115.30
1	A	21	LEU	CA-CB-CG	5.14	127.12	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	51	SER	Peptide

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	1094	0	1069	45	0
1	B	1094	0	1084	34	0
1	C	1091	0	1069	30	0
1	D	1146	0	1115	30	0
2	A	7	0	0	0	0
2	B	12	0	0	0	0
2	C	11	0	0	0	0
2	D	10	0	0	0	0
All	All	4465	0	4337	105	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 12.

The worst 5 of 105 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:149:ALA:O	1:B:152:ASN:ND2	2.11	0.81
1:D:56:PRO:HD2	1:D:59:LEU:HD12	1.67	0.77
1:A:134:ILE:H	1:A:137:ILE:HD12	1.56	0.70
1:A:155:VAL:HB	1:B:28:LEU:HD23	1.74	0.69
1:D:110:ARG:O	1:D:114:LYS:NZ	2.25	0.68

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	144/158 (91%)	128 (89%)	11 (8%)	5 (4%)	3 3
1	B	140/158 (89%)	132 (94%)	5 (4%)	3 (2%)	7 8

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	C	141/158 (89%)	132 (94%)	6 (4%)	3 (2%)	7 8
1	D	150/158 (95%)	133 (89%)	11 (7%)	6 (4%)	3 2
All	All	575/632 (91%)	525 (91%)	33 (6%)	17 (3%)	4 3

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	82	ARG
1	B	21	LEU
1	A	48	PHE
1	C	102	GLN
1	D	99	ASP

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	110/131 (84%)	104 (94%)	6 (6%)	21 35
1	B	111/131 (85%)	109 (98%)	2 (2%)	59 76
1	C	111/131 (85%)	105 (95%)	6 (5%)	22 36
1	D	116/131 (88%)	115 (99%)	1 (1%)	78 90
All	All	448/524 (86%)	433 (97%)	15 (3%)	38 57

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	111	ARG
1	C	128	ARG
1	C	8	ASP
1	D	115	ASP
1	C	96	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	102	GLN

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 monosaccharides 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	146/158 (92%)	-0.34	0	100	100	24, 39, 65, 76	0
1	B	144/158 (91%)	-0.41	0	100	100	25, 41, 59, 69	0
1	C	145/158 (91%)	-0.49	0	100	100	25, 40, 56, 63	0
1	D	152/158 (96%)	-0.33	2 (1%)	77	75	22, 39, 67, 85	0
All	All	587/632 (92%)	-0.39	2 (0%)	94	93	22, 40, 63, 85	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	80	TYR	3.9
1	D	85	SER	2.6

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.