Full wwPDB X-ray Structure Validation Report (i)

Feb 18, 2024 – 01:34 AM EST

PDB ID	:	417D
Title	:	A THYMINE-LIKE BASE ANALOGUE FORMS WOBBLE PAIRS WITH ADENINE
Authors	:	Lin, P.K.T.; Schuerman, M.H.; Moore, G.S.; Van Meervelt, L.; Loakes, D.; Brown, D.M.; Moore, M.H.
Deposited on	:	1998-07-15
Resolution	:	1.50 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

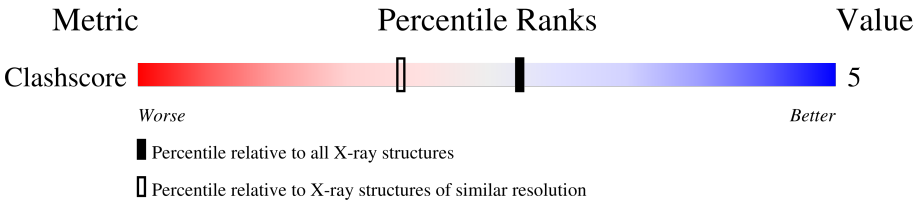
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.50 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))
Clashscore	141614	3144 (1.50-1.50)

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%

Mol	Chain	Length	Quality of chain
1	A	6	100%
1	B	6	50% 50%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 302 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(*CP*AP*CP*GP*(C46)P*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	6	Total	C	N	O	P	0	0	0
			122	59	24	34	5			
1	B	6	Total	C	N	O	P	0	0	0
			122	59	24	34	5			

- Molecule 2 is water.

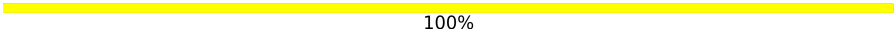
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	26	Total	O	0	0
			26	26		
2	B	32	Total	O	0	0
			32	32		

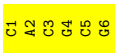
3 Residue-property plots (i)

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 (5'-D(*CP*AP*CP*GP*(C46)P*G)-3')

Chain A:





- Molecule 1: DNA (5'-D(*CP*AP*CP*GP*(C46)P*G)-3')

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	17.34Å 31.43Å 44.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	10.00 – 1.50 9.36 – 1.53	Depositor EDS
% Data completeness (in resolution range)	97.5 (10.00-1.50) 97.8 (9.36-1.53)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$	-	Xtriage
Refinement program	SHELXL-93	Depositor
R, R_{free}	0.155 , 0.250 0.210 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	9.2	Xtriage
Anisotropy	0.315	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.43 , 74.6	EDS
L-test for twinning[1]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	302	wwPDB-VP
Average B, all atoms (Å^2)	14.0	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 40.79 % 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 2.5990e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1]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: C46

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	2.59	7/111 (6.3%)	3.02	17/167 (10.2%)
1	B	2.67	10/111 (9.0%)	2.94	13/167 (7.8%)
All	All	2.63	17/222 (7.7%)	2.98	30/334 (9.0%)

All (17) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	2	DA	C3'-C2'	-8.94	1.41	1.52
1	B	10	DG	C3'-C2'	-7.88	1.42	1.52
1	B	12	DG	C3'-C2'	-7.70	1.43	1.52
1	A	4	DG	C3'-C2'	-7.41	1.43	1.52
1	B	9	DC	C3'-C2'	-7.28	1.43	1.52
1	B	8	DA	C3'-C2'	-6.97	1.43	1.52
1	A	6	DG	C3'-C2'	-6.88	1.44	1.52
1	A	3	DC	C3'-C2'	-6.41	1.44	1.52
1	B	7	DC	C4'-O4'	-6.12	1.39	1.45
1	B	7	DC	C3'-C2'	-5.82	1.45	1.52
1	A	1	DC	C3'-C2'	-5.61	1.45	1.52
1	A	4	DG	C4'-O4'	-5.42	1.39	1.45
1	B	8	DA	C1'-N9	-5.41	1.39	1.47
1	B	10	DG	C4'-O4'	-5.33	1.39	1.45
1	A	2	DA	C4'-O4'	-5.13	1.40	1.45
1	B	10	DG	C1'-N9	-5.03	1.40	1.47
1	B	8	DA	C4'-O4'	-5.00	1.40	1.45

All (30) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	8	DA	P-O3'-C3'	10.02	131.72	119.70
1	A	2	DA	O4'-C1'-N9	-9.28	101.51	108.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	9	DC	C6-N1-C2	9.15	123.96	120.30
1	A	2	DA	N1-C2-N3	-8.90	124.85	129.30
1	B	12	DG	O4'-C1'-N9	-8.83	101.82	108.00
1	A	2	DA	C5-C6-N1	-7.95	113.73	117.70
1	A	4	DG	O4'-C4'-C3'	-7.42	101.53	104.50
1	B	9	DC	OP1-P-OP2	-7.02	109.06	119.60
1	B	9	DC	P-O3'-C3'	6.97	128.06	119.70
1	A	3	DC	OP1-P-OP2	-6.84	109.34	119.60
1	B	8	DA	C5-C6-N1	-6.81	114.30	117.70
1	A	2	DA	C6-N1-C2	6.69	122.61	118.60
1	A	2	DA	O4'-C4'-C3'	-6.59	101.86	104.50
1	A	6	DG	C5-C6-N1	6.55	114.78	111.50
1	A	1	DC	C2-N3-C4	6.47	123.13	119.90
1	A	2	DA	C2-N3-C4	6.30	113.75	110.60
1	B	8	DA	O4'-C1'-N9	-6.14	103.70	108.00
1	A	6	DG	O4'-C4'-C3'	-6.03	102.09	104.50
1	B	9	DC	O4'-C1'-C2'	-6.01	101.09	105.90
1	A	1	DC	N3-C4-C5	-5.95	119.52	121.90
1	B	10	DG	O4'-C1'-N9	-5.87	103.89	108.00
1	B	10	DG	N9-C4-C5	5.47	107.59	105.40
1	A	4	DG	C8-N9-C4	-5.31	104.28	106.40
1	A	6	DG	O4'-C1'-N9	-5.26	104.31	108.00
1	A	3	DC	O4'-C4'-C3'	-5.21	102.41	104.50
1	A	6	DG	N3-C2-N2	5.20	123.54	119.90
1	A	2	DA	N9-C1'-C2'	5.16	122.40	112.60
1	B	7	DC	N1-C2-O2	5.14	121.98	118.90
1	B	7	DC	O4'-C1'-C2'	-5.11	101.81	105.90
1	B	10	DG	C8-N9-C4	-5.01	104.40	106.40

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	122	0	70	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	122	0	70	2	0
2	A	26	0	0	0	0
2	B	32	0	0	0	0
All	All	302	0	140	2	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 5.

All (2) 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:10:DG:C8	1:B:11:C46:H7C1	2.41	0.56
1:B:12:DG:N3	1:B:12:DG:H2'	2.32	0.44

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	C46	A	5	1	19,24,25	1.45	4 (21%)	24,34,37	2.24	7 (29%)
1	C46	B	11	1	19,24,25	1.54	3 (15%)	24,34,37	1.54	5 (20%)

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	C46	A	5	1	-	0/7/28/29	0/2/3/3
1	C46	B	11	1	-	0/7/28/29	0/2/3/3

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	11	C46	C2'-C3'	-3.57	1.43	1.52
1	B	11	C46	C4-N4	3.27	1.39	1.34
1	A	5	C46	C2'-C3'	-2.83	1.45	1.52
1	A	5	C46	C4-N4	2.61	1.38	1.34
1	A	5	C46	CM5-C5	2.34	1.55	1.50
1	A	5	C46	O5'-C5'	-2.25	1.39	1.44
1	B	11	C46	O5'-C5'	-2.21	1.39	1.44

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	5	C46	C5-C6-N1	-5.56	115.44	122.91
1	A	5	C46	C1'-N1-C6	-4.19	113.56	120.77
1	A	5	C46	C6-N1-C2	4.12	126.58	120.87
1	A	5	C46	C7-CM5-C5	-3.92	98.37	109.40
1	B	11	C46	C5-C6-N1	-3.85	117.74	122.91
1	A	5	C46	O2-C2-N3	3.24	127.60	122.33
1	B	11	C46	C6-N1-C2	3.00	125.02	120.87
1	B	11	C46	C1'-N1-C6	-2.51	116.45	120.77
1	A	5	C46	O4'-C1'-C2'	-2.49	101.54	106.25
1	B	11	C46	O4-C7-CM5	2.27	116.06	111.11
1	A	5	C46	CM5-C5-C4	2.17	120.34	115.46
1	B	11	C46	CM5-C5-C4	2.05	120.06	115.46

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	11	C46	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.