wwPDB NMR Structure Validation Summary Report (i)

Aug 17, 2022 – 01:50 PM EDT

PDB ID	:	2MHF
Title	:	Solution structure of the cyclic-nucleotide binding homology domain of a KCNH channel
Authors	:	Li, Q.; Ng, H.
Deposited on	:	2013-11-21

This is a wwPDB NMR Structure Validation Summary 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 types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

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

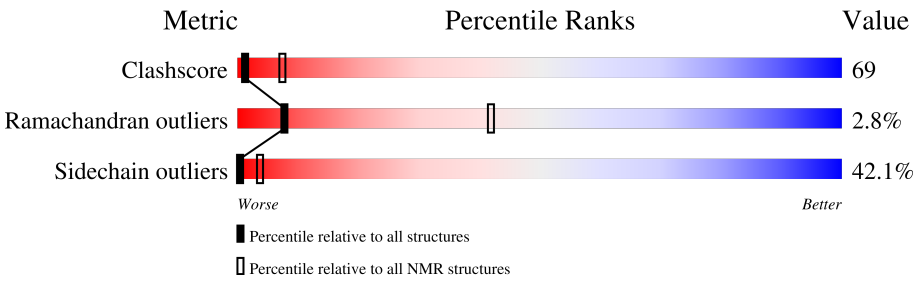
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.29
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 82%.

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	143	11% 53% 24% 6% 6%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:10-A:135 (126)	0.39	15

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 5 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 4, 11, 14, 15, 18, 20
2	2, 3, 5, 7, 8, 9, 16
3	6, 13
4	10, 19
5	12, 17

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2151 atoms, of which 1093 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Uncharacterized protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	135	2151	676	1093	176	200	6	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP F1QQX6
A	136	LEU	-	expression tag	UNP F1QQX6
A	137	GLU	-	expression tag	UNP F1QQX6
A	138	HIS	-	expression tag	UNP F1QQX6
A	139	HIS	-	expression tag	UNP F1QQX6
A	140	HIS	-	expression tag	UNP F1QQX6
A	141	HIS	-	expression tag	UNP F1QQX6
A	142	HIS	-	expression tag	UNP F1QQX6
A	143	HIS	-	expression tag	UNP F1QQX6

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Uncharacterized protein

Chain A: 11% 53% 24% 6% 6%

M1 N2 K3 E4 L5 L6 Q7 L8 P9 L10 F11 E12 S13 A14 S15 R16 G17 C18 L19 R20 S21 L22 S23 L24 I25 I26 K27 T28 S29 F30 C31 A32 P33 G34 E35 F36 L37 I38 R39 Q40 G41 D42 A43 L44 Q45 A46 I47 Y48 F49 V50 C51 S52 G53 S54 M55 E56 V57 L58 K59 D60

N61 T62 V63 L64 A65 I66 L67 G68 K69 G70 D71 L72 I73 G74 S75 D76 S77 L78 T79 K80 E81 Q82 V83 I84 K85 T86 N87 A88 N89 V90 K91 A92 L93 T94 Y95 C96 D97 L98 Q99 Y100 I101 S102 L103 K104 G105 L106 R107 E108 V109 L110 R111 L112 Y113 Y116 A117 Q118 K119 F120 V121

S122 E123 I124 Q125 H126 D127 L128 T129 Y130 N131 L132 R133 E134 G135 LEU GLU HIS HIS HIS HIS HIS HIS

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 15. Colouring as in section 4.1 above.

• Molecule 1: Uncharacterized protein

Chain A: 17% 41% 28% • 6% 6%

M1 N2 K3 E4 L5 L6 Q7 L8 P9 L10 F11 E12 S13 A14 S15 R16 G17 C18 L19 R20 S21 L22 S23 L24 I25 I26 K27 T28 S29 F30 C31 A32 P33 G34 E35 F36 L37 I38 R39 Q40 G41 D42 A43 L44 Q45 A46 I47 Y48 F49 V50 C51 S52 G53 S54 M55 E56 V57 L58 K59 D60

N61 T62 V63 L64 A65 I66 L67 G70 D71 L72 I73 G74 S75 D76 S77 L78 T79 K80 E81 Q82 V83 I84 K85 T86 N87 A88 N89 V90 K91 A92 L93 T94 Y95 C96 D97 L98 Q99 Y100 I101 S102 L103 K104 G105 L106 R107 E108 V109 L110 R111 L112 Y113 P114 E115 Y116 A117 Q118 K119 F120 V121

S122 E123 I124 Q125 H126 D127 L128 T129 Y130 N131 L132 G135 LEU GLU HIS HIS HIS HIS HIS HIS

5 Refinement protocol and experimental data overview (i)

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

Of the 60 calculated structures, 20 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
CYANA	structure solution	
X-PLOR-NIH	structure solution	
X-PLOR-NIH	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1524
Number of shifts mapped to atoms	1524
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

6 Model quality (i)

6.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.63±0.00	0±0/997 (0.0± 0.0%)	0.80±0.01	0±0/1343 (0.0± 0.0%)
All	All	0.63	0/19940 (0.0%)	0.80	4/26860 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	48	TYR	CB-CG-CD2	-6.91	116.85	121.00	10	4

There are no chirality outliers.

There are no planarity outliers.

6.2 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
1	A	984	1011	1010	137±9
All	All	19680	20220	20200	2739

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 69.

5 of 675 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:11:PHE:CD2	1:A:19:LEU:HD13	0.98	1.94	6	11

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:96:CYS:SG	1:A:98:LEU:HD21	0.98	1.98	2	1
1:A:73:ILE:HD12	1:A:74:GLY:N	0.95	1.74	7	11
1:A:79:THR:OG1	1:A:121:VAL:HG11	0.94	1.62	8	3
1:A:22:LEU:HD12	1:A:109:VAL:HG21	0.92	1.42	14	20

6.3 Torsion angles (i)

6.3.1 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	125/143 (87%)	116±2 (93±1%)	6±2 (5±1%)	3±1 (3±1%)	8	42
All	All	2500/2860 (87%)	2314 (93%)	117 (5%)	69 (3%)	8	42

5 of 8 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	40	GLN	20
1	A	81	GLU	17
1	A	76	ASP	7
1	A	71	ASP	7
1	A	78	LEU	6

6.3.2 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	109/126 (87%)	63±3 (58±2%)	46±3 (42±2%)	0	4
All	All	2180/2520 (87%)	1263 (58%)	917 (42%)	0	4

5 of 96 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	10	LEU	20
1	A	18	CYS	20
1	A	24	LEU	20
1	A	25	ILE	20
1	A	38	ILE	20

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 82% for the well-defined parts and 80% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

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	1524
Number of shifts mapped to atoms	1524
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

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	133	-0.27 ± 0.08	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	123	-0.05 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	129	0.10 ± 0.12	None needed (< 0.5 ppm)
^{15}N	130	0.42 ± 0.50	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 82%, i.e. 1277 atoms were assigned a chemical shift out of a possible 1548. 18 out of 26 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	623/626 (100%)	250/250 (100%)	249/252 (99%)	124/124 (100%)
Sidechain	625/821 (76%)	361/478 (76%)	258/307 (84%)	6/36 (17%)

Continued on next page...

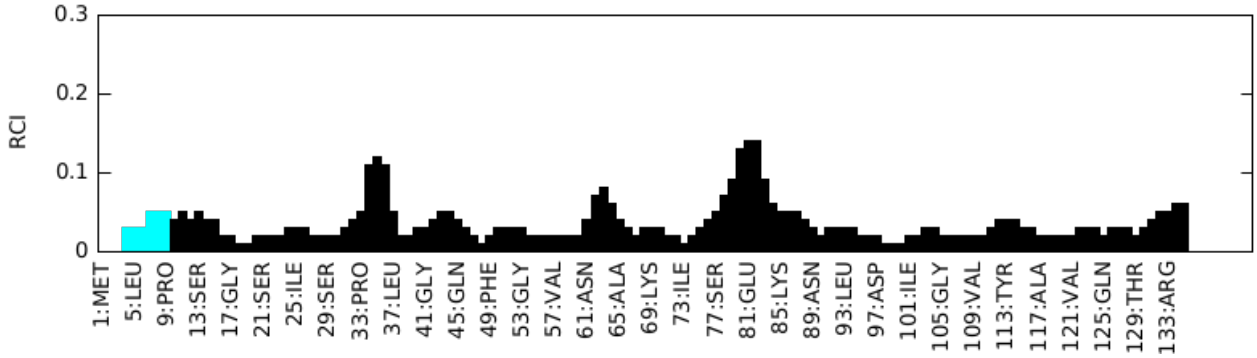
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	29/101 (29%)	18/53 (34%)	11/46 (24%)	0/2 (0%)
Overall	1277/1548 (82%)	629/781 (81%)	518/605 (86%)	130/162 (80%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 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: