Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:40 pm BST

PDB ID	:	2KH7
Title	:	Solution Structure of cis-5R,6S-thymine glycol opposite complementary guanine in duplex DNA
Authors	:	Brown, K.L.
Deposited on	:	2009-03-24

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

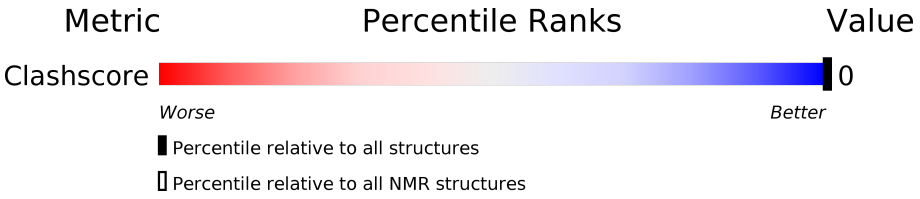
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 47%.

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 >=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 <=5%

Mol	Chain	Length	Quality of chain
1	A	12	8% 75% 17%
2	B	12	92% 8%

2 Ensemble composition and analysis (i)

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 are 2 unique types of molecules in this entry. The entry contains 765 atoms, of which 276 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	12	390	119	142	40	78	11	0

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

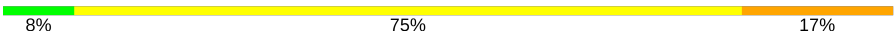
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	12	375	115	134	50	65	11	0

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: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

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

Chain B: 

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

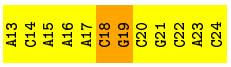
4.2.1 Score per residue for model 1

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

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

Chain B: 



4.2.2 Score per residue for model 2

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

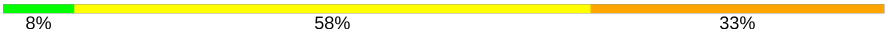
Chain A: 

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

Chain B: 

4.2.3 Score per residue for model 3

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

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

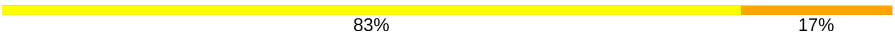
Chain B: 

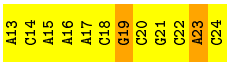
4.2.4 Score per residue for model 4

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

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

Chain B: 

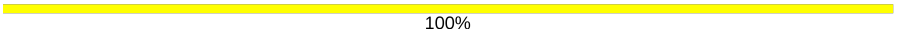


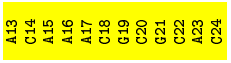
4.2.5 Score per residue for model 5

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

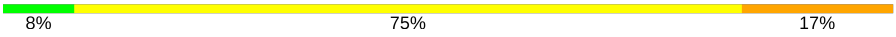
• Molecule 2: 5'-D(*AP*CP*AP*AP*AP*CP*GP*CP*GP*CP*AP*C)-3'

Chain B: 

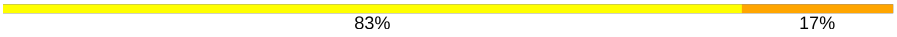


4.2.6 Score per residue for model 6

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

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

Chain B: 

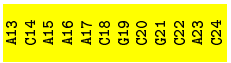
4.2.7 Score per residue for model 7

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

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

Chain B:



4.2.8 Score per residue for model 8

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

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

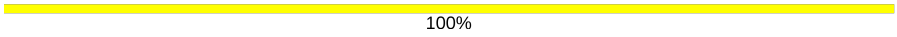
Chain B: 

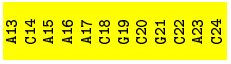
4.2.9 Score per residue for model 9

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

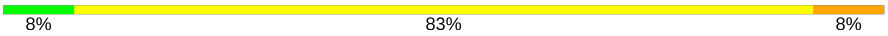
• Molecule 2: 5'-D(*AP*CP*AP*AP*AP*CP*GP*CP*GP*CP*AP*C)-3'

Chain B: 



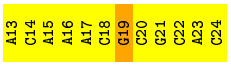
4.2.10 Score per residue for model 10

• Molecule 1: 5'-D(*GP*TP*GP*CP*GP*(CTG)P*GP*TP*TP*TP*GP*T)-3'

Chain A: 

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

Chain B: 



5 Refinement protocol and experimental data overview

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

Of the 10 calculated structures, 10 were deposited, based on the following criterion: *back calculated data agree with experimental NOESY spectrum.*

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

Software name	Classification	Version
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	230
Number of shifts mapped to atoms	230
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	47%

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	4890	2760	2594	-

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

5.2.1 Protein backbone

There are no protein molecules in this entry.

5.2.2 Protein sidechains

There are no protein molecules in this entry.

5.2.3 RNA

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
1	CTG	A	6	1	19,23,24	1.53±0.13	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
1	CTG	A	6	1	21,35,38	1.33±0.14	0±0 (0±0%)

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	CTG	A	6	1	-	0±0,7,45,46	0±0,2,2,2

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 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 47% for the well-defined parts and 47% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

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	230
Number of shifts mapped to atoms	230
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

6.1.2 Chemical shift referencing

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

6.1.3 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 47%, i.e. 214 atoms were assigned a chemical shift out of a possible 457. 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	214/457 (47%)	214/273 (78%)	0/154 (0%)	0/30 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 47%, i.e. 214 atoms were assigned a chemical shift out of a possible 457. 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	214/457 (47%)	214/273 (78%)	0/154 (0%)	0/30 (0%)

6.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

6.1.5 Random Coil Index (RCI) plots

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned_chem_shift_list_1). RCI is only applicable to proteins.