wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 01:27 AM BST

PDB ID	:	2LOO
Title	:	Backbone structure of human membrane protein TMEM14A from NOE data
Authors	:	Eichmann, C.; Klammt, C.; Maslennikov, I.; Riek, R.; Choe, S.
Deposited on	:	2012-01-26

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:

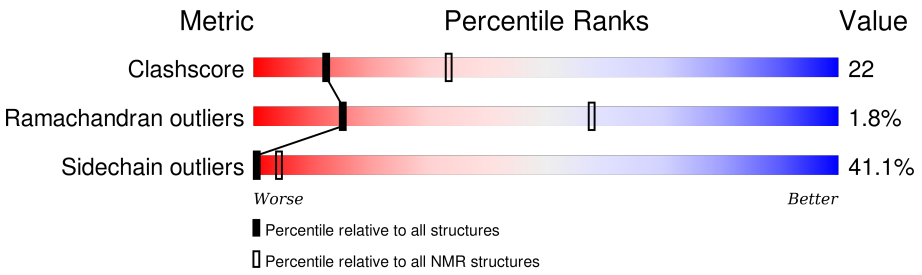
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 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	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	108	16% 32% • 40% 8%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 5 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:8-A:18 (11)	0.06	2
2	A:30-A:40, A:52-A:68, A:80-A:96 (45)	0.47	5

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 4 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1569 atoms, of which 817 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transmembrane protein 14A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	99	1569	500	817	127	119	6	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-8	MET	-	EXPRESSION TAG	UNP Q9Y6G1
A	-7	THR	-	EXPRESSION TAG	UNP Q9Y6G1
A	-6	SER	-	EXPRESSION TAG	UNP Q9Y6G1
A	-5	LEU	-	EXPRESSION TAG	UNP Q9Y6G1
A	-4	TYR	-	EXPRESSION TAG	UNP Q9Y6G1
A	-3	LYS	-	EXPRESSION TAG	UNP Q9Y6G1
A	-2	LYS	-	EXPRESSION TAG	UNP Q9Y6G1
A	-1	VAL	-	EXPRESSION TAG	UNP Q9Y6G1
A	0	GLY	-	EXPRESSION TAG	UNP Q9Y6G1

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: Transmembrane protein 14A

Chain A: 16% 32% • 40% 8%

MET THR SER LEU TYR LYS LYS VAL GLY M1 D2 L3 I4 G5 F6 G7 Y8 A9 A10 L11 V12 S16 G19 Y20 K21 R22 R23 G24 G25 V26 P27 S28 L29 I30 A31 G32 L33 F34 V35 G36 C37 L38 A39 G40 Y41 G42 A43 Y44 R45 V46 S47 N48 D49 K50 R51 D52 V53 K54

V55 S56 L57 F58 T59 A60 F61 F62 L63 A64 M67 G68 V69 R70 F71 K72 R73 S74 K75 K76 I77 M78 P79 A80 G81 L82 V83 A84 G85 L86 S87 L88 M89 M90 I91 L92 R93 L94 V95 L96 L97 L98 L99

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

The representative model is number 5. Colouring as in section 4.1 above.

- Molecule 1: Transmembrane protein 14A

Chain A: 27% 20% 5% 40% 8%

MET THR SER LEU TYR LYS LYS VAL GLY M1 D2 L3 I4 G5 F6 G7 Y8 V12 S16 G19 Y20 K21 R22 R23 G24 G25 V26 P27 S28 L29 I30 A31 G32 L33 F34 V35 G36 C37 L38 A39 G40 Y41 G42 A43 Y44 R45 V46 S47 N48 D49 K50 R51 K54 V55 S56 L57

F58 T59 A60 F61 F62 L63 A64 G68 V69 R70 F71 K72 R73 S74 K75 K76 I77 M78 P79 L82 V83 L86 S87 L88 L92 R93 L94 V95 L96 L97 L98 L99

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CYANA	structure solution	
CYANA	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)	2loo_cs.str
Number of chemical shift lists	1
Total number of shifts	819
Number of shifts mapped to atoms	819
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.

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	405	441	441	19±4
All	All	8100	8820	8820	374

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 22.

5 of 136 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:91:ILE:O	1:A:95:VAL:HG23	0.88	1.66	4	17
1:A:35:VAL:HG22	1:A:59:THR:CG2	0.88	1.99	13	1
1:A:30:ILE:HD13	1:A:30:ILE:O	0.87	1.70	16	3
1:A:52:ASP:O	1:A:55:VAL:HG12	0.82	1.75	2	3
1:A:60:ALA:HA	1:A:82:LEU:HD12	0.79	1.54	14	5

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	56/108 (52%)	45±2 (80±4%)	10±2 (18±4%)	1±1 (2±1%)	15 58
All	All	1120/2160 (52%)	900 (80%)	200 (18%)	20 (2%)	15 58

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	82	LEU	10
1	A	52	ASP	4
1	A	95	VAL	3
1	A	66	ILE	2
1	A	68	GLY	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	41/85 (48%)	24±2 (59±4%)	17±2 (41±4%)	0 4
All	All	820/1700 (48%)	483 (59%)	337 (41%)	0 4

5 of 32 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	86	LEU	20
1	A	92	LEU	20
1	A	16	SER	20
1	A	33	LEU	19
1	A	96	LEU	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)

The completeness of assignment taking into account all chemical shift lists is 70% for the well-defined parts and 61% for the entire structure.

7.1 Chemical shift list 1

File name: 2loo_cs.str

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	819
Number of shifts mapped to atoms	819
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
$^{13}C_{\alpha}$	98	0.04 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	83	0.65 ± 0.06	Should be applied
$^{13}C'$	81	2.18 ± 0.06	Should be applied
^{15}N	96	0.12 ± 0.19	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 70%, i.e. 444 atoms were assigned a chemical shift out of a possible 633. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	278/280 (99%)	112/112 (100%)	110/112 (98%)	56/56 (100%)
Sidechain	164/291 (56%)	63/165 (38%)	101/122 (83%)	0/4 (0%)

Continued on next page...

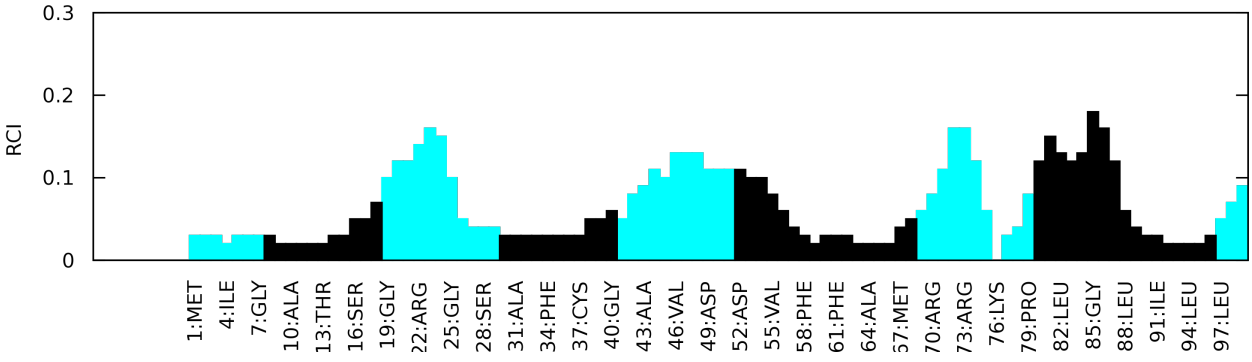
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	2/62 (3%)	2/34 (6%)	0/28 (0%)	0/0 (—%)
Overall	444/633 (70%)	177/311 (57%)	211/262 (81%)	56/60 (93%)

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: