



wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 02:59 am BST

PDB ID : 5J70
Title : The Chd1 DNA-binding domain in complex with 17mer DNA duplex
Authors : Bowman, G.D.; Jenkins, K.R.; Hauk, G.
Deposited on : 2016-04-05
Resolution : 2.96 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 ⓘ symbol.

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

MolProbity	:	4.02b-467
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

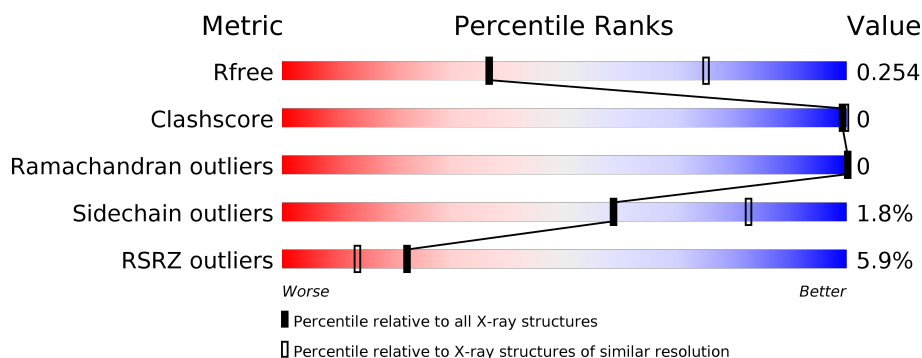
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.96 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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 $\leq 5\%$. 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	271	<div> <div>4%</div> <div> <div></div> <div>83%</div> <div>15%</div> </div> </div>
1	B	271	<div> <div>8%</div> <div> <div></div> <div>83%</div> <div>16%</div> </div> </div>
2	W	17	<div> <div></div> <div> <div></div> <div>88%</div> <div>12%</div> </div> </div>
2	Y	17	<div> <div></div> <div> <div></div> <div>100%</div> </div> </div>
3	X	17	<div> <div></div> <div> <div></div> <div>88%</div> <div>12%</div> </div> </div>
3	Z	17	<div> <div></div> <div> <div></div> <div>100%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9555 atoms, of which 4500 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 Chromo domain-containing protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	229	Total	C	H	N	O	S	0	0	0
			3741	1185	1889	315	349	3			
1	B	228	Total	C	H	N	O	S	0	0	0
			3660	1165	1839	311	342	3			

There are 4 discrepancies between the modelled and reference sequences:


Chain	Residue	Modelled	Actual	Comment	Reference
A	1004	GLY	-	expression tag	UNP P32657
A	1005	PRO	-	expression tag	UNP P32657
B	1004	GLY	-	expression tag	UNP P32657
B	1005	PRO	-	expression tag	UNP P32657

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*CP*TP*GP*GP*AP*AP*AP*TP*TP*TP*CP*CP*AP*GP*CP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	W	17	Total	C	H	N	O	P	0	0	0
			540	166	193	65	100	16			
2	Y	17	Total	C	H	N	O	P	0	0	0
			540	166	193	65	100	16			

- Molecule 3 is a DNA chain called DNA (5'-D(*CP*GP*CP*TP*GP*GP*AP*AP*AP*TP*TP*TP*CP*CP*AP*GP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	X	17	Total	C	H	N	O	P	0	0	0
			537	165	193	63	100	16			
3	Z	17	Total	C	H	N	O	P	0	0	0
			537	165	193	63	100	16			

Chain X:  88% 12%



- Molecule 3: DNA (5'-D(*CP*GP*CP*TP*GP*GP*AP*AP*AP*TP*TP*TP*CP*CP*AP*GP*C)-3')

Chain Z:  100%

There are no outlier residues recorded for this chain.

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	86.08Å 57.10Å 99.49Å 90.00° 103.90° 90.00°	Depositor
Resolution (Å)	46.98 – 2.96 49.15 – 2.96	Depositor EDS
% Data completeness (in resolution range)	98.1 (46.98-2.96) 87.2 (49.15-2.96)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.53 (at 2.96Å)	Xtriage
Refinement program	PHENIX (1.10.1 _2155: ???)	Depositor
R, R_{free}	0.231 , 0.252 0.232 , 0.254	Depositor DCC
R_{free} test set	1036 reflections (5.20%)	wwPDB-VP
Wilson B-factor (Å ²)	66.3	Xtriage
Anisotropy	0.796	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 44.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	9555	wwPDB-VP
Average B, all atoms (Å ²)	132.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.57% 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	0.26	0/1885	0.39	0/2528
1	B	0.26	0/1853	0.41	0/2489
2	W	0.71	0/389	0.93	0/599
2	Y	0.66	0/389	0.92	0/599
3	X	0.63	0/385	0.90	0/592
3	Z	0.60	0/385	0.88	0/592
All	All	0.41	0/5286	0.61	0/7399

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 [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	1852	1889	1889	0	0
1	B	1821	1839	1840	0	1
2	W	347	193	193	1	0
2	Y	347	193	193	0	0
3	X	344	193	193	1	1
3	Z	344	193	193	0	0
All	All	5055	4500	4501	2	1

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 0.

All (2) 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)
3:X:1:DC:H2'	3:X:2:DG:C8	2.48	0.49
2:W:1:DG:H2''	2:W:2:DC:H5''	1.96	0.47

All (1) 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:1089:LEU:O	3:X:1:DC:O5'[2_655]	2.16	0.04

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	225/271 (83%)	214 (95%)	11 (5%)	0	100	100
1	B	224/271 (83%)	212 (95%)	12 (5%)	0	100	100
All	All	449/542 (83%)	426 (95%)	23 (5%)	0	100	100

There are no Ramachandran outliers to report.

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	202/238 (85%)	198 (98%)	4 (2%)	55	80
1	B	195/238 (82%)	192 (98%)	3 (2%)	65	85
All	All	397/476 (83%)	390 (98%)	7 (2%)	59	82

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1151	TYR
1	B	1131	GLU
1	B	1028	LEU
1	A	1070	ASN
1	B	1070	ASN

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

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 ⓘ

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, 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	229/271 (84%)	0.42	10 (4%) 34 21	88, 116, 148, 164	0
1	B	228/271 (84%)	0.49	21 (9%) 9 5	99, 128, 149, 162	0
2	W	17/17 (100%)	0.03	0 100 100	77, 96, 155, 161	0
2	Y	17/17 (100%)	-0.42	0 100 100	93, 120, 162, 166	0
3	X	17/17 (100%)	0.02	0 100 100	81, 99, 137, 147	0
3	Z	17/17 (100%)	-0.45	0 100 100	119, 126, 143, 150	0
All	All	525/610 (86%)	0.37	31 (5%) 22 13	77, 122, 149, 166	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1098	ASN	10.7
1	B	1210	ILE	4.7
1	B	1246	LYS	4.1
1	B	1253	LEU	3.9
1	B	1211	PHE	3.4

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.