wwPDB NMR Structure Validation Summary Report

Apr 21, 2024 – 01:05 PM EDT

PDB ID	:	2MR3
BMRB ID	:	19219
Title	:	A subunit of 26S proteasome lid complex
Authors	:	Wu, Y.; Hu, Y.; Jin, C.
Deposited on	:	2014-06-30

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:

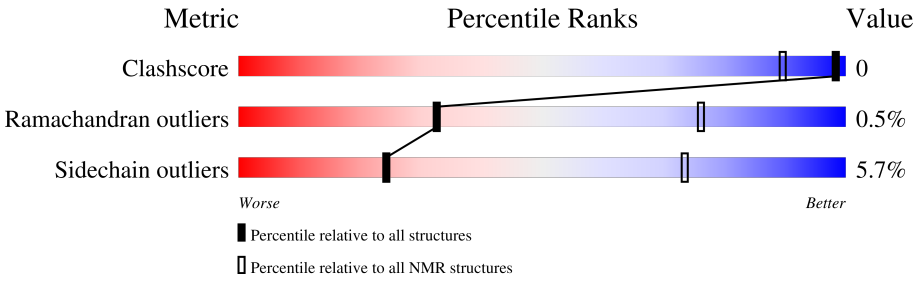
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 is 80%.

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	393	84% • 5% 9%

2 Ensemble composition and analysis

This entry contains 20 models. Model 8 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:6-A:110, A:124-A:356 (338)	2.74	8

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 and 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 5868 atoms, of which 2940 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called 26S proteasome regulatory subunit RPN9.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	356	5868	1889	2940	474	558	7	0

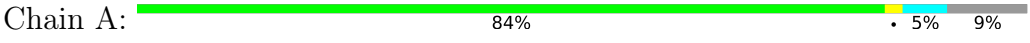
4 Residue-property plots

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: 26S proteasome regulatory subunit RPN9

Chain A:

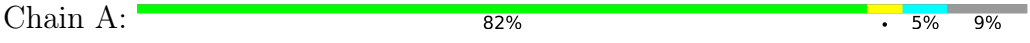


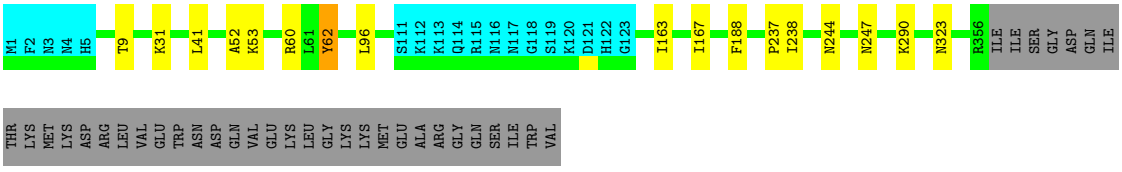
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 8. Colouring as in section 4.1 above.

• Molecule 1: 26S proteasome regulatory subunit RPN9

Chain A:





5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing.*

Of the 100 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
Amber	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	4165
Number of shifts mapped to atoms	3989
Number of unparsed shifts	0
Number of shifts with mapping errors	176
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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±0.00	0±0/2838 (0.0± 0.0%)	0.87±0.01	1±1/3838 (0.0± 0.0%)
All	All	0.64	0/56760 (0.0%)	0.87	23/76760 (0.0%)

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	1.6±1.2
All	All	0	31

There are no bond-length outliers.

5 of 13 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	356	ARG	NE-CZ-NH2	-7.38	116.61	120.30	16	2
1	A	70	TYR	CB-CG-CD2	-6.14	117.31	121.00	1	6
1	A	135	ARG	NE-CZ-NH2	-5.95	117.33	120.30	20	1
1	A	60	ARG	NE-CZ-NH2	-5.84	117.38	120.30	3	2
1	A	195	TYR	CB-CG-CD2	-5.71	117.58	121.00	7	1

There are no chirality outliers.

5 of 14 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	62	TYR	Sidechain	7
1	A	172	TYR	Sidechain	4
1	A	58	ARG	Sidechain	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	228	TYR	Sidechain	3
1	A	15	ARG	Sidechain	3

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	2782	2802	2802	1±1
All	All	55640	56040	56040	18

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.

5 of 12 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:172:TYR:CD1	1:A:191:THR:HG23	0.49	2.43	10	1
1:A:99:LEU:HB2	1:A:133:ILE:HD11	0.49	1.85	3	4
1:A:99:LEU:CB	1:A:133:ILE:HD11	0.46	2.40	20	3
1:A:228:TYR:CE1	1:A:290:LYS:HE3	0.44	2.48	11	1
1:A:337:LEU:CD1	1:A:350:ILE:HD12	0.43	2.44	12	2

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	337/393 (86%)	316±3 (94±1%)	19±3 (6±1%)	2±1 (1±0%)	32	76
All	All	6740/7860 (86%)	6328 (94%)	376 (6%)	36 (1%)	32	76

5 of 22 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	52	ALA	4
1	A	110	ASP	3
1	A	55	THR	3
1	A	162	SER	3
1	A	125	GLY	2

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	319/368 (87%)	301±4 (94±1%)	18±4 (6±1%)	24 73
All	All	6380/7360 (87%)	6014 (94%)	366 (6%)	24 73

5 of 115 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	238	ILE	20
1	A	9	THR	18
1	A	167	ILE	13
1	A	186	ASN	13
1	A	188	PHE	13

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

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 80% for the well-defined parts and 78% 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	4165
Number of shifts mapped to atoms	3989
Number of unparsed shifts	0
Number of shifts with mapping errors	176
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

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 176) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	357	ILE	H	7.88	.	1
1	A	357	ILE	HA	4.085	.	1
1	A	357	ILE	HB	1.777	.	1
1	A	357	ILE	HG12	1.424	.	2
1	A	357	ILE	HG13	1.134	.	2
1	A	357	ILE	HG21	0.82	.	1
1	A	357	ILE	HG22	0.82	.	1
1	A	357	ILE	HG23	0.82	.	1
1	A	357	ILE	HD11	0.805	.	1
1	A	357	ILE	HD12	0.805	.	1
1	A	357	ILE	HD13	0.805	.	1
1	A	357	ILE	C	175.681	.	1
1	A	357	ILE	CA	61.45	.	1
1	A	357	ILE	CB	38.41	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	357	ILE	CG1	27.38	.	1
1	A	357	ILE	CG2	17.712	.	1
1	A	357	ILE	CD1	13.2	.	1
1	A	357	ILE	N	121.3	.	1
1	A	358	ILE	H	8.031	.	1
1	A	358	ILE	HA	4.25	.	1
1	A	358	ILE	HB	1.838	.	1
1	A	358	ILE	HG12	1.137	.	2
1	A	358	ILE	HG13	1.408	.	2
1	A	358	ILE	HG21	0.864	.	1
1	A	358	ILE	HG22	0.864	.	1
1	A	358	ILE	HG23	0.864	.	1
1	A	358	ILE	HD11	0.819	.	1
1	A	358	ILE	HD12	0.819	.	1
1	A	358	ILE	HD13	0.819	.	1
1	A	358	ILE	C	176.014	.	1
1	A	358	ILE	CA	60.516	.	1
1	A	358	ILE	CB	38.8	.	1
1	A	358	ILE	CG1	27.216	.	1
1	A	358	ILE	CG2	17.57	.	1
1	A	358	ILE	CD1	12.89	.	1
1	A	358	ILE	N	124.022	.	1
1	A	359	SER	H	8.302	.	1
1	A	359	SER	HA	4.471	.	1
1	A	359	SER	HB2	3.882	.	1
1	A	359	SER	C	174.83	.	1
1	A	359	SER	CA	58.33	.	1
1	A	359	SER	CB	63.26	.	1
1	A	359	SER	N	119.83	.	1
1	A	360	GLY	H	8.556	.	1
1	A	360	GLY	HA2	3.978	.	1
1	A	360	GLY	C	173.714	.	1
1	A	360	GLY	CA	45.63	.	1
1	A	360	GLY	N	110.956	.	1
1	A	361	ASP	H	8.222	.	1
1	A	361	ASP	HA	4.589	.	1
1	A	361	ASP	HB2	2.586	.	2
1	A	361	ASP	HB3	2.658	.	2
1	A	361	ASP	C	176.0	.	1
1	A	361	ASP	CA	54.44	.	1
1	A	361	ASP	CB	41.235	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	361	ASP	N	120.5	.	1
1	A	362	GLN	H	8.265	.	1
1	A	362	GLN	HA	4.313	.	1
1	A	362	GLN	HB2	2.101	.	2
1	A	362	GLN	HB3	1.974	.	2
1	A	362	GLN	HG2	2.31	.	1
1	A	362	GLN	HE21	7.31	.	2
1	A	362	GLN	HE22	6.798	.	2
1	A	362	GLN	C	175.681	.	1
1	A	362	GLN	CA	55.82	.	1
1	A	362	GLN	CB	29.33	.	1
1	A	362	GLN	CG	33.73	.	1
1	A	362	GLN	N	120.28	.	1
1	A	362	GLN	NE2	110.84	.	1
1	A	363	ILE	H	8.199	.	1
1	A	363	ILE	HA	4.232	.	1
1	A	363	ILE	HB	1.902	.	1
1	A	363	ILE	HG12	1.169	.	2
1	A	363	ILE	HG13	1.479	.	2
1	A	363	ILE	HG21	0.906	.	1
1	A	363	ILE	HG22	0.906	.	1
1	A	363	ILE	HG23	0.906	.	1
1	A	363	ILE	HD11	0.822	.	1
1	A	363	ILE	HD12	0.822	.	1
1	A	363	ILE	HD13	0.822	.	1
1	A	363	ILE	C	176.426	.	1
1	A	363	ILE	CA	61.49	.	1
1	A	363	ILE	CB	38.65	.	1
1	A	363	ILE	CG1	27.24	.	1
1	A	363	ILE	CG2	17.688	.	1
1	A	363	ILE	CD1	12.89	.	1
1	A	363	ILE	N	123.14	.	1
1	A	372	GLU	H	7.85	.	1
1	A	372	GLU	CA	56.96	.	1
1	A	372	GLU	CB	28.22	.	1
1	A	372	GLU	N	120.22	.	1
1	A	373	TRP	H	8.275	.	1
1	A	373	TRP	HE1	10.078	.	1
1	A	373	TRP	CA	55.8	.	1
1	A	373	TRP	CB	29.86	.	1
1	A	373	TRP	N	121.592	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	373	TRP	NE1	129.74	.	1
1	A	374	ASN	H	8.446	.	1
1	A	374	ASN	CA	53.08	.	1
1	A	374	ASN	CB	38.93	.	1
1	A	374	ASN	N	118.9	.	1
1	A	375	ASP	H	7.981	.	1
1	A	375	ASP	CA	54.83	.	1
1	A	375	ASP	CB	40.16	.	1
1	A	375	ASP	N	124.65	.	1
1	A	376	GLN	H	8.312	.	1
1	A	376	GLN	CA	56.74	.	1
1	A	376	GLN	CB	30.08	.	1
1	A	376	GLN	N	120.33	.	1
1	A	377	VAL	H	8.089	.	1
1	A	377	VAL	CA	62.75	.	1
1	A	377	VAL	N	120.51	.	1
1	A	378	GLU	H	7.946	.	1
1	A	378	GLU	CA	56.44	.	1
1	A	378	GLU	CB	28.8	.	1
1	A	378	GLU	N	120.52	.	1
1	A	379	LYS	H	8.36	.	1
1	A	379	LYS	CA	56.69	.	1
1	A	379	LYS	CB	32.48	.	1
1	A	379	LYS	N	121.62	.	1
1	A	380	LEU	H	8.142	.	1
1	A	380	LEU	CA	55.74	.	1
1	A	380	LEU	CB	41.9	.	1
1	A	380	LEU	N	121.563	.	1
1	A	381	GLY	H	8.264	.	1
1	A	381	GLY	CA	45.54	.	1
1	A	381	GLY	N	108.71	.	1
1	A	382	LYS	H	8.012	.	1
1	A	382	LYS	CA	56.31	.	1
1	A	382	LYS	CB	32.08	.	1
1	A	382	LYS	N	120.78	.	1
1	A	383	LYS	H	8.142	.	1
1	A	383	LYS	CA	55.87	.	1
1	A	383	LYS	CB	31.11	.	1
1	A	383	LYS	N	121.34	.	1
1	A	384	MET	H	8.395	.	1
1	A	384	MET	CA	56.34	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	384	MET	CB	32.46	.	1
1	A	384	MET	N	121.55	.	1
1	A	385	GLU	H	8.41	.	1
1	A	385	GLU	CA	56.71	.	1
1	A	385	GLU	CB	30.14	.	1
1	A	385	GLU	N	122.36	.	1
1	A	386	ALA	H	8.312	.	1
1	A	386	ALA	CA	52.58	.	1
1	A	386	ALA	CB	18.91	.	1
1	A	386	ALA	N	125.16	.	1
1	A	387	ARG	H	8.314	.	1
1	A	387	ARG	CA	56.21	.	1
1	A	387	ARG	CB	30.37	.	1
1	A	387	ARG	N	120.34	.	1
1	A	388	GLY	H	8.422	.	1
1	A	388	GLY	CA	45.35	.	1
1	A	388	GLY	N	109.82	.	1
1	A	389	GLN	H	8.228	.	1
1	A	389	GLN	CA	55.34	.	1
1	A	389	GLN	CB	29.56	.	1
1	A	389	GLN	N	119.975	.	1
1	A	390	SER	H	8.358	.	1
1	A	390	SER	CA	58.31	.	1
1	A	390	SER	CB	63.58	.	1
1	A	390	SER	N	117.1	.	1
1	A	391	ILE	H	8.102	.	1
1	A	391	ILE	CA	61.69	.	1
1	A	391	ILE	CB	37.65	.	1
1	A	391	ILE	N	121.592	.	1
1	A	392	TRP	H	8.164	.	1
1	A	392	TRP	HE1	10.141	.	1
1	A	392	TRP	CA	56.71	.	1
1	A	392	TRP	CB	29.03	.	1
1	A	392	TRP	N	121.652	.	1
1	A	392	TRP	NE1	129.732	.	1
1	A	393	VAL	H	7.92	.	1
1	A	393	VAL	CA	63.91	.	1
1	A	393	VAL	CB	31.92	.	1
1	A	393	VAL	N	120.172	.	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	383	-0.64 ± 0.05	Should be checked
$^{13}C_{\beta}$	371	0.40 ± 0.05	None needed (< 0.5 ppm)
$^{13}C'$	344	-0.31 ± 0.04	None needed (< 0.5 ppm)
^{15}N	370	0.14 ± 0.16	None needed (< 0.5 ppm)

7.1.3 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 80%, i.e. 3859 atoms were assigned a chemical shift out of a possible 4842. 0 out of 69 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	1659/1678 (99%)	668/673 (99%)	663/676 (98%)	328/329 (100%)
Sidechain	2122/2740 (77%)	1404/1782 (79%)	698/870 (80%)	20/88 (23%)
Aromatic	78/424 (18%)	43/208 (21%)	32/207 (15%)	3/9 (33%)
Overall	3859/4842 (80%)	2115/2663 (79%)	1393/1753 (79%)	351/426 (82%)

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	349	THR	HG1	5.42	0.08 – 2.19	20.3
1	A	40	GLN	HB3	0.43	0.71 – 3.33	-6.0
1	A	20	PRO	HA	2.53	2.78 – 6.00	-5.8
1	A	346	GLU	HA	2.03	2.24 – 6.23	-5.5
1	A	36	LYS	HA	1.94	2.15 – 6.37	-5.5
1	A	238	ILE	H	11.83	4.90 – 11.63	5.3

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: