W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:31 PM GMT

PDB ID	:	4Y66
Title	:	Crystal structure of Giardia lamblia Hop2-Mnd1 complex
Authors	:	Kang, H.A.; Shin, H.C.; Oh, B.H.
Deposited on	:	2015-02-12
Resolution	:	3.20 Å(reported)

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

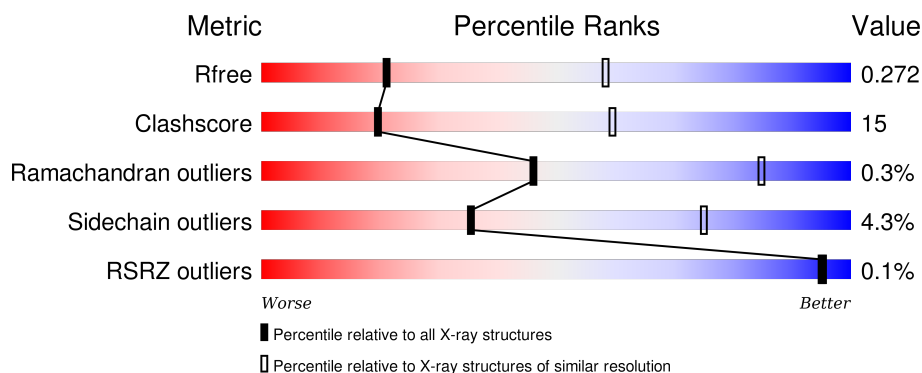|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.20 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	Percentile Ranks	Value
Rfree		0.272
Clashscore		15
Ramachandran outliers		0.3%
Sidechain outliers		4.3%
RSRZ outliers		0.1%

Worse   *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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	203	50% · 13% · 37%
1	C	203	64% · 32% · ·
1	E	203	60% · 12% · 26%
2	B	231	51% · 14% · 34%
2	D	231	60% · 27% · 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	231	55%　　　　　　16%　　·　　28%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7211 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 Mnd1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	128	Total	C	N	O	S	0	0	0
			892	566	155	168	3			
1	C	197	Total	C	N	O	S	0	0	0
			1521	957	259	299	6			
1	E	150	Total	C	N	O	S	0	0	0
			1027	638	179	208	2			

- Molecule 2 is a protein called Putative tbpip family protein.

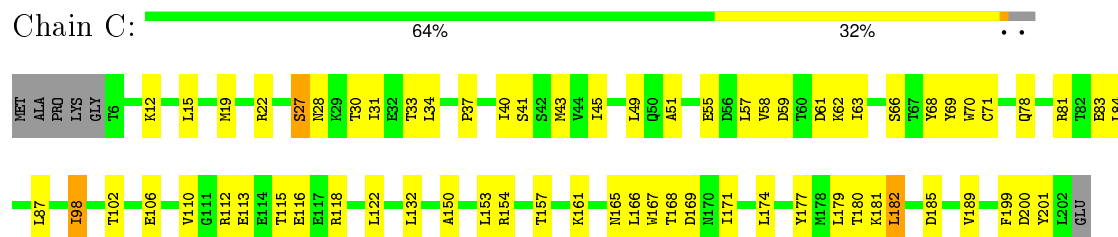| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	152	Total	C	N	O	S	0	0	0
			1062	659	189	207	7			
2	D	208	Total	C	N	O	S	0	0	0
			1547	955	272	310	10			
2	F	167	Total	C	N	O	S	0	0	0
			1162	727	199	230	6			

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 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.
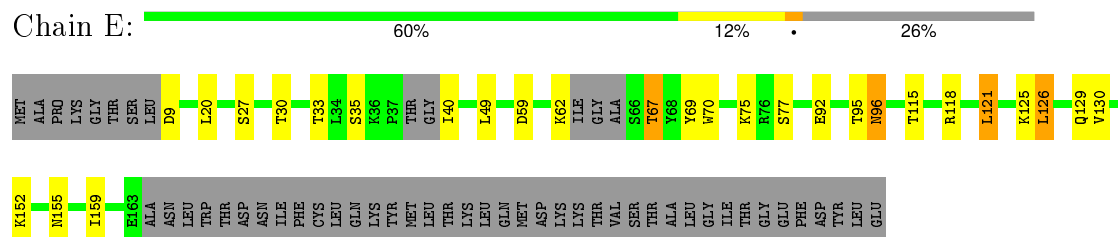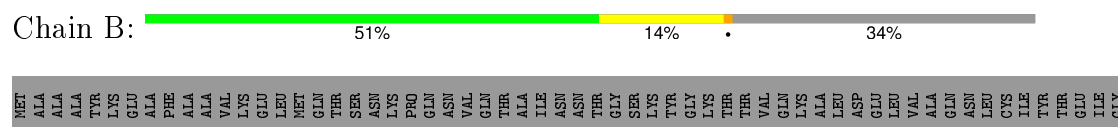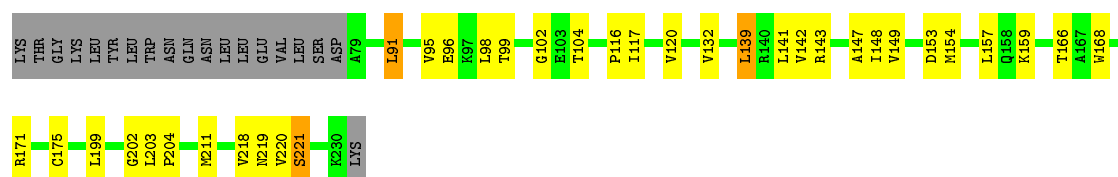
- Molecule 1: Mnd1
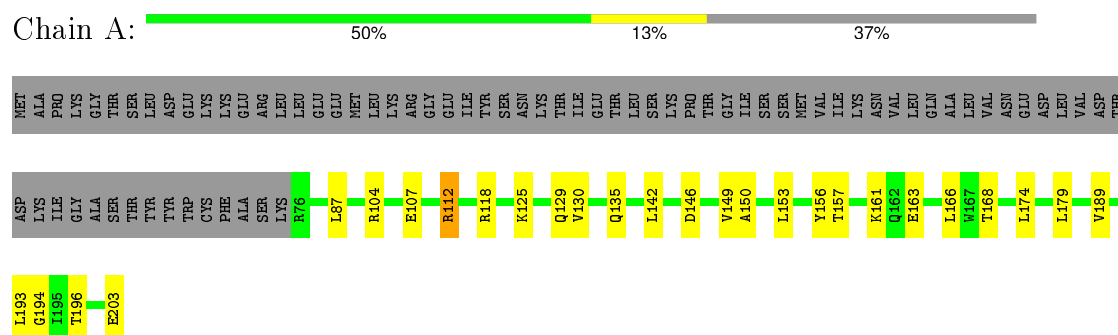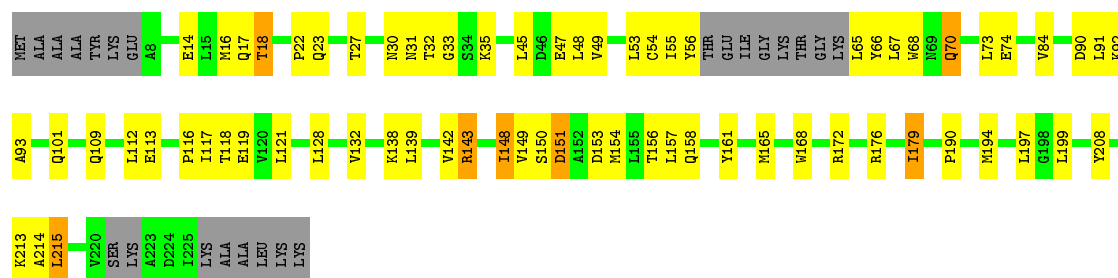

- Molecule 1: Mnd1

- Molecule 1: Mnd1

- Molecule 2: Putative tbpip family protein

LYS THR GLY LYS LEU TYR LEU TRP ASN GLN ASN LEU LEU GLU VAL LEU SER ASP A79 L91 V95 E96 K97 L98 T99 G102 E103 T104 P116 I117 V120 V132 L139 R140 L141 V142 R143 A147 I148 V149 D153 M154 L157 Q158 K159 T166 A167 W168

R171 C175 L199 G202 L203 P204 M211 V218 N219 V220 S221 K230 LYS

• Molecule 2: Putative tbpip family protein

Chain D:

60%   27%   •   10%

MET ALA ALA ALA TYR LYS GLU A8 E14 L15 M16 Q17 T18 P22 Q23 T27 N30 N31 T32 G33 S34 K35 L45 D46 E47 L48 V49 L53 C54 I55 Y56 THR GLU ILE GLY LYS THR GLY LYS L65 Y66 L67 W68 N69 Q70 L73 E74 V84 D90 L91 K92

A93 Q101 Q109 L112 E113 P116 I117 T118 E119 V120 L121 L128 V132 K138 L139 V142 R143 I148 V149 S150 D151 A152 D153 M154 L155 T156 L157 Q158 Y161 M165 W168 R172 R176 I179 P190 M194 L197 G198 L199 Y208

K213 A214 L215 V220 SER A223 D224 I225 LYS ALA ALA LEU LYS LYS

• Molecule 2: Putative tbpip family protein

Chain F:

55%   16%   •   28%

MET ALA A3 V12 K13 E14 L15 T18 S19 P22 Q23 Q26 I29 N30 N31 V41 L45 D46 E47 A50 Q51 E58 I59 G60 K61 T62 G63 K64 L65 Y66 L67 W68 L76 S77 D78 A79 Q80 L81 V95 Q100 E103 T104 L105 T118 L121

V142 ARG GLU SER ASN ALA ILE VAL SER ASP ALA D153 Y161 M165 THR ALA TRP ALA THR ARG ARG ALA LYS CYS ARG GLU VAL ILE ASP THR LEU SER GLU MET GLY LYS VAL LYS PRO SER ALA PHE MET ASP GLN LEU GLY LEU GLU GLU GLY LEU PRO M205 M211

L215 P216 P217 V218 ASN VAL SER LYS ALA ASP ILE LYS ALA ALA LEU LYS LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	117.85Å 69.07Å 292.14Å 90.00° 95.34° 90.00°	Depositor
Resolution (Å)	40.65 – 3.20 45.03 – 3.18	Depositor EDS
% Data completeness (in resolution range)	87.5 (40.65-3.20) 86.8 (45.03-3.18)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.08 (at 3.19Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.9_1692)	Depositor
R, R_{free}	0.239 , 0.286 0.262 , 0.272	Depositor DCC
R_{free} test set	1774 reflections (5.46%)	DCC
Wilson B-factor (Å^2)	70.0	Xtriage
Anisotropy	0.367	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 49.5	EDS
Estimated twinning fraction	0.055 for 1/2*h+3/2*k,1/2*h-1/2*k,-1/2*h-1/2*k-l 0.054 for 1/2*h-3/2*k,-1/2*h-1/2*k,-1/2*h+1/2*k-l	Xtriage
L-test for twinning [2]	$<|L|> = 0.42$, $<L^2> = 0.25$	Xtriage
Outliers	0 of 34750 reflections	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	7211	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.69% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0/900	0.62	0/1228
1	C	0.56	0/1537	0.76	1/2077 (0.0%)
1	E	0.47	0/1034	0.60	0/1409
2	B	0.52	0/1071	0.65	0/1461
2	D	0.56	0/1559	0.72	0/2114
2	F	0.56	1/1171 (0.1%)	0.60	0/1603
All	All	0.54	1/7272 (0.0%)	0.67	1/9892 (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
2	D	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	103	GLU	CG-CD	6.34	1.61	1.51

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	169	ASP	CB-CG-OD2	5.22	123.00	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	31	ASN	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	892	0	804	29	0
1	C	1521	0	1509	65	0
1	E	1027	0	857	22	0
2	B	1062	0	1000	34	0
2	D	1547	0	1499	75	0
2	F	1162	0	1038	35	0
All	All	7211	0	6707	207	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 207 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:87:LEU:HB3	2:B:91:LEU:HD11	1.40	1.03
1:C:112:ARG:HH21	2:D:116:PRO:C	1.69	0.94
2:D:70:GLN:HA	2:D:70:GLN:HE21	1.32	0.92
1:E:92:GLU:O	1:E:96:ASN:ND2	2.08	0.86
2:F:77:SER:CB	2:F:80:GLN:HG2	2.06	0.85

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	126/203 (62%)	121 (96%)	5 (4%)	0	100	100

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	195/203 (96%)	184 (94%)	10 (5%)	1 (0%)	34	78
1	E	144/203 (71%)	139 (96%)	5 (4%)	0	100	100
2	B	150/231 (65%)	145 (97%)	5 (3%)	0	100	100
2	D	202/231 (87%)	189 (94%)	12 (6%)	1 (0%)	34	78
2	F	161/231 (70%)	154 (96%)	6 (4%)	1 (1%)	30	75
All	All	978/1302 (75%)	932 (95%)	43 (4%)	3 (0%)	46	85

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	110	VAL
2	F	217	PRO
2	D	148	ILE

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	74/180 (41%)	73 (99%)	1 (1%)	74	92
1	C	160/180 (89%)	156 (98%)	4 (2%)	55	86
1	E	81/180 (45%)	74 (91%)	7 (9%)	13	46
2	B	98/196 (50%)	92 (94%)	6 (6%)	23	64
2	D	159/196 (81%)	151 (95%)	8 (5%)	30	71
2	F	102/196 (52%)	99 (97%)	3 (3%)	50	83
All	All	674/1128 (60%)	645 (96%)	29 (4%)	35	75

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	70	GLN
2	D	151	ASP
2	F	18	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	112	LEU
2	D	179	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	30	ASN
1	E	96	ASN
2	D	52	ASN
2	B	135	ASN
2	D	70	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)

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(Å²)	Q<0.9
1	A	128/203 (63%)	-0.47	0	100	100	28, 79, 137, 166	0
1	C	197/203 (97%)	-0.39	0	100	100	24, 48, 85, 129	0
1	E	150/203 (73%)	-0.51	0	100	100	31, 89, 152, 191	0
2	B	152/231 (65%)	-0.43	0	100	100	24, 69, 124, 163	0
2	D	208/231 (90%)	-0.33	0	100	100	26, 63, 110, 144	0
2	F	167/231 (72%)	-0.29	1 (0%)	90	84	29, 79, 151, 208	0
All	All	1002/1302 (76%)	-0.40	1 (0%)	95	95	24, 66, 138, 208	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	211	MET	2.3

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.