Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:49 pm BST

PDB ID	:	2KV6
Title	:	Tetrapeptide KWKK conjugated to oligonucleotide duplex by a trimethylene tether
Authors	:	Huang, H.; Kozekov, I.D.; Kozekova, A.; Rizzo, C.J.; McCullough, A.; LLoyd, R.S.; Stone, M.P.
Deposited on	:	2010-03-08

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:

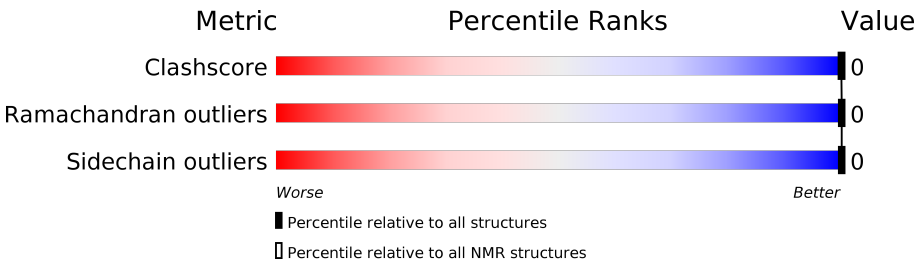
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 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
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	33% 67%
2	B	12	25% 75%
3	C	4	100%

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 858 atoms, of which 327 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	12	387	119	141	46	70	11	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	12	379	116	136	46	70	11	0

- Molecule 3 is a protein called KWKK Tetrapeptide.

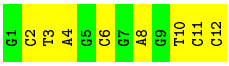
Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
3	C	4	92	29	50	8	5	0

4 Residue-property plots (i)

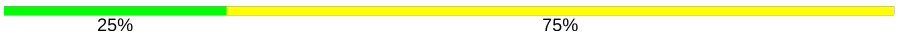
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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.

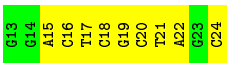
- Molecule 1: 5'-D(*GP*CP*TP*AP*GP*CP*GP*AP*GP*TP*CP*C)-3'

Chain A:



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

Chain B: 



- Molecule 3: KWKK Tetrapeptide

Chain C:

There are no outlier residues in this chain.

5 Refinement protocol and experimental data overview (i)

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

Of the 10 calculated structures, 1 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	7.0

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

5.1 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
All	All	531	327	325	-

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)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	C	2/4 (50%)	1 (50%)	1 (50%)	0 (0%)	100	100
All	All	2/4 (50%)	1 (50%)	1 (50%)	0 (0%)	100	100

There are no Ramachandran outliers.

5.2.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	C	4/4 (100%)	4 (100%)	0 (0%)	100	100
All	All	4/4 (100%)	4 (100%)	0 (0%)	100	100

There are no protein residues with a non-rotameric sidechain to report.

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 (i)

No chemical shift data were provided