wwPDB NMR Structure Validation Summary Report ⓘ

Nov 18, 2020 – 07:41 AM EST

PDB ID	:	6QWR
Title	:	Solid-state NMR structure of outer membrane protein AlkL in DMPC lipid bilayers
Authors	:	Schubeis, T.; Andreas, L.B.; Pintacuda, G.
Deposited on	:	2019-03-06

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

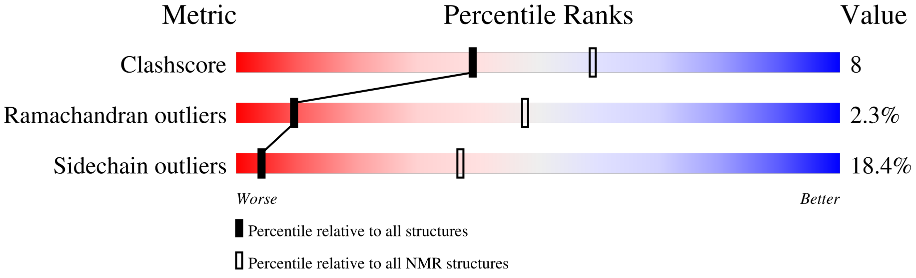
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.15.dev9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.15.dev9

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
SOLID-STATE NMR

The overall completeness of chemical shifts assignment is 67%.

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	219	

2 Ensemble composition and analysis ⓘ

This entry contains 20 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:14-A:33, A:45-A:108, A:116-A:179, A:184-A:205 (170)	0.55	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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3202 atoms, of which 1574 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Outer membrane protein AlkL.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	211	3202	1056	1574	258	312	2	0

There are 16 discrepancies between the modelled and reference sequences:

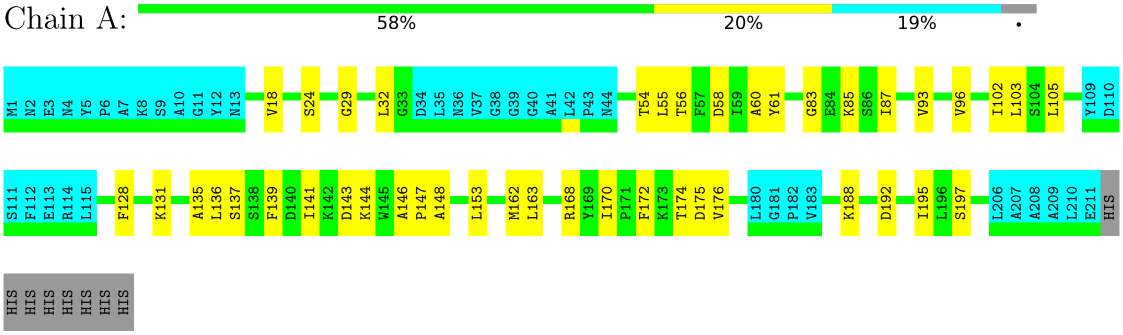
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP Q00595
A	205	LYS	-	expression tag	UNP Q00595
A	206	LEU	-	expression tag	UNP Q00595
A	207	ALA	-	expression tag	UNP Q00595
A	208	ALA	-	expression tag	UNP Q00595
A	209	ALA	-	expression tag	UNP Q00595
A	210	LEU	-	expression tag	UNP Q00595
A	211	GLU	-	expression tag	UNP Q00595
A	212	HIS	-	expression tag	UNP Q00595
A	213	HIS	-	expression tag	UNP Q00595
A	214	HIS	-	expression tag	UNP Q00595
A	215	HIS	-	expression tag	UNP Q00595
A	216	HIS	-	expression tag	UNP Q00595
A	217	HIS	-	expression tag	UNP Q00595
A	218	HIS	-	expression tag	UNP Q00595
A	219	HIS	-	expression tag	UNP Q00595

4 Residue-property plots ⓘ

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.

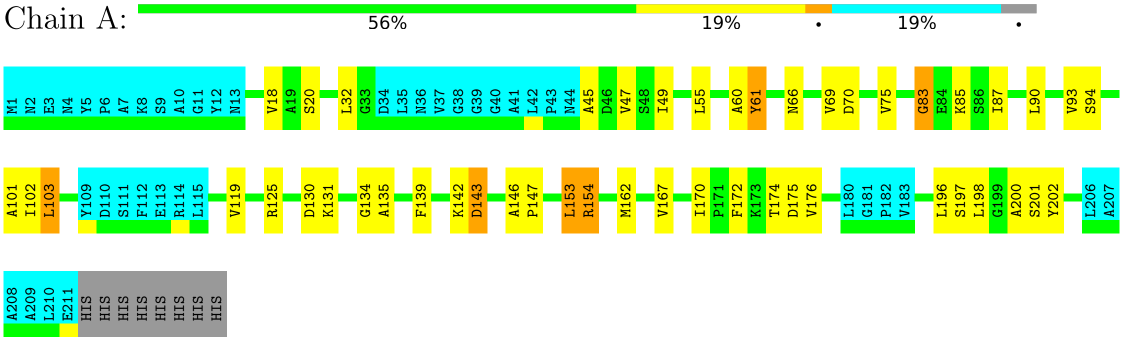
- Molecule 1: Outer membrane protein AlkL



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

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: Outer membrane protein AlkL



5 Refinement protocol and experimental data overview (i)

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

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 calculation	

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	1567
Number of shifts mapped to atoms	1567
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	67%

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

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	1325	1287	1287	20±5
All	All	26500	25740	25740	406

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 8.

5 of 245 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	**Atom-2**	**Clash(Å)**	**Distance(Å)**	**Models**	
				Worst	Total
1:A:87:ILE:HG21	1:A:135:ALA:HB2	0.78	1.52	1	9
1:A:146:ALA:HB1	1:A:147:PRO:HD2	0.78	1.54	14	20
1:A:87:ILE:HG22	1:A:93:VAL:HG22	0.78	1.56	9	18
1:A:141:ILE:HD13	1:A:176:VAL:HG13	0.76	1.57	12	1
1:A:28:VAL:HG23	1:A:191:VAL:HG11	0.75	1.57	16	2

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	170/219 (78%)	146±2 (86±1%)	20±2 (12±1%)	4±1 (2±1%)	9	48
All	All	3400/4380 (78%)	2925 (86%)	398 (12%)	77 (2%)	9	48

5 of 15 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	83	GLY	20
1	A	29	GLY	12
1	A	137	SER	11
1	A	143	ASP	8
1	A	134	GLY	7

6.3.2 Protein sidechains ⓘ

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	143/181 (79%)	117±3 (82±2%)	26±3 (18±2%)	4	37
All	All	2860/3620 (79%)	2334 (82%)	526 (18%)	4	37

5 of 97 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	175	ASP	18
1	A	143	ASP	17
1	A	139	PHE	17
1	A	103	LEU	17
1	A	105	LEU	16

6.3.3 RNA ⓘ

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 ⓘ

The completeness of assignment taking into account all chemical shift lists is 67% for the well-defined parts and 57% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_output*

7.1.1 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	1567
Number of shifts mapped to atoms	1567
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	11

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	173	0.40 ± 0.07	None needed (< 0.5 ppm)
$^{13}C_\beta$	146	-0.60 ± 0.14	Should be applied
$^{13}C'$	167	0.87 ± 0.06	Should be applied
^{15}N	166	-0.04 ± 0.34	None needed (< 0.5 ppm)

7.1.3 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 67%, i.e. 1343 atoms were assigned a chemical shift out of a possible 2016. 16 out of 32 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	774/836 (93%)	305/333 (92%)	316/340 (93%)	153/163 (94%)
Sidechain	566/939 (60%)	328/546 (60%)	233/360 (65%)	5/33 (15%)

Continued on next page...

Continued from previous page...

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Aromatic	3/241 (1%)	2/128 (2%)	0/109 (0%)	1/4 (25%)
Overall	1343/2016 (67%)	635/1007 (63%)	549/809 (68%)	159/200 (80%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

7.1.4 Statistically unusual chemical shifts (i)

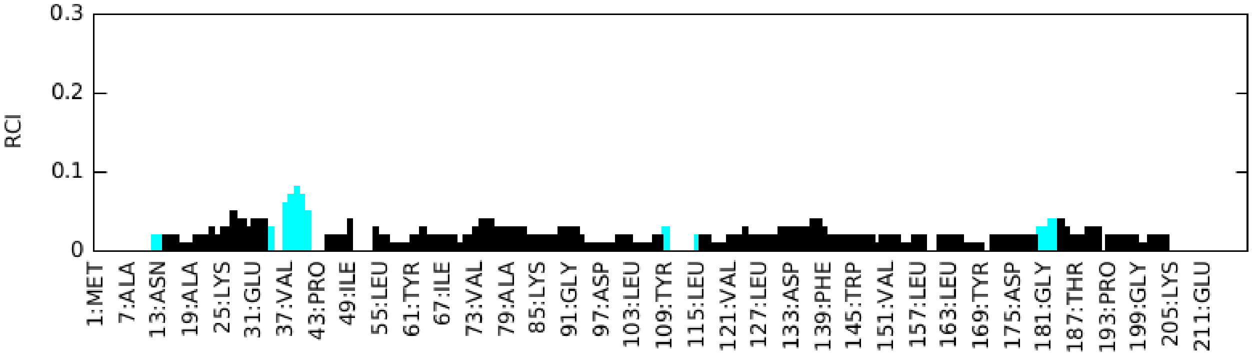
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	**Chain**	**Res**	**Type**	**Atom**	**Shift, *ppm***	**Expected range, *ppm***	**Z-score**
1	A	168	ARG	HD2	1.32	4.27 – 1.97	-7.8
1	A	68	ALA	HB1	-0.56	2.61 – 0.11	-7.7
1	A	68	ALA	HB3	-0.56	2.61 – 0.11	-7.7
1	A	68	ALA	HB2	-0.56	2.61 – 0.11	-7.7
1	A	76	PRO	HD2	1.15	5.45 – 1.85	-7.0
1	A	193	PRO	HD2	1.15	5.45 – 1.85	-6.9
1	A	188	LYS	HD3	2.88	2.75 – 0.45	5.6
1	A	94	SER	HB2	2.45	5.18 – 2.58	-5.5
1	A	191	VAL	HB	0.23	3.59 – 0.39	-5.5
1	A	168	ARG	NE	93.19	92.63 – 76.73	5.4
1	A	188	LYS	HD2	2.76	2.76 – 0.46	5.0

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:


RCI
0.3
0.2
0.1
0
1:MET
7:ALA
13:ASN
19:ALA
25:LYS
31:GLU
37:VAL
43:PRO
49:ILE
55:LEU
61:TYR
67:ILE
73:VAL
79:ALA
85:LYS
91:GLY
97:ASP
103:LEU
109:TYR
115:LEU
121:VAL
127:LEU
133:ASP
139:PHE
145:TRP
151:VAL
157:LEU
163:LEU
169:TYR
175:ASP
181:GLY
187:THR
193:PRO
199:GLY
205:LYS
211:GLU