



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 13, 2023 – 10:54 PM JST

PDB ID : 5YEH
Title : Crystal structure of CTCF ZFs4-8-eCBS
Authors : Yin, M.; Wang, J.; Wang, M.; Li, X.; Wang, Y.
Deposited on : 2017-09-17
Resolution : 2.33 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.36
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) : 2.36

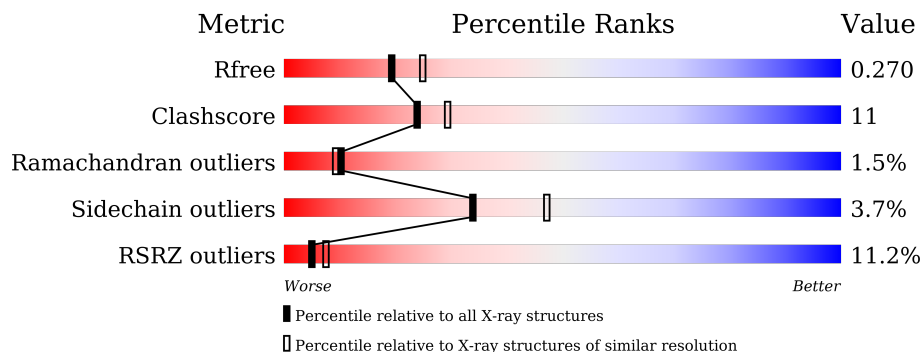
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.33 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	5974 (2.34-2.30)
Clashscore	141614	6604 (2.34-2.30)
Ramachandran outliers	138981	6523 (2.34-2.30)
Sidechain outliers	138945	6523 (2.34-2.30)
RSRZ outliers	127900	5855 (2.34-2.30)

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 of 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	142	 14% 73% 23% 5% 5%
1	B	142	 10% 72% 24% 5% 5%
2	C	20	 5% 55% 45%
2	E	20	 15% 50% 45% 5% 5%
3	D	20	 5% 65% 35%
3	F	20	 5% 70% 30%

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3853 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 Transcriptional repressor CTCF.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	139	Total 1072	C 663	N 209	O 186	S 14	0	0	0
1	B	139	Total 1076	C 666	N 210	O 186	S 14	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	20	Total 409	C 195	N 78	O 117	P 19	0	0	0
2	E	20	Total 409	C 195	N 78	O 117	P 19	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	D	20	Total 403	C 192	N 75	O 117	P 19	0	0	0
3	F	20	Total 403	C 192	N 75	O 117	P 19	0	0	0

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total 5	Zn 5	0	0
4	B	5	Total 5	Zn 5	0	0

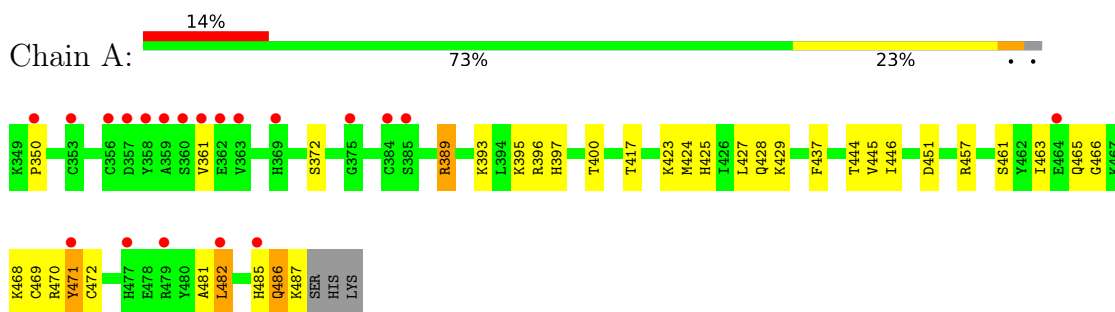
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	26	Total O 26 26	0	0
5	C	5	Total O 5 5	0	0
5	D	5	Total O 5 5	0	0
5	B	10	Total O 10 10	0	0
5	E	15	Total O 15 15	0	0
5	F	10	Total O 10 10	0	0

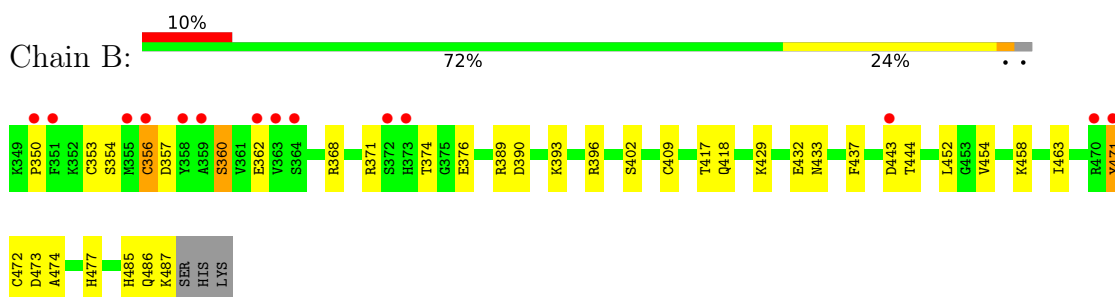
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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



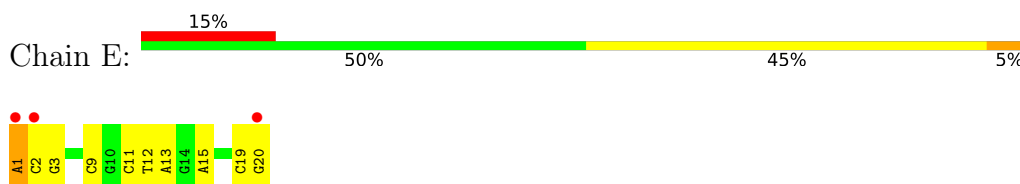
- Molecule 1: Transcriptional repressor CTCF



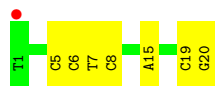
- Molecule 2: DNA (5'-D(*AP*CP*GP*GP*TP*TP*TP*CP*CP*GP*CP*TP*AP*GP*AP*GP*GP*GP*CP*G)-3')



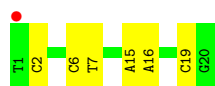
- Molecule 2: DNA (5'-D(*AP*CP*GP*GP*TP*TP*TP*CP*CP*GP*CP*TP*AP*GP*AP*GP*GP*GP*CP*G)-3')



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



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	45.84Å 54.88Å 66.94Å 81.17° 80.11° 79.72°	Depositor
Resolution (Å)	32.71 – 2.33 32.71 – 2.33	Depositor EDS
% Data completeness (in resolution range)	95.8 (32.71-2.33) 95.8 (32.71-2.33)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.70 (at 2.34Å)	Xtrriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, R_{free}	0.222 , 0.270 0.222 , 0.270	Depositor DCC
R_{free} test set	1291 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	45.6	Xtrriage
Anisotropy	0.391	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 40.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3853	wwPDB-VP
Average B, all atoms (Å ²)	63.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.08% 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](#)

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	0.52	0/1099	0.70	1/1480 (0.1%)
1	B	0.51	1/1103 (0.1%)	0.70	0/1484
2	C	1.02	1/459 (0.2%)	1.01	0/705
2	E	0.95	0/459	1.03	1/705 (0.1%)
3	D	0.92	0/451	1.01	0/693
3	F	0.92	1/451 (0.2%)	1.05	0/693
All	All	0.74	3/4022 (0.1%)	0.87	2/5760 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	409	CYS	CB-SG	7.31	1.94	1.82
2	C	6	DT	C3'-O3'	-6.18	1.35	1.44
3	F	16	DA	C3'-O3'	-5.05	1.37	1.44

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	482	LEU	CA-CB-CG	5.88	128.84	115.30
2	E	1	DA	C3'-C2'-C1'	-5.05	96.44	102.50

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	1072	0	977	32	0
1	B	1076	0	989	28	0
2	C	409	0	226	7	0
2	E	409	0	226	9	0
3	D	403	0	225	5	0
3	F	403	0	225	8	0
4	A	5	0	0	0	0
4	B	5	0	0	0	0
5	A	26	0	0	4	0
5	B	10	0	0	3	0
5	C	5	0	0	0	0
5	D	5	0	0	1	0
5	E	15	0	0	1	0
5	F	10	0	0	3	0
All	All	3853	0	2868	74	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 11.

The worst 5 of 74 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:356:CYS:SG	5:B:606:HOH:O	2.29	0.91
3:F:15:DA:OP1	5:F:101:HOH:O	1.89	0.90
1:B:432:GLU:OE1	5:B:601:HOH:O	1.94	0.86
1:A:451:ASP:OD2	5:A:601:HOH:O	1.95	0.82
1:B:418:GLN:HE22	2:E:13:DA:H62	1.24	0.82

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	137/142 (96%)	135 (98%)	1 (1%)	1 (1%)	22	26
1	B	137/142 (96%)	127 (93%)	7 (5%)	3 (2%)	6	5
All	All	274/284 (96%)	262 (96%)	8 (3%)	4 (2%)	10	9

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	486	GLN
1	B	473	ASP
1	B	486	GLN
1	B	444	THR

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	108/130 (83%)	104 (96%)	4 (4%)	34	47
1	B	109/130 (84%)	105 (96%)	4 (4%)	34	47
All	All	217/260 (84%)	209 (96%)	8 (4%)	34	47

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	471	TYR
1	B	443	ASP
1	B	356	CYS
1	A	472	CYS
1	B	360	SER

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

Mol	Chain	Res	Type
1	B	369	HIS
1	B	418	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	428	GLN
1	A	428	GLN
1	A	425	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 10 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	139/142 (97%)	0.80	20 (14%) 2 3	29, 58, 91, 106	0
1	B	139/142 (97%)	0.89	14 (10%) 7 10	26, 54, 105, 112	0
2	C	20/20 (100%)	0.32	1 (5%) 28 36	36, 58, 114, 140	0
2	E	20/20 (100%)	0.39	3 (15%) 2 3	34, 66, 123, 142	0
3	D	20/20 (100%)	0.22	1 (5%) 28 36	45, 55, 113, 135	0
3	F	20/20 (100%)	0.20	1 (5%) 28 36	50, 59, 109, 129	0
All	All	358/364 (98%)	0.72	40 (11%) 5 7	26, 57, 105, 142	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	363	VAL	4.9
1	B	358	TYR	4.7
1	A	359	ALA	4.3
1	A	375	GLY	4.1
1	A	361	VAL	3.9

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 monosaccharides 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	A	505	1/1	0.88	0.08	86,86,86,86	0
4	ZN	B	505	1/1	0.92	0.12	92,92,92,92	0
4	ZN	B	501	1/1	0.94	0.06	88,88,88,88	0
4	ZN	A	501	1/1	0.96	0.05	97,97,97,97	0
4	ZN	A	503	1/1	0.98	0.11	42,42,42,42	0
4	ZN	A	502	1/1	0.98	0.10	46,46,46,46	0
4	ZN	B	502	1/1	0.99	0.10	43,43,43,43	0
4	ZN	B	503	1/1	1.00	0.12	41,41,41,41	0
4	ZN	B	504	1/1	1.00	0.16	45,45,45,45	0
4	ZN	A	504	1/1	1.00	0.12	38,38,38,38	0

6.5 Other polymers [i](#)

There are no such residues in this entry.