wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 12:25 AM BST

PDB ID	:	2L4R
Title	:	NMR solution structure of the N-terminal PAS domain of hERG
Authors	:	Gayen, N.; Li, Q.; Chen, A.S.; Huang, Q.; Raida, M.; Kang, C.
Deposited on	:	2010-10-13

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
http://wwpdb.org/validation/2016/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:

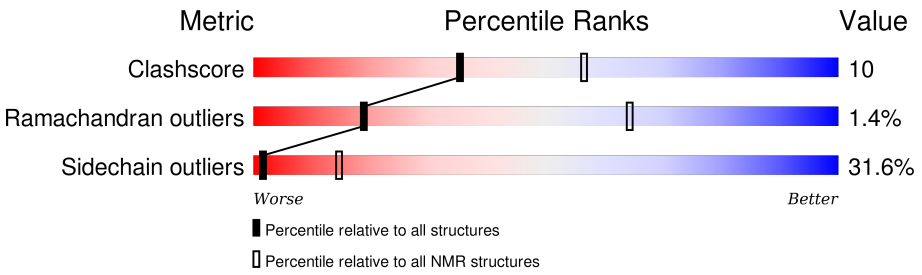
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	135	46% 32% 22%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 2 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:27-A:34, A:39-A:135 (105)	0.52	2

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, 5, 7, 9, 10, 12, 13, 14, 15, 16, 18, 20
2	4, 8, 11, 19
3	3, 6
Single-model clusters	17

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2137 atoms, of which 1071 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Potassium voltage-gated channel subfamily H member 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	135	2137	672	1071	193	189	12	0

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 H member 2

Chain A: 46% 32% 22%

M1 P2 V3 R4 R5 G6 H7 V8 A9 P10 Q11 N12 T13 F14 L15 D16 T17 I18 I19 R20 K21 F22 E23 G24 Q25 S26 R27 K28 F29 I30 I31 A32 N33 A34 R35 V36 E37 N38 C39 A40 V41 I42 Y43 E50 L51 Y54 S55 Q61 R62 P63 C64 T65 L69 R73

R76 R77 A78 Q81 I82 L86 E91 R92 K93 V94 E95 I96 S104 C105 F106 L109 V110 N117 E118 D119 I123 M124 F125 I126 L127 N128 F129 M133 E134 K135

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

The representative model is number 2. Colouring as in section 4.1 above.

- Molecule 1: Potassium voltage-gated channel subfamily H member 2

Chain A: 43% 33% • 22%

M1 P2 V3 R4 R5 G6 H7 V8 A9 P10 Q11 N12 T13 F14 L15 D16 T17 I18 I19 R20 K21 F22 E23 G24 Q25 S26 R27 K28 F29 I30 I31 A32 N33 A34 R35 V36 E37 N38 V41 I42 Y43 C44 F48 C49 E50 Y54 S55 R56 V59 M60 Q61 C64 T65 F68

L69 R73 R77 A78 Q81 I82 L86 K93 V94 E95 I96 A97 F98 K101 S104 C105 F106 L107 C108 L109 V110 N117 E118 D119 G120 A121 M124 F125 I126 L127 E130 M133 E134 K135

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *TORSION ANGLE DYNAMICS*.

Of the 100 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	2.1
CYANA	refinement	2.1

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

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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	821	824	824	17±3
All	All	16420	16480	16480	343

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 10.

5 of 156 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:65:THR:HG23	1:A:82:ILE:HG21	0.91	1.41	9	6
1:A:110:VAL:HG13	1:A:127:LEU:HD13	0.86	1.46	14	1
1:A:29:PHE:CE2	1:A:126:ILE:HD11	0.72	2.19	7	6
1:A:78:ALA:HB1	1:A:96:ILE:CG1	0.70	2.16	11	19
1:A:31:ILE:O	1:A:41:VAL:HG22	0.69	1.86	7	3

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	104/135 (77%)	95±1 (92±1%)	7±1 (7±1%)	2±1 (1±1%)	19 64
All	All	2080/2700 (77%)	1907 (92%)	143 (7%)	30 (1%)	19 64

All 5 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	61	GLN	17
1	A	27	ARG	9
1	A	91	GLU	2
1	A	34	ALA	1
1	A	54	TYR	1

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	87/114 (76%)	60±3 (68±4%)	27±3 (32±4%)	1 14
All	All	1740/2280 (76%)	1191 (68%)	549 (32%)	1 14

5 of 68 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	133	MET	20
1	A	86	LEU	20
1	A	104	SER	20
1	A	119	ASP	20
1	A	50	GLU	19

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 carbohydrates 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)

No chemical shift data were provided