



# Full wwPDB NMR Structure Validation Report ⓘ

May 28, 2020 – 10:34 pm BST

PDB ID : 2KFL  
Title : Tammar Wallaby Prion Protein (121-230)  
Authors : Christen, B.; Hornemann, S.; Damberger, F.F.; Wuthrich, K.  
Deposited on : 2009-02-24

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

Cyrange : Kirchner and Güntert (2011)  
NmrClust : Kelley et al. (1996)  
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 : 2.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

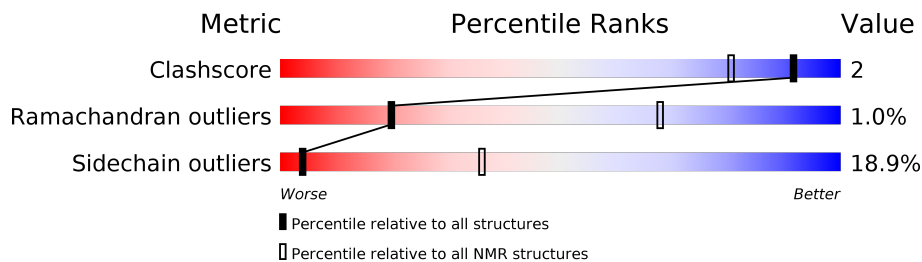
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 89%.

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	112	 74% 15% • 10%

## 2 Ensemble composition and analysis i

This entry contains 20 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: *fewest violations*.

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:126-A:226 (101)	0.33	6

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. No single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1776 atoms, of which 856 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Major prion protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	112	1776	571	856	160	181	8	0

There are 2 discrepancies between the modelled and reference sequences:

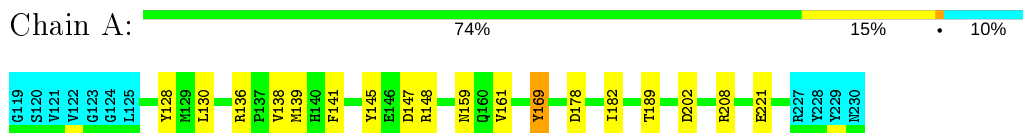
Chain	Residue	Modelled	Actual	Comment	Reference
A	119	GLY	-	expression tag	UNP Q58YZ3
A	120	SER	-	expression tag	UNP Q58YZ3

## 4 Residue-property plots

### 4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: Major prion protein

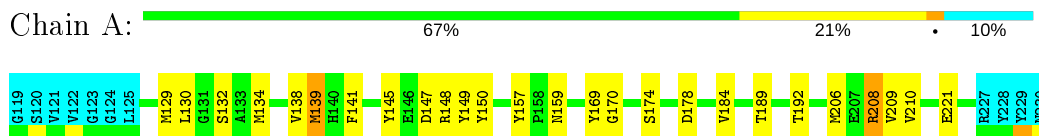


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

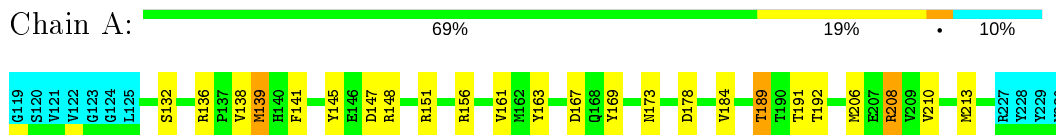
#### 4.2.1 Score per residue for model 1

- Molecule 1: Major prion protein



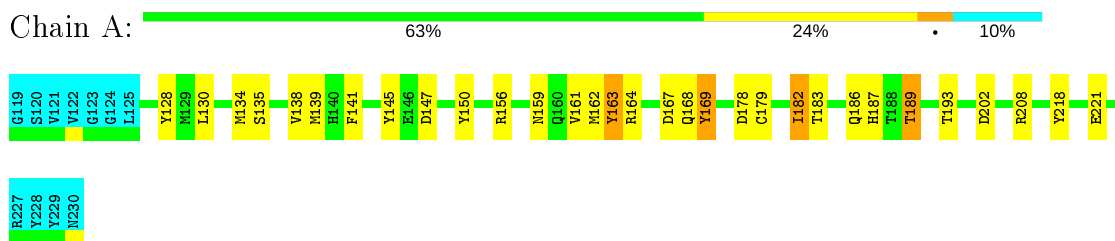
#### 4.2.2 Score per residue for model 2

- Molecule 1: Major prion protein



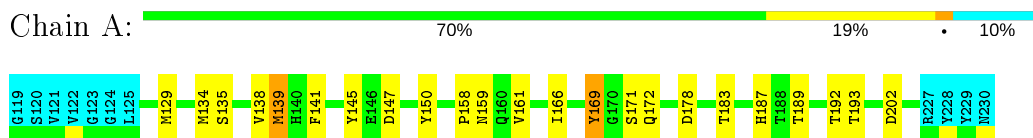
### 4.2.3 Score per residue for model 3

- Molecule 1: Major prion protein



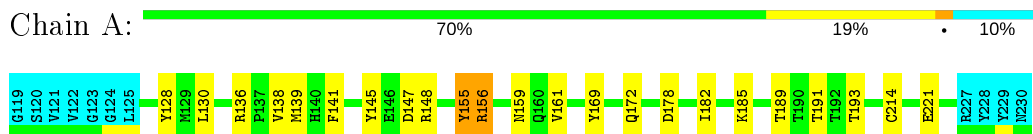
### 4.2.4 Score per residue for model 4

- Molecule 1: Major prion protein



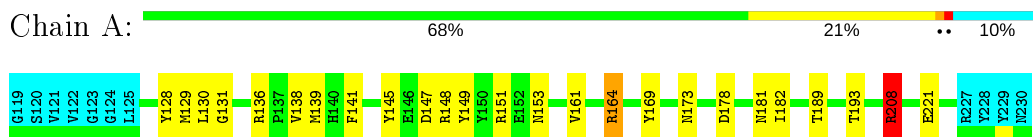
### 4.2.5 Score per residue for model 5

- Molecule 1: Major prion protein



### 4.2.6 Score per residue for model 6 (medoid)

- Molecule 1: Major prion protein



### 4.2.7 Score per residue for model 7

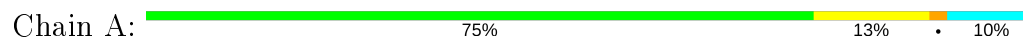
- Molecule 1: Major prion protein





#### 4.2.8 Score per residue for model 8

- Molecule 1: Major prion protein



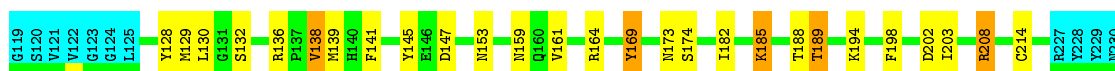
#### 4.2.9 Score per residue for model 9

- Molecule 1: Major prion protein



#### 4.2.10 Score per residue for model 10

- Molecule 1: Major prion protein



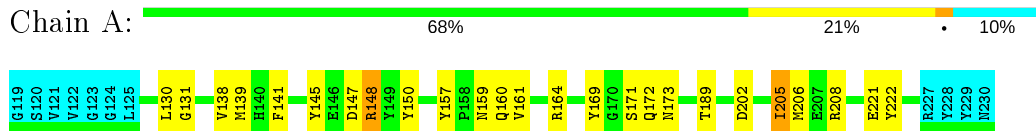
#### 4.2.11 Score per residue for model 11

- Molecule 1: Major prion protein



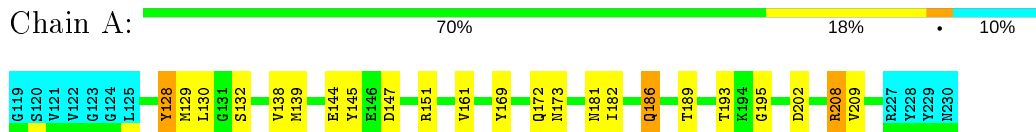
#### 4.2.12 Score per residue for model 12

- Molecule 1: Major prion protein



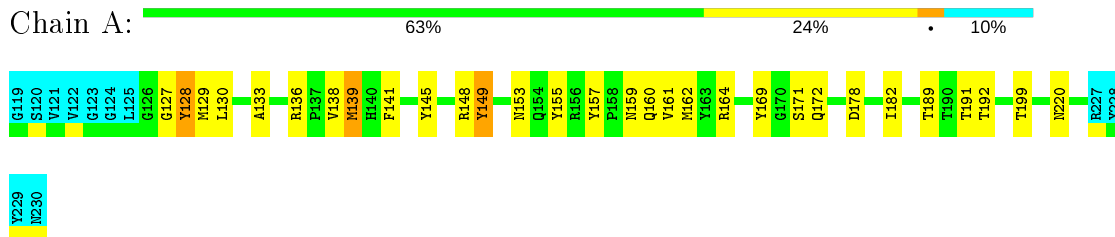
#### 4.2.13 Score per residue for model 13

- Molecule 1: Major prion protein



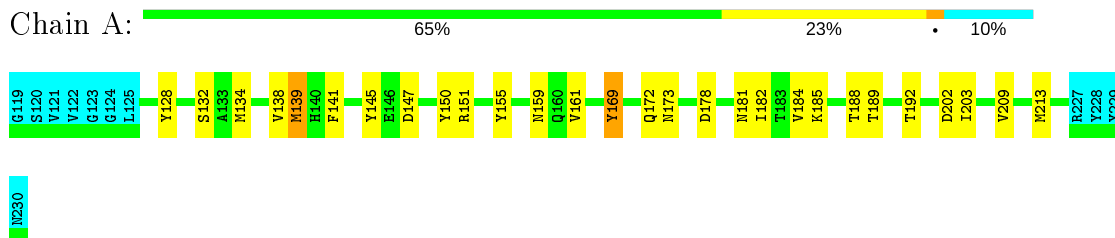
#### 4.2.14 Score per residue for model 14

- Molecule 1: Major prion protein



#### 4.2.15 Score per residue for model 15

- Molecule 1: Major prion protein



#### 4.2.16 Score per residue for model 16

- Molecule 1: Major prion protein







#### 4.2.17 Score per residue for model 17

- Molecule 1: Major prion protein



#### 4.2.18 Score per residue for model 18

- Molecule 1: Major prion protein



#### 4.2.19 Score per residue for model 19

- Molecule 1: Major prion protein



#### 4.2.20 Score per residue for model 20

- Molecule 1: Major prion protein



## 5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
DYANA	structure solution	1.0.3
OPALp	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1387
Number of shifts mapped to atoms	1387
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	89%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

### 5.1 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	837	776	776	2±2
All	All	16740	15520	15520	50

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:139:MET:HG2	1:A:141:PHE:CZ	0.57	2.34	5	7
1:A:188:THR:HG21	1:A:203:ILE:HD11	0.53	1.80	10	1
1:A:130:LEU:HD11	1:A:160:GLN:CB	0.52	2.34	12	1
1:A:130:LEU:HD11	1:A:160:GLN:HB2	0.51	1.82	12	1
1:A:139:MET:HG2	1:A:141:PHE:CE2	0.50	2.41	11	1
1:A:206:MET:O	1:A:210:VAL:HG22	0.48	2.09	2	2
1:A:139:MET:SD	1:A:141:PHE:CZ	0.47	3.07	1	5
1:A:139:MET:CG	1:A:141:PHE:CZ	0.47	2.98	20	9
1:A:188:THR:CG2	1:A:203:ILE:HD11	0.46	2.40	10	2
1:A:128:TYR:CD2	1:A:182:ILE:HG21	0.46	2.45	5	8
1:A:158:PRO:HG3	1:A:187:HIS:CD2	0.45	2.47	16	2
1:A:130:LEU:HD11	1:A:160:GLN:HB3	0.44	1.88	20	1
1:A:141:PHE:CZ	1:A:205:ILE:HG12	0.44	2.47	12	1
1:A:148:ARG:HH11	1:A:148:ARG:CB	0.43	2.26	12	1
1:A:133:ALA:HA	1:A:160:GLN:HG2	0.43	1.90	14	1
1:A:133:ALA:HB1	1:A:159:ASN:ND2	0.43	2.28	14	1
1:A:139:MET:HB3	1:A:150:TYR:CE2	0.43	2.48	15	1
1:A:185:LYS:O	1:A:189:THR:HG22	0.42	2.14	10	1
1:A:183:THR:O	1:A:187:HIS:CD2	0.42	2.73	4	2
1:A:180:VAL:O	1:A:184:VAL:HG12	0.41	2.15	17	1
1:A:166:ILE:HD11	1:A:226:GLN:HE22	0.41	1.75	19	1

## 5.2 Torsion angles [i](#)

### 5.2.1 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	101/112 (90%)	87±2 (86±2%)	13±2 (13±2%)	1±1 (1±1%)	20	68
All	All	2020/2240 (90%)	1746 (86%)	254 (13%)	20 (1%)	20	68

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	132	SER	8
1	A	131	GLY	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	170	GLY	2
1	A	169	TYR	2
1	A	198	PHE	2
1	A	127	GLY	1
1	A	128	TYR	1
1	A	195	GLY	1
1	A	184	VAL	1

## 5.2.2 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	92/100 (92%)	75±2 (81±2%)	17±2 (19±2%)	4	36
All	All	1840/2000 (92%)	1493 (81%)	347 (19%)	4	36

All 56 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	189	THR	20
1	A	145	TYR	20
1	A	138	VAL	20
1	A	169	TYR	20
1	A	147	ASP	16
1	A	130	LEU	15
1	A	161	VAL	14
1	A	202	ASP	12
1	A	178	ASP	12
1	A	139	MET	12
1	A	221	GLU	11
1	A	208	ARG	11
1	A	159	ASN	10
1	A	173	ASN	9
1	A	129	MET	9
1	A	193	THR	7
1	A	134	MET	7
1	A	172	GLN	7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	209	VAL	6
1	A	135	SER	6
1	A	148	ARG	6
1	A	214	CYS	6
1	A	171	SER	6
1	A	181	ASN	5
1	A	192	THR	5
1	A	191	THR	5
1	A	136	ARG	5
1	A	153	ASN	5
1	A	164	ARG	5
1	A	184	VAL	4
1	A	185	LYS	4
1	A	155	TYR	4
1	A	199	THR	3
1	A	213	MET	3
1	A	167	ASP	3
1	A	182	ILE	3
1	A	186	GLN	3
1	A	149	TYR	2
1	A	160	GLN	2
1	A	174	SER	2
1	A	179	CYS	2
1	A	194	LYS	2
1	A	206	MET	2
1	A	220	ASN	2
1	A	162	MET	2
1	A	226	GLN	2
1	A	205	ILE	1
1	A	151	ARG	1
1	A	166	ILE	1
1	A	212	GLN	1
1	A	196	GLU	1
1	A	219	GLN	1
1	A	144	GLU	1
1	A	218	TYR	1
1	A	163	TYR	1
1	A	175	PHE	1

### 5.2.3 RNA ⓘ

There are no RNA molecules in this entry.

### 5.3 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.4 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 5.5 Ligand geometry [i](#)

There are no ligands in this entry.

### 5.6 Other polymers [i](#)

There are no such molecules in this entry.

### 5.7 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Chemical shift validation [i](#)

The completeness of assignment taking into account all chemical shift lists is 89% for the well-defined parts and 89% for the entire structure.

### 6.1 Chemical shift list 1

File name: input\_cs.cif

Chemical shift list name: *assigned\_chem\_shift\_list\_1*

#### 6.1.1 Bookkeeping [i](#)

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	1387
Number of shifts mapped to atoms	1387
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

#### 6.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	112	$-0.29 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	103	$0.30 \pm 0.11$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	111	$2.23 \pm 0.09$	Should be applied
$^{15}\text{N}$	107	$0.15 \pm 0.23$	None needed ( $< 0.5$ ppm)

#### 6.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 89%, i.e. 1140 atoms were assigned a chemical shift out of a possible 1278. 8 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	489/499 (98%)	191/199 (96%)	201/202 (100%)	97/98 (99%)
Sidechain	558/651 (86%)	351/384 (91%)	184/229 (80%)	23/38 (61%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	93/128 (73%)	61/67 (91%)	32/58 (55%)	0/3 (0%)
Overall	1140/1278 (89%)	603/650 (93%)	417/489 (85%)	120/139 (86%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 89%, i.e. 1252 atoms were assigned a chemical shift out of a possible 1404. 11 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	542/554 (98%)	212/221 (96%)	223/224 (100%)	107/109 (98%)
Sidechain	605/706 (86%)	379/416 (91%)	201/248 (81%)	25/42 (60%)
Aromatic	105/144 (73%)	69/75 (92%)	36/66 (55%)	0/3 (0%)
Overall	1252/1404 (89%)	660/712 (93%)	460/538 (86%)	132/154 (86%)

#### 6.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

#### 6.1.5 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.

Random coil index (RCI) for chain A:

