wwPDB NMR Structure Validation Summary Report ⓘ

May 7, 2024 – 04:12 pm BST

PDB ID	:	4AR0
BMRB ID	:	18459
Title	:	N0 domain of Neisseria meningitidis Pilus assembly protein PilQ
Authors	:	Phelan, M.M.; Berry, J.L.; Derrick, J.P.; Lian, L.Y.
Deposited on	:	2012-04-20

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

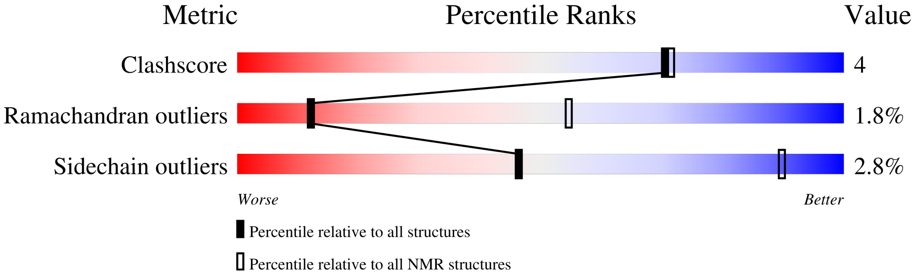
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 87%.

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	128	

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:345-A:421 (77)	0.84	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 4 clusters and 5 single-model clusters were found.

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

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 2026 atoms, of which 1011 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TYPE IV PILUS BIOGENESIS AND COMPETENCE PROTEIN PILQ.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	128	2026	637	1011	177	194	7	0

There are 29 discrepancies between the modelled and reference sequences:

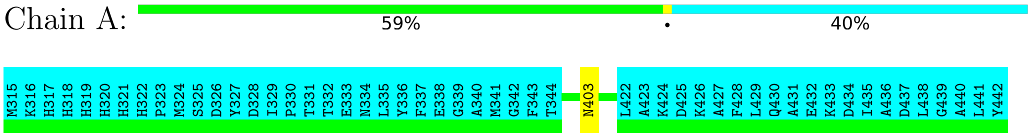
Chain	Residue	Modelled	Actual	Comment	Reference
A	315	MET	-	expression tag	UNP Q70M91
A	316	LYS	-	expression tag	UNP Q70M91
A	317	HIS	-	expression tag	UNP Q70M91
A	318	HIS	-	expression tag	UNP Q70M91
A	319	HIS	-	expression tag	UNP Q70M91
A	320	HIS	-	expression tag	UNP Q70M91
A	321	HIS	-	expression tag	UNP Q70M91
A	322	HIS	-	expression tag	UNP Q70M91
A	323	PRO	-	expression tag	UNP Q70M91
A	324	MET	-	expression tag	UNP Q70M91
A	325	SER	-	expression tag	UNP Q70M91
A	326	ASP	-	expression tag	UNP Q70M91
A	327	TYR	-	expression tag	UNP Q70M91
A	328	ASP	-	expression tag	UNP Q70M91
A	329	ILE	-	expression tag	UNP Q70M91
A	330	PRO	-	expression tag	UNP Q70M91
A	331	THR	-	expression tag	UNP Q70M91
A	332	THR	-	expression tag	UNP Q70M91
A	333	GLU	-	expression tag	UNP Q70M91
A	334	ASN	-	expression tag	UNP Q70M91
A	335	LEU	-	expression tag	UNP Q70M91
A	336	TYR	-	expression tag	UNP Q70M91
A	337	PHE	-	expression tag	UNP Q70M91
A	338	GLU	-	expression tag	UNP Q70M91
A	339	GLY	-	expression tag	UNP Q70M91
A	340	ALA	-	expression tag	UNP Q70M91
A	341	MET	-	expression tag	UNP Q70M91
A	342	GLY	-	expression tag	UNP Q70M91
A	428	PHE	LEU	variant	UNP Q70M91

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.

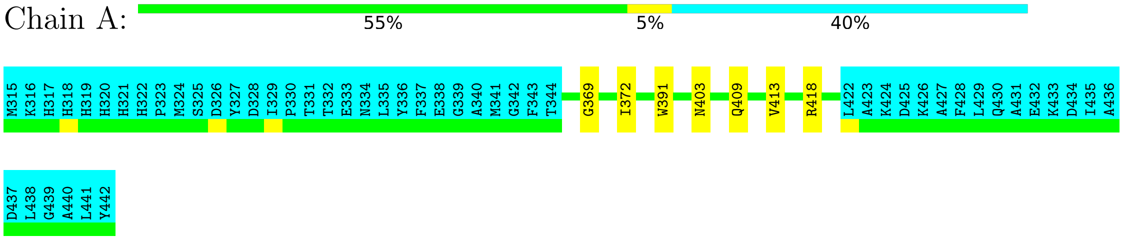
- Molecule 1: TYPE IV PILUS BIOGENESIS AND COMPETENCE PROTEIN PILQ



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: TYPE IV PILUS BIOGENESIS AND COMPETENCE PROTEIN PILQ



5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *CYANA2.1 AND CNS1.2 WATER REFINEMENT*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *LOWEST ENERGY*.

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

Software name	**Classification**	**Version**
CYANA	refinement	
TopSpin	structure solution	
CcpNmr Analysis	structure solution	ANALYSIS
CYANA	structure solution	
CNS	structure solution	
TALOS+	structure solution	

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	1520
Number of shifts mapped to atoms	1412
Number of unparsed shifts	0
Number of shifts with mapping errors	108
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.2±0.4
All	All	0	4

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	402	ARG	Sidechain	2
1	A	407	ARG	Sidechain	1
1	A	358	ARG	Sidechain	1

6.2 Too-close contacts ⓘ

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	602	620	620	4±3
All	All	12040	12400	12400	89

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 4.

5 of 54 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:382:MET:SD	1:A:384:LEU:HG	0.65	2.31	17	2
1:A:395:LEU:HG	1:A:399:MET:SD	0.60	2.36	10	1
1:A:395:LEU:O	1:A:398:VAL:HG22	0.58	1.99	2	2
1:A:362:GLN:HE21	1:A:362:GLN:HA	0.57	1.60	8	1
1:A:375:SER:O	1:A:378:VAL:HG12	0.55	2.01	7	3

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	77/128 (60%)	67±15 (87±20%)	5±2 (6±2%)	1±1 (2±1%)	13	56
All	All	1463/2560 (57%)	1344 (92%)	93 (6%)	26 (2%)	12	54

5 of 8 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	403	ASN	18
1	A	368	SER	2
1	A	353	GLN	1
1	A	347	LYS	1
1	A	374	ALA	1

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	68/110 (62%)	63±14 (92±21%)	2±1 (3±2%)	49	91
All	All	1292/2200 (59%)	1256 (97%)	36 (3%)	46	90

5 of 18 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	387	LYS	5
1	A	384	LEU	5
1	A	409	GLN	4
1	A	364	LEU	3
1	A	346	ARG	2

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 87% for the well-defined parts and 80% 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	1520
Number of shifts mapped to atoms	1412
Number of unparsed shifts	0
Number of shifts with mapping errors	108
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 108) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	323	PRO	HB1	2.195	0.007	2
1	A	324	MET	HB1	2.061	0.006	2
1	A	325	SER	HB1	3.77	0.002	2
1	A	326	ASP	HB1	2.497	0.003	2
1	A	327	TYR	HB1	2.997	0.005	2
1	A	327	TYR	HD3	6.916	0.002	3
1	A	328	ASP	HB1	2.574	0.000	2
1	A	330	PRO	HB1	2.238	0.013	2
1	A	333	GLU	HB1	1.961	0.007	2
1	A	334	ASN	HB1	2.684	0.005	2
1	A	335	LEU	HB1	1.392	0.007	2
1	A	336	TYR	HB1	2.787	0.000	2
1	A	336	TYR	HD3	7.032	0.001	3
1	A	337	PHE	HB1	3.02	0.001	2

Continued on next page...

Continued from previous page...

List ID	**Chain**	**Res**	**Type**	**Atom**	**Shift Data**		
					Value	Uncertainty	Ambiguity
1	A	337	PHE	HD3	7.193	0.003	3
1	A	338	GLU	HB1	1.982	0.007	2
1	A	340	ALA	3HB	1.273	0.001	2
1	A	341	MET	HB1	1.936	0.010	2
1	A	343	PHE	HB1	2.63	0.000	2
1	A	346	ARG	HB1	1.932	0.002	2
1	A	347	LYS	HB1	1.59	0.005	2
1	A	347	LYS	HE1	2.869	0.000	2
1	A	349	SER	HB1	3.609	0.006	2
1	A	350	LEU	HB1	1.206	0.001	2
1	A	351	ASP	HB1	2.256	0.002	2
1	A	352	PHE	HB1	2.91	0.002	2
1	A	352	PHE	HD3	7.144	0.000	3
1	A	353	GLN	HB1	1.884	0.009	2
1	A	354	ASP	HB1	2.864	0.007	2
1	A	356	GLU	HB1	1.925	0.004	2
1	A	358	ARG	HB1	1.78	0.003	2
1	A	361	LEU	HB1	1.811	0.004	2
1	A	362	GLN	HB1	2.247	0.003	2
1	A	364	LEU	HB1	1.241	0.000	2
1	A	365	ALA	3HB	1.594	0.003	2
1	A	366	LYS	HB1	1.955	0.000	2
1	A	366	LYS	HE1	2.894	0.000	2
1	A	367	GLU	HB1	1.632	0.000	2
1	A	368	SER	HB1	2.145	0.001	2
1	A	370	MET	HB1	1.542	0.000	2
1	A	371	ASN	HB1	2.819	0.004	2
1	A	374	ALA	3HB	1.155	0.001	2
1	A	375	SER	HB1	3.772	0.000	2
1	A	376	ASP	HB1	2.641	0.015	2
1	A	377	SER	HB1	3.817	0.000	2
1	A	379	ASN	HB1	2.806	0.000	2
1	A	381	LYS	HB1	1.396	0.002	2
1	A	381	LYS	HE1	2.832	0.000	2
1	A	382	MET	HB1	1.662	0.003	2
1	A	384	LEU	HB1	1.613	0.005	2
1	A	385	SER	HB1	3.652	0.008	2
1	A	386	LEU	HB1	1.635	0.013	2
1	A	387	LYS	HB1	1.632	0.008	2
1	A	387	LYS	HE1	2.838	0.000	2
1	A	388	ASP	HB1	2.733	0.001	2

Continued on next page...

Continued from previous page...

List ID	**Chain**	**Res**	**Type**	**Atom**	**Shift Data**		
					Value	Uncertainty	Ambiguity
1	A	390	PRO	HB1	2.416	0.000	2
1	A	391	TRP	HB1	3.084	0.000	2
1	A	391	TRP	HD	7.582	0.000	1
1	A	391	TRP	HH	7.038	0.000	1
1	A	392	ASP	HB1	1.001	0.000	2
1	A	393	GLN	HB1	1.084	0.000	2
1	A	394	ALA	3HB	1.294	0.003	2
1	A	395	LEU	HB1	1.416	0.000	2
1	A	396	ASP	HB1	2.513	0.001	2
1	A	397	LEU	HB1	1.882	0.004	2
1	A	399	MET	HB1	2.263	0.000	2
1	A	400	GLN	HB1	2.121	0.000	2
1	A	401	ALA	3HB	1.474	0.008	2
1	A	402	ARG	HB1	2.116	0.004	2
1	A	403	ASN	HB1	2.665	0.003	2
1	A	404	LEU	HB1	1.498	0.000	2
1	A	405	ASP	HB1	2.476	0.002	2
1	A	406	MET	HB1	1.692	0.000	2
1	A	407	ARG	HB1	1.693	0.015	2
1	A	408	GLN	HB1	1.968	0.004	2
1	A	409	GLN	HB1	1.969	0.000	2
1	A	411	ASN	HB1	2.849	0.005	2
1	A	414	ASN	HB1	2.523	0.000	2
1	A	416	ALA	3HB	1.25	0.003	2
1	A	417	PRO	HB1	1.748	0.004	2
1	A	418	ARG	HB1	1.616	0.006	2
1	A	419	ASP	HB1	2.532	0.000	2
1	A	420	GLU	HB1	1.982	0.006	2
1	A	421	LEU	HB1	1.633	0.005	2
1	A	422	LEU	HB1	1.602	0.013	2
1	A	423	ALA	3HB	1.373	0.001	2
1	A	424	LYS	HB1	1.808	0.004	2
1	A	424	LYS	HE1	2.926	0.011	2
1	A	425	ASP	HB1	2.738	0.000	2
1	A	426	LYS	HB1	1.767	0.000	2
1	A	426	LYS	HE1	2.888	0.000	2
1	A	427	ALA	3HB	1.277	0.002	2
1	A	428	PHE	HB1	3.036	0.003	2
1	A	428	PHE	HD3	6.974	0.000	3
1	A	429	LEU	HB1	1.476	0.003	2
1	A	430	GLN	HB1	2.017	0.004	2

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	431	ALA	3HB	1.324	0.001	2
1	A	432	GLU	HB1	1.863	0.003	2
1	A	433	LYS	HB1	1.718	0.013	2
1	A	433	LYS	HE1	2.883	0.000	2
1	A	434	ASP	HB1	2.529	0.008	2
1	A	436	ALA	3HB	1.301	0.000	2
1	A	437	ASP	HB1	2.534	0.001	2
1	A	438	LEU	HB1	1.539	0.002	2
1	A	440	ALA	3HB	1.244	0.000	2
1	A	441	LEU	HB1	1.37	0.003	2
1	A	442	TYR	HB1	3.01	0.000	2
1	A	442	TYR	HD3	7.004	0.001	3

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	120	-0.16 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_\beta$	113	0.24 ± 0.04	None needed (< 0.5 ppm)
$^{13}C'$	106	0.28 ± 0.15	None needed (< 0.5 ppm)
^{15}N	116	-0.42 ± 0.40	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 87%, i.e. 935 atoms were assigned a chemical shift out of a possible 1076. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	379/385 (98%)	156/156 (100%)	148/154 (96%)	75/75 (100%)
Sidechain	537/669 (80%)	346/434 (80%)	177/205 (86%)	14/30 (47%)
Aromatic	19/22 (86%)	8/11 (73%)	10/10 (100%)	1/1 (100%)
Overall	935/1076 (87%)	510/601 (85%)	335/369 (91%)	90/106 (85%)

7.1.4 Statistically unusual chemical shifts

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	368	SER	HB2	2.32	2.61 – 5.13	-6.2

7.1.5 Random Coil Index (RCI) plots ⓘ

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: