Full wwPDB X-ray Structure Validation Report (i)

May 29, 2020 – 08:05 am BST

PDB ID	:	5T0U
Title	:	CTCF ZnF2-7 and DNA complex structure
Authors	:	Hashimoto, H.; Wang, D.; Cheng, X.
Deposited on	:	2016-08-16
Resolution	:	3.20 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references (1)) were used in the production of this report:

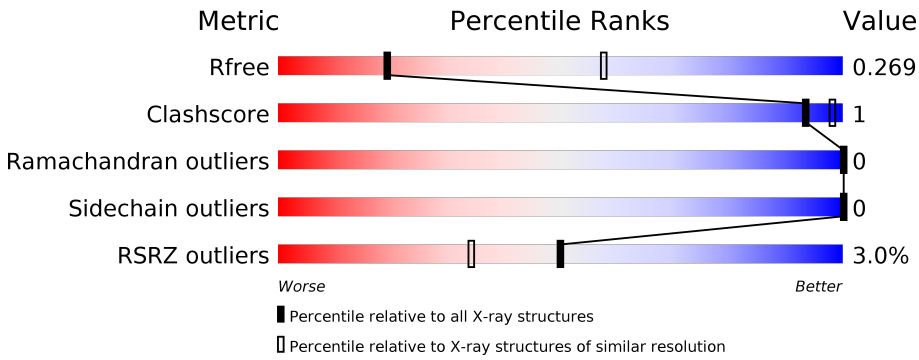
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	177	90% 6% •
1	D	177	7% 93% • 5%
2	B	23	91% 9%
2	E	23	100%
3	C	23	87% 13%
3	F	23	96% •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8388 atoms, of which 3725 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 Transcriptional repressor CTCF.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	170	2736	857	1350	275	237	17	0	0	0
1	D	169	2722	853	1343	274	235	17	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	289	PRO	-	expression tag	UNP P49711
A	290	LEU	-	expression tag	UNP P49711
A	291	GLY	-	expression tag	UNP P49711
A	292	SER	-	expression tag	UNP P49711
A	293	THR	-	expression tag	UNP P49711
D	289	PRO	-	expression tag	UNP P49711
D	290	LEU	-	expression tag	UNP P49711
D	291	GLY	-	expression tag	UNP P49711
D	292	SER	-	expression tag	UNP P49711
D	293	THR	-	expression tag	UNP P49711

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	B	23	716	218	259	79	138	22	0	0	0
2	E	23	716	218	259	79	138	22	0	0	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
3	C	23	737	225	257	99	134	22	0	0	0
3	F	23	737	225	257	99	134	22	0	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
4	A	6	6	6	0	0
4	D	6	6	6	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	8	8	8	0	0
5	B	1	1	1	0	0
5	C	1	1	1	0	0
5	D	2	2	2	0	0

3 Residue-property plots (i)

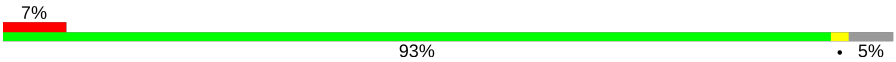
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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.

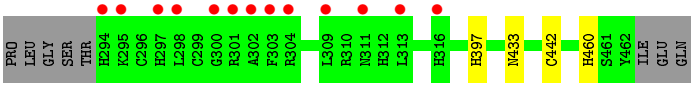
• Molecule 1: Transcriptional repressor CTCF

Chain A:

• Molecule 1: Transcriptional repressor CTCF

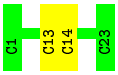
Chain D:





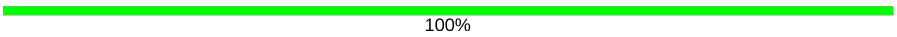
• Molecule 2: DNA (5'-D(*CP*CP*TP*CP*AP*CP*TP*AP*GP*CP*GP*CP*CP*CP*CP*CP*TP*GP*CP*TP*GP*GP*C)-3')

Chain B:



• Molecule 2: DNA (5'-D(*CP*CP*TP*CP*AP*CP*TP*AP*GP*CP*GP*CP*CP*CP*CP*CP*TP*GP*CP*TP*GP*GP*C)-3')

Chain E:



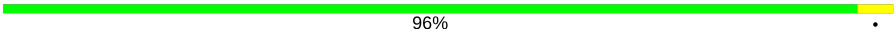
There are no outlier residues recorded for this chain.

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

Chain C:

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

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	101.63Å 40.96Å 106.67Å 90.00° 92.21° 90.00°	Depositor
Resolution (Å)	53.30 – 3.20 74.98 – 3.20	Depositor EDS
% Data completeness (in resolution range)	97.5 (53.30-3.20) 89.1 (74.98-3.20)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.22 (at 3.19Å)	Xtriage
Refinement program	PHENIX (1.11_2558: ???)	Depositor
R, R_{free}	0.230 , 0.266 0.231 , 0.269	Depositor DCC
R_{free} test set	747 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	41.8	Xtriage
Anisotropy	0.313	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 15.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.025 for l,k,-h 0.040 for h,-k,-l 0.075 for l,-k,h	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	8388	wwPDB-VP
Average B, all atoms (Å^2)	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.71% 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)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.23	0/1423	0.41	0/1910
1	D	0.24	0/1416	0.40	0/1900
2	B	0.50	0/509	0.90	0/781
2	E	0.51	0/509	0.88	0/781
3	C	0.49	0/541	0.79	0/836
3	F	0.52	0/541	0.81	0/836
All	All	0.37	0/4939	0.65	0/7044

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	1386	1350	1350	8	0
1	D	1379	1343	1343	3	0
2	B	457	259	259	1	0
2	E	457	259	259	0	0
3	C	480	257	257	3	0
3	F	480	257	257	1	0
4	A	6	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	6	0	0	0	0
5	A	8	0	0	0	0
5	B	1	0	0	0	0
5	C	1	0	0	0	0
5	D	2	0	0	0	0
All	All	4663	3725	3725	12	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 1.

All (12) 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:293:THR:OG1	1:A:303:PHE:O	2.07	0.72
1:A:339:ARG:NH2	3:C:14:DG:N7	2.42	0.68
1:A:339:ARG:NH1	1:A:362:GLU:OE2	2.28	0.66
1:A:389:ARG:NH1	3:C:9:DG:OP2	2.33	0.61
1:D:442:CYS:SG	1:D:460:HIS:CE1	2.94	0.60
1:A:456:LEU:O	1:A:462:TYR:N	2.43	0.51
1:A:433:ASN:O	1:A:433:ASN:ND2	2.43	0.51
1:A:371:ARG:O	1:A:375:GLY:N	2.42	0.45
1:D:397:HIS:ND1	3:F:7:DA:OP1	2.41	0.45
2:B:13:DC:H2'	2:B:14:DC:C6	2.53	0.44
1:D:433:ASN:O	1:D:433:ASN:ND2	2.51	0.43
1:A:365:LYS:NZ	3:C:12:DG:O6	2.47	0.42

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	168/177 (95%)	161 (96%)	7 (4%)	0	100	100
1	D	167/177 (94%)	158 (95%)	9 (5%)	0	100	100
All	All	335/354 (95%)	319 (95%)	16 (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	155/162 (96%)	155 (100%)	0	100	100
1	D	154/162 (95%)	154 (100%)	0	100	100
All	All	309/324 (95%)	309 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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)

Of 12 ligands modelled in this entry, 12 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	170/177 (96%)	-0.33	0 100 100	45, 63, 94, 122	0
1	D	169/177 (95%)	0.22	13 (7%) 13 7	59, 82, 167, 192	0
2	B	23/23 (100%)	-0.34	0 100 100	49, 53, 114, 126	0
2	E	23/23 (100%)	-0.26	0 100 100	57, 79, 138, 158	0
3	C	23/23 (100%)	-0.38	0 100 100	40, 47, 131, 142	0
3	F	23/23 (100%)	-0.36	0 100 100	53, 67, 142, 155	0
All	All	431/446 (96%)	-0.12	13 (3%) 50 34	40, 72, 158, 192	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	313	LEU	6.6
1	D	298	LEU	5.8
1	D	301	ARG	4.2
1	D	309	LEU	3.8
1	D	300	GLY	3.3
1	D	316	HIS	3.1
1	D	302	ALA	2.7
1	D	311	ASN	2.6
1	D	294	HIS	2.6
1	D	295	LYS	2.5
1	D	303	PHE	2.3
1	D	297	HIS	2.3
1	D	304	ARG	2.2

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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	ZN	D	506	1/1	0.90	0.07	98,98,98,98	0
4	ZN	D	502	1/1	0.94	0.15	115,115,115,115	0
4	ZN	D	503	1/1	0.95	0.15	66,66,66,66	0
4	ZN	D	501	1/1	0.95	0.05	172,172,172,172	0
4	ZN	A	504	1/1	0.95	0.17	56,56,56,56	0
4	ZN	A	506	1/1	0.98	0.12	79,79,79,79	0
4	ZN	A	501	1/1	0.99	0.11	72,72,72,72	0
4	ZN	A	502	1/1	0.99	0.15	73,73,73,73	0
4	ZN	D	505	1/1	0.99	0.12	92,92,92,92	0
4	ZN	A	503	1/1	0.99	0.14	58,58,58,58	0
4	ZN	A	505	1/1	0.99	0.13	53,53,53,53	0
4	ZN	D	504	1/1	0.99	0.11	69,69,69,69	0

6.5 Other polymers (i)

There are no such residues in this entry.