wwPDB NMR Structure Validation Summary Report (i)

May 29, 2020 – 09:12 am BST

PDB ID	:	5UK6
Title	:	Structure of Anabaena Sensory Rhodopsin Determined by Solid State NMR Spectroscopy and DEER
Authors	:	Milikisiyants, S.; Wang, S.; Munro, R.A.; Donohue, M.; Ward, M.E.; Brown, L.S.; Smirnova, T.I.; Ladizhansky, V.; Smirnov, A.I.
Deposited on	:	2017-01-20

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

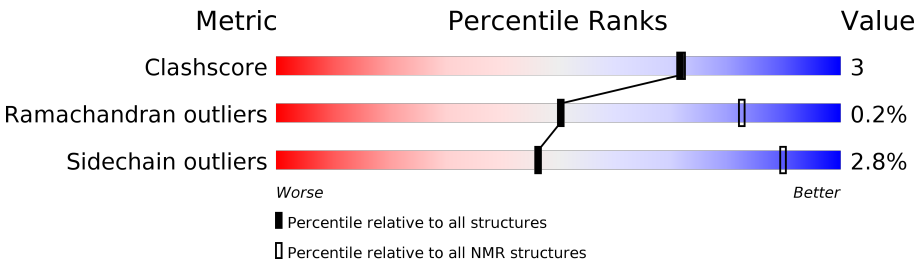
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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:
SOLID-STATE NMR

The overall completeness of chemical shifts assignment is 13%.

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	235	73% 8% 17% •
1	B	235	76% 5% 17% •
1	C	235	72% 7% 18% •

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 3 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:6-A:148, A:164-A:185, A:197-A:209, A:211-A:221, B:6-B:148, B:164-B:186, B:197-B:209, B:211-B:221, C:6-C:29, C:33-C:148, C:164-C:185, C:196-C:209, C:211-C:221 (566)	0.60	3

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 11223 atoms, of which 5559 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Bacteriorhodopsin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	229	3741	1268	1853	298	311	11	0
1	B	229	3741	1268	1853	298	311	11	0
1	C	229	3741	1268	1853	298	311	11	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	230	HIS	-	expression tag	UNP Q8YSC4
A	231	HIS	-	expression tag	UNP Q8YSC4
A	232	HIS	-	expression tag	UNP Q8YSC4
A	233	HIS	-	expression tag	UNP Q8YSC4
A	234	HIS	-	expression tag	UNP Q8YSC4
A	235	HIS	-	expression tag	UNP Q8YSC4
B	230	HIS	-	expression tag	UNP Q8YSC4
B	231	HIS	-	expression tag	UNP Q8YSC4
B	232	HIS	-	expression tag	UNP Q8YSC4
B	233	HIS	-	expression tag	UNP Q8YSC4
B	234	HIS	-	expression tag	UNP Q8YSC4
B	235	HIS	-	expression tag	UNP Q8YSC4
C	230	HIS	-	expression tag	UNP Q8YSC4
C	231	HIS	-	expression tag	UNP Q8YSC4
C	232	HIS	-	expression tag	UNP Q8YSC4
C	233	HIS	-	expression tag	UNP Q8YSC4
C	234	HIS	-	expression tag	UNP Q8YSC4
C	235	HIS	-	expression tag	UNP Q8YSC4

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

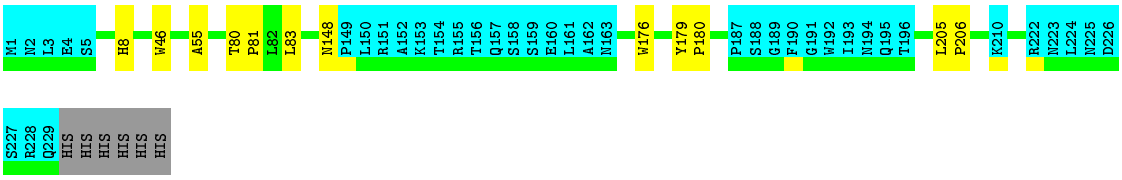
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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: Bacteriorhodopsin

Chain A:

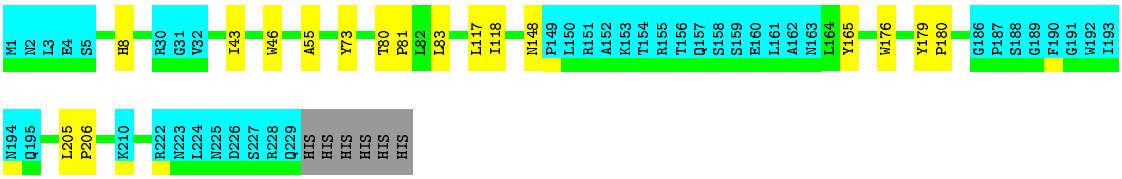
- Molecule 1: Bacteriorhodopsin

Chain B:



- Molecule 1: Bacteriorhodopsin

Chain C:

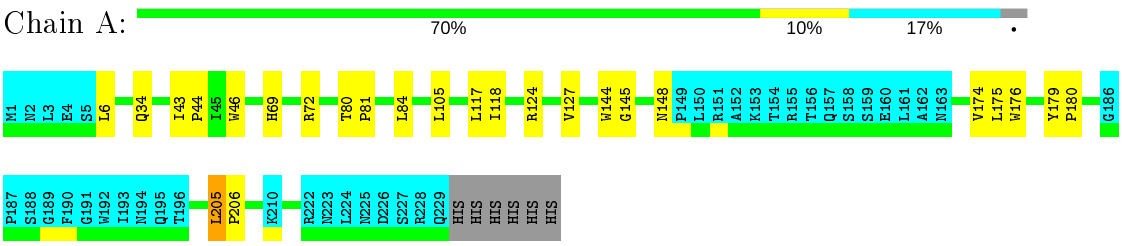


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 3. Colouring as in section 4.1 above.

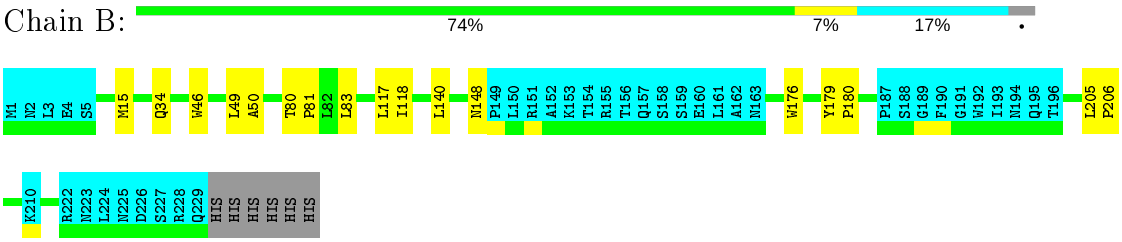
- Molecule 1: Bacteriorhodopsin

Chain A:



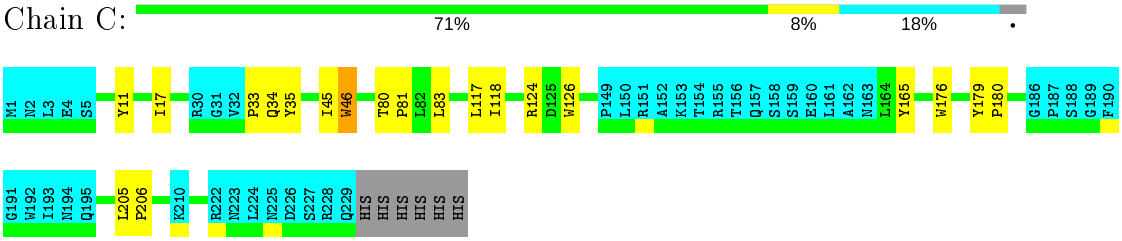
- Molecule 1: Bacteriorhodopsin

Chain B:



- Molecule 1: Bacteriorhodopsin

Chain C:



5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 10 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
TopSpin	refinement	3.0
CNS	structure calculation	

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	2
Total number of shifts	1185
Number of shifts mapped to atoms	1185
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	13%

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.

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	1553	1515	1563	12±4
1	B	1557	1518	1566	9±3
1	C	1538	1497	1545	11±4
All	All	46480	45300	46740	325

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 3.

5 of 171 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:145:GLY:HA2	1:A:148:ASN:ND2	1.27	1.44	2	3
1:C:145:GLY:HA2	1:C:148:ASN:ND2	1.23	1.48	10	1
1:B:145:GLY:HA2	1:B:148:ASN:ND2	1.22	1.45	10	2
1:B:145:GLY:HA2	1:B:148:ASN:HD21	1.01	1.14	8	2
1:B:145:GLY:CA	1:B:148:ASN:ND2	1.00	2.24	10	2

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	189/235 (80%)	179±3 (95±1%)	9±2 (5±1%)	1±1 (0±0%)	44	80
1	B	190/235 (81%)	178±4 (94±2%)	11±3 (6±2%)	0±0 (0±0%)	54	85
1	C	187/235 (80%)	177±2 (95±1%)	9±2 (5±1%)	1±1 (0±0%)	44	80
All	All	5660/7050 (80%)	5351 (95%)	295 (5%)	14 (0%)	50	82

5 of 11 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	C	29	PRO	2
1	B	29	PRO	2
1	A	29	PRO	2
1	A	184	ILE	1
1	A	98	ASP	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	163/203 (80%)	159±2 (97±1%)	4±2 (3±1%)	50	91
1	B	163/203 (80%)	159±2 (97±1%)	5±2 (3±1%)	46	90
1	C	162/203 (80%)	157±2 (97±1%)	5±2 (3±1%)	43	88
All	All	4880/6090 (80%)	4744 (97%)	136 (3%)	46	90

5 of 55 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	C	165	TYR	8
1	C	176	TRP	8
1	C	83	LEU	7
1	B	46	TRP	7
1	B	176	TRP	7

5.2.3 RNA (i)

There are no RNA molecules in this entry.

MODRES-GEOMETRY INFOmissingINFO

5.3 Carbohydrates (i)

There are no carbohydrates in this entry.

5.4 Ligand geometry (i)

There are no ligands in this entry.

5.5 Other polymers (i)

There are no such molecules in this entry.

5.6 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 13% for the well-defined parts and 12% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 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	1142
Number of shifts mapped to atoms	1142
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

6.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	205	-1.35 ± 0.19	Should be applied
$^{13}C_{\beta}$	176	0.53 ± 0.21	Should be applied
$^{13}C'$	203	-0.44 ± 0.14	None needed (< 0.5 ppm)
^{15}N	205	1.39 ± 0.27	Should be applied

6.1.3 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 13%, i.e. 964 atoms were assigned a chemical shift out of a possible 7193. 0 out of 119 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	539/2794 (19%)	0/1114 (0%)	363/1132 (32%)	176/548 (32%)
Sidechain	362/3283 (11%)	0/1908 (0%)	357/1300 (27%)	5/75 (7%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	63/1116 (6%)	0/576 (0%)	59/480 (12%)	4/60 (7%)
Overall	964/7193 (13%)	0/3598 (0%)	779/2912 (27%)	185/683 (27%)

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.

6.1.4 Statistically unusual chemical shifts (i)

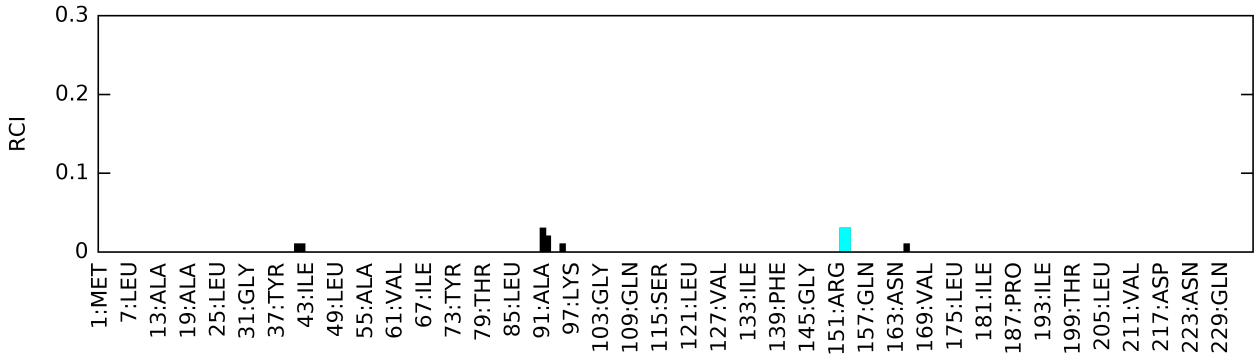
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	109	GLN	NE2	98.30	120.91 – 102.81	-7.5
1	B	128	ARG	CD	37.40	47.57 – 38.77	-6.6

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 B:



6.2 Chemical shift list 2

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1_dup*

6.2.1 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	43
Number of shifts mapped to atoms	43
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

6.2.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

6.2.3 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 0%, i.e. 34 atoms were assigned a chemical shift out of a possible 7193. 0 out of 119 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	18/2794 (1%)	0/1114 (0%)	12/1132 (1%)	6/548 (1%)
Sidechain	12/3283 (0%)	0/1908 (0%)	12/1300 (1%)	0/75 (0%)
Aromatic	4/1116 (0%)	0/576 (0%)	4/480 (1%)	0/60 (0%)
Overall	34/7193 (0%)	0/3598 (0%)	28/2912 (1%)	6/683 (1%)

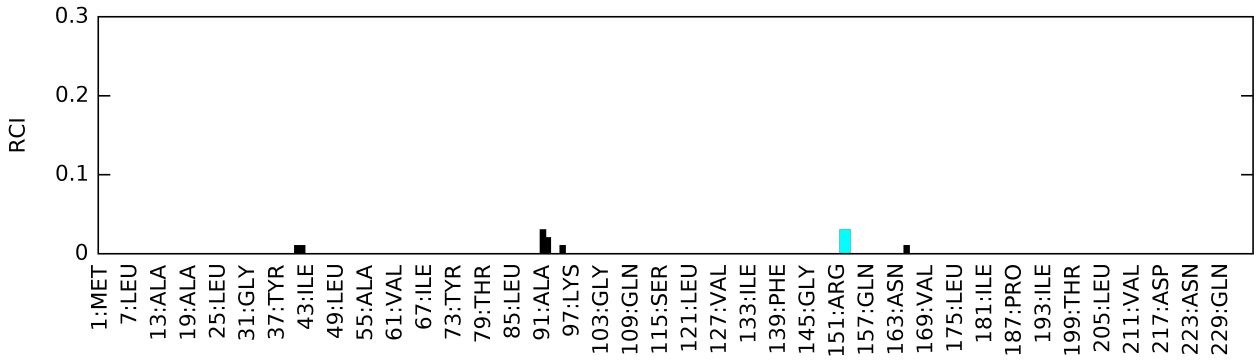
Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

6.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

6.2.5 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 B: