wwPDB NMR Structure Validation Summary Report

Nov 7, 2023 – 04:43 pm GMT

PDB ID	:	7QS6
BMRB ID	:	34699
Title	:	Solution structure of thanatin-like derivative 7 in complex with E.coli LptA
Authors	:	Moehle, K.; Zerbe, O.
Deposited on	:	2022-01-12

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

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

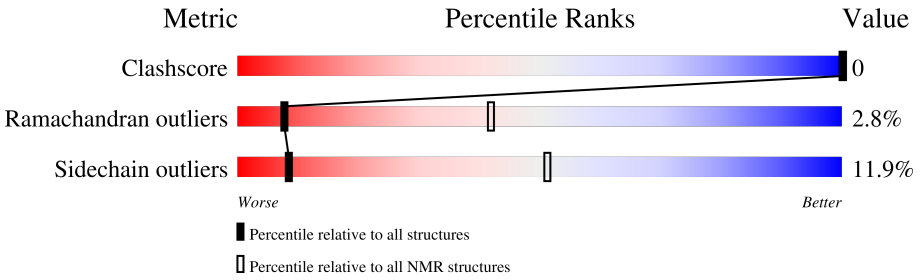
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 83%.

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	118	92% 5% • •
2	B	16	44% 19% 38%

2 Ensemble composition and analysis

This entry contains 20 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *target function*.

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:30-A:145, B:208-B:210, B:212-B:213, B:215-B:215, B:217-B:218, B:220-B:221 (126)	1.46	2

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 7 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2073 atoms, of which 1031 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Lipopolysaccharide export system protein LptA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	118	1787	564	884	152	184	3	0

- Molecule 2 is a protein called Thanatin-like derivative.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	16	286	85	147	29	23	2	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	209	THR	ILE	engineered mutation	UNP P55788
B	211	LE1	CYS	modified residue	UNP P55788
B	214	DAB	ARG	modified residue	UNP P55788
B	216	4FO	GLY	modified residue	UNP P55788
B	219	DAB	GLN	modified residue	UNP P55788
B	221	TYR	MET	engineered mutation	UNP P55788

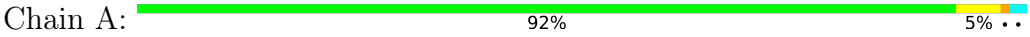
4 Residue-property plots (i)

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.

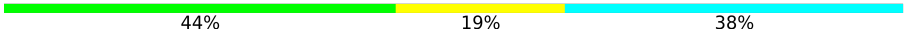
• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:



• Molecule 2: Thanatin-like derivative

Chain B:

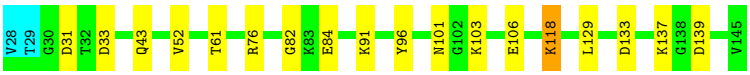


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 2. Colouring as in section 4.1 above.

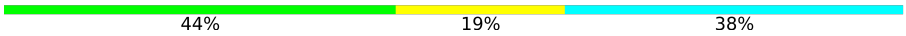
• Molecule 1: Lipopolysaccharide export system protein LptA

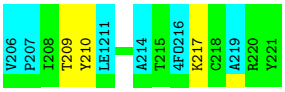
Chain A:



• Molecule 2: Thanatin-like derivative

Chain B:





5 Refinement protocol and experimental data overview

The models were refined using the following method: *molecular dynamics*.

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	3.98
MOE	refinement	2019.01

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	1462
Number of shifts mapped to atoms	1462
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	83%

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: EU0, LE1, 4FO, HYP, DAB

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.67±0.01	0±0/903 (0.0± 0.0%)	1.02±0.03	1±1/1223 (0.1± 0.1%)
2	B	0.93±0.03	0±0/89 (0.0± 0.0%)	1.35±0.10	1±1/110 (1.0± 0.5%)
All	All	0.69	0/19840 (0.0%)	1.05	40/26660 (0.2%)

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	2.6±1.6
2	B	0.0±0.0	0.8±0.9
All	All	0	67

There are no bond-length outliers.

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
1	A	76	ARG	NE-CZ-NH1	11.21	125.90	120.30	5	13
2	B	220	ARG	NE-CZ-NH1	8.34	124.47	120.30	8	9
2	B	213	ARG	NE-CZ-NH1	8.04	124.32	120.30	9	12
1	A	128	TYR	CB-CG-CD2	-7.24	116.66	121.00	14	1
1	A	76	ARG	CD-NE-CZ	5.78	131.69	123.60	5	1

There are no chirality outliers.

5 of 28 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	118	LYS	Peptide	10
2	B	220	ARG	Sidechain,Peptide	7
1	A	76	ARG	Sidechain	5
1	A	103	LYS	Peptide	5
2	B	210	TYR	Sidechain	5

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	889	866	866	0±0
2	B	92	92	92	0±0
All	All	19620	19160	19160	1

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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:91:LYS:HE2	1:A:108:HIS:CE1	0.48	2.43	1	1

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	115/118 (97%)	99±4 (86±4%)	13±5 (11±4%)	3±2 (3±1%)	7 41
2	B	9/16 (56%)	8±1 (86±9%)	1±1 (13±8%)	0±0 (1±3%)	18 66
All	All	2480/2680 (93%)	2128 (86%)	283 (11%)	69 (3%)	8 42

5 of 29 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	118	LYS	10
1	A	117	ALA	5
1	A	91	LYS	4
1	A	126	ASN	4
1	A	101	ASN	4

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	98/100 (98%)	86±3 (88±3%)	12±3 (12±3%)	9 52
2	B	10/10 (100%)	9±1 (86±9%)	1±1 (14±9%)	6 46
All	All	2160/2200 (98%)	1902 (88%)	258 (12%)	8 51

5 of 65 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	43	GLN	10
1	A	137	LYS	10
1	A	80	GLU	10
1	A	118	LYS	10
1	A	140	LYS	10

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

6 non-standard protein/DNA/RNA residues are modelled in this entry.

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
2	EU0	B	206	2	8,9,10	1.19±0.08	0±0 (6±6%)
2	HYP	B	207	2	6,8,9	0.48±0.04	0±0 (0±0%)
2	DAB	B	214	2	5,6,7	0.60±0.08	0±0 (0±0%)
2	LE1	B	211	2	3,7,8	0.58±0.03	0±0 (0±0%)
2	DAB	B	219	2	5,6,7	0.68±0.11	0±0 (0±0%)
2	4FO	B	216	2	5,6,7	0.54±0.05	0±0 (0±0%)

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
2	EU0	B	206	2	10,11,13	1.39±0.24	1±1 (13±8%)
2	HYP	B	207	2	5,10,12	0.99±0.10	0±0 (4±8%)
2	DAB	B	214	2	1,6,8	0.26±0.16	0±0 (0±0%)
2	LE1	B	211	2	3,10,12	0.96±0.07	0±0 (0±0%)
2	DAB	B	219	2	1,6,8	0.26±0.18	0±0 (0±0%)
2	4FO	B	216	2	1,6,8	0.14±0.10	0±0 (0±0%)

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
2	EU0	B	206	2	-	0±0,9,10,12	-
2	4FO	B	216	2	-	0±0,4,5,7	-
2	DAB	B	219	2	-	0±0,4,5,7	-
2	LE1	B	211	2	-	0±0,4,8,10	-
2	HYP	B	207	2	-	0±0,0,11,13	0±0,1,1,1
2	DAB	B	214	2	-	0±0,4,5,7	-

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
2	B	206	EU0	CA-N	2.62	1.50	1.46	12	2
2	B	206	EU0	CT-N	2.24	1.29	1.33	14	8

5 of 8 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
2	B	206	EU0	CB-CA-N	3.48	106.63	111.17	5	3
2	B	206	EU0	CG1-CB-CA	3.48	116.53	111.21	5	2
2	B	206	EU0	CA-N-CT	3.12	132.27	123.19	5	5
2	B	206	EU0	CG2-CB-CA	3.05	115.88	111.21	11	7
2	B	206	EU0	N-CT-NT1	2.50	124.95	120.59	3	2

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 83% 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: *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	1462
Number of shifts mapped to atoms	1462
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
$^{13}C_{\alpha}$	118	-0.05 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	105	-0.13 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	111	0.43 ± 0.10	None needed (< 0.5 ppm)
^{15}N	110	-0.55 ± 0.47	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. 1396 atoms were assigned a chemical shift out of a possible 1672. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	591/635 (93%)	257/261 (98%)	225/252 (89%)	109/122 (89%)
Sidechain	754/923 (82%)	524/596 (88%)	220/291 (76%)	10/36 (28%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	51/114 (45%)	37/55 (67%)	14/56 (25%)	0/3 (0%)
Overall	1396/1672 (83%)	818/912 (90%)	459/599 (77%)	119/161 (74%)

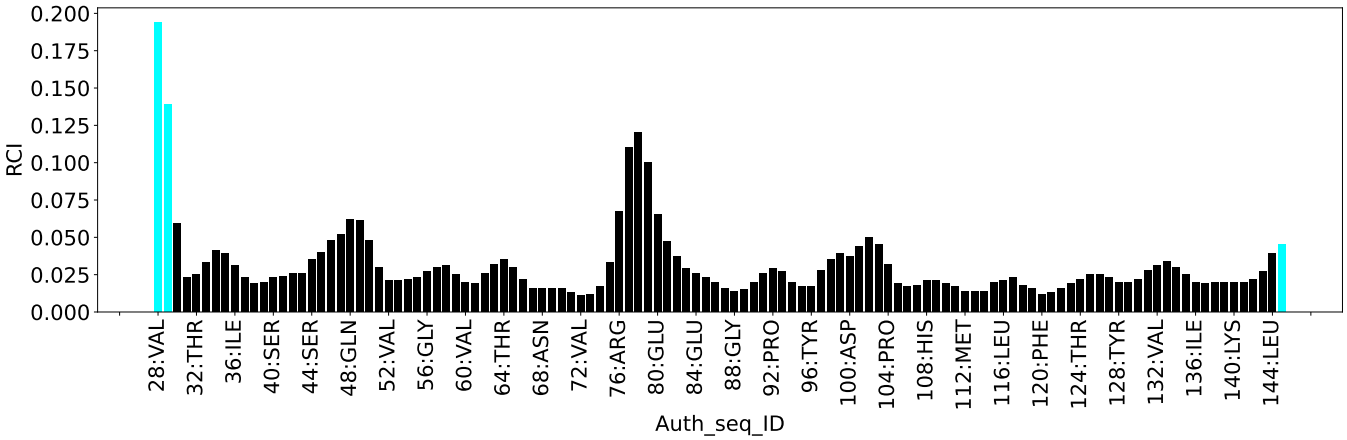
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

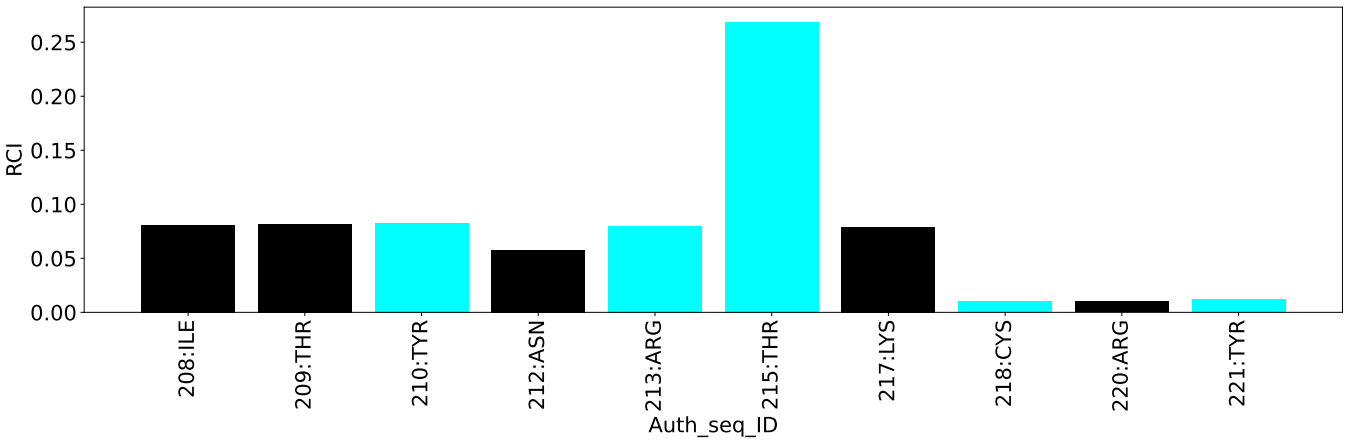
7.1.5 Random Coil Index (RCI) plots (i)

The 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:



Random coil index (RCI) for chain B:


RCI
0.25
0.20
0.15
0.10
0.05
0.00
208:ILE
209:THR
210:TYR
212:ASN
213:ARG
215:THR
217:LYS
218:CYS
220:ARG
221:TYR
Auth_seq_ID