wwPDB NMR Structure Validation Summary Report

Nov 5, 2023 – 10:13 pm GMT

PDB ID	:	8C5J
BMRB ID	:	16707
Title	:	Spatial structure of Lch-alpha peptide from two-component lantibiotic system Lichenicidin VK21
Authors	:	Mineev, K.S.; Paramonov, A.S.; Arseniev, A.S.; Ovchinnikova, T.V.; Shenkarev, Z.O.
Deposited on	:	2023-01-09

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:

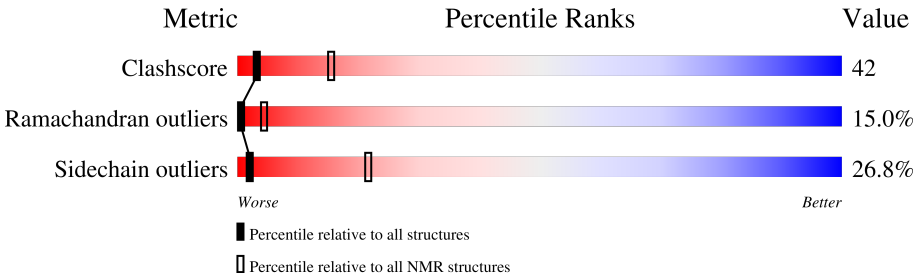
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	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

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

The overall completeness of chemical shifts assignment is 81%.

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	32	34% 44% 19% .

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
1	A	ABA	12	20	-
1	A	ABA	31	20	-
1	A	ABA	33	20	-

2 Ensemble composition and analysis (i)

This entry contains 20 models.

Cyrange was unable to find well-defined residues.

Error message: Only domains with < 8 residues could be identified.

NmrClust was unable to cluster the ensemble.

Error message: Wrapper check: not enough residues in core to run NmrClust

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 447 atoms, of which 224 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Lantibiotic lichenicidin VK21 A1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	32	447	142	224	36	40	5	0

There are 5 discrepancies between the modelled and reference sequences:

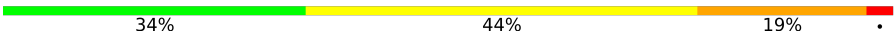
Chain	Residue	Modelled	Actual	Comment	Reference
A	10	2KT	THR	conflict	UNP P86475
A	12	ABA	THR	conflict	UNP P86475
A	20	DAL	SER	conflict	UNP P86475
A	31	ABA	THR	conflict	UNP P86475
A	33	ABA	THR	conflict	UNP P86475

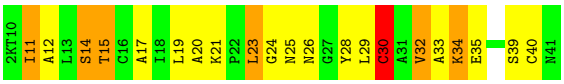
4 Residue-property plots (i)

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: Lantibiotic lichenicidin VK21 A1

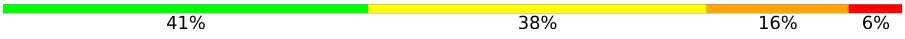
Chain A: 

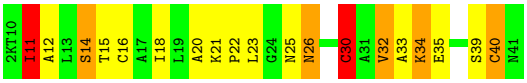


4.2 Residue scores for the representative (author defined) model from the NMR ensemble

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

• Molecule 1: Lantibiotic lichenicidin VK21 A1

Chain A: 



5 Refinement protocol and experimental data overview

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

Of the 400 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
CYANA	refinement	3.98.13
CYANA	structure calculation	3.98.13

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	321
Number of shifts mapped to atoms	321
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	81%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 2KT, DBU, ABA, DHA, DAL

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	3.0±0.0	3.0±0.0
All	All	60	60

There are no bond-length outliers.

There are no bond-angle outliers.

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

Mol	Chain	Res	Type	Atoms	Models (Total)
1	A	12	ABA	CA	20
1	A	31	ABA	CA	20
1	A	33	ABA	CA	20

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	11	ILE	Peptide	20
1	A	30	CYS	Peptide	20
1	A	32	VAL	Peptide	20

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	223	224	219	18±4
All	All	4460	4480	4357	369

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 42.

5 of 150 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:37:MET:SD	1:A:40:CYS:SG	1.26	2.32	16	1
1:A:16:CYS:HB2	1:A:36:CYS:SG	1.17	1.78	13	1
1:A:19:LEU:O	1:A:30:CYS:SG	1.12	2.07	7	3
1:A:16:CYS:CB	1:A:36:CYS:SG	1.08	2.40	13	1
1:A:16:CYS:O	1:A:30:CYS:SG	1.02	2.18	11	1

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	24/32 (75%)	14±2 (57±9%)	7±2 (28±8%)	4±1 (15±5%)	0 4
All	All	480/640 (75%)	275 (57%)	133 (28%)	72 (15%)	0 4

5 of 13 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	26	ASN	11
1	A	34	LYS	11
1	A	40	CYS	8
1	A	23	LEU	7
1	A	11	ILE	6

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	22/22 (100%)	16±1 (73±5%)	6±1 (27±5%)	2 21
All	All	440/440 (100%)	322 (73%)	118 (27%)	2 21

5 of 16 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	30	CYS	20
1	A	34	LYS	17
1	A	23	LEU	15
1	A	40	CYS	13
1	A	19	LEU	12

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
1	ABA	A	31	1	4,5,6	0.96±0.01	0±0 (0±0%)
1	DHA	A	14	1	4,4,5	1.21±0.01	1±0 (25±0%)
1	DBU	A	15	1	4,5,6	2.28±0.01	1±0 (25±0%)
1	ABA	A	33	1	4,5,6	0.96±0.01	0±0 (0±0%)
1	ABA	A	12	1	4,5,6	0.97±0.01	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
1	ABA	A	31	1	1,5,7	0.52±0.01	0±0 (0±0%)
1	DHA	A	14	1	2,4,6	2.40±0.01	2±0 (100±0%)
1	DBU	A	15	1	2,5,7	4.61±0.01	2±0 (100±0%)
1	ABA	A	33	1	1,5,7	0.52±0.01	0±0 (0±0%)
1	ABA	A	12	1	1,5,7	0.51±0.01	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
1	ABA	A	31	1	1±0,1,1,2	0±0,3,4,6	-
1	ABA	A	12	1	1±0,1,1,2	0±0,3,4,6	-
1	ABA	A	33	1	1±0,1,1,2	0±0,3,4,6	-
1	DBU	A	15	1	-	0±0,1,4,6	-
1	DHA	A	14	1	-	0±0,0,2,4	-

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
1	A	15	DBU	C-CA	4.49	1.52	1.45	16	20
1	A	14	DHA	C-CA	2.12	1.41	1.45	16	20

All 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	15	DBU	CG-CB-CA	5.70	119.00	126.38	12	20
1	A	15	DBU	O-C-CA	3.26	121.25	125.39	20	20
1	A	14	DHA	CB-CA-N	2.47	119.97	125.81	17	20
1	A	14	DHA	O-C-CA	2.38	121.10	125.54	4	20

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

Mol	Chain	Res	Type	Atoms	Models (Total)
1	A	12	ABA	CA	20
1	A	31	ABA	CA	20
1	A	33	ABA	CA	20

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 81% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

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	321
Number of shifts mapped to atoms	321
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	25	2.35 ± 0.61	Should be checked
$^{13}C_{\beta}$	23	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	0	—	None (insufficient data)

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 81%, i.e. 264 atoms were assigned a chemical shift out of a possible 324. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	75/123 (61%)	50/50 (100%)	25/50 (50%)	0/23 (0%)
Sidechain	183/192 (95%)	127/127 (100%)	56/60 (93%)	0/5 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	6/9 (67%)	4/4 (100%)	2/5 (40%)	0/0 (—%)
Overall	264/324 (81%)	181/181 (100%)	83/115 (72%)	0/28 (0%)

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	35	GLU	CG	25.07	30.20 – 42.01	-9.3

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: