



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2026 – 02:07 PM UTC

PDB ID : 5YEL / pdb_00005yel
Title : Crystal structure of CTCF ZFs6-11-gb7CSE
Authors : Yin, M.; Wang, J.; Wang, M.; Li, X.; Wang, Y.
Deposited on : 2017-09-18
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 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

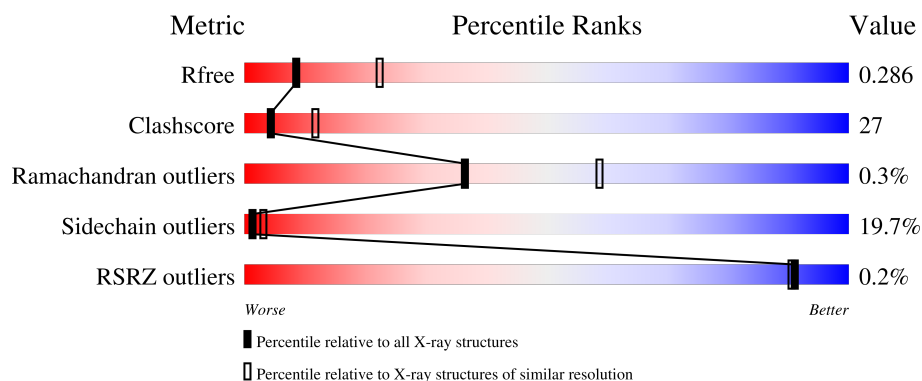
1 Overall quality at a glance

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}	180053	1130 (2.98-2.94)
Clashscore	190562	1157 (2.98-2.94)
Ramachandran outliers	187476	1101 (2.98-2.94)
Sidechain outliers	187428	1101 (2.98-2.94)
RSRZ outliers	180081	1130 (2.98-2.94)

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	D	26	46% 54%
1	F	26	38% 62%
2	A	176	46% 37% 11% 6%
2	B	176	% 43% 39% 13% 5%
3	C	26	65% 35%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	26	 62% 38%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4772 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 DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	26	Total	C	N	O	P	0	0	0
			526	253	98	150	25			
1	F	26	Total	C	N	O	P	0	0	0
			526	253	98	150	25			

- Molecule 2 is a protein called Transcriptional repressor CTCF.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	166	Total	C	N	O	S	0	0	0
			1315	816	259	221	19			
2	B	168	Total	C	N	O	S	0	0	0
			1327	823	264	221	19			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	504	SER	CYS	engineered mutation	UNP P49711
B	504	SER	CYS	engineered mutation	UNP P49711

- Molecule 3 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	26	Total	C	N	O	P	0	0	0
			533	256	95	157	25			
3	E	26	Total	C	N	O	P	0	0	0
			533	256	95	157	25			

- Molecule 4 is ZINC ION (CCD ID: ZN) (formula: Zn).

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

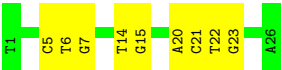
Continued on next page...

Continued from previous page...

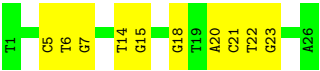
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	6	Total	Zn	0	0
			6	6		



● Molecule 3: DNA (26-MER)



● Molecule 3: DNA (26-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	190.99Å 69.63Å 84.44Å 90.00° 116.25° 90.00°	Depositor
Resolution (Å)	37.87 – 2.96 37.87 – 2.96	Depositor EDS
% Data completeness (in resolution range)	98.1 (37.87-2.96) 98.0 (37.87-2.96)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.08 (at 2.95Å)	Xtriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, R_{free}	0.268 , 0.286 0.267 , 0.286	Depositor DCC
R_{free} test set	1551 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å ²)	68.6	Xtriage
Anisotropy	0.520	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 65.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.477 for -h-2*k,-k,l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	4772	wwPDB-VP
Average B, all atoms (Å ²)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.76% 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

5.1 Standard geometry

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	D	0.23	0/590	0.52	0/906
1	F	0.23	0/590	0.52	0/906
2	A	0.62	1/1349 (0.1%)	0.98	8/1807 (0.4%)
2	B	0.79	5/1361 (0.4%)	1.08	12/1824 (0.7%)
3	C	0.28	0/597	0.57	0/921
3	E	0.28	0/597	0.56	0/921
All	All	0.55	6/5084 (0.1%)	0.82	20/7285 (0.3%)

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
2	A	0	1
2	B	0	1
All	All	0	2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	405	LYS	C-N	5.87	1.40	1.33
2	B	548	PRO	N-CD	5.85	1.55	1.47
2	B	551	VAL	C-N	5.77	1.41	1.33
2	A	548	PRO	N-CD	5.37	1.55	1.47
2	B	522	PRO	N-CD	5.33	1.55	1.47

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	411	ILE	N-CA-C	9.47	119.44	110.53
2	B	554	ALA	N-CA-C	8.37	120.18	111.14
2	A	554	ALA	N-CA-C	8.06	119.84	111.14
2	B	410	TYR	N-CA-C	7.36	119.08	111.14
2	B	435	ALA	N-CA-C	7.11	119.73	109.14

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	A	546	HIS	Peptide
2	B	546	HIS	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	D	526	0	294	14	0
1	F	526	0	294	14	0
2	A	1315	0	1193	89	3
2	B	1327	0	1205	101	3
3	C	533	0	297	10	0
3	E	533	0	297	18	0
4	A	6	0	0	0	0
4	B	6	0	0	0	0
All	All	4772	0	3580	224	3

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 27.

The worst 5 of 224 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)
2:B:497:CYS:HB2	2:B:510:MET:CE	1.39	1.52
2:B:497:CYS:CB	2:B:510:MET:HE1	1.50	1.40
2:A:568:ASN:ND2	3:E:22:DT:OP2	1.78	1.16

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:516:THR:HA	2:A:533:ARG:NH1	1.62	1.12
2:B:414:ALA:CA	2:B:415:ARG:HB2	1.82	1.09

All (3) 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)
2:A:413:HIS:NE2	2:B:413:HIS:NE2[1_566]	2.01	0.19
2:A:413:HIS:CE1	2:B:413:HIS:CE1[1_566]	2.05	0.15
2:A:413:HIS:CE1	2:B:413:HIS:NE2[1_566]	2.19	0.01

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	
2	A	158/176 (90%)	149 (94%)	9 (6%)	0	100	100
2	B	160/176 (91%)	150 (94%)	9 (6%)	1 (1%)	21	45
All	All	318/352 (90%)	299 (94%)	18 (6%)	1 (0%)	36	59

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	415	ARG

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	
2	A	132/158 (84%)	109 (83%)	23 (17%)	2	5
2	B	132/158 (84%)	103 (78%)	29 (22%)	1	2
All	All	264/316 (84%)	212 (80%)	52 (20%)	1	3

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	421	THR
2	B	464	GLU
2	B	568	ASN
2	B	425	HIS
2	B	450	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	A	455	HIS
2	A	534	GLN
2	A	568	ASN
2	B	465	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	D	26/26 (100%)	-0.18	0 100 100	44, 79, 101, 105	0
1	F	26/26 (100%)	0.04	0 100 100	47, 80, 95, 98	0
2	A	166/176 (94%)	-0.18	0 100 100	35, 75, 103, 128	0
2	B	168/176 (95%)	-0.20	1 (0%) 85 82	15, 78, 110, 136	0
3	C	26/26 (100%)	-0.10	0 100 100	53, 70, 102, 107	0
3	E	26/26 (100%)	-0.10	0 100 100	53, 68, 101, 109	0
All	All	438/456 (96%)	-0.17	1 (0%) 91 90	15, 76, 107, 136	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	551	VAL	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 oligosaccharides 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($\text{\AA}^2$)	Q<0.9
4	ZN	B	601	1/1	0.95	0.16	248,248,248,248	0
4	ZN	A	604	1/1	0.96	0.12	42,42,42,42	0
4	ZN	A	601	1/1	0.97	0.11	50,50,50,50	0
4	ZN	A	605	1/1	0.99	0.04	37,37,37,37	0
4	ZN	A	606	1/1	0.99	0.06	31,31,31,31	0
4	ZN	A	602	1/1	0.99	0.09	9,9,9,9	0
4	ZN	B	604	1/1	0.99	0.06	110,110,110,110	0
4	ZN	B	605	1/1	0.99	0.04	60,60,60,60	0
4	ZN	B	606	1/1	0.99	0.03	61,61,61,61	0
4	ZN	A	603	1/1	1.00	0.02	23,23,23,23	0
4	ZN	B	602	1/1	1.00	0.02	68,68,68,68	0
4	ZN	B	603	1/1	1.00	0.01	82,82,82,82	0

6.5 Other polymers [i](#)

There are no such residues in this entry.