Full wwPDB NMR Structure Validation Report (i)

May 31, 2020 – 11:21 am BST

PDB ID	:	6MIE
Title	:	Solution NMR structure of the KCNQ1 voltage-sensing domain
Authors	:	Taylor, K.C.; Kuenze, G.; Smith, J.A.; Meiler, J.; McFeeters, R.L.; Sanders, C.R.
Deposited on	:	2018-09-19

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:

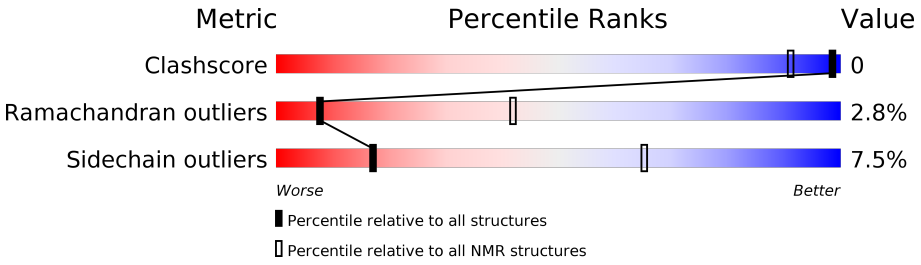
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	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 is 70%.

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	159	74% 12% • 7% 6%

2 Ensemble composition and analysis

This entry contains 10 models. Model 1 is the overall representative, medoid model (most similar to other models).

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:103-A:241 (139)	0.68	1

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 5
2	6, 7
3	9, 10
Single-model clusters	8

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2471 atoms, of which 1263 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Potassium voltage-gated channel subfamily KQT member 1.

Mol	Chain	Residues	Atoms						Trace
1	A	150	Total	C	H	N	O	S	0
			2471	801	1263	210	190	7	

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	91	MET	-	expression tag	UNP P51787
A	92	GLY	-	expression tag	UNP P51787
A	93	HIS	-	expression tag	UNP P51787
A	94	HIS	-	expression tag	UNP P51787
A	95	HIS	-	expression tag	UNP P51787
A	96	HIS	-	expression tag	UNP P51787
A	97	HIS	-	expression tag	UNP P51787
A	98	HIS	-	expression tag	UNP P51787
A	99	GLY	-	expression tag	UNP P51787

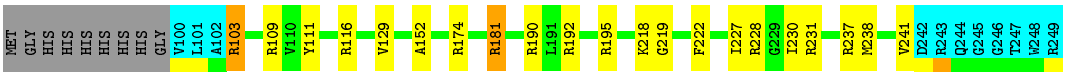
4 Residue-property plots (i)

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: Potassium voltage-gated channel subfamily KQT member 1

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 (medoid)

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

Chain A:

4.2.2 Score per residue for model 2

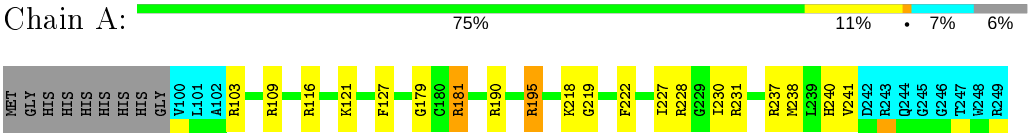
• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

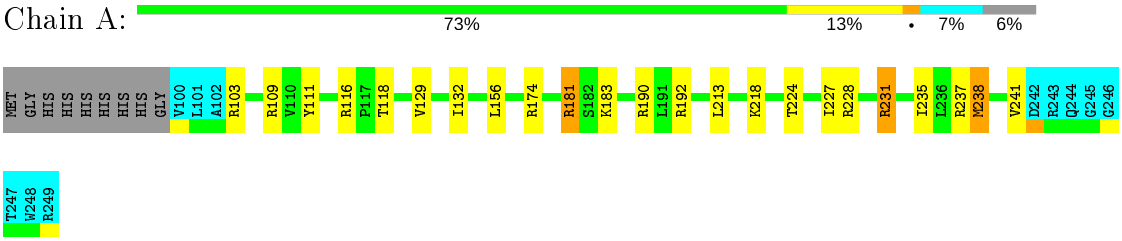
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

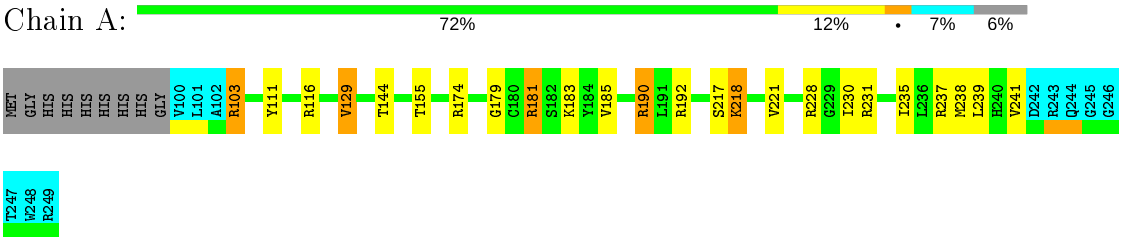
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

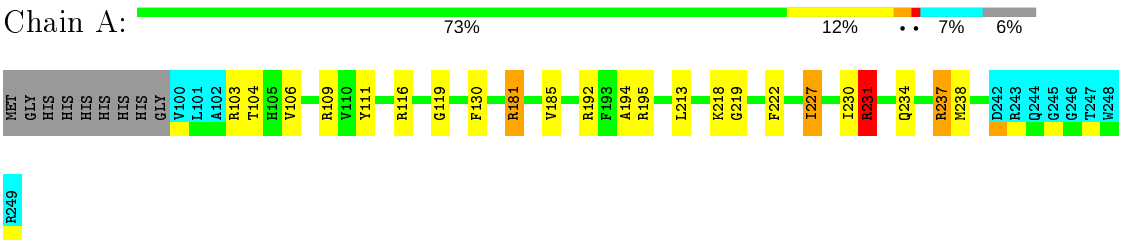
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

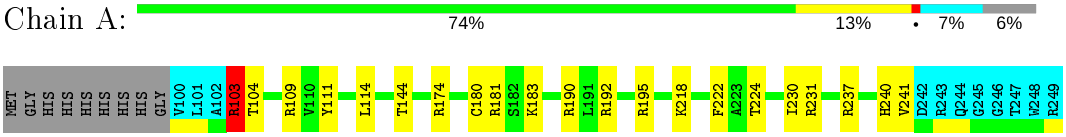
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

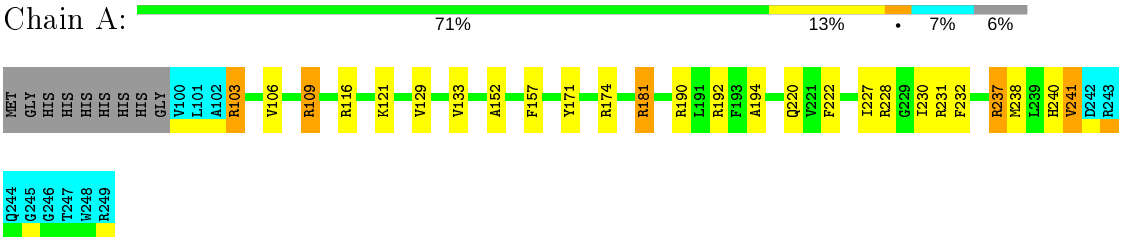
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

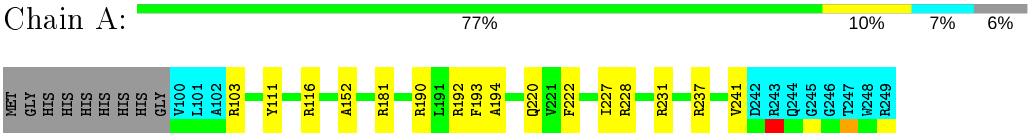
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

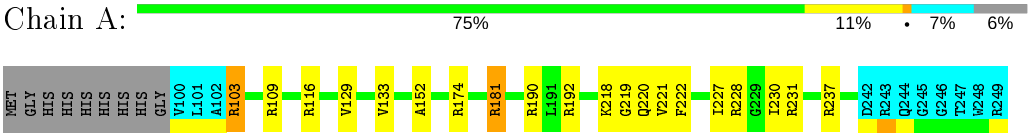
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Potassium voltage-gated channel subfamily KQT member 1

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 150 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
Amber	refinement	16
XPLOR-NIH	structure calculation	2.48

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	1581
Number of shifts mapped to atoms	1581
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	70%

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	1119	1177	1177	0±0
All	All	11190	11770	11770	2

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:219:GLY:H	1:A:221:VAL:H	0.43	1.56	2	1
1:A:103:ARG:HA	1:A:106:VAL:HG12	0.40	1.93	8	1

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	
1	A	139/159 (87%)	127±4 (91±3%)	9±3 (6±2%)	4±2 (3±1%)	8	42
All	All	1390/1590 (87%)	1265 (91%)	86 (6%)	39 (3%)	8	42

All 19 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	227	ILE	8
1	A	238	MET	4
1	A	241	VAL	4
1	A	219	GLY	4
1	A	217	SER	3
1	A	194	ALA	2
1	A	179	GLY	2
1	A	182	SER	1
1	A	218	LYS	1
1	A	180	CYS	1
1	A	104	THR	1
1	A	119	GLY	1
1	A	220	GLN	1
1	A	221	VAL	1
1	A	183	LYS	1
1	A	185	VAL	1
1	A	106	VAL	1
1	A	156	LEU	1
1	A	193	PHE	1

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
1	A	120/135 (89%)	111±3 (92±3%)	9±3 (8±3%)	17 65
All	All	1200/1350 (89%)	1110 (92%)	90 (8%)	17 65

All 36 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	181	ARG	9
1	A	230	ILE	8
1	A	222	PHE	7
1	A	103	ARG	6
1	A	224	THR	4
1	A	129	VAL	4
1	A	231	ARG	4
1	A	235	ILE	4
1	A	213	LEU	3
1	A	218	LYS	3
1	A	133	VAL	3
1	A	183	LYS	3
1	A	104	THR	2
1	A	185	VAL	2
1	A	238	MET	2
1	A	157	PHE	2
1	A	232	PHE	2
1	A	121	LYS	2
1	A	109	ARG	2
1	A	240	HIS	2
1	A	227	ILE	1
1	A	155	THR	1
1	A	144	THR	1
1	A	114	LEU	1
1	A	130	PHE	1
1	A	195	ARG	1
1	A	187	LEU	1
1	A	127	PHE	1
1	A	106	VAL	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	156	LEU	1
1	A	221	VAL	1
1	A	131	LEU	1
1	A	132	ILE	1
1	A	234	GLN	1
1	A	220	GLN	1
1	A	118	THR	1

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

There are no ligands in this entry.

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

The completeness of assignment taking into account all chemical shift lists is 70% for the well-defined parts and 70% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *starch_output*

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	1581
Number of shifts mapped to atoms	1581
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

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	135	-0.38 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	113	-0.07 ± 0.06	None needed (< 0.5 ppm)
$^{13}C'$	129	0.01 ± 0.11	None needed (< 0.5 ppm)
^{15}N	139	0.42 ± 0.13	None needed (< 0.5 ppm)

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 70%, i.e. 1245 atoms were assigned a chemical shift out of a possible 1790. 29 out of 37 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	621/691 (90%)	248/276 (90%)	245/278 (88%)	128/137 (93%)
Sidechain	581/891 (65%)	357/517 (69%)	223/332 (67%)	1/42 (2%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	43/208 (21%)	39/111 (35%)	0/90 (0%)	4/7 (57%)
Overall	1245/1790 (70%)	644/904 (71%)	468/700 (67%)	133/186 (72%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 70%, i.e. 1345 atoms were assigned a chemical shift out of a possible 1931. 31 out of 39 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	673/746 (90%)	270/298 (91%)	264/300 (88%)	139/148 (94%)
Sidechain	626/965 (65%)	385/560 (69%)	240/356 (67%)	1/49 (2%)
Aromatic	46/220 (21%)	41/117 (35%)	0/95 (0%)	5/8 (62%)
Overall	1345/1931 (70%)	696/975 (71%)	504/751 (67%)	145/205 (71%)

6.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

6.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: