



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 3, 2024 – 05:25 PM EST

PDB ID : 1MLX  
Title : Crystal Structure Analysis of a 2'-O-[2-(Methylthio)-ethyl]-Modified Oligodeoxynucleotide Duplex  
Authors : Prakash, T.P.; Manoharan, M.; Kawasaki, A.M.; Fraser, A.S.; Lesnik, E.A.; Sioufi, N.; Leeds, J.M.; Teplova, M.; Egli, M.  
Deposited on : 2002-08-31  
Resolution : 1.25 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 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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

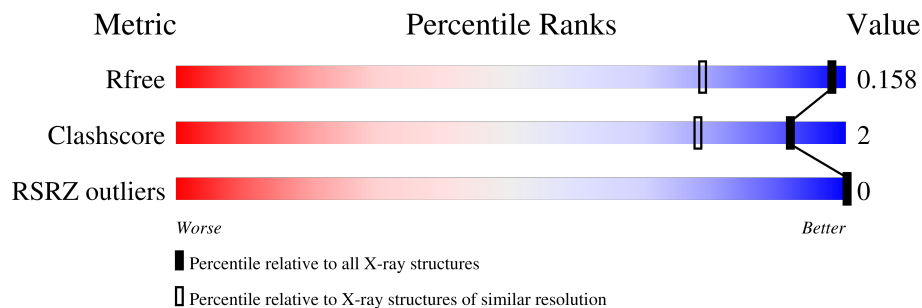
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.25 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	1023 (1.28-1.24)
Clashscore	141614	1060 (1.28-1.24)
RSRZ outliers	127900	1004 (1.28-1.24)

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  $\geq 3$ , 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	10	100%
1	B	10	10% (green), 80% (yellow), 10% (orange)

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 781 atoms, of which 224 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 5'-D(\*GP\*CP\*GP\*TP\*AP\*SMTP\*AP\*CP\*GP\*C)-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace	
			Total	C	H	N	O	P				
1	A	10	316	98	112	38	59	9	0	0	0	
1	B	10	Total	C	H	N	O	P	S	0	0	0
			319	100	112	38	59	9	1			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	84	Total	O	0	0
			84	84		
2	B	62	Total	O	0	0
			62	62		

### 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: 5'-D(\*GP\*CP\*GP\*TP\*AP\*SMTP\*AP\*CP\*GP\*C)-3'

Chain A:  100%

G1	C2	G3	T4	A5	T6	A7	C8	G9	C10
----	----	----	----	----	----	----	----	----	-----

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

Chain B:  10% 80% 10%

G111	C112	G113	T114	A115	T116	A117	G118	G119	C120
------	------	------	------	------	------	------	------	------	------

## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	24.73Å 44.47Å 46.23Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 1.25 19.58 – 1.25	Depositor EDS
% Data completeness (in resolution range)	(Not available) (15.00-1.25) 93.3 (19.58-1.25)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.32 (at 1.25Å)	Xtrriage
Refinement program	SHELXL-97	Depositor
R, $R_{free}$	0.141 , (Not available) 0.145 , 0.158	Depositor DCC
$R_{free}$ test set	1395 reflections (9.90%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.1	Xtrriage
Anisotropy	0.198	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 57.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.020 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	781	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.26% 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 i

### 5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: SMT

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	2.83	18/203 (8.9%)	2.99	31/309 (10.0%)
1	B	2.87	18/203 (8.9%)	2.94	29/309 (9.4%)
All	All	2.85	36/406 (8.9%)	2.97	60/618 (9.7%)

All (36) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	114	DT	C5-C6	-9.91	1.27	1.34
1	B	111	DG	N9-C8	8.87	1.44	1.37
1	A	8	DC	C5-C6	-7.99	1.27	1.34
1	A	4	DT	C5-C7	-7.82	1.45	1.50
1	B	115	DA	N1-C2	-7.74	1.27	1.34
1	A	3	DG	O4'-C1'	7.03	1.50	1.42
1	A	9	DG	C6-N1	6.87	1.44	1.39
1	B	119	DG	C2'-C1'	-6.75	1.45	1.52
1	A	8	DC	C4-N4	-6.68	1.27	1.33
1	B	117	DA	C8-N7	6.61	1.36	1.31
1	B	117	DA	C6-N1	6.56	1.40	1.35
1	B	111	DG	N3-C4	-6.51	1.30	1.35
1	B	119	DG	N7-C5	6.51	1.43	1.39
1	A	5	DA	N1-C2	6.50	1.40	1.34
1	A	2	DC	C4-C5	6.26	1.48	1.43
1	A	7	DA	C6-N1	-6.26	1.31	1.35
1	A	10	DC	C2-N3	6.24	1.40	1.35
1	B	120	DC	C4-C5	6.24	1.48	1.43
1	A	10	DC	C2-O2	-6.18	1.18	1.24
1	B	113	DG	C2-N3	-6.01	1.27	1.32
1	A	9	DG	C8-N7	5.75	1.34	1.30
1	A	7	DA	C6-N6	5.66	1.38	1.33
1	B	113	DG	C8-N7	-5.65	1.27	1.30
1	B	111	DG	N7-C5	5.64	1.42	1.39

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1	DG	O4'-C1'	5.63	1.49	1.42
1	B	119	DG	N9-C4	-5.62	1.33	1.38
1	B	113	DG	N1-C2	5.61	1.42	1.37
1	B	115	DA	N7-C5	5.59	1.42	1.39
1	B	111	DG	O4'-C1'	5.58	1.49	1.42
1	A	1	DG	N9-C4	-5.53	1.33	1.38
1	B	119	DG	O4'-C1'	5.44	1.48	1.42
1	A	3	DG	C2'-C1'	-5.29	1.47	1.52
1	A	9	DG	C2-N3	-5.21	1.28	1.32
1	A	9	DG	C2-N2	5.08	1.39	1.34
1	B	115	DA	C6-N1	5.06	1.39	1.35
1	A	9	DG	N9-C4	-5.03	1.33	1.38

All (60) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	112	DC	C6-N1-C2	-14.21	114.61	120.30
1	A	9	DG	N1-C2-N3	13.09	131.75	123.90
1	B	112	DC	C5-C6-N1	12.20	127.10	121.00
1	A	9	DG	C6-N1-C2	-10.09	119.05	125.10
1	B	111	DG	C2-N3-C4	9.95	116.88	111.90
1	A	3	DG	N1-C2-N3	9.93	129.85	123.90
1	A	5	DA	N1-C2-N3	-9.22	124.69	129.30
1	A	8	DC	O4'-C4'-C3'	-8.66	100.81	106.00
1	A	9	DG	N3-C2-N2	-8.32	114.08	119.90
1	B	112	DC	C4-C5-C6	-8.12	113.34	117.40
1	A	9	DG	N1-C6-O6	-7.93	115.14	119.90
1	A	3	DG	C6-N1-C2	-7.92	120.35	125.10
1	A	1	DG	C6-N1-C2	-7.92	120.35	125.10
1	A	1	DG	C5-N7-C8	-7.70	100.45	104.30
1	B	113	DG	O4'-C4'-C3'	-7.62	101.43	106.00
1	B	114	DT	C5-C6-N1	7.60	128.26	123.70
1	B	117	DA	C8-N9-C4	7.56	108.82	105.80
1	A	1	DG	N1-C2-N3	7.25	128.25	123.90
1	A	3	DG	C2-N3-C4	-7.22	108.29	111.90
1	A	1	DG	P-O3'-C3'	7.08	128.19	119.70
1	B	114	DT	O4'-C4'-C3'	-7.02	101.69	104.50
1	A	3	DG	C8-N9-C4	6.98	109.19	106.40
1	B	113	DG	N7-C8-N9	6.91	116.56	113.10
1	A	3	DG	N7-C8-N9	-6.87	109.67	113.10
1	B	111	DG	N1-C2-N3	-6.85	119.79	123.90
1	B	113	DG	C8-N9-C4	-6.84	103.66	106.40

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1	DG	N7-C8-N9	6.84	116.52	113.10
1	B	111	DG	O4'-C1'-N9	-6.83	103.22	108.00
1	B	119	DG	P-O5'-C5'	6.65	131.54	120.90
1	A	5	DA	N7-C8-N9	-6.54	110.53	113.80
1	A	1	DG	N3-C4-C5	-6.35	125.42	128.60
1	A	5	DA	C5-N7-C8	6.33	107.06	103.90
1	A	7	DA	C6-N1-C2	6.29	122.37	118.60
1	B	111	DG	C6-C5-N7	-6.26	126.64	130.40
1	B	117	DA	N9-C4-C5	-6.11	103.36	105.80
1	A	3	DG	O4'-C1'-N9	-6.05	103.77	108.00
1	A	9	DG	C5-C6-O6	6.01	132.20	128.60
1	B	111	DG	C5-C6-N1	-5.98	108.51	111.50
1	B	119	DG	C1'-O4'-C4'	-5.85	104.25	110.10
1	B	119	DG	O4'-C4'-C3'	-5.85	102.16	104.50
1	A	8	DC	C2-N3-C4	-5.84	116.98	119.90
1	A	3	DG	C5-C6-N1	5.80	114.40	111.50
1	A	1	DG	O4'-C1'-N9	-5.79	103.94	108.00
1	A	3	DG	C5-N7-C8	5.76	107.18	104.30
1	B	114	DT	C6-N1-C2	-5.69	118.45	121.30
1	A	7	DA	C2-N3-C4	-5.69	107.76	110.60
1	B	112	DC	N1-C2-N3	5.63	123.14	119.20
1	A	7	DA	O4'-C4'-C3'	-5.54	102.29	104.50
1	B	111	DG	N7-C8-N9	-5.46	110.37	113.10
1	B	112	DC	N1-C2-O2	-5.44	115.63	118.90
1	A	4	DT	C4-C5-C6	-5.43	114.74	118.00
1	B	117	DA	N3-C4-C5	5.42	130.59	126.80
1	B	115	DA	N9-C4-C5	5.27	107.91	105.80
1	B	114	DT	N3-C4-O4	-5.23	116.76	119.90
1	B	119	DG	N1-C2-N3	-5.21	120.77	123.90
1	A	9	DG	C4-C5-N7	-5.20	108.72	110.80
1	A	9	DG	O4'-C1'-N9	-5.19	104.36	108.00
1	B	119	DG	C5-C6-O6	5.14	131.69	128.60
1	B	117	DA	C2-N3-C4	-5.14	108.03	110.60
1	B	119	DG	N1-C6-O6	-5.11	116.84	119.90

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	204	112	111	0	0
1	B	207	112	120	1	0
2	A	84	0	0	0	0
2	B	62	0	0	0	0
All	All	557	224	231	1	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 2.

All (1) 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:116:SMT:SC	1:B:116:SMT:CD	2.13	1.35

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 [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
1	SMT	A	6	1	20,23,27	1.67	4 (20%)	28,33,39	1.98	6 (21%)
1	SMT	B	116	1	23,26,27	2.08	5 (21%)	31,36,39	3.27	8 (25%)

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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	SMT	A	6	1	-	0/9/27/31	0/2/2/2
1	SMT	B	116	1	-	1/12/30/31	0/2/2/2

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	116	SMT	CD-SC	5.71	2.13	1.78
1	B	116	SMT	O2'-CA'	4.56	1.55	1.43
1	A	6	SMT	O2'-CA'	3.93	1.56	1.42
1	B	116	SMT	C2-N3	-3.63	1.31	1.38
1	A	6	SMT	C6-N1	-2.81	1.33	1.38
1	A	6	SMT	C2-N3	-2.53	1.33	1.38
1	B	116	SMT	O4'-C4'	2.25	1.50	1.45
1	A	6	SMT	O4'-C1'	2.16	1.47	1.42
1	B	116	SMT	C6-C5	-2.04	1.31	1.34

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	116	SMT	CD-SC-CB	15.80	154.67	100.40
1	A	6	SMT	C6-C5-C4	-5.20	113.68	118.03
1	A	6	SMT	C5-C6-N1	5.15	128.64	123.34
1	A	6	SMT	O2'-C2'-C1'	4.21	117.29	109.08
1	B	116	SMT	O2'-CA'-CB	-3.45	95.19	109.54
1	B	116	SMT	C6-N1-C2	-3.43	117.83	121.30
1	B	116	SMT	N3-C2-N1	3.36	119.35	114.89
1	B	116	SMT	C5-C6-N1	3.29	126.72	123.34
1	A	6	SMT	C5A-C5-C4	2.62	121.66	118.77

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	116	SMT	C4-N3-C2	-2.40	124.25	127.35
1	B	116	SMT	C6-C5-C4	-2.35	116.07	118.03
1	A	6	SMT	C1'-N1-C6	2.25	124.87	121.12
1	A	6	SMT	C6-N1-C2	-2.10	119.17	121.30
1	B	116	SMT	C5A-C5-C6	2.09	125.63	122.85

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	B	116	SMT	CA'-CB-SC-CD

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	116	SMT	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 [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	9/10 (90%)	-0.96	0 <a href="#">100</a> <a href="#">100</a>	15, 17, 19, 22	0
1	B	9/10 (90%)	-0.90	0 <a href="#">100</a> <a href="#">100</a>	16, 19, 20, 24	0
All	All	18/20 (90%)	-0.93	0 <a href="#">100</a> <a href="#">100</a>	15, 18, 22, 24	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	SMT	A	6	22/26	0.97	0.05	14,20,29,29	0
1	SMT	B	116	25/26	0.98	0.05	14,18,27,51	0

### 6.3 Carbohydrates [i](#)

There are no monosaccharides 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.