Full wwPDB NMR Structure Validation Report (i)

Mar 5, 2022 – 08:20 AM EST

PDB ID	:	2K3N
Title	:	Solution structure of the type 1 repetitive domain (TUSP1-RP1) of the egg case silk from Nephila Antipodiana
Authors	:	Lin, Z.; Huang, W.; Fan, J.; Yang, D.
Deposited on	:	2008-05-14

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.

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

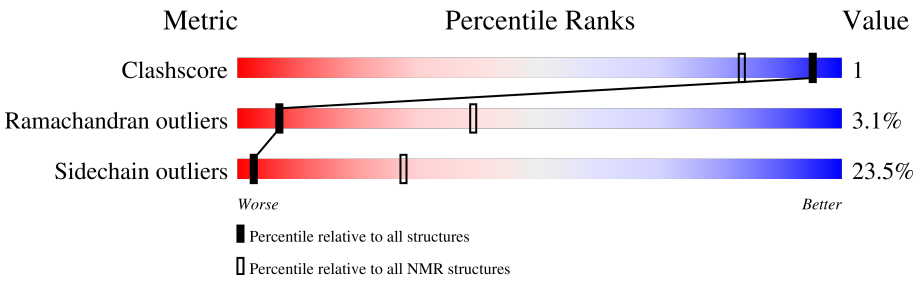
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 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	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	160	69% 10% 21%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 6 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:18-A:105, A:109-A:133 (113)	0.34	6
2	A:134-A:147 (14)	0.96	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. No single-model clusters were found.

Cluster number	Models
1	2, 3, 7, 8, 10
2	4, 6, 9
3	1, 5

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1956 atoms, of which 901 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TuSp1.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	160	1956	634	901	194	227	0

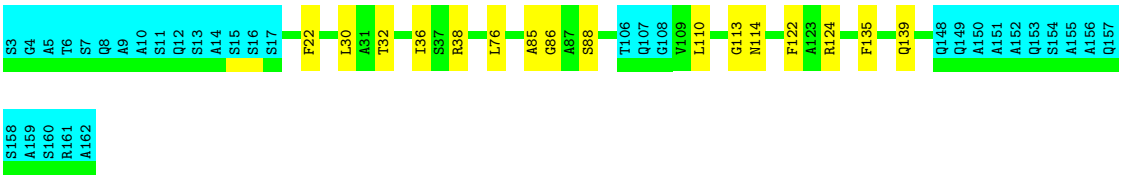
4 Residue-property plots (i)

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: TuSp1

Chain A: 69% 10% 21%



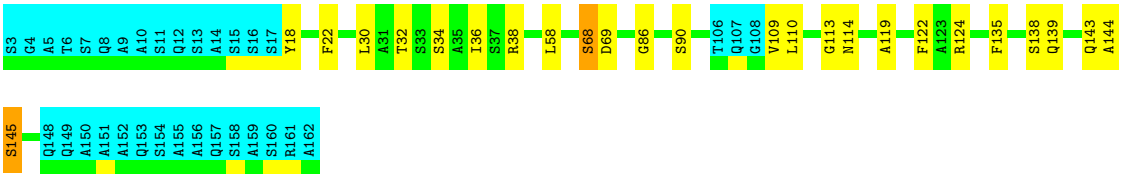
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: TuSp1

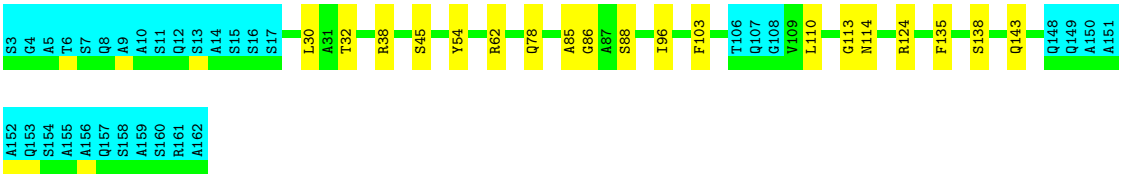
Chain A: 64% 14% • 21%



4.2.2 Score per residue for model 2

• Molecule 1: TuSp1

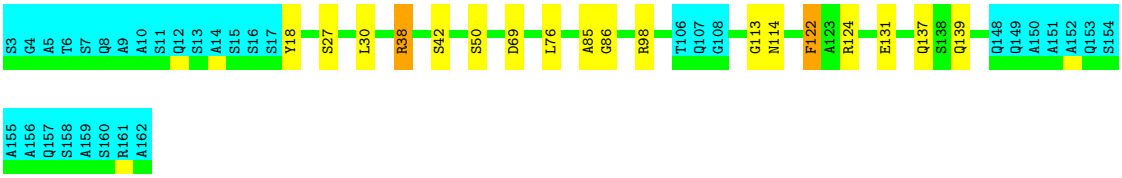
Chain A: 68% 12% 21%



4.2.3 Score per residue for model 3

- Molecule 1: TuSp1

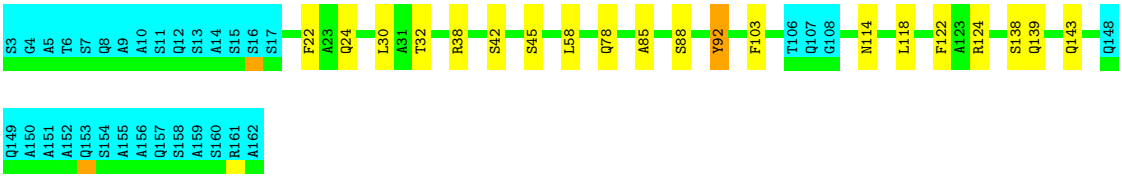
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: TuSp1

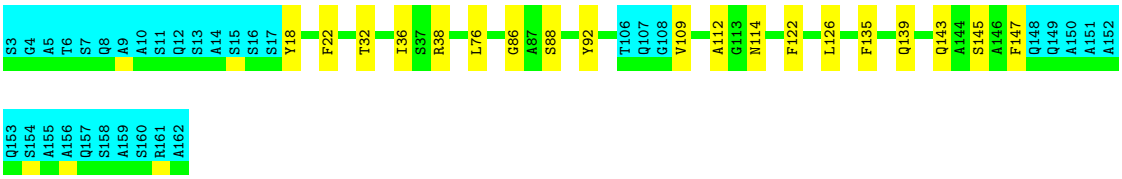
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: TuSp1

Chain A:



4.2.6 Score per residue for model 6 (medoid)

- Molecule 1: TuSp1

Chain A: 61% 18% • 21%

S3 G4 A5 T6 S7 Q8 A9 A10 S11 Q12 S13 A14 S15 S16 S17 F22 S27 S28 A29 L30 A31 T32 F40 V43 S44 S47 S50 L64 G84 A85 G86 A87 S88 N94 T106 Q107 G108 V109 L110 G113 N114 L118 A119 F122 A123 R124 A125

L126 S127 F135 Q139 S145 A146 F147 Q148 Q149 A150 A151 A152 Q153 S154 A155 A156 Q157 S158 A159 S160 R161 A162

4.2.7 Score per residue for model 7

- Molecule 1: TuSp1

Chain A: 58% 19% • 21%

S3 G4 A5 T6 S7 Q8 A9 A10 S11 Q12 S13 A14 S15 S16 S17 S27 L30 A31 T32 I36 S37 R38 V43 S44 S47 S50 A53 Y54 N55 L58 L72 L76 A77 Q78 A79 V80 A85 S88 Y92 Q102 F103 L104 A105 T106 Q107 G108

V109 L110 G113 N114 L118 F122 A123 R124 F135 Q139 Q143 A144 S145 A146 F147 Q148 Q149 A150 A151 A152 Q153 S154 A155 A156 Q157 S158 A159 S160 R161 A162

4.2.8 Score per residue for model 8

- Molecule 1: TuSp1

Chain A: 61% 18% • 21%

S3 G4 A5 T6 S7 Q8 A9 A10 S11 Q12 S13 A14 S15 S16 S17 F22 L30 S33 I36 S37 R38 S47 S50 A53 Y54 N55 L58 L76 A77 Q78 A79 V80 A85 G86 A87 S88 A89 S90 A91 Y92 F103 T106 Q107 G108 V109 L110 G113

N114 L118 R124 E131 S132 Q133 S134 F135 Q139 F147 Q148 Q149 A150 A151 A152 Q153 S154 A155 A156 Q157 S158 A159 S160 R161 A162

4.2.9 Score per residue for model 9

- Molecule 1: TuSp1

Chain A: 65% 13% • 21%

S3 G4 A5 T6 S7 Q8 A9 A10 S11 Q12 S13 A14 S15 S16 S17 F22 L30 A31 T32 S33 I36 S37 R38 A39 F40 A41 S42 S45 Y54 L58 A85 S88 G101 T106 Q107 G108 V109 L110 G113 N114 L118 F122 A123 R124 E131

S138 Q142 Q148 Q149 A150 A151 A152 Q153 S154 A155 A156 Q157 S158 A159 S160 R161 A162

4.2.10 Score per residue for model 10

- Molecule 1: TuSp1

Chain A:

5 Refinement protocol and experimental data overview (i)

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	7.0
NMRPipe	structure solution	2004
NMRView	structure solution	5.2.2
CYANA	structure solution	1.05

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	0.71±0.08	0±0/854 (0.0± 0.0%)	0.98±0.04	1±2/1160 (0.1± 0.2%)
All	All	0.71	1/8540 (0.0%)	0.98	7/11600 (0.1%)

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	0.4±0.5
All	All	0	4

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	131	GLU	CD-OE2	5.19	1.31	1.25	3	1

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	124	ARG	NE-CZ-NH1	6.20	123.40	120.30	3	1
1	A	69	ASP	CB-CG-OD1	5.98	123.68	118.30	3	1
1	A	38	ARG	NE-CZ-NH1	5.38	122.99	120.30	3	1
1	A	122	PHE	CB-CG-CD2	-5.26	117.12	120.80	9	2
1	A	98	ARG	NE-CZ-NH1	5.11	122.86	120.30	3	1
1	A	18	TYR	CB-CG-CD2	-5.01	117.99	121.00	3	1

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	92	TYR	Sidechain	4

6.2 Too-close contacts (i)

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	844	734	819	2±2
All	All	8440	7340	8190	17

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:30:LEU:HD23	1:A:123:ALA:HB1	0.51	1.82	10	1
1:A:53:ALA:HB1	1:A:76:LEU:HD23	0.50	1.84	7	2
1:A:22:PHE:CD2	1:A:119:ALA:HB2	0.49	2.42	1	2
1:A:53:ALA:CB	1:A:80:VAL:HG11	0.48	2.39	8	2
1:A:76:LEU:HD12	1:A:99:ALA:HB3	0.47	1.87	10	1
1:A:53:ALA:HB2	1:A:80:VAL:HG11	0.46	1.85	7	1
1:A:64:LEU:HD21	1:A:122:PHE:HA	0.46	1.86	6	1
1:A:22:PHE:CD2	1:A:110:LEU:HD11	0.46	2.45	6	1
1:A:18:TYR:CD1	1:A:112:ALA:HB2	0.44	2.47	5	1
1:A:22:PHE:CZ	1:A:101:GLY:HA3	0.43	2.49	9	1
1:A:104:LEU:HD21	1:A:122:PHE:CE2	0.43	2.49	7	1
1:A:30:LEU:CD1	1:A:123:ALA:HB1	0.41	2.45	7	1
1:A:40:PHE:CE2	1:A:89:ALA:HB2	0.41	2.50	10	1
1:A:30:LEU:CD2	1:A:123:ALA:HB1	0.40	2.47	10	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	127/160 (79%)	112±1 (88±1%)	11±2 (9±1%)	4±1 (3±1%)	7 39
All	All	1270/1600 (79%)	1122 (88%)	109 (9%)	39 (3%)	7 39

All 9 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	113	GLY	8
1	A	88	SER	8
1	A	86	GLY	7
1	A	85	ALA	7
1	A	144	ALA	3
1	A	145	SER	3
1	A	68	SER	1
1	A	139	GLN	1
1	A	84	GLY	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	74/94 (79%)	57±5 (76±6%)	17±5 (24±6%)	3 27
All	All	740/940 (79%)	566 (76%)	174 (24%)	3 27

All 52 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	114	ASN	10
1	A	30	LEU	9
1	A	32	THR	8
1	A	38	ARG	8
1	A	124	ARG	8
1	A	110	LEU	7
1	A	135	PHE	7
1	A	139	GLN	7
1	A	36	ILE	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	122	PHE	6
1	A	58	LEU	5
1	A	138	SER	5
1	A	143	GLN	5
1	A	50	SER	5
1	A	118	LEU	5
1	A	78	GLN	4
1	A	27	SER	4
1	A	76	LEU	4
1	A	22	PHE	4
1	A	147	PHE	4
1	A	45	SER	3
1	A	103	PHE	3
1	A	42	SER	3
1	A	126	LEU	3
1	A	47	SER	3
1	A	55	ASN	3
1	A	90	SER	2
1	A	109	VAL	2
1	A	145	SER	2
1	A	54	TYR	2
1	A	40	PHE	2
1	A	43	VAL	2
1	A	44	SER	2
1	A	33	SER	2
1	A	131	GLU	2
1	A	18	TYR	1
1	A	34	SER	1
1	A	68	SER	1
1	A	69	ASP	1
1	A	62	ARG	1
1	A	96	ILE	1
1	A	137	GLN	1
1	A	24	GLN	1
1	A	92	TYR	1
1	A	28	SER	1
1	A	94	ASN	1
1	A	127	SER	1
1	A	72	LEU	1
1	A	102	GLN	1
1	A	104	LEU	1
1	A	133	GLN	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	142	GLN	1

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 (i)

No chemical shift data were provided