



# Full wwPDB X-ray Structure Validation Report ⓘ

May 25, 2020 – 09:21 am BST

PDB ID : 4DKZ  
Title : Hairpin structure of a DNA dodecamer containing North-methanocarpa-2'-deoxythymidine  
Authors : Egli, M.; Pallan, P.S.  
Deposited on : 2012-02-05  
Resolution : 1.80 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](mailto: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 following versions of software and data (see [references ⓘ](#)) 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 : 2.11  
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) : 2.11

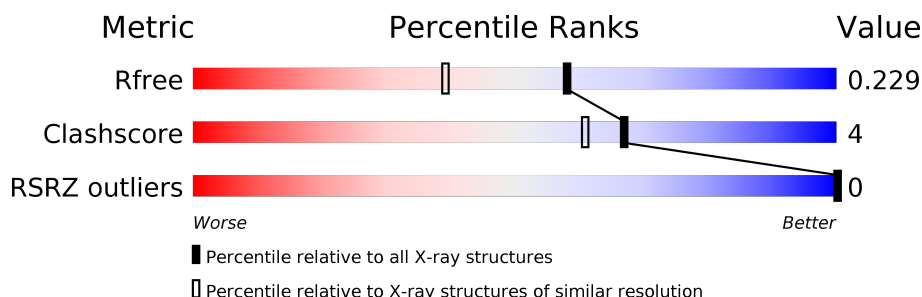
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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  $\geq 3$ , 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	12	 92% 8%
1	B	12	 75% 25%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
1	0KZ	A	107	X	-	-	-
1	0KZ	A	108	X	-	-	-
1	0KZ	B	207	X	-	-	-
1	0KZ	B	208	X	-	-	-

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 553 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 Synthetic DNA dodecamer containg North-methanocarba-2'-deoxythymidine (N-MCdT).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	12	Total	C	N	O	P	0	0	0
			245	120	46	68	11			
1	B	12	Total	C	N	O	P	0	0	0
			245	120	46	68	11			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	29	Total	O	0	0
			29	29		
2	B	34	Total	O	0	0
			34	34		

### 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: Synthetic DNA dodecamer containing North-methanocarpa-2'-deoxythymidine (N-MCdT)

Chain A: 



- Molecule 1: Synthetic DNA dodecamer containing North-methanocarpa-2'-deoxythymidine (N-MCdT)

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	24.30Å 35.28Å 36.71Å 90.00° 92.84° 90.00°	Depositor
Resolution (Å)	50.00 – 1.80 25.42 – 1.80	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-1.80) 98.8 (25.42-1.80)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	20.31 (at 1.80Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.234 , 0.252 0.220 , 0.229	Depositor DCC
$R_{free}$ test set	265 reflections (4.57%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.9	Xtriage
Anisotropy	0.727	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 39.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.117 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	553	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.49% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: 0KZ

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	0.48	0/227	0.75	0/346
1	B	0.46	0/227	0.78	0/346
All	All	0.47	0/454	0.77	0/692

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
1	A	2	0
1	B	2	0
All	All	4	0

There are no bond length outliers.

There are no bond angle outliers.

All (4) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	107	0KZ	C4'
1	A	108	0KZ	C4'
1	B	207	0KZ	C4'
1	B	208	0KZ	C4'

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	245	0	141	1	0
1	B	245	0	142	2	0
2	A	29	0	0	0	0
2	B	34	0	0	0	0
All	All	553	0	283	3	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 4.

All (3) 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:206:DA:H3'	1:B:207:OKZ:H3	1.94	0.50
1:A:112:DG:H2'	1:A:112:DG:N3	2.28	0.48
1:B:212:DG:N3	1:B:212:DG:H2'	2.31	0.45

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

There are no protein molecules in this entry.

### 5.3.2 Protein sidechains [i](#)

There are no protein molecules in this entry.

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

4 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
1	0KZ	A	108	-	15,23,24	2.17	7 (46%)	13,36,39	5.66	6 (46%)
1	0KZ	A	107	1	15,23,24	2.12	6 (40%)	13,36,39	6.03	6 (46%)
1	0KZ	B	207	1	15,23,24	2.31	8 (53%)	13,36,39	5.82	7 (53%)
1	0KZ	B	208	-	15,23,24	2.14	7 (46%)	13,36,39	5.72	7 (53%)

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
1	0KZ	A	108	-	1/1/5/5	1/4/32/33	0/3/3/3
1	0KZ	A	107	1	1/1/5/5	1/4/32/33	0/3/3/3
1	0KZ	B	207	1	1/1/5/5	1/4/32/33	0/3/3/3
1	0KZ	B	208	-	1/1/5/5	1/4/32/33	0/3/3/3

All (28) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	108	0KZ	C4-N3	5.12	1.42	1.33
1	B	207	0KZ	C4-N3	5.05	1.41	1.33
1	B	208	0KZ	C4-N3	4.94	1.41	1.33
1	A	107	0KZ	C4-N3	4.70	1.41	1.33
1	B	207	0KZ	C6'-C1'	3.71	1.59	1.53
1	A	107	0KZ	C6'-C1'	2.98	1.58	1.53
1	B	208	0KZ	C7'-C6'	2.90	1.56	1.50
1	A	107	0KZ	C7'-C6'	2.89	1.56	1.50
1	B	207	0KZ	C7'-C6'	2.85	1.56	1.50
1	B	208	0KZ	C6'-C1'	2.84	1.58	1.53
1	A	108	0KZ	C6'-C1'	2.76	1.58	1.53

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	108	0KZ	C7'-C6'	2.73	1.56	1.50
1	A	107	0KZ	C4-C5	2.65	1.47	1.41
1	A	108	0KZ	C4-C5	2.64	1.47	1.41
1	B	207	0KZ	C4'-C6'	2.59	1.56	1.51
1	A	107	0KZ	C4'-C6'	2.56	1.56	1.51
1	B	208	0KZ	C4-C5	2.52	1.46	1.41
1	A	108	0KZ	C5'-C4'	2.46	1.59	1.53
1	B	207	0KZ	C2-N3	2.44	1.43	1.38
1	A	107	0KZ	C6-C5	-2.40	1.33	1.40
1	B	208	0KZ	C4'-C6'	2.28	1.55	1.51
1	B	207	0KZ	C5'-C4'	2.27	1.59	1.53
1	B	207	0KZ	C4-C5	2.23	1.46	1.41
1	B	208	0KZ	C5'-C4'	2.22	1.58	1.53
1	B	207	0KZ	C6-C5	-2.15	1.34	1.40
1	A	108	0KZ	C6-C5	-2.12	1.34	1.40
1	B	208	0KZ	C2-N3	2.07	1.42	1.38
1	A	108	0KZ	C4'-C6'	2.01	1.55	1.51

All (26) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	208	0KZ	C4-N3-C2	14.44	127.34	115.14
1	A	108	0KZ	C4-N3-C2	14.42	127.31	115.14
1	A	107	0KZ	C4-N3-C2	14.17	127.10	115.14
1	B	207	0KZ	C4-N3-C2	14.09	127.03	115.14
1	A	107	0KZ	C2'-C1'-N1	-11.44	98.10	113.39
1	B	207	0KZ	C2'-C1'-N1	-10.04	99.96	113.39
1	A	108	0KZ	C2'-C1'-N1	-8.33	102.25	113.39
1	B	208	0KZ	C2'-C1'-N1	-8.21	102.42	113.39
1	B	208	0KZ	C7'-C4'-C5'	8.13	134.90	117.84
1	A	108	0KZ	C7'-C4'-C5'	8.00	134.64	117.84
1	A	107	0KZ	C4'-C6'-C1'	-7.72	102.37	108.39
1	B	207	0KZ	C7'-C4'-C5'	7.60	133.80	117.84
1	A	107	0KZ	C7'-C4'-C5'	7.38	133.34	117.84
1	B	208	0KZ	C4'-C6'-C1'	-7.16	102.80	108.39
1	A	108	0KZ	C4'-C6'-C1'	-6.87	103.03	108.39
1	B	207	0KZ	C4'-C6'-C1'	-6.60	103.24	108.39
1	B	207	0KZ	C5-C6-N1	4.24	126.76	122.19
1	B	208	0KZ	C5-C6-N1	3.82	126.32	122.19
1	A	108	0KZ	C5-C6-N1	3.69	126.18	122.19
1	A	107	0KZ	C5-C6-N1	3.26	125.70	122.19
1	A	107	0KZ	C7'-C6'-C4'	-2.63	57.79	59.62

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	207	0KZ	C7'-C6'-C4'	-2.41	57.94	59.62
1	B	207	0KZ	O3'-C3'-C2'	2.38	116.96	111.28
1	B	208	0KZ	C7'-C6'-C4'	-2.34	57.99	59.62
1	B	208	0KZ	O3'-C3'-C2'	2.17	116.46	111.28
1	A	108	0KZ	C7'-C6'-C4'	-2.15	58.12	59.62

All (4) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	108	0KZ	C4'
1	A	107	0KZ	C4'
1	B	207	0KZ	C4'
1	B	208	0KZ	C4'

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	108	0KZ	C7'-C4'-C5'-O5'
1	B	207	0KZ	C7'-C4'-C5'-O5'
1	B	208	0KZ	C7'-C4'-C5'-O5'
1	A	107	0KZ	C7'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	207	0KZ	1	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates 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 ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	10/12 (83%)	-0.39	0 100 100	17, 20, 24, 25	0
1	B	10/12 (83%)	-0.49	0 100 100	18, 20, 24, 24	0
All	All	20/24 (83%)	-0.44	0 100 100	17, 20, 24, 25	0

There are no RSRZ outliers to report.

### 6.2 Non-standard residues in protein, DNA, RNA chains [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	0KZ	B	208	21/22	0.90	0.15	23,25,27,28	0
1	0KZ	B	207	21/22	0.91	0.14	23,25,27,28	0
1	0KZ	A	107	21/22	0.91	0.15	22,24,26,27	0
1	0KZ	A	108	21/22	0.93	0.12	22,24,26,27	0

### 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.