



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 14, 2018 – 05:14 PM EDT

PDB ID : 6ES2
Title : Structure of CDX2-DNA(CAA)
Authors : Morgunova, E.; Yin, Y.; Jolma, A.; Popov, A.; Taipale, J.
Deposited on : 2017-10-19
Resolution : 2.95 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

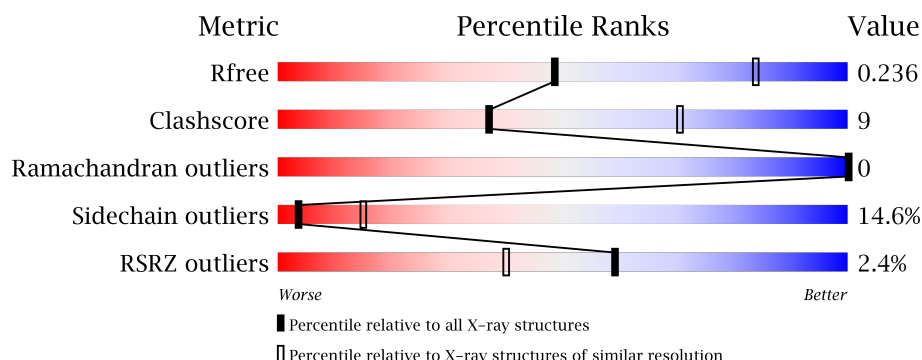
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.95 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}	100719	2395 (3.00-2.92)
Clashscore	112137	2773 (3.00-2.92)
Ramachandran outliers	110173	2680 (3.00-2.92)
Sidechain outliers	110143	2683 (3.00-2.92)
RSRZ outliers	101464	2421 (3.00-2.92)

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. 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	18	
1	B	18	
2	K	71	
2	L	71	
3	D	18	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	18	 A horizontal bar chart showing the quality of chain E. The bar is divided into three segments: green (56%), yellow (39%), and orange (6%). The percentages are labeled below the bar. <div><div></div><div>56%</div><div>39%</div><div>6%</div></div>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 2730 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 (5'-D(P*TP*TP*GP*TP*GP*TP*TP*TP*AP*TP*TP*GP*CP*CP*TP*CP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	18	Total	C	N	O	P	0	0	0
			363	176	52	117	18			
1	B	18	Total	C	N	O	P	0	0	0
			363	176	52	117	18			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	K	70	Total	C	N	O	0	1	0
			626	393	126	107			
2	L	70	Total	C	N	O	0	0	0
			618	389	123	106			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	18	Total	C	N	O	P	0	0	0
			375	177	81	99	18			
3	E	18	Total	C	N	O	P	0	0	0
			375	177	81	99	18			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total	O	0	0
			3	3		
4	K	2	Total	O	0	0
			2	2		
4	B	1	Total	O	0	0
			1	1		

Continued on next page...

Continued from previous page...

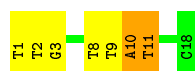
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	L	2	Total	O	0	0
			2	2		
4	D	1	Total	O	0	0
			1	1		
4	E	1	Total	O	0	0
			1	1		

3 Residue-property plots

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: DNA (5'-D(P*TP*TP*GP*TP*GP*TP*TP*TP*TP*AP*TP*TP*GP*CP*CP*TP*CP*C)-3')

Chain A: 



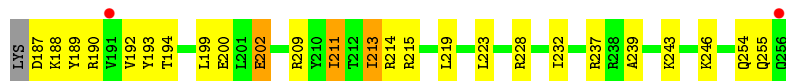
- Molecule 1: DNA (5'-D(P*TP*TP*GP*TP*GP*TP*TP*TP*TP*AP*TP*TP*GP*CP*CP*TP*CP*C)-3')

Chain B: 




- Molecule 2: Homeobox protein CDX-2

Chain K: 



- Molecule 2: Homeobox protein CDX-2

Chain L: 



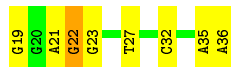
- Molecule 3: DNA (5'-D(P*GP*GP*AP*GP*GP*CP*AP*AP*TP*AP*AP*AP*AP*CP*AP*C P*AP*A)-3')

Chain D: 



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

Chain E: 56% 39% 6%



4 Data and refinement statistics

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	70.25Å 46.70Å 128.63Å 90.00° 101.40° 90.00°	Depositor
Resolution (Å)	66.19 – 2.95 55.96 – 2.95	Depositor EDS
% Data completeness (in resolution range)	96.5 (66.19-2.95) 96.5 (55.96-2.95)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.15 (at 2.96Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.208 , 0.244 0.209 , 0.236	Depositor DCC
R_{free} test set	418 reflections (5.18%)	DCC
Wilson B-factor (Å ²)	83.8	Xtriage
Anisotropy	0.624	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 64.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2730	wwPDB-VP
Average B, all atoms (Å ²)	101.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.77% 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

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	0.72	0/402	1.09	2/618 (0.3%)
1	B	0.70	0/402	1.11	2/618 (0.3%)
2	K	1.06	2/638 (0.3%)	1.10	1/848 (0.1%)
2	L	1.02	2/627 (0.3%)	1.06	1/833 (0.1%)
3	D	0.75	0/424	1.20	3/652 (0.5%)
3	E	0.70	0/424	1.20	5/652 (0.8%)
All	All	0.87	4/2917 (0.1%)	1.12	14/4221 (0.3%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	K	202	GLU	CD-OE2	6.67	1.32	1.25
2	L	202	GLU	CD-OE2	5.57	1.31	1.25
2	K	239	ALA	C-O	5.53	1.33	1.23
2	L	239	ALA	C-O	5.06	1.32	1.23

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	27	DT	C1'-O4'-C4'	-7.60	102.50	110.10
3	E	27	DT	C1'-O4'-C4'	-7.54	102.56	110.10
3	D	32	DC	C1'-O4'-C4'	-6.71	103.39	110.10
3	E	32	DC	C1'-O4'-C4'	-6.64	103.45	110.10
3	E	19	DG	O5'-P-OP1	5.95	117.84	110.70
1	B	10	DA	C1'-O4'-C4'	-5.82	104.28	110.10
1	A	10	DA	C1'-O4'-C4'	-5.74	104.36	110.10
3	D	22	DG	C1'-O4'-C4'	-5.67	104.43	110.10
1	A	11	DT	C1'-O4'-C4'	-5.59	104.51	110.10
1	B	10	DA	O5'-P-OP2	5.55	117.36	110.70
3	E	22	DG	O5'-P-OP2	5.54	117.35	110.70
3	E	22	DG	C1'-O4'-C4'	-5.43	104.67	110.10
2	L	190	ARG	NE-CZ-NH1	5.16	122.88	120.30
2	K	190	ARG	NE-CZ-NH1	5.04	122.82	120.30

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	363	0	209	7	0
1	B	363	0	209	9	0
2	K	626	0	651	11	0
2	L	618	0	643	9	0
3	D	375	0	200	7	0
3	E	375	0	200	6	0
4	A	3	0	0	0	0
4	B	1	0	0	0	0
4	D	1	0	0	0	0
4	E	1	0	0	0	0
4	K	2	0	0	0	0
4	L	2	0	0	0	0
All	All	2730	0	2112	41	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 9.

All (41) 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:K:228:ARG:O	2:K:232:ILE:HG12	2.02	0.59
3:D:22:DG:H2''	3:D:23:DG:O5'	2.03	0.59
2:L:228:ARG:O	2:L:232:ILE:HG12	2.03	0.59
3:E:22:DG:H2''	3:E:23:DG:O5'	2.04	0.57
1:A:11:DT:OP2	2:K:228:ARG:NH2	2.37	0.57
2:L:202:GLU:OE1	2:L:237:ARG:NE	2.36	0.57
3:D:21:DA:H2''	3:D:22:DG:O5'	2.05	0.56
3:E:21:DA:H2''	3:E:22:DG:O5'	2.06	0.56
2:K:202:GLU:OE1	2:K:237:ARG:NE	2.36	0.56
1:B:9:DT:H5'	2:L:193:TYR:CE2	2.42	0.55
3:E:35:DA:H2''	3:E:36:DA:C8	2.42	0.55

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1:DT:H2''	1:A:2:DT:C6	2.45	0.52
1:B:2:DT:H3	3:E:35:DA:H62	1.58	0.52
2:K:211:ILE:CD1	2:K:215:ARG:HB3	2.40	0.51
2:L:211:ILE:CD1	2:L:215:ARG:HB3	2.40	0.51
1:B:1:DT:H2''	1:B:2:DT:C6	2.45	0.51
3:E:35:DA:C2'	3:E:36:DA:C8	2.96	0.49
1:A:9:DT:H5'	2:K:193:TYR:CE2	2.48	0.49
3:D:35:DA:H2''	3:D:36:DA:C8	2.47	0.49
3:D:35:DA:C2'	3:D:36:DA:C8	2.96	0.48
3:E:21:DA:H2'	3:E:22:DG:C8	2.49	0.48
1:B:11:DT:OP2	2:L:228:ARG:NH2	2.47	0.47
3:D:21:DA:H2'	3:D:22:DG:C8	2.50	0.47
2:K:189:TYR:HB3	3:D:30:DA:H5'	1.97	0.47
1:B:9:DT:H5'	2:L:193:TYR:CZ	2.51	0.46
1:A:9:DT:H2''	1:A:10:DA:O5'	2.17	0.44
1:B:2:DT:H1'	1:B:3:DG:C8	2.53	0.44
2:L:211:ILE:HD11	2:L:215:ARG:HB3	2.00	0.44
1:A:2:DT:H1'	1:A:3:DG:C8	2.54	0.43
1:A:9:DT:H5'	2:K:193:TYR:CZ	2.54	0.43
1:A:8:DT:C6	1:A:9:DT:H72	2.53	0.43
1:B:17:DC:H5'	1:B:18:DC:OP2	2.19	0.43
2:K:200:GLU:HG3	2:K:223:LEU:HD21	2.01	0.43
1:B:9:DT:H2''	1:B:10:DA:O5'	2.19	0.42
3:D:19:DG:H2''	3:D:20:DG:OP1	2.20	0.42
2:L:219:LEU:HA	2:L:219:LEU:HD23	1.83	0.41
1:B:8:DT:C6	1:B:9:DT:H72	2.55	0.41
2:K:211:ILE:HD11	2:K:215:ARG:HB3	2.02	0.41
2:K:219:LEU:HA	2:K:219:LEU:HD23	1.84	0.41
2:L:213:ILE:HD12	2:L:214:ARG:H	1.86	0.40
2:K:213:ILE:HD12	2:K:214:ARG:H	1.85	0.40

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	
2	K	69/71 (97%)	67 (97%)	2 (3%)	0	100	100
2	L	68/71 (96%)	66 (97%)	2 (3%)	0	100	100
All	All	137/142 (96%)	133 (97%)	4 (3%)	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	
2	K	66/66 (100%)	53 (80%)	13 (20%)	1	7
2	L	65/66 (98%)	58 (89%)	7 (11%)	7	27
All	All	131/132 (99%)	111 (85%)	20 (15%)	3	13

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	K	187	ASP
2	K	188	LYS
2	K	192	VAL
2	K	194	THR
2	K	199	LEU
2	K	209[A]	ARG
2	K	209[B]	ARG
2	K	211	ILE
2	K	213	ILE
2	K	243	LYS
2	K	246	LYS
2	K	254	GLN
2	K	255	GLN
2	L	192	VAL
2	L	194	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	L	199	LEU
2	L	211	ILE
2	L	213	ILE
2	L	243	LYS
2	L	246	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
2	L	235	GLN
2	L	236	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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(Å ²)	Q<0.9
1	A	18/18 (100%)	-0.71	0	100	100	70, 98, 145, 154	0
1	B	18/18 (100%)	-0.60	0	100	100	74, 99, 148, 163	0
2	K	70/71 (98%)	0.14	2 (2%)	52	33	63, 96, 138, 152	0
2	L	70/71 (98%)	0.05	3 (4%)	36	22	62, 100, 139, 148	0
3	D	18/18 (100%)	-0.70	0	100	100	75, 92, 130, 140	0
3	E	18/18 (100%)	-0.63	0	100	100	75, 96, 136, 151	0
All	All	212/214 (99%)	-0.16	5 (2%)	59	40	62, 98, 141, 163	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	K	256	GLN	2.8
2	K	191	VAL	2.5
2	L	250	GLN	2.1
2	L	225	LEU	2.1
2	L	253	GLN	2.1

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.