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 (i)

May 17, 2020 – 05:27 am BST

PDB ID	:	6AAY
Title	:	the Cas13b binary complex
Authors	:	Bo, Z.; Weiwei, Y.; Yangmiao, Y.; Ouyang, S.Y.
Deposited on	:	2018-07-19
Resolution	:	2.79 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

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 2.79 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.
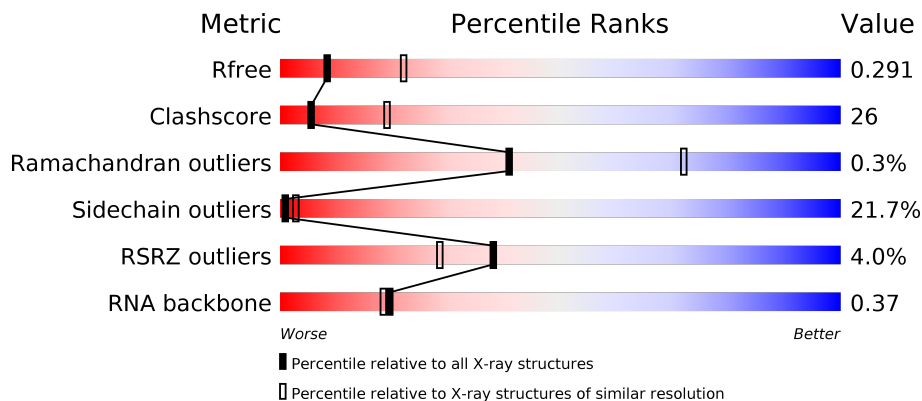
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)
RNA backbone	3102	1227 (3.10-2.50)

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 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	1232	4% / 51% / 35% / 11% / · ·
2	B	59	3% / 34% / 32% / 20% / · / 12%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11262 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 Bergeyella zoohelcum Cas13b (R1177A) mutant.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	1199	10107	6494	1706	1888	7	12	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	initiating methionine	UNP K1LVU1
A	1177	ALA	ARG	engineered mutation	UNP K1LVU1
A	1225	LEU	-	expression tag	UNP K1LVU1
A	1226	GLU	-	expression tag	UNP K1LVU1
A	1227	HIS	-	expression tag	UNP K1LVU1
A	1228	HIS	-	expression tag	UNP K1LVU1
A	1229	HIS	-	expression tag	UNP K1LVU1
A	1230	HIS	-	expression tag	UNP K1LVU1
A	1231	HIS	-	expression tag	UNP K1LVU1
A	1232	HIS	-	expression tag	UNP K1LVU1

- Molecule 2 is a RNA chain called RNA (52-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	52	1118	501	210	355	52	0	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	36	36	36	0	0
3	B	1	1	1	0	0

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 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: Bergeyella zoohelcum Cas13b (R1177A) mutant

S848 E849 N850 F853 E857 N858 E859 C860 F861 K862 K866 Q867 T870 H871 K872 E873 L874 V878 I881 I886 F887 L888 E889 F892 M893 T898 I899 I900 K901 G902 F905 K906 G907 N908 E909 A910 L911 R917 K920 E921 Y922 Q923 N924 F928 Y929 D930 T931

E932 N933 L936 E940 K941 K942 Q943 A944 D945 R946 K947 R948 K951 I952 Q955 K956 T962 L963 L964 M965 A966 K967 H968 I969 F974 K975 Q976 D977 S978 I979 D980 Q981 E985 D986 L987 Y988 Q989 E992 E993 R994 L995 G996 N997 R1000 A1001 R1002 Q1003 T1004 G1005 E1006 R1007

N1008 T1009 N1010 Y1011 I1012 W1013 N1014 K1015 T1016 V1017 D1018 L1019 K1020 D1023 G1024 K1025 I1026 T1027 V1028 E1029 K1032 L1033 K1034 N1035 V1036 G1037 D1038 F1039 I1040 K1041 D1045 Q1046 R1047 V1048 Q1049 A1050 F1051 L1052 K1053 Y1054 E1055 E1056 N1057 I1058 E1059 W1060 Q1061 L1064 I1065 LYS GLU SER LYS E1070 E1071 R1072

Y1074 P1075 Y1076 V1077 V1078 E1079 R1080 E1081 I1082 E1083 Q1084 Y1085 E1086 K1087 V1088 R1089 R1090 E1091 E1092 L1093 L1094 K1095 E1096 Y1103 I1104 L1105 K1106 K1107 V1108 K1109 D1110 K1111 E1112 I1113 L1114 D1118 N1119 K1123 N1128 G1129 L1130 Q1133 L1134 K1135 N1136 E1137 D1138 V1139 E1140 S1141 Y1142 K1143 V1144 L1147

E1152 D1153 V1154 N1155 I1156 N1157 Q1158 L1159 K1160 Q1161 E1162 A1163 T1164 D1165 L1166 E1167 Q1168 K1169 L1173 T1174 Y1175 K1179 Q1184 L1185 P1186 K1187 K1188 E1189 Y1193 C1194 Q1195 E1196 K1197 Y1198 K1205 T1206 Y1207 A1208 E1209 K1217 E1218 L1222 I1223 K1224 L1225 GLU HIS HIS HIS HIS HIS

- Molecule 2: RNA (52-MER)

Chain B:

3%

34%   32%   20%   12%

A1 A2 A3 A4 A5 G6 G7 G U U U A A A A15 A16 A17 U18 G19 A20 A21 A22 G23 U24 U25 G26 G27 C30 U31 G32 C33 A38 U39 U40 U41 U42 G43 G44 G47 G48 U49 A50 A51 U52 C53 A54 C55 A59

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	196.74Å 196.74Å 96.23Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	65.02 – 2.79 64.93 – 2.79	Depositor EDS
% Data completeness (in resolution range)	99.9 (65.02-2.79) 99.9 (64.93-2.79)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.82 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.8.0222	Depositor
R, R_{free}	0.221 , 0.276 0.232 , 0.291	Depositor DCC
R_{free} test set	2390 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	72.3	Xtriage
Anisotropy	0.096	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 52.4	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	11262	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.84% 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 (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	0.66	0/10294	0.77	6/13778 (0.0%)
2	B	0.84	7/1253 (0.6%)	0.81	0/1949
All	All	0.68	7/11547 (0.1%)	0.77	6/15727 (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	23

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	22	A	O3'-P	-7.36	1.52	1.61
2	B	23	G	O3'-P	-7.17	1.52	1.61
2	B	47	G	O3'-P	-6.09	1.53	1.61
2	B	24	U	O3'-P	-5.94	1.54	1.61
2	B	25	U	O3'-P	-5.80	1.54	1.61

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	15	PRO	N-CA-CB	-11.34	89.69	103.30
1	A	499	PRO	CA-N-CD	-11.29	95.70	111.50
1	A	498	PHE	CB-CA-C	7.75	125.91	110.40
1	A	15	PRO	N-CD-CG	-7.01	92.68	103.20
1	A	16	PHE	CB-CA-C	6.22	122.83	110.40

There are no chirality outliers.

5 of 23 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	186	ARG	Sidechain
1	A	193	ARG	Sidechain
1	A	196	ARG	Sidechain
1	A	45	ARG	Sidechain
1	A	85	ARG	Sidechain

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	10107	0	10188	556	0
2	B	1118	0	562	36	0
3	A	36	0	0	0	0
3	B	1	0	0	0	0
All	All	11262	0	10750	567	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 26.

The worst 5 of 567 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:947:LYS:HE3	1:A:951:LYS:CE	1.52	1.39
1:A:15:PRO:HA	1:A:24:PHE:CZ	1.70	1.26
1:A:1144:VAL:CG2	1:A:1163:ALA:HB2	1.65	1.25
1:A:979:ILE:CD1	1:A:1015:LYS:HD3	1.64	1.25
1:A:1084:GLN:O	1:A:1088:VAL:HG23	1.37	1.23

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	1191/1232 (97%)	1140 (96%)	48 (4%)	3 (0%)	41 72

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	15	PRO
1	A	499	PRO
1	A	1135	LYS

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	1104/1119 (99%)	864 (78%)	240 (22%)	1 3

5 of 240 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	570	GLN
1	A	769	ASP
1	A	1139	VAL
1	A	620	LYS
1	A	660	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	650	GLN
1	A	666	GLN
1	A	1057	ASN
1	A	544	GLN
1	A	997	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	50/59 (84%)	13 (26%)	3 (6%)

5 of 13 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	2	A
2	B	7	G
2	B	16	A
2	B	19	G
2	B	21	A

All (3) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	16	A
2	B	21	A
2	B	50	A

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	1187/1232 (96%)	0.44	48 (4%)	38	28	30, 68, 110, 163	0
2	B	52/59 (88%)	0.14	2 (3%)	40	30	42, 59, 139, 148	0
All	All	1239/1291 (95%)	0.43	50 (4%)	38	28	30, 68, 113, 163	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	402	GLU	5.7
1	A	1071	GLU	5.5
1	A	198	ARG	5.2
1	A	669	ASN	4.8
1	A	181	LEU	4.4

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.