



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 8, 2022 – 04:03 pm GMT

PDB ID : 7Q4N
Title : transcription factor CDX2 bound to hydroxymethylated DNA
Authors : Morgunova, E.; Yin, Y.; Popov, A.; Taipale, J.
Deposited on : 2021-11-01
Resolution : 3.20 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
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.31.2
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.31.2

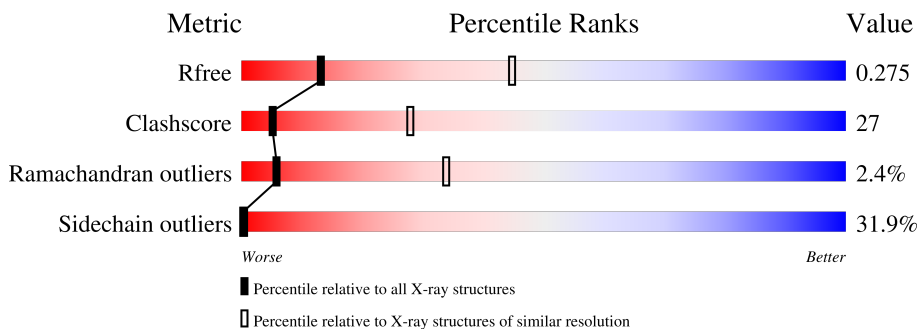
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 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)

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	18	28% (green), 56% (yellow), 17% (orange)
1	B	18	28% (green), 50% (yellow), 22% (orange)
2	D	18	17% (green), 28% (yellow), 56% (orange)
2	E	18	17% (green), 44% (yellow), 39% (orange)
3	C	67	36% (green), 36% (yellow), 21% (orange), 6% (grey)
3	K	67	37% (green), 48% (yellow), 13% (orange), 2% (grey)

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 2634 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 (18-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	18	368	177	57	116	18	0	0	0
1	B	18	368	177	57	116	18	0	0	0

- Molecule 2 is a DNA chain called DNA (18-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	18	373	177	76	102	18	0	0	0
2	E	18	373	177	76	102	18	0	0	0

- Molecule 3 is a protein called Homeobox protein CDX-2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	K	66	582	368	115	99	0	0	0
3	C	63	556	353	110	93	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	2	Total O 2 2	0	0
4	D	1	Total O 1 1	0	0
4	K	4	Total O 4 4	0	0
4	B	2	Total O 2 2	0	0

Continued on next page...

Continued from previous page...

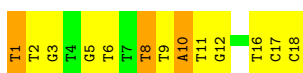
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	2	Total O 2 2	0	0
4	C	3	Total O 3 3	0	0

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. 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. 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: DNA (18-MER)

Chain A: 



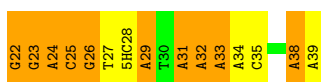
- Molecule 1: DNA (18-MER)

Chain B: 

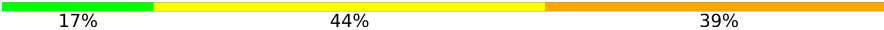


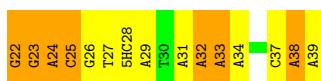
- Molecule 2: DNA (18-MER)

Chain D: 



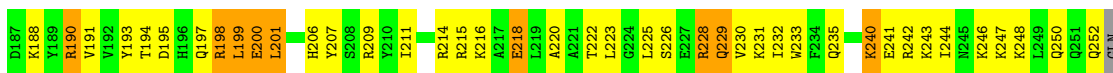
- Molecule 2: DNA (18-MER)

Chain E: 



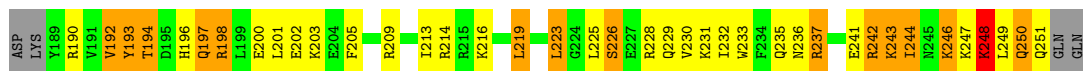
- Molecule 3: Homeobox protein CDX-2

Chain K: 



- Molecule 3: Homeobox protein CDX-2

Chain C: 



4 Data and refinement statistics

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	69.90Å 46.63Å 120.05Å 90.00° 98.25° 90.00°	Depositor
Resolution (Å)	43.45 – 3.20 43.41 – 3.03	Depositor EDS
% Data completeness (in resolution range)	98.2 (43.45-3.20) 96.7 (43.41-3.03)	Depositor EDS
R_{merge}	0.27	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.03 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.238 , 0.279 0.238 , 0.275	Depositor DCC
R_{free} test set	385 reflections (5.20%)	wwPDB-VP
Wilson B-factor (Å ²)	40.7	Xtriage
Anisotropy	1.578	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	2634	wwPDB-VP
Average B, all atoms (Å ²)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.33% 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: 5HC

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	1.22	1/409 (0.2%)	1.65	8/628 (1.3%)
1	B	1.26	3/409 (0.7%)	1.62	7/628 (1.1%)
2	D	1.10	0/396	1.85	17/606 (2.8%)
2	E	1.14	2/396 (0.5%)	1.67	10/606 (1.7%)
3	C	0.75	0/565	0.95	0/752
3	K	0.72	1/591 (0.2%)	1.00	0/786
All	All	1.02	7/2766 (0.3%)	1.47	42/4006 (1.0%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1	DT	OP3-P	-7.26	1.52	1.61
2	E	22	DG	P-OP1	6.85	1.60	1.49
1	B	1	DT	OP3-P	-6.81	1.52	1.61
1	B	18	DC	P-O5'	6.62	1.66	1.59
2	E	37	DC	P-O5'	5.41	1.65	1.59

The worst 5 of 42 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	22	DG	P-O3'-C3'	10.99	132.89	119.70
2	D	32	DA	P-O3'-C3'	-9.74	108.02	119.70
2	D	23	DG	O4'-C1'-N9	9.25	114.47	108.00
2	E	32	DA	P-O3'-C3'	-8.72	109.23	119.70
1	A	16	DT	P-O3'-C3'	-8.40	109.62	119.70

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	368	0	208	18	0
1	B	368	0	208	17	1
2	D	373	0	203	14	1
2	E	373	0	203	16	0
3	C	556	0	581	42	0
3	K	582	0	606	29	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
4	C	3	0	0	1	0
4	D	1	0	0	0	0
4	E	2	0	0	0	0
4	K	4	0	0	0	0
All	All	2634	0	2009	122	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 27.

The worst 5 of 122 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:18:DC:N4	2:E:22:DG:O6	1.87	1.07
3:C:202:GLU:OE2	3:C:237:ARG:HD2	1.63	0.98
2:D:22:DG:H2'	2:D:23:DG:C8	2.10	0.86
2:E:22:DG:OP1	2:E:22:DG:C5	2.33	0.81
1:A:1:DT:H2'	1:A:2:DT:C6	2.19	0.78

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)
2:D:23:DG:OP1	1:B:18:DC:O3'[2_465]	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	
3	C	61/67 (91%)	51 (84%)	7 (12%)	3 (5%)	2	17
3	K	64/67 (96%)	56 (88%)	8 (12%)	0	100	100
All	All	125/134 (93%)	107 (86%)	15 (12%)	3 (2%)	6	34

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	193	TYR
3	C	248	LYS
3	C	250	GLN

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	
3	C	58/62 (94%)	40 (69%)	18 (31%)	0	0
3	K	61/62 (98%)	41 (67%)	20 (33%)	0	0
All	All	119/124 (96%)	81 (68%)	38 (32%)	0	0

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	219	LEU
3	C	244	ILE
3	C	223	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	C	237	ARG
3	C	248	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	250	GLN
3	C	251	GLN
3	K	250	GLN
3	C	196	HIS
3	C	197	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

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	$\# Z > 2$
2	5HC	E	28	1,2	18,22,23	0.36	0	23,31,34	0.52	0
2	5HC	D	28	1,2	18,22,23	0.41	0	23,31,34	0.49	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
2	5HC	E	28	1,2	-	4/9/23/24	0/2/2/2

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	5HC	D	28	1,2	-	4/9/23/24	0/2/2/2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	28	5HC	O4'-C4'-C5'-O5'
2	D	28	5HC	C3'-C4'-C5'-O5'
2	D	28	5HC	C4-C5-C5M-O5
2	D	28	5HC	C6-C5-C5M-O5
2	E	28	5HC	O4'-C4'-C5'-O5'

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	28	5HC	3	0
2	D	28	5HC	1	0

5.5 Carbohydrates [i](#)

There are no monosaccharides 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 [i](#)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.