W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB NMR Structure Validation Summary Report (i)

Jun 12, 2024 – 11:50 AM EDT

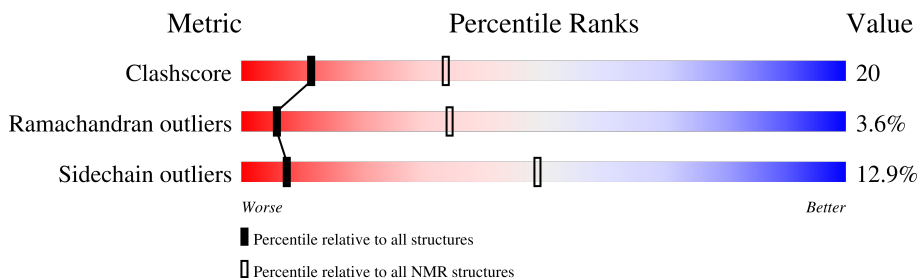|		
PDB ID	:	2KPZ
BMRB ID	:	16574
Title	:	Human NEDD4 3RD WW Domain Complex with The Human T-cell Leukemia virus 1 GAG-Pro poliprotein Derived Peptide SDPQIPPPYVEP
Authors	:	Iglesias-Bexiga, M.; Macias, M.; Bonet, R.; Blanco, F.J.; Cobos, E.S.; Luque, I.
Deposited on	:	2009-10-23

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 74%.

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		20
Ramachandran outliers		3.6%
Sidechain outliers		12.9%

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	49	41% 24% • • 31%
2	B	12	42% 17% 8% 8% 25%

2   Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 4 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:13-A:45, B:112-B:119 (41)	0.62	4

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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 718 atoms, of which 353 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called E3 ubiquitin-protein ligase NEDD4.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	34	573	189	282	54	48	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P46934
A	2	ALA	-	expression tag	UNP P46934
A	3	MET	-	expression tag	UNP P46934
A	4	GLY	-	expression tag	UNP P46934

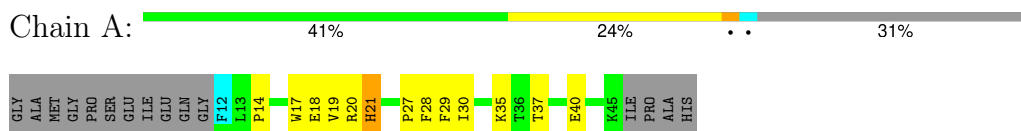- Molecule 2 is a protein called 12-mer from Gag-Pro polyprotein.

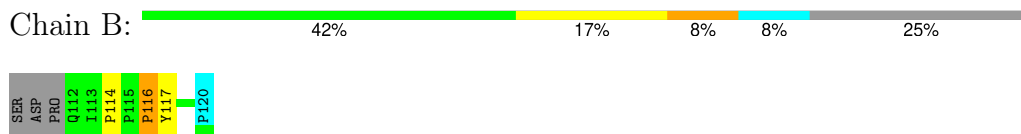| Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	9	145	50	71	10	14	0

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.

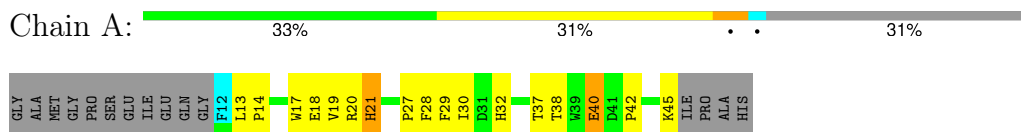• Molecule 1: E3 ubiquitin-protein ligase NEDD4

• Molecule 2: 12-mer from Gag-Pro polyprotein

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

The representative model is number 4. Colouring as in section 4.1 above.

• Molecule 1: E3 ubiquitin-protein ligase NEDD4

• Molecule 2: 12-mer from Gag-Pro polyprotein

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing.*

Of the 120 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
CNSSOLVE	structure solution	
CNSSOLVE	refinement	

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	567
Number of shifts mapped to atoms	448
Number of unparsed shifts	0
Number of shifts with mapping errors	119
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	74%

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	0.44±0.01	0±0/291 (0.0± 0.0%)	0.57±0.02	0±0/396 (0.0± 0.0%)
2	B	0.88±0.02	0±0/70 (0.0± 0.0%)	0.93±0.01	1±0/100 (0.9± 0.2%)
All	All	0.55	0/7220 (0.0%)	0.66	19/9920 (0.2%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models Worst	Total
2	B	116	PRO	CA-C-N	-5.22	105.71	117.20	18	19

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	280	273	271	13±2
2	B	66	64	63	4±1
All	All	6920	6740	6680	271

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 20.

5 of 61 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:37:THR:HB	2:B:114:PRO:HB2	0.96	1.33	12	16
1:A:30:ILE:HB	1:A:32:HIS:CE1	0.88	2.02	9	2
1:A:41:ASP:HB3	1:A:44:LEU:HD13	0.81	1.51	15	1
1:A:30:ILE:HG21	2:B:117:TYR:HB3	0.75	1.58	1	17
1:A:13:LEU:HD11	1:A:19:VAL:HG23	0.74	1.58	1	2

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	32/49 (65%)	29±1 (91±3%)	3±1 (8±2%)	0±0 (1±1%)	24	71
2	B	7/12 (58%)	6±0 (82±6%)	0±0 (1±4%)	1±0 (16±5%)	0	3
All	All	780/1220 (64%)	699 (90%)	53 (7%)	28 (4%)	6	34

All 3 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	116	PRO	20
1	A	35	LYS	5
2	B	113	ILE	3

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	30/41 (73%)	26±2 (85±6%)	4±2 (15±6%)	6	45
2	B	8/12 (67%)	8±0 (94±6%)	0±0 (6±6%)	21	70
All	All	760/1060 (72%)	662 (87%)	98 (13%)	7	49

5 of 19 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	21	HIS	18
1	A	28	PHE	17
1	A	40	GLU	12
1	A	35	LYS	7
1	A	36	THR	7

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 ⓘ

The completeness of assignment taking into account all chemical shift lists is 74% for the well-defined parts and 72% 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	567
Number of shifts mapped to atoms	448
Number of unparsed shifts	0
Number of shifts with mapping errors	119
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	9

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 119) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	5	PRO	HA	4.288	0.100	1
1	A	5	PRO	HB2	2.162	0.100	1
1	A	5	PRO	HB3	1.99475	0.100	1
1	A	5	PRO	HD2	3.775	0.100	2
1	A	5	PRO	HD3	3.54575	0.100	2
1	A	5	PRO	HG2	1.911	0.100	2
1	A	5	PRO	HG3	1.88725	0.100	2
1	A	5	PRO	CA	61.932	0.300	1
1	A	5	PRO	CB	29.847	0.300	1
1	A	5	PRO	CD	49.292	0.300	1
1	A	5	PRO	CG	24.986	0.300	1
1	A	6	SER	H	8.4615	0.100	1
1	A	6	SER	HA	4.35825	0.100	1
1	A	6	SER	HB2	3.835	0.100	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	6	SER	HB3	3.8135	0.100	1
1	A	6	SER	CA	57.07	0.300	1
1	A	6	SER	CB	61.932	0.300	1
1	A	6	SER	N	116.3945	0.300	1
1	A	7	GLU	H	8.3955	0.100	1
1	A	7	GLU	HA	4.177	0.100	1
1	A	7	GLU	HB2	1.886	0.100	1
1	A	7	GLU	HB3	1.895	0.100	1
1	A	7	GLU	HG2	2.177	0.100	2
1	A	7	GLU	HG3	2.0935	0.100	2
1	A	7	GLU	N	122.7715	0.300	1
1	A	8	ILE	H	8.034	0.100	1
1	A	8	ILE	HA	3.99975	0.100	1
1	A	8	ILE	HB	1.74125	0.100	1
1	A	8	ILE	HD11	0.6975	0.100	1
1	A	8	ILE	HD12	0.6975	0.100	1
1	A	8	ILE	HD13	0.6975	0.100	1
1	A	8	ILE	HG12	1.37025	0.100	2
1	A	8	ILE	HG13	1.07925	0.100	2
1	A	8	ILE	HG21	0.78425	0.100	1
1	A	8	ILE	HG22	0.78425	0.100	1
1	A	8	ILE	HG23	0.78425	0.100	1
1	A	8	ILE	CA	59.987	0.300	1
1	A	8	ILE	CB	36.653	0.300	1
1	A	8	ILE	CD1	10.402	0.300	1
1	A	8	ILE	CG1	25.958	0.300	1
1	A	8	ILE	CG2	15.263	0.300	1
1	A	8	ILE	N	120.66	0.300	1
1	A	9	GLU	H	8.202	0.100	1
1	A	9	GLU	HA	4.1545	0.100	1
1	A	9	GLU	HB2	1.918	0.100	1
1	A	9	GLU	HB3	1.8855	0.100	1
1	A	9	GLU	HG2	2.26	0.100	2
1	A	9	GLU	HG3	2.1785	0.100	2
1	A	9	GLU	N	124.2855	0.300	1
1	A	10	GLN	H	8.20325	0.100	1
1	A	10	GLN	HA	4.15175	0.100	1
1	A	10	GLN	HB2	2.01575	0.100	1
1	A	10	GLN	HB3	1.85525	0.100	1
1	A	10	GLN	HE21	7.1045	0.100	2
1	A	10	GLN	HE22	7.179	0.100	2

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	10	GLN	HG2	2.264	0.100	2
1	A	10	GLN	HG3	2.21975	0.100	2
1	A	10	GLN	CA	54.154	0.300	1
1	A	10	GLN	CB	27.902	0.300	1
1	A	10	GLN	CG	33.736	0.300	1
1	A	10	GLN	N	120.6115	0.300	1
1	A	10	GLN	NE2	112.7875	0.300	1
1	A	11	GLY	H	8.135	0.100	1
1	A	11	GLY	HA2	3.765	0.100	2
1	A	11	GLY	HA3	3.7525	0.100	2
1	A	11	GLY	CA	42.486	0.300	1
1	A	11	GLY	N	108.6245	0.300	1
1	A	46	ILE	H	7.9155	0.100	1
1	A	46	ILE	HA	4.28225	0.100	1
1	A	46	ILE	HB	1.732	0.100	1
1	A	46	ILE	HD11	0.78675	0.100	1
1	A	46	ILE	HD12	0.78675	0.100	1
1	A	46	ILE	HD13	0.78675	0.100	1
1	A	46	ILE	HG12	1.3955	0.100	2
1	A	46	ILE	HG13	1.0045	0.100	2
1	A	46	ILE	HG21	0.94175	0.100	1
1	A	46	ILE	HG22	0.94175	0.100	1
1	A	46	ILE	HG23	0.94175	0.100	1
1	A	46	ILE	CA	56.098	0.300	1
1	A	46	ILE	CB	34.708	0.300	1
1	A	46	ILE	N	124.489	0.300	1
1	A	47	PRO	HA	9.0145	0.100	1
1	A	47	PRO	HB2	2.151	0.100	1
1	A	47	PRO	HB3	1.9215	0.100	1
1	A	47	PRO	HD2	3.3295	0.100	2
1	A	47	PRO	HD3	3.40025	0.100	2
1	A	47	PRO	HG2	1.966	0.100	2
1	A	47	PRO	HG3	1.84975	0.100	2
1	A	47	PRO	CB	31.913	0.300	1
1	A	47	PRO	CD	49.292	0.300	1
1	A	48	ALA	H	8.34475	0.100	1
1	A	48	ALA	HA	4.154	0.100	1
1	A	48	ALA	HB1	1.324	0.100	1
1	A	48	ALA	HB2	1.324	0.100	1
1	A	48	ALA	HB3	1.324	0.100	1
1	A	48	ALA	CA	51.237	0.300	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	48	ALA	CB	15.263	0.300	1
1	A	48	ALA	N	125.376	0.300	1
1	A	49	HIS	H	7.84675	0.100	1
1	A	49	HIS	HA	4.36175	0.100	1
1	A	49	HIS	HB2	3.09175	0.100	1
1	A	49	HIS	HB3	2.9935	0.100	1
1	A	49	HIS	CA	55.126	0.300	1
1	A	49	HIS	CB	27.902	0.300	1
1	A	49	HIS	N	122.9415	0.300	1
1	B	109	SER	H	8.227	0.100	1
1	B	109	SER	HA	4.315	0.100	1
1	B	109	SER	HB2	3.723	0.100	1
1	B	109	SER	HB3	3.686	0.100	1
1	B	110	ASP	H	8.35	0.100	1
1	B	110	ASP	HA	4.771	0.100	1
1	B	110	ASP	HB2	2.652	0.100	1
1	B	110	ASP	HB3	2.392	0.100	1
1	B	111	PRO	HA	4.32	0.100	1
1	B	111	PRO	HB2	2.17	0.100	1
1	B	111	PRO	HB3	1.927	0.100	1
1	B	111	PRO	HD2	3.775	0.100	2
1	B	111	PRO	HD3	3.686	0.100	2
1	B	111	PRO	HG3	2.656	0.100	2

7.1.2   Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	40	2.66 ± 0.26	Should be applied
$^{13}C_{\beta}$	39	2.43 ± 0.47	Should be applied
$^{13}C'$	0	—	None (insufficient data)
^{15}N	39	0.38 ± 0.71	None needed (< 0.5 ppm)

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 74%, i.e. 437 atoms were assigned a chemical shift out of a possible 592. 0 out of 4 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

WORLDWIDE
PDB
PROTEIN DATA BANK

	Total	^{1}H	^{13}C	^{15}N
Backbone	137/193 (71%)	77/77 (100%)	31/82 (38%)	29/34 (85%)
Sidechain	269/330 (82%)	199/212 (94%)	66/103 (64%)	4/15 (27%)
Aromatic	31/69 (45%)	29/34 (85%)	0/29 (0%)	2/6 (33%)
Overall	437/592 (74%)	305/323 (94%)	97/214 (45%)	35/55 (64%)

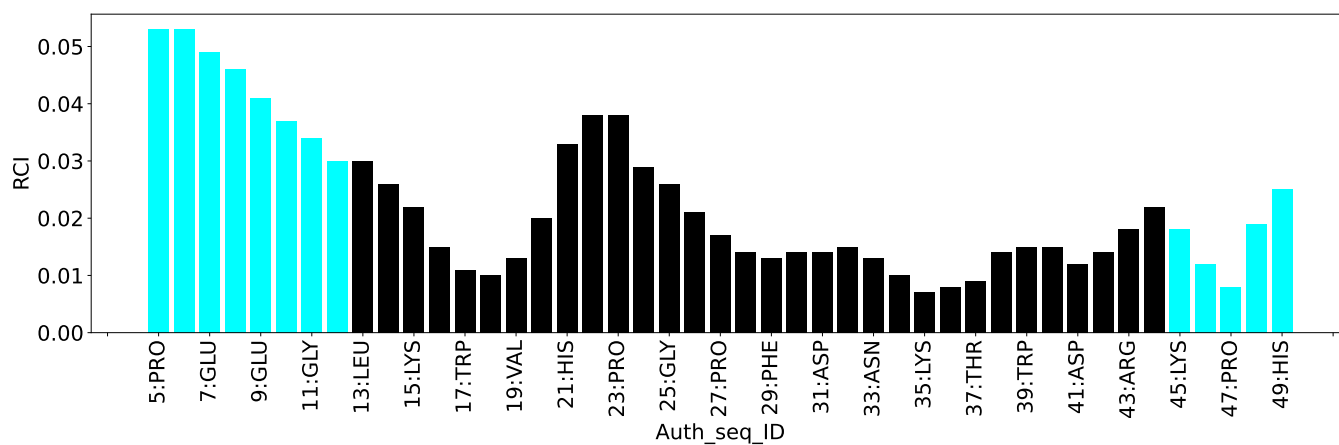### 7.1.4　Statistically unusual chemical shifts ⓘ

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	20	ARG	NE	112.90	76.53 − 92.65	17.6
1	A	43	ARG	NE	108.46	76.53 − 92.65	14.8
1	A	47	PRO	HA	9.01	2.78 − 6.00	14.4
1	A	31	ASP	HB3	-0.13	1.32 − 4.00	-10.4
1	A	40	GLU	CG	26.93	30.20 − 42.01	-7.8
1	A	41	ASP	HA	2.30	3.04 − 6.12	-7.4
1	A	26	ARG	HD3	1.54	1.81 − 4.39	-6.0
1	A	35	LYS	HA	2.04	2.15 − 6.37	-5.2
1	A	42	PRO	HG3	0.29	0.33 − 3.48	-5.1

7.1.5　Random Coil Index (RCI) plots ⓘ

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: