Full wwPDB X-ray Structure Validation Report (i)

Aug 12, 2024 – 09:31 PM EDT

PDB ID	:	8TLF
Title	:	CDCA7 (Mouse) Binds Non-B-form DNA oligo 36-mer (sg C2-Form 2)
Authors	:	Horton, J.R.; Ren, R.; Cheng, X.
Deposited on	:	2023-07-26
Resolution	:	1.64 Å(reported)

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

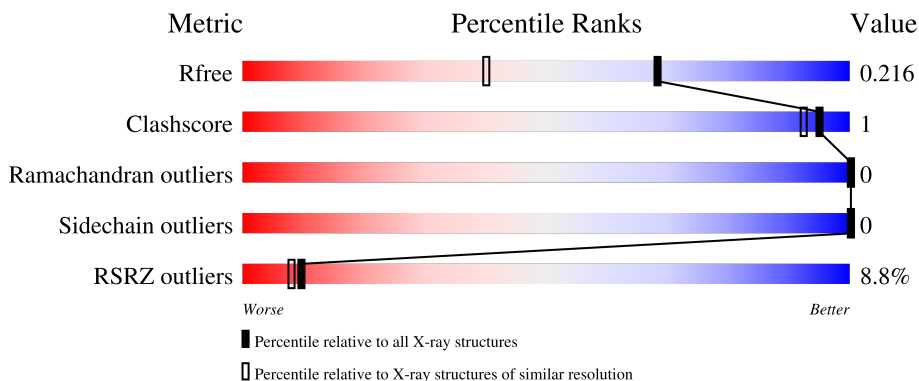|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.64 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	Percentile Ranks	Value
Rfree		0.216
Clashscore		1
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		8.8%

Worse   *Better*

▌ Percentile relative to all X-ray structures
▌ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3122 (1.66-1.62)
Clashscore	141614	3268 (1.66-1.62)
Ramachandran outliers	138981	3215 (1.66-1.62)
Sidechain outliers	138945	3215 (1.66-1.62)
RSRZ outliers	127900	3079 (1.66-1.62)

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 <=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	144	6% / 71% · 28%
1	B	144	10% / 68% · 28%
2	X	36	67% 6% 28%
2	Y	36	67% 6% 28%

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 3007 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 Cell division cycle-associated protein 7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	104	Total	C	N	O	S	0	0	0
			830	501	161	155	13			
1	B	104	Total	C	N	O	S	0	1	0
			844	509	167	155	13			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	239	GLY	-	expression tag	UNP Q9D0M2
A	240	PRO	-	expression tag	UNP Q9D0M2
A	241	LEU	-	expression tag	UNP Q9D0M2
B	239	GLY	-	expression tag	UNP Q9D0M2
B	240	PRO	-	expression tag	UNP Q9D0M2
B	241	LEU	-	expression tag	UNP Q9D0M2

- Molecule 2 is a DNA chain called ssDNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	26	Total	C	N	O	P	0	0	0
			534	253	95	160	26			
2	Y	26	Total	C	N	O	P	0	0	0
			518	243	93	156	26			

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 6	C 3	O 3	0	0
3	X	1	Total 6	C 3	O 3	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 4	C 2	O 2	0	0
4	X	1	Total 4	C 2	O 2	0	0

Continued on next page...

W O R L D W I D E

PDB

PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	Y	1	Total 4	C 2	O 2	0	0

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	3	Total 3	Zn 3	0	0
5	B	3	Total 3	Zn 3	0	0

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	X	3	Total 3	Mg 3	0	0
6	Y	1	Total 1	Mg 1	0	0

- Molecule 7 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	53	Total 53	O 53	0	0
7	B	42	Total 42	O 42	0	0
7	X	86	Total 86	O 86	0	0
7	Y	66	Total 66	O 66	0	0

3 Residue-property plots ⓘ

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.
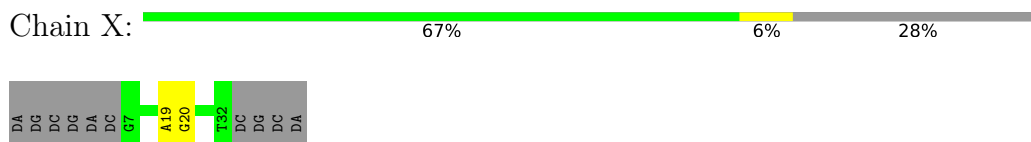
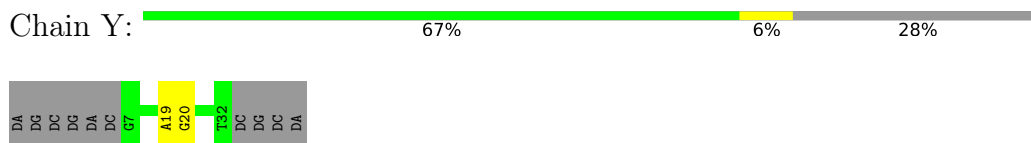• Molecule 1: Cell division cycle-associated protein 7
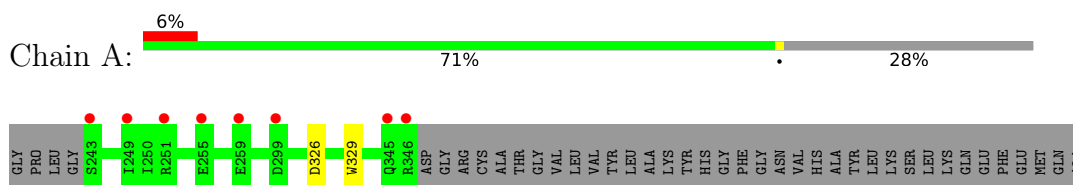
Chain A:

• Molecule 1: Cell division cycle-associated protein 7

Chain B:

• Molecule 2: ssDNA

Chain X:

• Molecule 2: ssDNA

Chain Y:
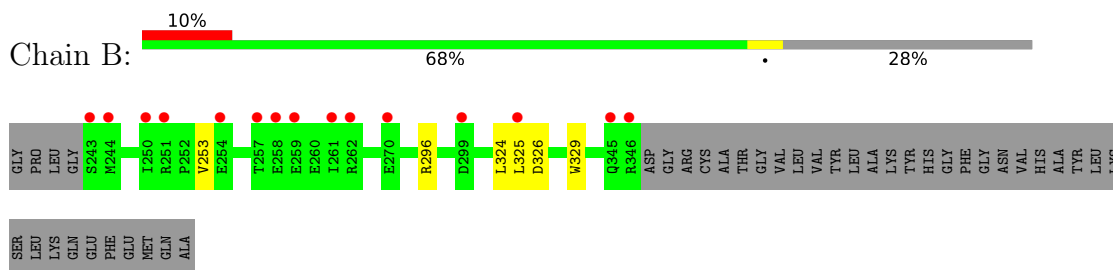
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	89.40Å 31.49Å 133.32Å 90.00° 95.75° 90.00°	Depositor
Resolution (Å)	29.90 – 1.64 29.90 – 1.64	Depositor EDS
% Data completeness (in resolution range)	81.8 (29.90-1.64) 77.1 (29.90-1.64)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.98 (at 1.64Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.192 , 0.218 0.193 , 0.216	Depositor DCC
R_{free} test set	1815 reflections (4.81%)	wwPDB-VP
Wilson B-factor (Å^2)	15.5	Xtriage
Anisotropy	0.926	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 44.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3007	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 34.48 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.7804e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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 ⓘ

5.1　Standard geometry ⓘ

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

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	0/847	0.50	0/1144
1	B	0.23	0/864	0.52	0/1165
2	X	0.48	0/597	0.91	0/920
2	Y	0.47	0/579	0.90	0/893
All	All	0.35	0/2887	0.71	0/4122

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 ⓘ

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	830	0	785	1	0
1	B	844	0	809	3	0
2	X	534	0	294	1	0
2	Y	518	0	281	1	0
3	A	6	0	8	0	0
3	X	6	0	8	0	0
4	A	4	0	6	0	0
4	X	4	0	6	0	0
4	Y	4	0	6	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	3	0	0	0	0
5	B	3	0	0	0	0
6	X	3	0	0	0	0
6	Y	1	0	0	0	0
7	A	53	0	0	0	0
7	B	42	0	0	0	0
7	X	86	0	0	0	0
7	Y	66	0	0	0	0
All	All	3007	0	2203	6	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 (6) 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:326:ASP:HB3	1:B:329:TRP:HB2	1.94	0.49
1:A:326:ASP:HB3	1:A:329:TRP:HB2	1.97	0.47
2:X:19:DA:H2'	2:X:20:DG:C8	2.49	0.46
2:Y:19:DA:H2'	2:Y:20:DG:C8	2.51	0.46
1:B:296:ARG:NH2	1:B:324:LEU:O	2.50	0.43
1:B:253:VAL:HG13	1:B:325:LEU:HA	2.01	0.42

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	102/144 (71%)	102 (100%)	0	0	100	100
1	B	103/144 (72%)	103 (100%)	0	0	100	100
All	All	205/288 (71%)	205 (100%)	0	0	100	100

There are no Ramachandran outliers to report.

5.3.2   Protein sidechains ⓘ

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	96/127 (76%)	96 (100%)	0	100	100
1	B	98/127 (77%)	98 (100%)	0	100	100
All	All	194/254 (76%)	194 (100%)	0	100	100

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

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

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 monosaccharides in this entry.

5.6   Ligand geometry ⓘ

Of 15 ligands modelled in this entry, 10 are monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
4	EDO	A	502	-	3,3,3	0.46	0	2,2,2	0.33	0
3	GOL	A	501	-	5,5,5	0.89	0	5,5,5	0.99	0
3	GOL	X	102	-	5,5,5	0.89	0	5,5,5	0.99	0
4	EDO	Y	102	-	3,3,3	0.46	0	2,2,2	0.33	0
4	EDO	X	101	-	3,3,3	0.46	0	2,2,2	0.34	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	EDO	A	502	-	-	0/1/1/1	-
3	GOL	A	501	-	-	0/4/4/4	-
3	GOL	X	102	-	-	0/4/4/4	-
4	EDO	Y	102	-	-	0/1/1/1	-
4	EDO	X	101	-	-	0/1/1/1	-

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 ⓘ

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 (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(Å²)	Q<0.9
1	A	104/144 (72%)	0.49	8 (7%)	13	11	12, 24, 49, 69	0
1	B	104/144 (72%)	0.81	15 (14%)	2	2	15, 29, 58, 87	0
2	X	26/36 (72%)	-0.10	0	100	100	16, 22, 40, 44	0
2	Y	26/36 (72%)	-0.09	0	100	100	22, 29, 34, 52	0
All	All	260/360 (72%)	0.50	23 (8%)	10	8	12, 27, 55, 87	0

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	346	ARG	6.1
1	B	345	GLN	5.8
1	A	345	GLN	4.5
1	A	346	ARG	4.4
1	B	325	LEU	3.3
1	B	243	SER	3.3
1	B	258	GLU	3.3
1	B	261	ILE	3.1
1	A	243	SER	3.0
1	A	259	GLU	2.7
1	B	254	GLU	2.7
1	B	259	GLU	2.6
1	B	262	ARG	2.6
1	B	299	ASP	2.5
1	A	299	ASP	2.4
1	B	257	THR	2.4
1	B	244	MET	2.4
1	A	251	ARG	2.2
1	B	270	GLU	2.1
1	B	250	ILE	2.1
1	A	255	GLU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	251	ARG	2.0
1	A	249	ILE	2.0

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 monosaccharides in this entry.

6.4 Ligands ⓘ

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(Å²)	Q<0.9
3	GOL	X	102	6/6	0.32	0.35	47,50,54,66	0
4	EDO	Y	102	4/4	0.50	0.31	40,44,46,50	0
3	GOL	A	501	6/6	0.72	0.21	29,46,46,58	0
4	EDO	X	101	4/4	0.85	0.18	27,32,33,40	0
4	EDO	A	502	4/4	0.87	0.16	27,32,36,47	0
6	MG	Y	101	1/1	0.91	0.05	39,39,39,39	0
6	MG	X	105	1/1	0.95	0.06	37,37,37,37	0
6	MG	X	103	1/1	0.98	0.06	14,14,14,14	0
6	MG	X	104	1/1	0.99	0.04	17,17,17,17	0
5	ZN	B	403	1/1	0.99	0.05	24,24,24,24	0
5	ZN	A	503	1/1	0.99	0.07	14,14,14,14	0
5	ZN	A	505	1/1	1.00	0.05	20,20,20,20	0
5	ZN	B	401	1/1	1.00	0.06	16,16,16,16	0
5	ZN	B	402	1/1	1.00	0.06	18,18,18,18	0
5	ZN	A	504	1/1	1.00	0.05	15,15,15,15	0

6.5 Other polymers ⓘ

There are no such residues in this entry.