

# Full wwPDB NMR Structure Validation Report (i)

### May 28, 2020 – 08:39 pm BST

PDB ID : 1SSJ

Title : A DNA DUPLEX CONTAINING A CHOLESTEROL ADDUCT (BETA-

ANOMER)

Authors: Gomez-Pinto, I.; Cubero, E.; Kalko, S.G.; Monaco, V.; van der Marel, G.; van

Boom, J.H.; Orozco, M.; Gonzalez, C.

Deposited on : 2004-03-24

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 (1)) 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)

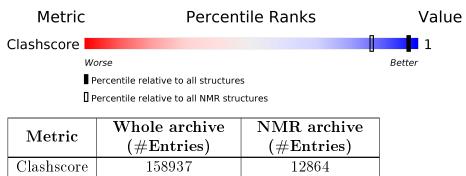
Validation Pipeline (wwPDB-VP) : 2.11

# 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 was not calculated.

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.



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	10	10%	10% 70% 20%				
2	В	10		70%	30%			



# 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 (i)

There are 2 unique types of molecules in this entry. The entry contains 706 atoms, of which 275 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
1	A	10	Total	С	Н	N	О	Р	0
		A   10	386	122	160	38	57	9	U

• Molecule 2 is a DNA chain called 5'-D(GP\*TP\*TP\*CP\*GP\*GP\*GP\*TP\*GP\*G)-3'.

Mol	Chain	Residues	Atoms						Trace
2	В	10	Total	С	Н	N	О	Р	0
		В 10	320	98	115	37	61	9	U

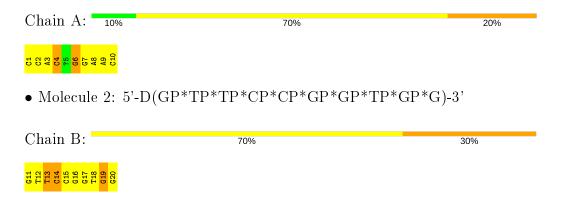


# 4 Residue-property plots (i)

### 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(\*CP\*CP\*AP\*CP\*(HOB)P\*GP\*GP\*AP\*AP\*C)-3'

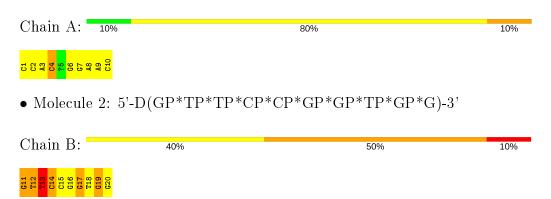


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

### 4.2.1 Score per residue for model 1

• Molecule 1: 5'-D(\*CP\*CP\*AP\*CP\*(HOB)P\*GP\*GP\*AP\*AP\*C)-3'





### 4.2.2 Score per residue for model 2

• Molecule 1: 5'-D(\*CP\*CP\*AP\*CP\*(HOB)P\*GP\*GP\*AP\*AP\*C)-3'

Chain A: 10% 50% 40%

### C1 C2 A3 C4 C4 C6 C6 C6 C6 C7 A8 A8

• Molecule 2: 5'-D(GP\*TP\*TP\*CP\*CP\*GP\*GP\*TP\*GP\*G)-3'

Chain B: 50% 50%

### 611 T12 C14 C15 G16 G16 G17 T18 G19 G20

### 4.2.3 Score per residue for model 3

• Molecule 1: 5'-D(\*CP\*CP\*AP\*CP\*(HOB)P\*GP\*GP\*AP\*AP\*C)-3'

Chain A: 20% 70% 10%

### C1 C2 A3 C4 C4 C4 C6 G6 G7 A8 A8 A9

• Molecule 2: 5'-D(GP\*TP\*TP\*CP\*CP\*GP\*GP\*TP\*GP\*G)-3'

Chain B: 10% 50% 40%

### G11 T12 T13 C14 C15 G16 G17 T18 G20

### 4.2.4 Score per residue for model 4

Chain A: 10% 60% 30%

• Molecule 2: 5'-D(GP\*TP\*TP\*CP\*CP\*GP\*GP\*TP\*GP\*G)-3'

Chain B: 10% 60% 30%

G11 C12 C14 C15 G16 G16 G17 T18 T18 G19 G19



### 4.2.5 Score per residue for model 5

• Molecule 1: 5'-D(\*CP\*CP\*AP\*CP\*(HOB)P\*GP\*GP\*AP\*AP\*C)-3'

Chain A: 10% 50% 40%

C1 C2 A3 C4 C4 C6 G6 G6 G7 A8 A9

• Molecule 2: 5'-D(GP\*TP\*TP\*CP\*CP\*GP\*GP\*TP\*GP\*G)-3'

Chain B: 80% 20%

G11 T12 T13 C14 C15 G16 G17 T18 G19

### 4.2.6 Score per residue for model 6

• Molecule 1: 5'-D(\*CP\*CP\*AP\*CP\*(HOB)P\*GP\*GP\*AP\*AP\*C)-3'

Chain A: 10% 80% 10%

C1 C2 C4 C4 C4 C6 G6 G7 A8 A8

Chain B: 20% 40% 30% 10%

611 T12 C14 C15 G16 G17 T18 G19 G19

### 4.2.7 Score per residue for model 7

Chain A: 10% 50% 40%

C1 C2 A3 C4 C4 C6 C6 C6 C7 A8 A9

• Molecule 2: 5'-D(GP\*TP\*TP\*CP\*CP\*GP\*GP\*TP\*GP\*G)-3'

Chain B: 70% 30%

G11 G16 G16 G17 G10 G10 G10 G10 G10



### 4.2.8 Score per residue for model 8

• Molecule 1: 5'-D(\*CP\*CP\*AP\*CP\*(HOB)P\*GP\*GP\*AP\*AP\*C)-3'

Chain A: 10% 50% 40%

## CC1 CC2 CC4 CC4 CC4 CG6 CG7 CG7 CG7 CG7 CG7 CG7 CG7 CGC

• Molecule 2: 5'-D(GP\*TP\*TP\*CP\*CP\*GP\*GP\*TP\*GP\*G)-3'

Chain B: 60% 40%

### 611 T12 T13 C14 C15 G16 G17 T18 G20

### 4.2.9 Score per residue for model 9

• Molecule 1: 5'-D(\*CP\*CP\*AP\*CP\*(HOB)P\*GP\*GP\*AP\*AP\*C)-3'

Chain A: 20% 50% 30%

### C1 C2 A3 C4 C4 C4 C6 G6 G7 A8 A8 A9

Chain B: 10% 20% 70%

### G11 T12 T13 C14 C15 G16 G17 T18 G19

### 4.2.10 Score per residue for model 10

Chain A: 10% 70% 20%

• Molecule 2: 5'-D(GP\*TP\*TP\*CP\*CP\*GP\*GP\*TP\*GP\*G)-3'

Chain B: 40% 60%

### 611 712 713 713 713 714 616 617 718 619 620



### 5 Refinement protocol and experimental data overview (i)



The models were refined using the following method: relaxation matrix restrained molecular dynamics.

Of the 10 calculated structures, 10 were deposited, based on the following criterion: all calculated structures submitted.

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

Software name	Classification	Version		
DYANA	structure solution	5.1		
AMBER	refinement	5.0		

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

### 5.1Too-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	
2	B 205		115	115	0±0	
All	All	4310	2750	2722	4	

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 1.

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

Atom-1	Atom-2	Clash(Å)	$\mathbf{Distance}(\mathbf{\mathring{A}})$	Models	
Atom-1	Atom-2	Clash(A)	Distance(A)	Worst	Total
2:B:12:DT:C2'	2:B:13:DT:H71	0.50	2.37	2	4

### Torsion angles (i) 5.2

### 5.2.1Protein 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.

### 5.3 Non-standard residues in protein, DNA, RNA chains (i)

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

## 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 (i)

No chemical shift data were provided

