Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:35 pm BST

PDB ID	:	2KQG
Title	:	A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics
Authors	:	Hsu, S.-T.D.; Varnai, P.; Bugaut, A.; Reszka, A.P.; Neidle, S.; Balasubramanian, S.
Deposited on	:	2009-11-05

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:

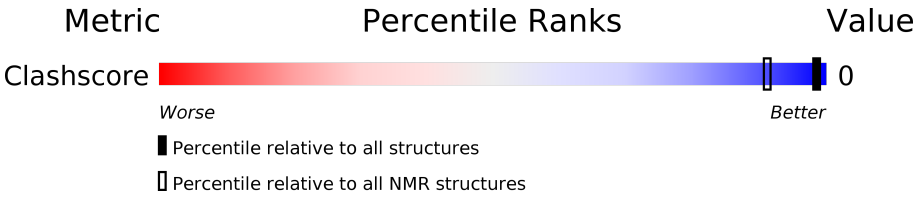
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	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.



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	21	5% 76% 19%

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 678 atoms, of which 234 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	21	676	206	234	94	122	20	0

- Molecule 2 is POTASSIUM ION (three-letter code: K) (formula: K).

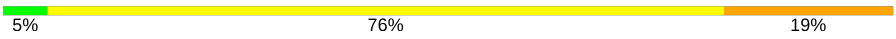
Mol	Chain	Residues	Atoms	
			Total	K
2	A	2	2	2

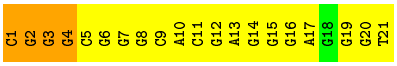
4 Residue-property plots

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: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

Chain A: 



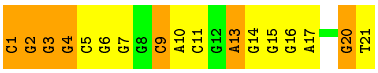
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

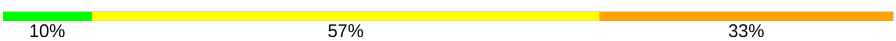
• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

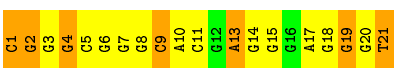
Chain A: 



4.2.2 Score per residue for model 2

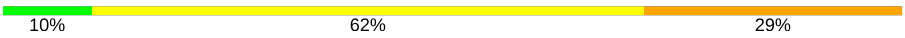
• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

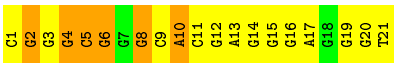
Chain A: 



4.2.3 Score per residue for model 3

• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

Chain A: 



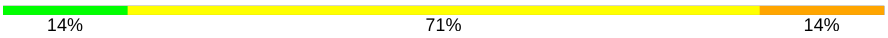
4.2.4 Score per residue for model 4

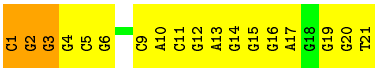
• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

Chain A: 

4.2.5 Score per residue for model 5

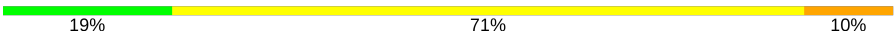
• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

Chain A: 



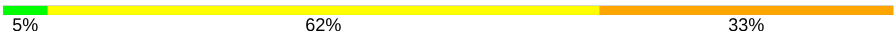
4.2.6 Score per residue for model 6

• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

Chain A: 

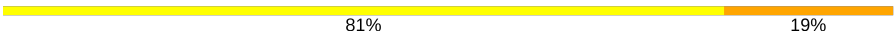
4.2.7 Score per residue for model 7

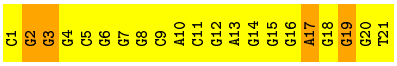
• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

Chain A: 

4.2.8 Score per residue for model 8

• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

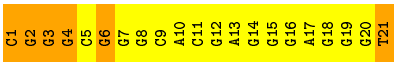
Chain A: 



4.2.9 Score per residue for model 9

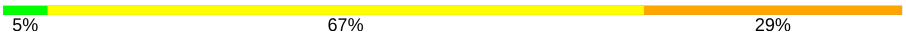
• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

Chain A: 



4.2.10 Score per residue for model 10

• Molecule 1: 5'-D(*CP*GP*GP*GP*CP*GP*GP*GP*CP*AP*CP*GP*AP*GP*GP*GP*AP*GP*GP*GP*T)-3'

Chain A: 

5 Refinement protocol and experimental data overview

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

Of the 100 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
AMBER	structure solution	
AMBER	refinement	

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

5.1 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
1	A	442	234	234	0±0
All	All	4440	2340	2340	1

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:19:DG:N2	1:A:21:DT:H71	0.42	2.30	2	1

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

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

5.4 Carbohydrates

There are no carbohydrates in this entry.

5.5 Ligand geometry

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

5.6 Other polymers

There are no such molecules in this entry.

5.7 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

No chemical shift data were provided