

wwPDB X-ray Structure Validation Summary Report (i)

Mar 13, 2023 – 02:01 PM EDT

PDB ID : 1D78

Title : HIGH RESOLUTION REFINEMENT OF THE HEXAGONAL A-DNA OC-

TAMER D(GTGTACAC) AT 1.4 ANGSTROMS RESOLUTION

Authors: Thota, N.; Li, X.H.; Bingman, C.A.; Sundaralingam, M.

Deposited on : 1992-06-12

Resolution : 1.40 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 (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 (1)) were used in the production of this report:

MolProbity : 4.02b-467 Xtriage (Phenix) : 1.13

EDS : 2.32.1

Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)

 $Refmac \quad : \quad 5.8.0158$

CCP4 : 7.0.044 (Gargrove)

Ideal geometry (proteins) : Engh & Huber (2001) Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

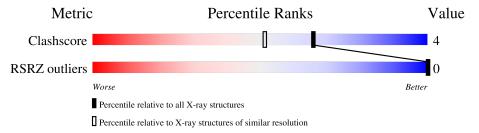
Validation Pipeline (wwPDB-VP) : 2.32.1

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 1.40 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 $(\# \mathrm{Entries})$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries},\ {\rm resolution\ range}({\rm \AA})) \end{array}$
Clashscore	141614	1812 (1.40-1.40)
RSRZ outliers	127900	1674 (1.40-1.40)

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% 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	8	12%	75%	12%



2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 198 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(*GP*TP*GP*TP*AP*CP*AP*C)-3').

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	А	8	Total	С	N	О	Р	0	0	0
	11		161	78	30	46	7			

• Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	37	Total O 37 37	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 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(*GP*TP*GP*TP*AP*CP*AP*C)-3')





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants	32.18Å 32.18Å 78.51Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 120.00°	Depositor
Resolution (Å)	5.00 - 1.40	Depositor
Resolution (A)	5.17 - 1.40	EDS
% Data completeness	(Not available) $(5.00-1.40)$	Depositor
(in resolution range)	89.8 (5.17-1.40)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	3.25 (at 1.40Å)	Xtriage
Refinement program	NUCLSQ	Depositor
P.P.	0.198 , (Not available)	Depositor
R, R_{free}	0.196 , (Not available)	DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å ²)	14.6	Xtriage
Anisotropy	0.380	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.97, 138.2	EDS
L-test for twinning ²	$ < L > = 0.47, < L^2> = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	198	wwPDB-VP
Average B, all atoms (Å ²)	20.0	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 64.21 % 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 8.5801e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.

²Theoretical 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.



¹Intensities estimated from amplitudes.

5 Model quality (i)

5.1 Standard geometry (i)

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	Chain	Bon	d lengths	Bond angles	
	Chain	RMSZ	# Z > 5	RMSZ # Z > 5	
1	A	1.98	6/180 (3.3%)	2.63	20/276 (7.2%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(\text{\AA})$	$\operatorname{Ideal}(ext{\AA})$
1	A	8	DC	P-O5'	6.67	1.66	1.59
1	A	6	DC	P-O5'	-6.42	1.53	1.59
1	A	3	DG	P-O5'	6.07	1.65	1.59
1	A	6	DC	P-OP1	5.47	1.58	1.49
1	A	3	DG	O3'-P	-5.22	1.54	1.61

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$Observed(^o)$	$Ideal(^{o})$
1	A	5	DA	N1-C2-N3	-9.91	124.35	129.30
1	A	7	DA	C6-N1-C2	8.30	123.58	118.60
1	A	7	DA	N1-C2-N3	-7.58	125.51	129.30
1	A	7	DA	C5-C6-N1	-7.54	113.93	117.70
1	A	2	DT	O5'-P-OP2	7.37	119.55	110.70

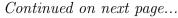
There are no chirality outliers.

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)	$\mathbf{H}(\mathbf{added})$	Clashes	Symm-Clashes
1	A	161	0	92	1	0





Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	37	0	0	1	0
All	All	198	0	92	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 4.

All (1) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{aligned} & ext{Interatomic} \ & ext{distance} \ & ext{(Å)} \end{aligned}$	Clash overlap (Å)
1:A:5:DA:H1'	2:A:26:HOH:O	1.94	0.66

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)

There are no non-standard protein/DNA/RNA residues in this entry.

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^{th} 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 $>$	$\#\text{RSRZ}{>}2$	$OWAB(Å^2)$	Q<0.9
1	A	8/8 (100%)	-0.53	0 100 100	16, 17, 19, 23	0

There are no RSRZ outliers to report.

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 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.

