



# wwPDB NMR Structure Validation Summary Report

Jun 7, 2020 – 04:30 am BST

PDB ID : 6I4O  
Title : Dodecamer DNA containing the synthetic base pair P-Z  
Authors : Padroni, G.; Reichenbach, L.F.; Parkinson, J.A.; Burley, G.A.  
Deposited on : 2018-11-10

This is a wwPDB NMR Structure Validation Summary 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/NMRValidationReportHelp>  
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:

Cyrange : Kirchner and Güntert (2011)  
NmrClust : Kelley et al. (1996)  
MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : 2.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

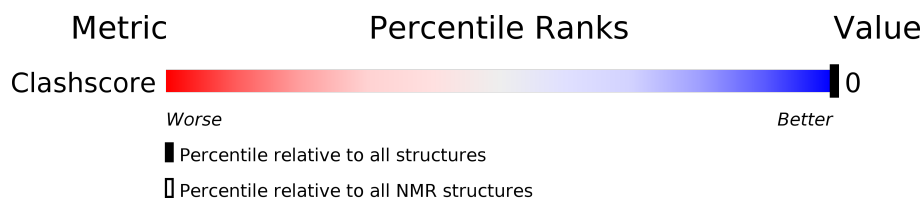
## 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

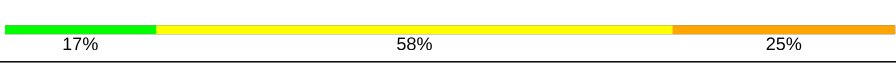
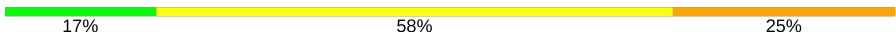
The overall completeness of chemical shifts assignment is 30%.

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)	NMR archive (#Entries)
Clashscore	158937	12864

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	12	 17% 58% 25%
1	B	12	 17% 58% 25%

## 2 Ensemble composition and analysis

This entry contains 10 models. This entry does not contain polypeptide chains, therefore identification of well-defined residues and clustering analysis are not possible. All residues are included in the validation scores.

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 766 atoms, of which 274 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called (DCZ)(DG)(DA)(DT)(DP)(DT)(DA)(DZ)(DA)(DT)(DC)(DG3).

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		P
1	A	12	Total 383	118	137	45	72	11	0
1	B	12	Total 383	118	137	45	72	11	0

## 4 Residue-property plots

### 4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: (DCZ)(DG)(DA)(DT)(DP)(DT)(DA)(DZ)(DA)(DT)(DC)(DG3)

Chain A: 



- Molecule 1: (DCZ)(DG)(DA)(DT)(DP)(DT)(DA)(DZ)(DA)(DT)(DC)(DG3)

Chain B: 

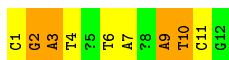


### 4.2 Residue scores for the representative (author defined) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: (DCZ)(DG)(DA)(DT)(DP)(DT)(DA)(DZ)(DA)(DT)(DC)(DG3)

Chain A: 



- Molecule 1: (DCZ)(DG)(DA)(DT)(DP)(DT)(DA)(DZ)(DA)(DT)(DC)(DG3)

Chain B: 



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing, molecular dynamics*.

Of the 5000 calculated structures, 10 were deposited, based on the following criterion: *clustering of MD*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
Amber	refinement	
MARDIGRAS	refinement	
Amber	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	175
Number of shifts mapped to atoms	158
Number of unparsed shifts	0
Number of shifts with mapping errors	14
Number of shifts with mapping warnings	3
Assignment completeness (well-defined parts)	30%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

### 5.1 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
All	All	4920	2740	2300	-

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 -.

There are no clashes.

## 5.2 Torsion angles [i](#)

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

There are no protein molecules in this entry.

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

There are no protein molecules in this entry.

### 5.2.3 RNA [i](#)

There are no RNA molecules in this entry.

MODRES-GEOMETRY INFOmissingINFO

## 5.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.4 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.5 Other polymers [i](#)

There are no such molecules in this entry.

## 5.6 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 30% for the well-defined parts and 30% for the entire structure.

### 6.1 Chemical shift list 1

File name: input\_cs.cif

Chemical shift list name: *starch\_output*

#### 6.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	175
Number of shifts mapped to atoms	158
Number of unparsed shifts	0
Number of shifts with mapping errors	14
Number of shifts with mapping warnings	3
Number of shift outliers (ShiftChecker)	20

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atoms found in structure. All 3 occurrences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	4	DT	H7	1.437	0.002	1
A	6	DT	H7	1.407	0.001	1
A	10	DT	H7	1.328	0.001	1

- Residue not found in structure. First 5 (of 14) occurrences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	1	DCZ	C1'	85.219	0.0	1
A	1	DCZ	H2'1	1.898	0.001	1
A	1	DCZ	H6	7.609	0.001	1
A	1	DCZ	H5	5.883	0.0	1
A	1	DCZ	C6	140.456	0.0	1



### 6.1.2 Chemical shift referencing [i](#)

No chemical shift referencing corrections were calculated (not enough data).

### 6.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 30%, i.e. 119 atoms were assigned a chemical shift out of a possible 392. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)
Sidechain	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)
Aromatic	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)
Overall	119/392 (30%)	67/232 (29%)	52/136 (38%)	0/24 (0%)

### 6.1.4 Statistically unusual chemical shifts [i](#)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	10	DT	C7	11.29	15.29 – 13.69	-20.0
1	A	4	DT	C7	11.32	15.29 – 13.69	-19.8
1	A	6	DT	C7	11.53	15.29 – 13.69	-18.5
1	A	4	DT	C6	135.72	141.81 – 136.91	-7.4
1	A	10	DT	C6	135.89	141.81 – 136.91	-7.1
1	A	4	DT	C1'	82.23	90.81 – 83.51	-6.8
1	A	10	DT	C1'	82.49	90.81 – 83.51	-6.4
1	A	6	DT	C1'	82.59	90.81 – 83.51	-6.3
1	A	6	DT	C6	136.53	141.81 – 136.91	-5.8
1	A	11	DC	C2'	36.58	44.32 – 37.12	-5.7
1	A	6	DT	C2'	36.44	43.42 – 36.92	-5.7
1	A	10	DT	C2'	36.64	43.42 – 36.92	-5.4
1	A	3	DA	C8	138.62	144.55 – 138.85	-5.4
1	A	7	DA	C1'	82.25	88.48 – 82.48	-5.4
1	A	2	DG	C1'	81.34	89.32 – 81.62	-5.4
1	A	9	DA	C2	151.44	158.48 – 151.68	-5.4
1	A	3	DA	C1'	82.35	88.48 – 82.48	-5.2
1	A	4	DT	C2'	36.87	43.42 – 36.92	-5.1
1	A	7	DA	C5'	65.59	70.89 – 65.59	-5.0
1	A	9	DA	C4'	84.60	90.20 – 84.60	-5.0

### 6.1.5 Random Coil Index (RCI) plots

No *random coil index* (RCI) plot could be generated from the current chemical shift list (starch\_output). RCI is only applicable to proteins.