



Full wwPDB NMR Structure Validation Report ⓘ

Sep 7, 2017 – 06:59 PM EDT

PDB ID : 5UZD
Title : Insights into Watson-Crick/Hoogsteen Breathing Dynamics and Damage Repair from the Solution Structure and Dynamic Ensemble of DNA Duplexes containing m1A - A2-DNA structure
Authors : Sathyamoorthy, B.; Shi, H.; Xue, Y.; Al-Hashimi, H.M.
Deposited on : unknown

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/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
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : rb-20029824
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

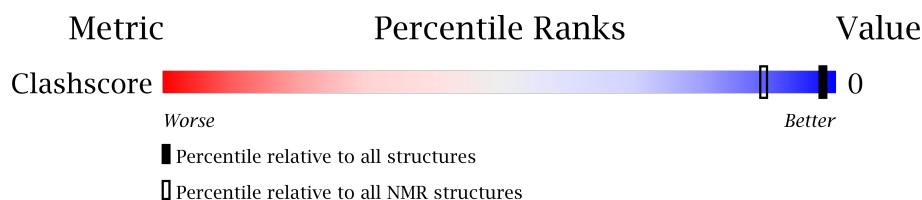
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR



The overall completeness of chemical shifts assignment is 54%.

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	125131	11601

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 ≥ 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	 75% 25%
2	B	12	 83% 17%

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 [i](#)

There are 2 unique types of molecules in this entry. The entry contains 755 atoms, of which 269 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(*GP*CP*AP*TP*CP*GP*AP*TP*TP*GP*GP*C)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	12	379	117	135	45	71	11	0

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3').

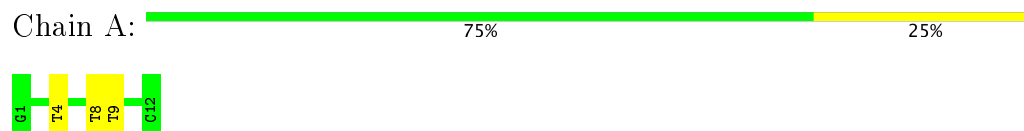
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	12	376	116	134	46	69	11	0

4 Residue-property plots [i](#)

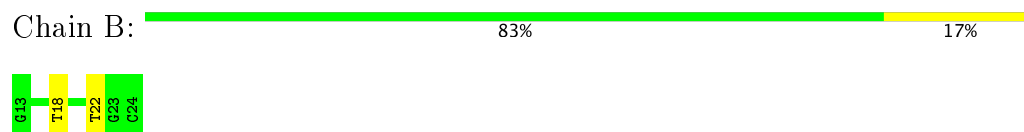
4.1 Average score per residue in the NMR ensemble

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: DNA (5'-D(*GP*CP*AP*TP*CP*GP*AP*TP*TP*GP*GP*C)-3')



- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')



4.2 Scores per residue for each member of the ensemble

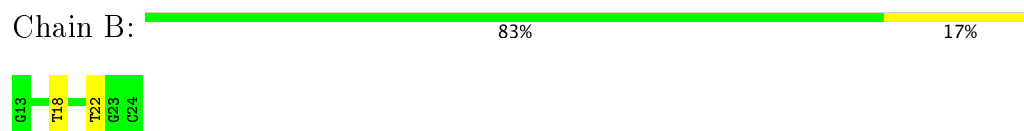
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

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



- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')



4.2.2 Score per residue for model 2

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

Chain A:  58% 42%



- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')

Chain B:  67% 33%




4.2.3 Score per residue for model 3

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

Chain A:  75% 25%



- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')

Chain B:  83% 17%




4.2.4 Score per residue for model 4

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

Chain A:  75% 25%



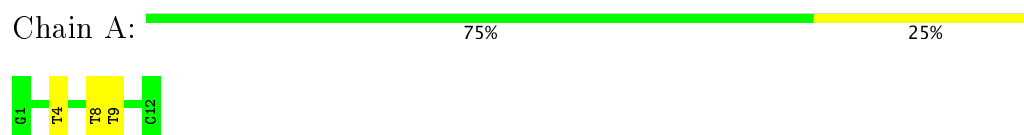
- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')

Chain B:  83% 17%

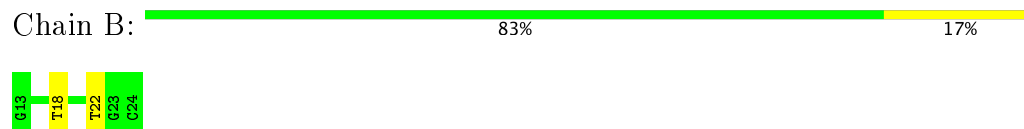


4.2.5 Score per residue for model 5

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

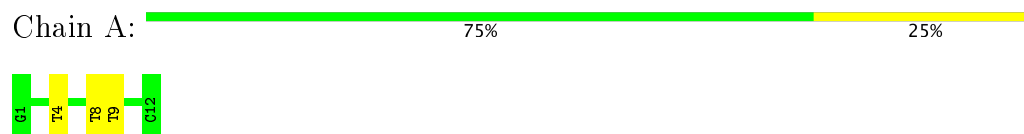


- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')

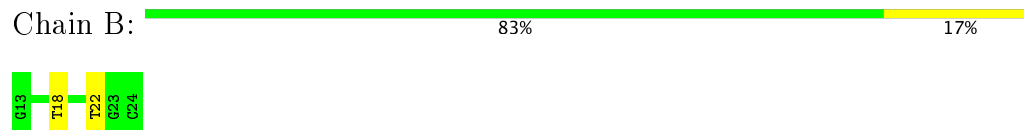


4.2.6 Score per residue for model 6

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

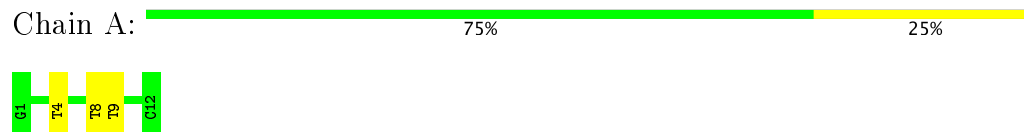


- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')

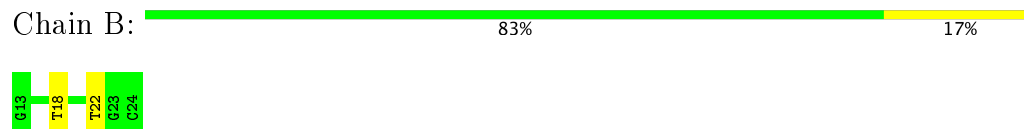


4.2.7 Score per residue for model 7

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

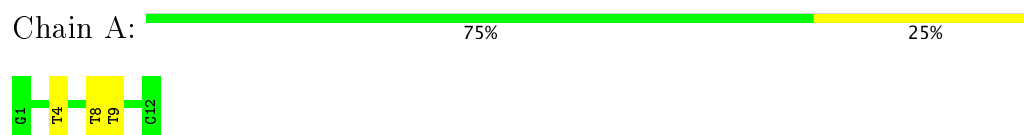


- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')

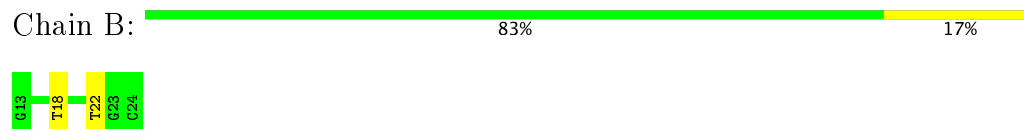


4.2.8 Score per residue for model 8

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

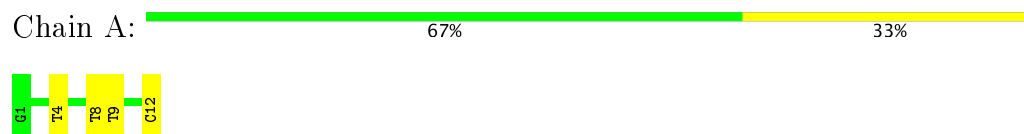


- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')

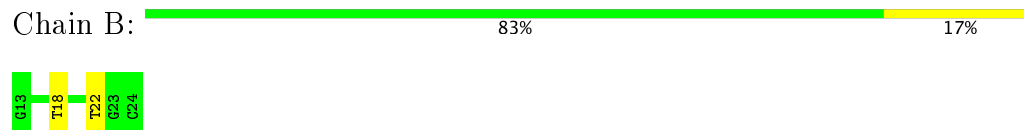


4.2.9 Score per residue for model 9

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

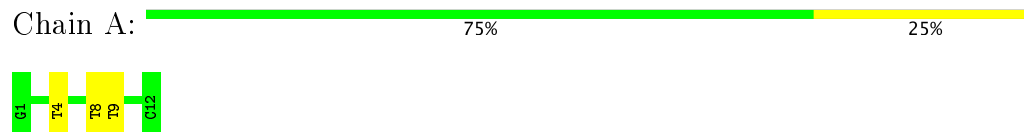


- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')

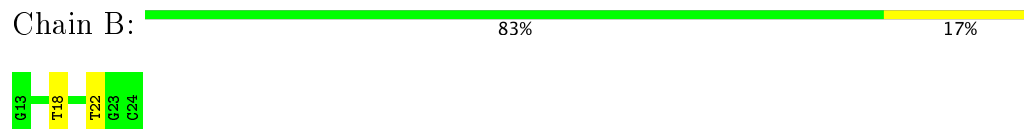


4.2.10 Score per residue for model 10

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



- Molecule 2: DNA (5'-D(*GP*CP*CP*AP*AP*TP*CP*GP*AP*TP*GP*C)-3')



5 Refinement protocol and experimental data overview

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

Of the 256 calculated structures, 10 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
X-PLOR NIH	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	5uzd_cs.cif
Number of chemical shift lists	1
Total number of shifts	300
Number of shifts mapped to atoms	300
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	54%

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

6 Model quality [i](#)

6.1 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 (average) 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	1.11±0.02	0±0/273 (0.0±0.0%)	1.52±0.01	4±1/420 (1.0±0.2%)
2	B	1.00±0.02	0±0/271 (0.0±0.0%)	1.45±0.01	3±0/416 (0.6±0.1%)
All	All	1.05	0/5440 (0.0%)	1.48	69/8360 (0.8%)

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
2	B	18	DT	C4-C5-C6	5.44	121.26	118.00	2	10
1	A	8	DT	C4-C5-C6	5.37	121.22	118.00	2	10
2	B	22	DT	C4-C5-C6	5.35	121.21	118.00	6	10
1	A	9	DT	C4-C5-C6	5.34	121.21	118.00	5	10
1	A	4	DT	C4-C5-C6	5.30	121.18	118.00	9	10
1	A	9	DT	C6-C5-C7	-5.08	119.85	122.90	8	3
2	B	18	DT	C6-C5-C7	-5.07	119.86	122.90	2	4
1	A	8	DT	C6-C5-C7	-5.07	119.86	122.90	1	6
2	B	22	DT	C6-C5-C7	-5.04	119.87	122.90	7	3
1	A	12	DC	C1'-O4'-C4'	-5.02	105.08	110.10	9	1
1	A	4	DT	C6-C5-C7	-5.01	119.89	122.90	10	2

There are no chirality outliers.

There are no planarity outliers.

6.2 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 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
1	A	244	135	137	0±0
2	B	242	134	136	0±0
All	All	4860	2690	2730	3

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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:11:DG:C5	1:A:12:DC:C5	0.48	3.01	2	2
2:B:15:DC:C2'	2:B:16:DA:C8	0.41	3.04	2	1

6.3 Torsion angles [i](#)

6.3.1 Protein backbone [i](#)

There are no protein molecules in this entry.

6.3.2 Protein sidechains [i](#)

There are no protein molecules in this entry.

6.3.3 RNA [i](#)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.6 Ligand geometry [i](#)

There are no ligands in this entry.

6.7 Other polymers [i](#)

There are no such molecules in this entry.

6.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 54% for the well-defined parts and 54% for the entire structure.

7.1 Chemical shift list 1

File name: 5uzd_cs.cif

Chemical shift list name: *A2-DNA_CS.final*

7.1.1 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	300
Number of shifts mapped to atoms	300
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing

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

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 54%, i.e. 258 atoms were assigned a chemical shift out of a possible 477. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ 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	258/477 (54%)	140/285 (49%)	108/161 (67%)	10/31 (32%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 54%, i.e. 258 atoms were assigned a chemical shift out of a possible 477. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹H	¹³C	¹⁵N
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	258/477 (54%)	140/285 (49%)	108/161 (67%)	10/31 (32%)

7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots ⓘ

No *random coil index* (RCI) plot could be generated from the current chemical shift list (A2-DNA_CS.final). RCI is only applicable to proteins.