wwPDB NMR Structure Validation Summary Report

Jun 26, 2024 – 05:55 AM EDT

PDB ID	:	7K7F
BMRB ID	:	30800
Title	:	Solution Structure of the Corynebacterium diphtheriae SpaA Pilin-Signal Peptide Complex
Authors	:	McConnell, S.A.; Clubb, R.T.
Deposited on	:	2020-09-22

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 ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

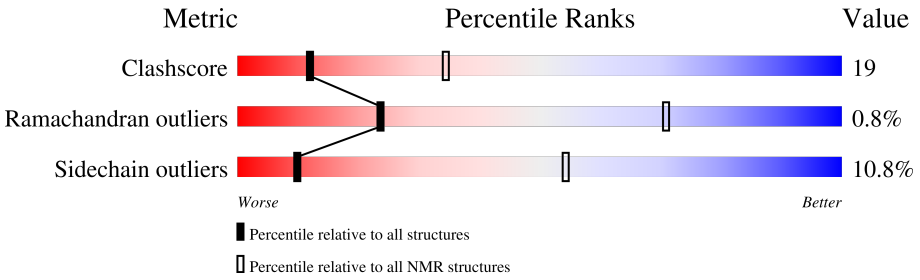
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 85%.

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	143	67% 29% • •
2	B	10	40% 20% 20% 20%

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:54-A:193, B:487-B:494 (148)	0.60	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 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2328 atoms, of which 1167 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative surface-anchored fimbrial subunit.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	143	2172	687	1087	185	211	2	0

- Molecule 2 is a protein called SpaA sorting signal peptide.

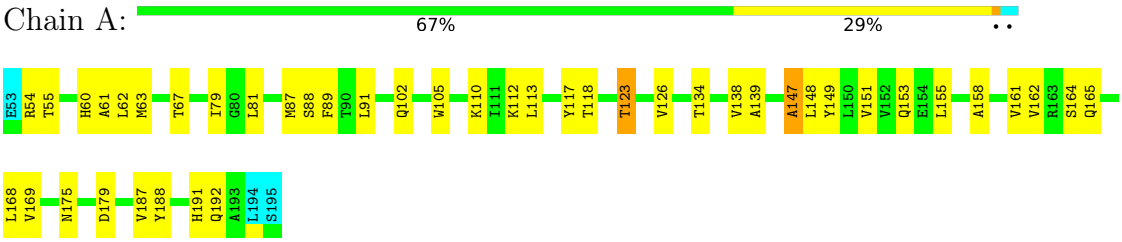
Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	10	156	50	80	12	14	0

4 Residue-property plots

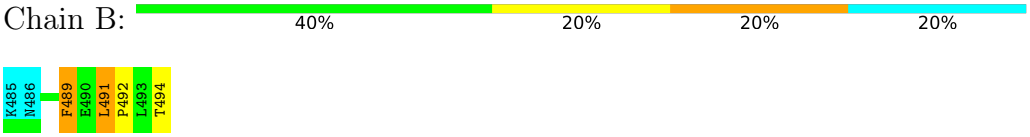
4.1 Average score per residue in the NMR ensemble

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: Putative surface-anchored fimbrial subunit



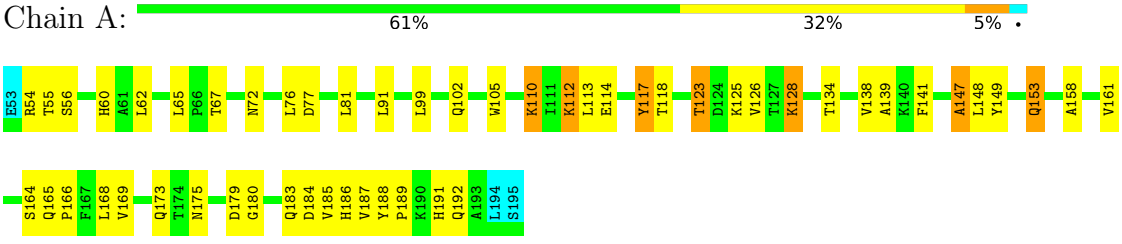
• Molecule 2: SpaA sorting signal peptide



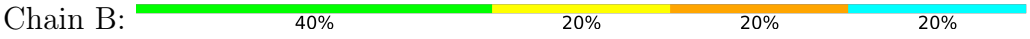
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

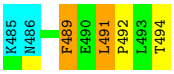
The representative model is number 1. Colouring as in section 4.1 above.

• Molecule 1: Putative surface-anchored fimbrial subunit



• Molecule 2: SpaA sorting signal peptide




K485
N486
F489
E490
L491
P492
L493
T494

5 Refinement protocol and experimental data overview

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

Of the 200 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
X-PLOR NIH	refinement	2.37
X-PLOR NIH	structure calculation	2.37
UNIO	structure calculation	
TALOS-N	geometry optimization	
MOLMOL	refinement	
PROCHECK / PROCHECK-NMR	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)	working_cs.cif
Number of chemical shift lists	2
Total number of shifts	1691
Number of shifts mapped to atoms	1691
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

6 Model quality

6.1 Standard geometry

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

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	1061	1063	1063	40±6
2	B	59	59	59	6±2
All	All	22400	22440	22440	869

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 19.

5 of 290 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:79:ILE:HG22	1:A:81:LEU:HD13	0.82	1.51	15	9
1:A:113:LEU:HD12	1:A:113:LEU:O	0.80	1.77	18	1
2:B:491:LEU:N	2:B:491:LEU:HD23	0.80	1.92	7	20
1:A:173:GLN:NE2	1:A:185:VAL:HG13	0.80	1.92	6	2
1:A:99:LEU:HD12	1:A:147:ALA:O	0.78	1.78	8	4

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	140/143 (98%)	132±1 (94±1%)	7±1 (5±1%)	1±0 (1±0%)	24 71
2	B	7/10 (70%)	4±1 (64±8%)	2±1 (35±8%)	0±0 (1±4%)	15 61
All	All	2940/3060 (96%)	2727 (93%)	189 (6%)	24 (1%)	24 71

All 4 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	147	ALA	20
2	B	487	ALA	2
1	A	157	GLY	1
1	A	158	ALA	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	114/117 (97%)	104±3 (91±2%)	10±3 (9±2%)	13 60
2	B	6/8 (75%)	3±0 (52±5%)	3±0 (48±5%)	0 1
All	All	2400/2500 (96%)	2140 (89%)	260 (11%)	10 54

5 of 42 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)
2	B	491	LEU	20
2	B	494	THR	20
1	A	118	THR	18
1	A	168	LEU	18
2	B	489	PHE	18

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)

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 85% for the well-defined parts and 84% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *protein.str.txt*

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	1630
Number of shifts mapped to atoms	1630
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	9

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	142	-0.02 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	130	0.03 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	139	0.26 ± 0.11	None needed (< 0.5 ppm)
^{15}N	126	-0.54 ± 0.57	None needed (imprecise)

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 83%, i.e. 1605 atoms were assigned a chemical shift out of a possible 1945. 0 out of 29 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	673/733 (92%)	272/299 (91%)	277/296 (94%)	124/138 (90%)
Sidechain	871/1084 (80%)	576/711 (81%)	286/341 (84%)	9/32 (28%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	61/128 (48%)	31/64 (48%)	28/58 (48%)	2/6 (33%)
Overall	1605/1945 (83%)	879/1074 (82%)	591/695 (85%)	135/176 (77%)

7.1.4 Statistically unusual chemical shifts (i)

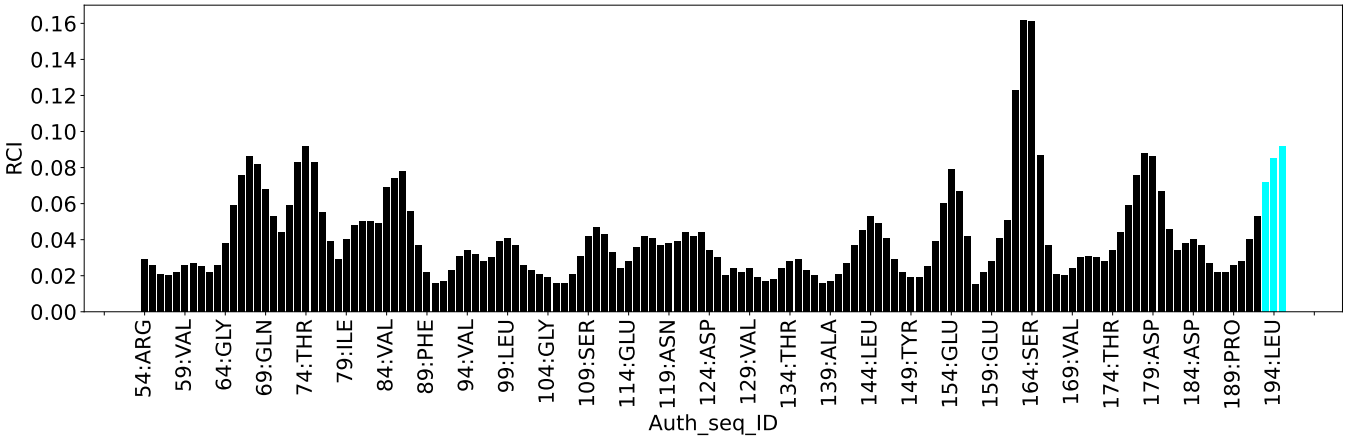
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	78	SER	CB	32.43	56.28 – 71.32	-20.9
1	A	157	GLY	CA	61.19	38.93 – 51.79	12.3
1	A	135	GLU	HG2	0.80	1.24 – 3.30	-7.1
1	A	141	PHE	CD1	122.71	125.33 – 137.83	-7.1
1	A	60	HIS	CE1	121.66	126.08 – 149.12	-6.9
1	A	151	VAL	HB	0.24	0.43 – 3.54	-5.6
1	A	77	ASP	CB	32.43	32.98 – 48.76	-5.3
1	A	91	LEU	HB3	-0.31	-0.26 – 3.31	-5.2
1	A	125	LYS	CD	34.53	23.50 – 34.42	5.1

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.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:



7.2 Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *peptide.str.txt*

7.2.1 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	61
Number of shifts mapped to atoms	61
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

7.2.2 Chemical shift referencing

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

7.2.3 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 3%, i.e. 53 atoms were assigned a chemical shift out of a possible 1945. 0 out of 29 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	15/733 (2%)	15/299 (5%)	0/296 (0%)	0/138 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Sidechain	35/1084 (3%)	35/711 (5%)	0/341 (0%)	0/32 (0%)
Aromatic	3/128 (2%)	3/64 (5%)	0/58 (0%)	0/6 (0%)
Overall	53/1945 (3%)	53/1074 (5%)	0/695 (0%)	0/176 (0%)

7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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

Random coil index (RCI) for chain B: