WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Mar 5, 2022 – 11:29 AM EST

PDB ID	:	2K3P
Title	:	Solution structure of the C-terminal domain (TUSP1-C) of the egg case silk from Nephila antipodiana
Authors	:	Lin, Z.; Huang, W.; Fan, J.; Yang, D.
Deposited on	:	2008-05-15

This is a Full wwPDB NMR Structure Validation 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
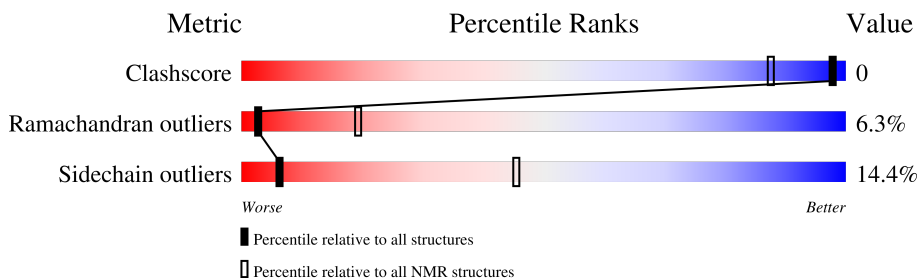with specific help available everywhere you see the (i) symbol.

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 was not calculated.

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		0
Ramachandran outliers		6.3%
Sidechain outliers		14.4%

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	161	76% 9% 15%

2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 9 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:3-A:21 (19)	1.75	1
2	A:44-A:161 (118)	1.22	9

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 2 clusters and 2 single-model clusters were found.

Cluster number	Models
1	1, 4, 5, 6, 8
2	2, 9, 10
Single-model clusters	3; 7

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2264 atoms, of which 1150 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TuSp1.

Mol	Chain	Residues	Atoms					Trace
1	A	161	Total	C	H	N	O	0
			2264	692	1150	190	232	

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.
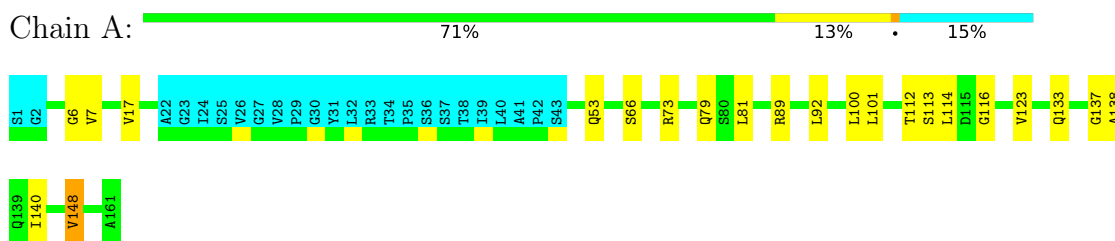
• Molecule 1: TuSp1

Chain A:
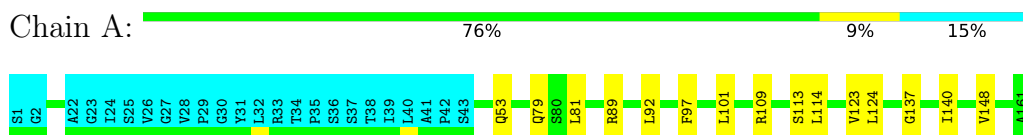
4.2 Scores per residue for each member of the ensemble
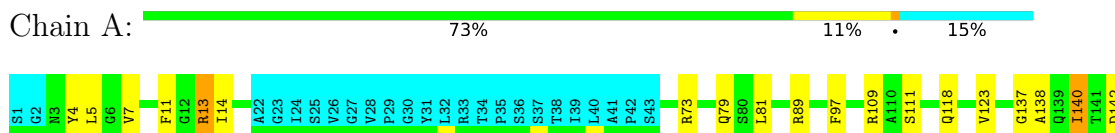
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: TuSp1

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: TuSp1

Chain A:

A161

4.2.3 Score per residue for model 3

• Molecule 1: TuSp1

Chain A:
71%   14%   15%

S1 G2 V7 R13 G20 T21 A22 G23 I24 S25 V26 G27 V28 P29 G30 Y31 L32 R33 T34 P35 S36 S37 T38 I39 L40 A41 P42 S43 N44 Q53 L56 L65 R89 L101 S113 L114 D115 G116 V123 L124 L132 Q133 N136 G137 S146 S147 V148

V151 N152 A161
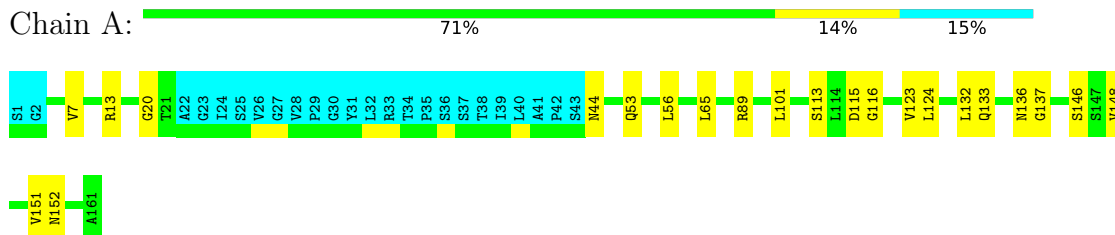
4.2.4 Score per residue for model 4

• Molecule 1: TuSp1

Chain A:
67%   17%   •   15%

S1 G2 L5 G20 T21 A22 G23 I24 S25 V26 G27 V28 P29 G30 Y31 L32 R33 T34 P35 S36 S37 T38 I39 L40 A41 P42 S43 T54 S67 R73 L77 A78 Q79 S83 L92 T96 F97 L101 I104 I108 S111 D115 V123 L124

L132 Q133 V134 G137 A138 Q139 I140 T141 D142 V148 P149 N152 L155 V160 A161
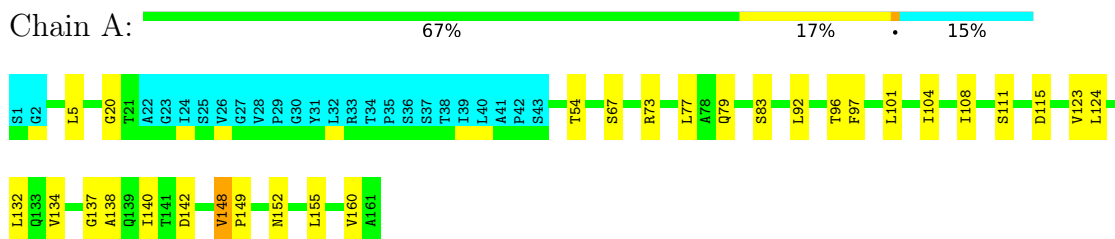
4.2.5 Score per residue for model 5

• Molecule 1: TuSp1

Chain A:
70%   16%   15%

S1 G2 L5 G6 V7 N10 I14 T18 A22 G23 I24 S25 V26 G27 V28 P29 G30 Y31 L32 R33 T34 P35 S36 S37 T38 I39 L40 A41 P42 S43 Q53 Q79 S80 L81 A82 S83 R89 L92 F97 L101 S105 T112 S113 L114 V123 L124

V134 G137 I140 P149 S150 L155 A161
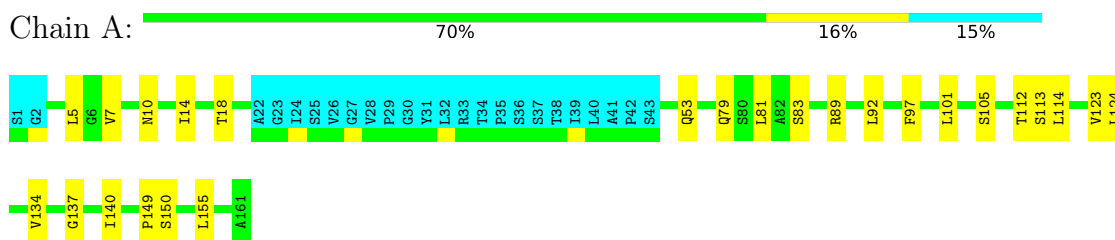
4.2.6 Score per residue for model 6

• Molecule 1: TuSp1

Chain A:
67%   16%   •   15%

S1 G2 N3 Y4 I14 A15 P16 G19 G20 T21 A22 G23 I24 S25 V26 G27 V28 P29 G30 Y31 L32 R33 T34 P35 S36 S37 T38 I39 L40 A41 P42 S43 N44 I47 Q53 T54 S76 L77 A78 Q79 S80 L81 S88 R89 L92 F97 L101 V123 L124

L132 G137 I140 S147 V148 P149 S150 V160 A161
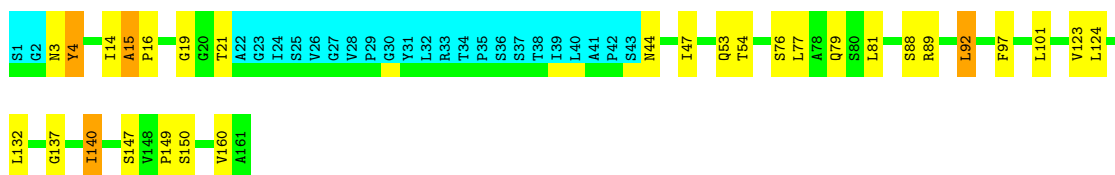
4.2.7 Score per residue for model 7

● Molecule 1: TuSp1

Chain A: 71% 14% 15%

S1 G2 V17 A22 G23 I24 S25 V26 G27 V28 P29 G30 Y31 L32 R33 T34 P35 S36 S37 T38 I39 L40 A41 P42 S43 L50 G51 L52 Q53 T54 L65 L81 S88 R89 L92 L101 R109 S113 L114 D115 V123 L124 L125 L132 Q133 G137

V148 N152 A161
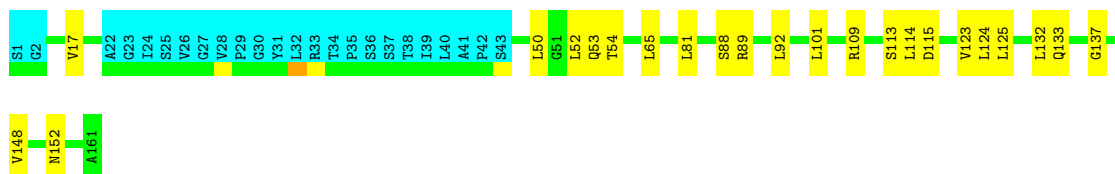
4.2.8 Score per residue for model 8

● Molecule 1: TuSp1

Chain A: 70% 14% · 15%

S1 G2 N3 Y4 L5 R13 G20 T21 A22 G23 I24 S25 V26 G27 V28 P29 G30 Y31 L32 R33 T34 P35 S36 S37 T38 I39 L40 A41 P42 S43 I47 L56 S63 Q64 L65 R73 Q79 S80 L81 L92 F97 L100 L101 R109 T112 S113 L114 D115

V134 G137 S147 V148 P149 S150 A161

4.2.9 Score per residue for model 9 (medoid)

● Molecule 1: TuSp1

Chain A: 69% 14% · 15%

S1 G2 L5 G6 V7 I14 T18 A22 G23 I24 S25 V26 G27 V28 P29 G30 Y31 L32 R33 T34 P35 S36 S37 T38 I39 L40 A41 P42 S43 Q53 T54 T55 L56 S66 S67 L77 A78 Q79 S80 L81 R89 L92 F97 L101 R109 S113 L114

V123 L124 G137 D142 V148 P149 L159 V160 A161
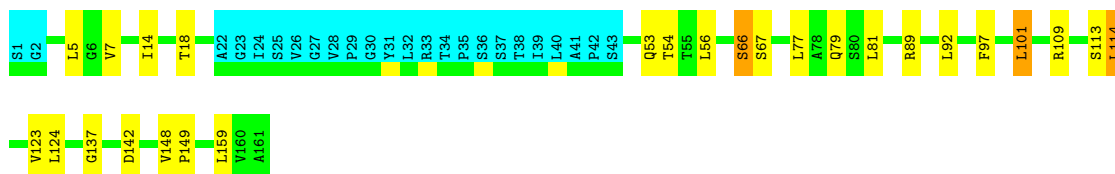
4.2.10 Score per residue for model 10

● Molecule 1: TuSp1

Chain A:
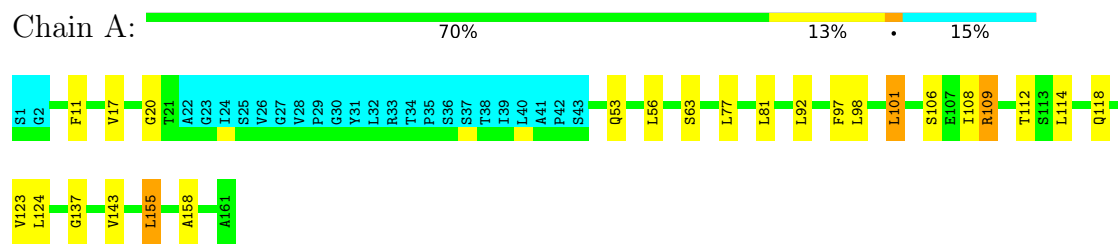
5 Refinement protocol and experimental data overview ⓘ

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

Of the 100 calculated structures, 10 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
Amber	refinement	7

No chemical shift data was provided.

6   Model quality ⓘ

6.1   Standard geometry ⓘ

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	0.64±0.01	0±0/959 (0.0± 0.0%)	1.04±0.03	0±0/1311 (0.0± 0.0%)
All	All	0.64	0/9590 (0.0%)	1.04	3/13110 (0.0%)

There are no bond-length outliers.

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	73	ARG	NE-CZ-NH1	8.97	124.78	120.30	1	2
1	A	13	ARG	NE-CZ-NH1	5.03	122.82	120.30	2	1

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	954	987	994	1±1
All	All	9540	9870	9940	7

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:97:PHE:O	1:A:101:LEU:HD22	0.57	1.99	9	3
1:A:106:SER:HA	1:A:112:THR:HG21	0.54	1.79	10	1
1:A:15:ALA:HB1	1:A:16:PRO:HD2	0.51	1.82	6	1
1:A:15:ALA:HB1	1:A:16:PRO:CD	0.42	2.45	6	1
1:A:155:LEU:HA	1:A:158:ALA:HB3	0.40	1.94	10	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	136/161 (84%)	101±3 (74±2%)	26±4 (19±3%)	9±3 (6±2%)	3	19
All	All	1360/1610 (84%)	1010 (74%)	264 (19%)	86 (6%)	3	19

All 34 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	137	GLY	10
1	A	113	SER	6
1	A	7	VAL	5
1	A	140	ILE	5
1	A	149	PRO	5
1	A	148	VAL	4
1	A	14	ILE	4
1	A	20	GLY	4
1	A	17	VAL	3
1	A	138	ALA	3
1	A	92	LEU	3
1	A	66	SER	2
1	A	116	GLY	2
1	A	111	SER	2
1	A	115	ASP	2
1	A	67	SER	2
1	A	108	ILE	2
1	A	18	THR	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	89	ARG	2
1	A	4	TYR	2
1	A	147	SER	2
1	A	109	ARG	2
1	A	6	GLY	1
1	A	5	LEU	1
1	A	112	THR	1
1	A	150	SER	1
1	A	3	ASN	1
1	A	15	ALA	1
1	A	19	GLY	1
1	A	21	THR	1
1	A	44	ASN	1
1	A	63	SER	1
1	A	114	LEU	1
1	A	143	VAL	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	109/127 (86%)	93±3 (86±2%)	16±3 (14±2%)	6	45
All	All	1090/1270 (86%)	933 (86%)	157 (14%)	6	45

All 53 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	101	LEU	9
1	A	123	VAL	9
1	A	81	LEU	8
1	A	53	GLN	7
1	A	79	GLN	7
1	A	124	LEU	7
1	A	92	LEU	6
1	A	89	ARG	5
1	A	114	LEU	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	5	LEU	4
1	A	97	PHE	4
1	A	109	ARG	4
1	A	56	LEU	4
1	A	132	LEU	4
1	A	54	THR	4
1	A	77	LEU	4
1	A	133	GLN	3
1	A	148	VAL	3
1	A	13	ARG	3
1	A	142	ASP	3
1	A	65	LEU	3
1	A	152	ASN	3
1	A	134	VAL	3
1	A	155	LEU	3
1	A	100	LEU	2
1	A	112	THR	2
1	A	4	TYR	2
1	A	11	PHE	2
1	A	73	ARG	2
1	A	118	GLN	2
1	A	140	ILE	2
1	A	83	SER	2
1	A	115	ASP	2
1	A	160	VAL	2
1	A	47	ILE	2
1	A	88	SER	2
1	A	150	SER	2
1	A	44	ASN	1
1	A	136	ASN	1
1	A	146	SER	1
1	A	151	VAL	1
1	A	96	THR	1
1	A	104	ILE	1
1	A	10	ASN	1
1	A	105	SER	1
1	A	76	SER	1
1	A	50	LEU	1
1	A	52	LEU	1
1	A	125	LEU	1
1	A	66	SER	1
1	A	159	LEU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	63	SER	1
1	A	98	LEU	1

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 (i)

No chemical shift data were provided