WORLDWIDE PDB PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report ⓘ

Sep 16, 2021 – 01:13 pm BST

PDB ID	:	7NQC
Title	:	Calmodulin extracts the Ras family protein RalA from lipid bilayers by engagement with two membrane targeting motifs
Authors	:	Chamberlain, S.G.; Owen, D.; Mott, H.R.
Deposited on	:	2021-03-01

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

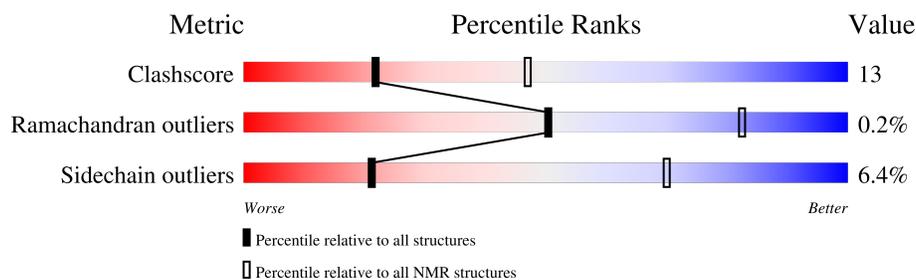|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.1
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.1

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 77%.

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	Percentile Ranks	Value
Clashscore		13
Ramachandran outliers		0.2%
Sidechain outliers		6.4%

Worse — Better

■ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	165	64% 18% • 7% 10%
2	B	18	6% 17% 6% 72%

2 Ensemble composition and analysis (i)

This entry contains 30 models. Model 28 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:1-A:74 (74)	0.26	28
2	A:84-A:146, B:195-B:199 (68)	0.31	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 and 1 single-model cluster was found.

Cluster number	Models
1	1, 6, 7, 18, 20, 21, 23, 24, 26, 27, 29
2	3, 4, 14, 16, 25, 28, 30
3	5, 10, 15, 17, 19, 22
4	2, 11, 13
5	9, 12
Single-model clusters	8

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2651 atoms, of which 1304 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Calmodulin-1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	148	2260	714	1095	188	254	9	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	MET	-	initiating methionine	UNP P0DP29
A	-11	ARG	-	expression tag	UNP P0DP29
A	-10	GLY	-	expression tag	UNP P0DP29
A	-9	SER	-	expression tag	UNP P0DP29
A	-8	HIS	-	expression tag	UNP P0DP29
A	-7	HIS	-	expression tag	UNP P0DP29
A	-6	HIS	-	expression tag	UNP P0DP29
A	-5	HIS	-	expression tag	UNP P0DP29
A	-4	HIS	-	expression tag	UNP P0DP29
A	-3	HIS	-	expression tag	UNP P0DP29
A	-2	GLY	-	expression tag	UNP P0DP29
A	-1	SER	-	expression tag	UNP P0DP29
A	149	GLY	-	expression tag	UNP P0DP29
A	150	GLY	-	expression tag	UNP P0DP29
A	151	SER	-	expression tag	UNP P0DP29
A	152	LEU	-	expression tag	UNP P0DP29

- Molecule 2 is a protein called PRO-ASN-GLY-LYS-LYS-LYS-ARG-LYS-SER-LEU-ALA-L YS-ARG-ILE-ARG-GLU-ARG-CMF.
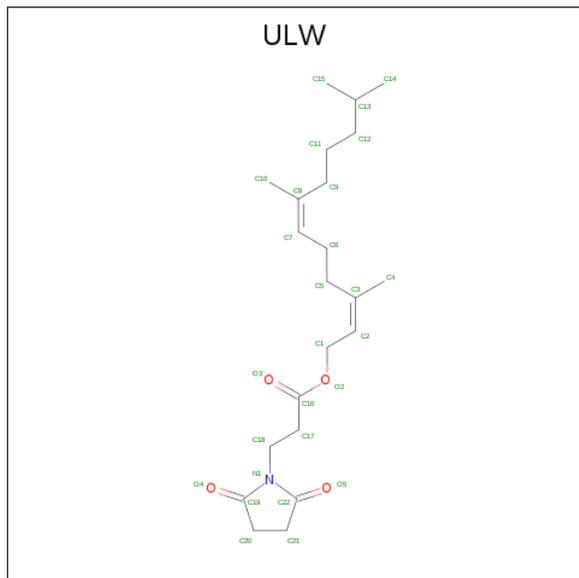
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	18	328	91	177	36	23	1	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
			Total	Ca
3	A	4	4	4

- Molecule 4 is [(2 {Z},6 {Z})-3,7,11-trimethyldodeca-2,6-dienyl] 3-[2,5-bis(oxidanylidene)pyr

rolidin-1-yl]propanoate (three-letter code: ULW) (formula: $C_{22}H_{35}NO_4$) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms				
			Total	C	H	N	O
4	B	1	59	22	32	1	4

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: Calmodulin-1

- Molecule 2: PRO-ASN-GLY-LYS-LYS-LYS-ARG-LYS-SER-LEU-ALA-LYS-ARG-ILE-ARG-GLU-ARG-CMF

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

The representative model is number 28. Colouring as in section 4.1 above.

- Molecule 1: Calmodulin-1

- Molecule 2: PRO-ASN-GLY-LYS-LYS-LYS-ARG-LYS-SER-LEU-ALA-LYS-ARG-ILE-ARG-GLU-ARG-CMF

Chain B:

17% 11% 72%

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 30 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
CNS	refinement	
ARIA	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	1751
Number of shifts mapped to atoms	1751
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	77%

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ULW, CA

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	1078	1010	1011	28±4
2	B	41	53	53	5±2
4	B	27	32	0	0±1
All	All	34500	32850	31920	858

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 13.

5 of 229 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
4:B:301:ULW:O2	4:B:301:ULW:C1	1.44	1.63	18	4
1:A:55:VAL:HG11	1:A:71:MET:HB2	1.03	1.27	4	17
1:A:4:LEU:HB3	1:A:8:GLN:HE21	0.84	1.33	13	1
4:B:301:ULW:C1	4:B:301:ULW:C16	0.82	2.56	10	4
1:A:104:GLU:HA	1:A:107:HIS:CD2	0.81	2.10	15	28

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	136/165 (82%)	132±1 (97±1%)	4±1 (3±1%)	0±0 (0±0%)	100	100
2	B	5/18 (28%)	4±0 (80±5%)	1±0 (15±9%)	0±0 (5±9%)	3	23
All	All	4230/5490 (77%)	4076 (96%)	145 (3%)	9 (0%)	50	82

All 2 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	195	LEU	8
1	A	3	GLN	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	116/139 (83%)	109±2 (94±2%)	7±2 (6±2%)	23	72
2	B	4/16 (25%)	3±1 (82±17%)	1±1 (18±17%)	4	37
All	All	3600/4650 (77%)	3369 (94%)	231 (6%)	21	70

5 of 38 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	4	LEU	28
1	A	133	ASP	21
1	A	95	ASP	15
1	A	60	ASN	14
2	B	195	LEU	14

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)

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

| Mol | Type | Chain | Res | Link | Bond lengths | | |
					Counts	RMSZ	#Z>2
4	ULW	B	301	2	27,27,27	1.22±0.22	3±0 (10±1%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

| Mol | Type | Chain | Res | Link | Bond angles | | |
					Counts	RMSZ	#Z>2
4	ULW	B	301	2	34,34,34	1.10±0.05	2±1 (5±2%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means

no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	ULW	B	301	2	-	0±0,23,36,36	0±0,1,1,1

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
4	B	301	ULW	O2-C1	7.76	1.68	1.46	27	25
4	B	301	ULW	C22-N1	3.71	1.44	1.38	28	29
4	B	301	ULW	C12-C13	2.41	1.35	1.51	9	30

5 of 9 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
4	B	301	ULW	C10-C8-C9	3.08	120.44	115.27	27	30
4	B	301	ULW	C17-C18-N1	2.40	117.49	111.73	28	2
4	B	301	ULW	C4-C3-C5	2.39	119.29	115.27	15	1
4	B	301	ULW	O5-C22-N1	2.36	126.53	123.92	28	1
4	B	301	ULW	C11-C12-C13	2.22	126.42	115.98	23	9

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand ULW B 301

Bond lengths

Bond angles

Torsions

Rings
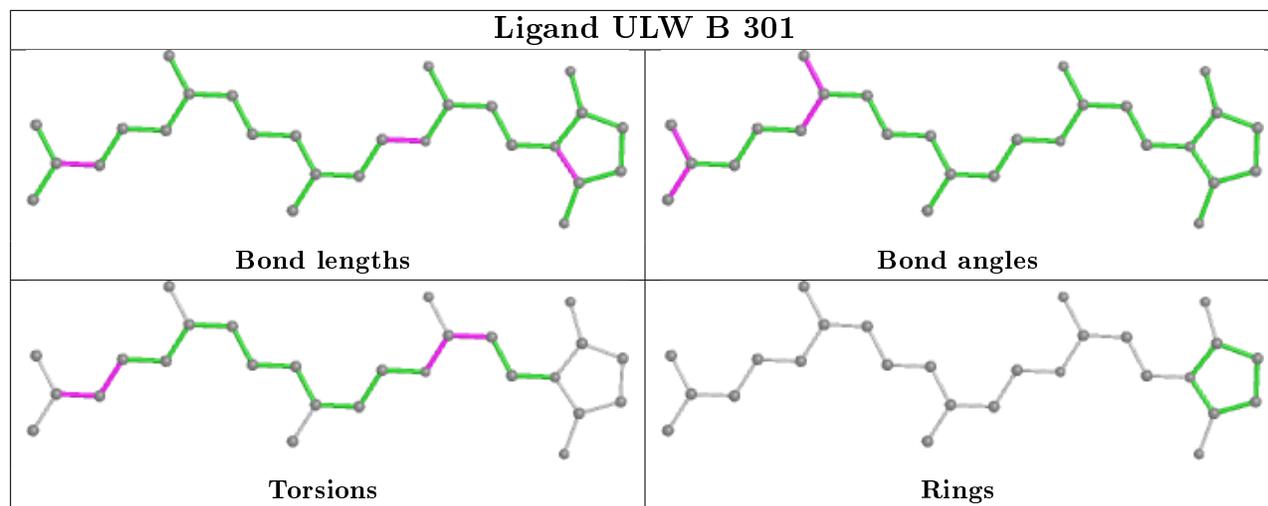
6.7 Other polymers (i)

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 77% for the well-defined parts and 74% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list*

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	1751
Number of shifts mapped to atoms	1751
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 ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}\mathrm{C}_\alpha$	158	-0.15 ± 0.11	None needed (< 0.5 ppm)
$^{13}\mathrm{C}_\beta$	141	0.17 ± 0.06	None needed (< 0.5 ppm)
$^{13}\mathrm{C}'$	0	—	None (insufficient data)
$^{15}\mathrm{N}$	137	0.17 ± 0.09	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 77%, i.e. 1311 atoms were assigned a chemical shift out of a possible 1704. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\mathrm{H}$	$^{13}\mathrm{C}$	$^{15}\mathrm{N}$
Backbone	530/706 (75%)	274/282 (97%)	134/284 (47%)	122/140 (87%)
Sidechain	693/902 (77%)	427/520 (82%)	263/343 (77%)	3/39 (8%)

Continued on next page...

Continued from previous page...

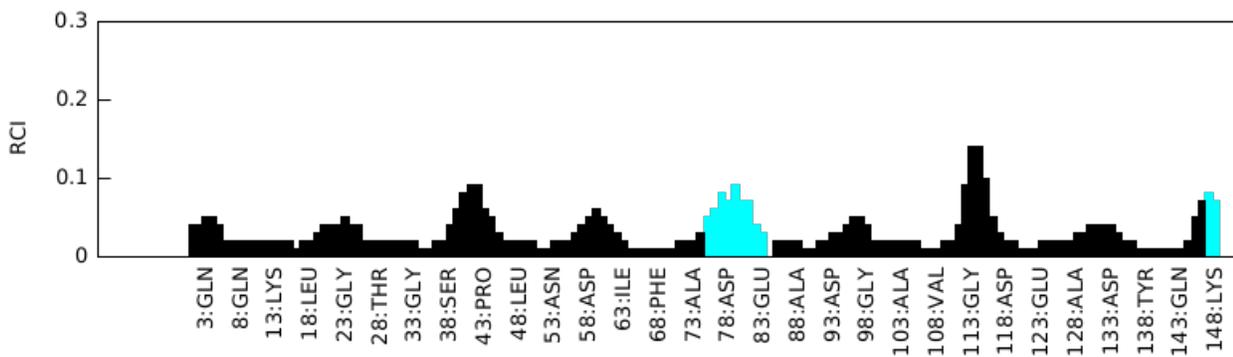|	Total	^{1}H	^{13}C	^{15}N
Aromatic	88/96 (92%)	48/52 (92%)	40/42 (95%)	0/2 (0%)
Overall	1311/1704 (77%)	749/854 (88%)	437/669 (65%)	125/181 (69%)

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 images below report *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:

Random coil index (RCI) for chain B: