



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 30, 2017 – 07:08 PM EST

PDB ID : 5WTJ
Title : Crystal structure of an endonuclease
Authors : Liu, L.; Wang, Y.
Deposited on : 2016-12-13
Resolution : 3.50 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 ⓘ symbol.

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

MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
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-20028442

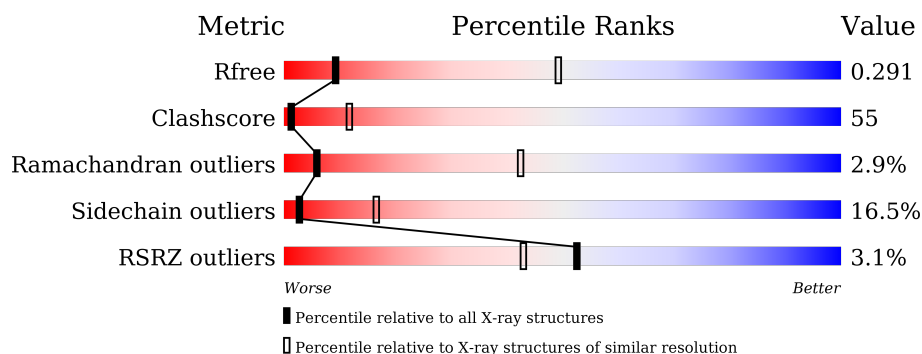
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.50 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	1051 (3.60-3.40)
Clashscore	102246	1157 (3.60-3.40)
Ramachandran outliers	100387	1120 (3.60-3.40)
Sidechain outliers	100360	1121 (3.60-3.40)
RSRZ outliers	91569	1058 (3.60-3.40)

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 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 $\leq 5\%$. 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	1397	<div> <div> <div></div> <div>30%</div> <div>33%</div> <div>9%</div> <div>•</div> <div>27%</div> </div> </div>
1	B	1397	<div> <div> <div>3%</div> <div>32%</div> <div>31%</div> <div>9%</div> <div>•</div> <div>28%</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 15952 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 CRISPR-associated endoribonuclease C2c2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1019	Total	C	N	O	S	Se	0	0	0
			8013	5109	1350	1538	5	11			
1	B	1012	Total	C	N	O	S	Se	0	0	0
			7938	5055	1326	1541	6	10			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1390	LEU	-	expression tag	UNP P0DOC6
A	1391	GLU	-	expression tag	UNP P0DOC6
A	1392	HIS	-	expression tag	UNP P0DOC6
A	1393	HIS	-	expression tag	UNP P0DOC6
A	1394	HIS	-	expression tag	UNP P0DOC6
A	1395	HIS	-	expression tag	UNP P0DOC6
A	1396	HIS	-	expression tag	UNP P0DOC6
A	1397	HIS	-	expression tag	UNP P0DOC6
B	1390	LEU	-	expression tag	UNP P0DOC6
B	1391	GLU	-	expression tag	UNP P0DOC6
B	1392	HIS	-	expression tag	UNP P0DOC6
B	1393	HIS	-	expression tag	UNP P0DOC6
B	1394	HIS	-	expression tag	UNP P0DOC6
B	1395	HIS	-	expression tag	UNP P0DOC6
B	1396	HIS	-	expression tag	UNP P0DOC6
B	1397	HIS	-	expression tag	UNP P0DOC6

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	O	0	0
			1	1		



H1345	K1242	Q1145	F1072	V995	Q920	H847	K776	K711	D841	F574	H501	I431
M1346	F1243	M1146	I1075	I999	I921	S848	V777	E712	V642	F574	T502	Y432
D1347	F1244	K1147	Y1076	I999	E824	H849	I778	L713	D503	D576	D503	R433
I1348		M1148	Y1077	D1003	I925	H850		Y714	F644	I575	D504	Y434
L1349	Y1249	Y1149	G1077	I1006	Y829	H851	I782	K715	K645	K577	F505	L435
E1350	K1250		L1078		Y830	H852	G783	K716	D646	K578	S506	K436
H1351	K1251	F1152	I1079		Y831	H853	Y784	L717	K647	N579	E507	G437
L1352	F1252				Y832	H854		I718	K648	L508	E508	R438
M1353	E1253	D1155	I1083	L1011	E931	H855	M788	L719	M649	T582	H509	I439
K1354	K1254	Y1156	K1084		E931	H856		E720	I650	N583	K511	I440
P1355	I1255			I1015	K934		L792	L723	I651	T582	E512	I442
K1356	C1256	V1159	K1089		K935		F793	E724	I654	I586	E513	K443
K1357		S1160	F1090	L1021	N936		D794	E725	I655	K588	E514	VAL
V1358	F1259	E1161	L1091	K1022	I937	L866	F795	N726	D856	F589	D515	ARG
S1359	G1260	K1162	F1092	K1023	I938		S796	E727	D857	I657	L516	LEU
V1360	I1261	M1163	M1093		T939	S869	D797	K728	K658	K591	L519	LVS
L1361	D1262	K1164	I1094	K1026	E940		PHE	K729	I659	I592	T520	NSE
E1362	L1263	I1165	D1095	E1027	F941	H873	LVS			G593	F521	GLU
L1363	S1264	R1166	G1096	I1028	K942	H874	NSE		E862	T594		LVS
E1364	R1265			D1029	D943	H875	ASN	F732	M663	N595		
S1365	M1266	E1170		M1030	D944	H876	ILE	L733	M664	I600	S524	E457
Y1366	S1267	F1171		I1032	N946	H877	Q803	E735	N665	L601		E458
M1367	E1268			I1033			E805	L736	D666	L602	M527	
S1368	I1269	L1174	K1103	E1033	D949	E880	K806	K737	I667	E528		K461
	M1270	M1175	I1104	D1034	V950	H881	K807	I738	I667			I462
I1371	K1271	K1176	S1105	M1035		H882	Q808	I739	Y669	K531		L463
	P1272	I1177	E1106	E1036	K953	H883	Q809	K740	K606	T532		M464
L1374	N1273	E1178		S1037	L884	H884	K810	G741	E607	F533		E465
	N1274			E1038	K954	H885	D811	N742	R608	S534		S466
				M1039		H886	H812	I743	D609	R535		I467
L1378		M1185	L1111	K1039		H887	I812	E744	L610	E536		L468
				E1040	VAL		K816	D745	Q611		N537	
K1381	I1277	L1188	L1113	I1041	ILE	H889	T817	E746	G612			L472
I1382	R1278		L1114	K1042	PHE	E890	Y818	D747	T613	N539		L473
E1383		Q1191	M1115	E1043	ASP	C891		E748	Q614	N540		
ASN	M1289	M1192	K1117	Q1044	ASP	H892			D615			V476
THR	F1291	A1193	L1118	E1045	GLU	T893	I821	N749	D616			K477
ASN	R1292	R1194	L1119	I1046	THR		T822	I750	D617			Q478
ASP	D1293	F1195	G1120	Y1047	LVS		V823	I751	Y617			Q479
THR	Y1294	E1196	Y1121	Y1048	PHE	H897	K824	E752	N618			T480
LEU		R1197	S1122	P1049		L898	T825	N753	K619			R551
GLU	I1300	D1301	K1123	E1050	E968	H899	S826	Y754	R684			L481
HIS	R1302	M1199	E1124	E1051	I969	L900	D827	V755	R685			E482
HIS	V1303	H1200	E1124	K1052	D970	E901	K828	K756	M686			E483
HIS		Y1201	E1127	K1053		E902	T829	I757	N690			I484
HIS		I1202	K1128	N1054	I975	E903	I830	Y758	D694			M485
HIS	L1306	V1203	Y1129	E1055	L976	F903	V831	Q759	T695			Y486
HIS	L1307	N1204	I1130	L1056	Q977	I904	I832	I760	L627			L487
	S1308			Y1057			I833	I761	L628			G488
	Y1309	R1207	L1133	I1058	Q980	H907	D834	S761	K628			K489
		E1208	K1134	Y1059	R981	K908	D835	N766	S630			L490
			E1135	K1060	K982		D836	N767	D631			R491
	V1321	I1212	M1136	K1061	K983	E911	F836	K768	K700			H492
	F1322		D1137	N1062	S984	K912	E837	A769	I701			M493
		M1218	D1138	L1063	I986	F914	Y838	I770	S635			M494
	V1329	T1219	F1139	I1067	N987	D915	I842	K771	K636			I495
	D1332			G1068		D916	F843	K772	A637			D496
		R1223	A1141	M1069	D990	F917	A844	K773	L638			
	K1341	A1224	M1143	P1070	L991	K918	L845	Q774	I569			T499
	L1342		I1144	N1071		I919	L846	K775	I570			V500

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	91.66Å 94.23Å 338.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.38 – 3.50 48.37 – 3.50	Depositor EDS
% Data completeness (in resolution range)	81.0 (48.38-3.50) 75.3 (48.37-3.50)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.71 (at 3.48Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.263 , 0.291 0.262 , 0.291	Depositor DCC
R_{free} test set	1547 reflections (5.05%)	DCC
Wilson B-factor (Å ²)	18.6	Xtriage
Anisotropy	0.648	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , -0.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	0.053 for k,h,-l	Xtriage
F_o, F_c correlation	0.78	EDS
Total number of atoms	15952	wwPDB-VP
Average B, all atoms (Å ²)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.41% 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 [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	0.48	2/8117 (0.0%)	0.67	16/10925 (0.1%)
1	B	0.42	2/8038 (0.0%)	0.66	11/10824 (0.1%)
All	All	0.45	4/16155 (0.0%)	0.66	27/21749 (0.1%)

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	2
1	B	0	7
All	All	0	9

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	671	PRO	N-CD	5.24	1.55	1.47
1	B	1038	GLU	CB-CG	-5.17	1.42	1.52
1	A	1272	PRO	N-CD	5.09	1.54	1.47
1	B	671	PRO	N-CD	5.09	1.54	1.47

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	634	VAL	CB-CA-C	-11.15	90.21	111.40
1	A	997	GLN	CB-CA-C	6.94	124.28	110.40
1	B	1021	LEU	CB-CG-CD1	-6.59	99.80	111.00
1	B	403	ILE	CB-CA-C	-6.57	98.46	111.60
1	A	758	ALA	CB-CA-C	-6.33	100.60	110.10

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1051	GLU	Peptide
1	A	1356	LYS	Peptide
1	B	412	PHE	Peptide
1	B	605	SER	Peptide
1	B	644	PHE	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	8013	0	7662	834	3
1	B	7938	0	7580	876	1
2	A	1	0	0	0	0
All	All	15952	0	15242	1710	4

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 55.

The worst 5 of 1710 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:976:LEU:HD23	1:B:977:GLN:N	1.22	1.45
1:A:696:ILE:CB	1:A:699:GLU:HB3	1.55	1.35
1:B:817:THR:CG2	1:B:818:TYR:HA	1.54	1.35
1:B:914:PHE:HA	1:B:915:ASP:CB	1.47	1.34
1:B:914:PHE:CA	1:B:915:ASP:HB2	1.54	1.32

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:426:GLU:OE1	1:B:1037:SER:OG[4_446]	1.51	0.69
1:A:776:LYS:CB	1:A:1348:ILE:CD1[3_655]	1.91	0.29
1:A:398:THR:CB	1:A:907:MSE:CE[3_655]	2.08	0.12

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:776:LYS:CD	1:A:1348:ILE:CD1[3_655]	2.10	0.10

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	1013/1397 (72%)	908 (90%)	79 (8%)	26 (3%)	7	45
1	B	1004/1397 (72%)	888 (88%)	84 (8%)	32 (3%)	5	40
All	All	2017/2794 (72%)	1796 (89%)	163 (8%)	58 (3%)	6	42

5 of 58 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	392	LYS
1	A	410	VAL
1	A	458	GLU
1	A	460	GLU
1	A	461	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	824/1316 (63%)	681 (83%)	143 (17%)	2	14
1	B	822/1316 (62%)	693 (84%)	129 (16%)	3	19

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1646/2632 (62%)	1374 (84%)	272 (16%)	3 16

5 of 272 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1185	ASN
1	B	438	ARG
1	B	1084	LYS
1	A	1244	PHE
1	A	1368	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	464	ASN
1	B	501	ASN
1	B	1175	ASN
1	B	492	HIS
1	B	539	ASN

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 [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(Å ²)	Q<0.9
1	A	1008/1397 (72%)	-0.36	14 (1%)	78 68	2, 42, 106, 153	0
1	B	1002/1397 (71%)	0.08	48 (4%)	34 27	12, 67, 134, 199	0
All	All	2010/2794 (71%)	-0.14	62 (3%)	52 43	2, 55, 128, 199	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	726	ASN	8.0
1	B	744	ASP	6.9
1	B	728	SER	6.7
1	B	743	ILE	6.7
1	B	727	GLU	6.0

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.