wwPDB NMR Structure Validation Summary Report

Jun 12, 2024 – 07:56 AM EDT

PDB ID	:	1QJT
BMRB ID	:	4491
Title	:	SOLUTION STRUCTURE OF THE APO EH1 DOMAIN OF MOUSE EPIDERMAL GROWTH FACTOR RECEPTOR SUBSTRATE 15, EPS15
Authors	:	Whitehead, B.; Tessari, M.; Carotenuto, A.; van Bergen en Henegouwen, P.M.; Vuister, G.W.
Deposited on	:	1999-07-02

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:

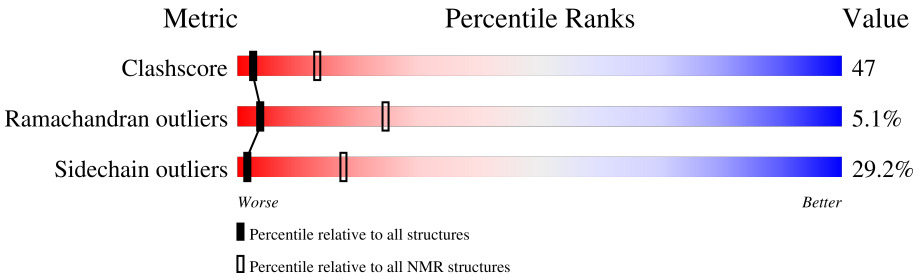
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 90%.

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	99	28% 56% 14% .

2 Ensemble composition and analysis (i)

This entry contains 30 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:9-A:105 (97)	0.65	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 4 clusters and 3 single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 6, 7, 8, 11, 12, 13, 14, 21, 22, 23, 25, 26, 28, 29
2	15, 16, 17, 19
3	5, 20, 27
4	18, 24
Single-model clusters	9; 10; 30

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1509 atoms, of which 757 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called EPIDERMAL GROWTH FACTOR RECEPTOR SUBSTRATE SUBSTRATE 15, EPS15.

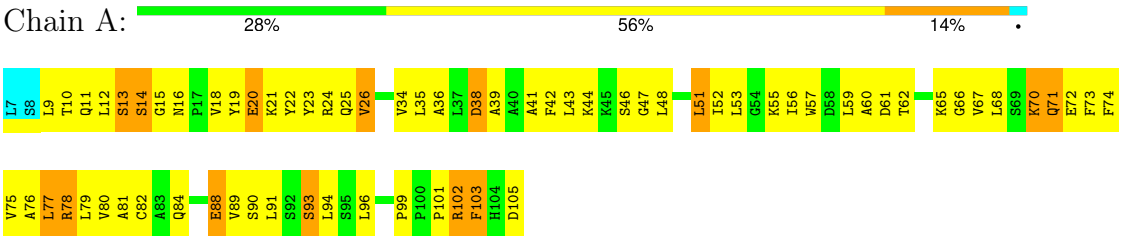
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	99	1509	483	757	127	141	1	0

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.

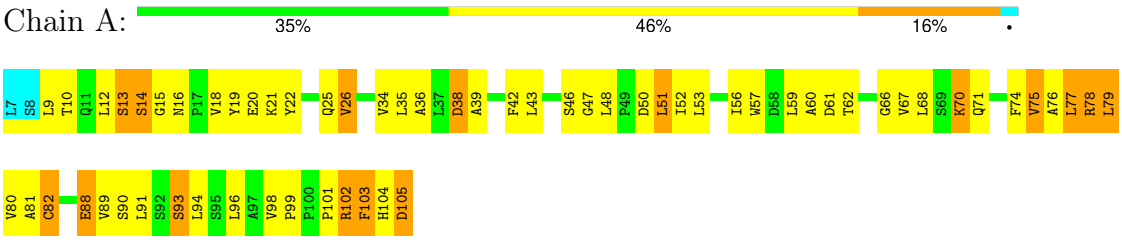
• Molecule 1: EPIDERMAL GROWTH FACTOR RECEPTOR SUBSTRATE SUBSTRATE 15, EPS15



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

The representative model is number 2. Colouring as in section 4.1 above.

• Molecule 1: EPIDERMAL GROWTH FACTOR RECEPTOR SUBSTRATE SUBSTRATE 15, EPS15



5 Refinement protocol and experimental data overview

The models were refined using the following method: *TAD*.

Of the 100 calculated structures, 30 were deposited, based on the following criterion: *LOWEST OVERALL TARGET FUNCTION*.

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

Software name	Classification	Version
DYANA	refinement	1.5
NMRPipe	structure solution	
XEASY	structure solution	
DYANA	structure solution	1.5

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	1399
Number of shifts mapped to atoms	1206
Number of unparsed shifts	0
Number of shifts with mapping errors	193
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

6 Model quality

6.1 Standard geometry

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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	738	741	753	70±5
All	All	22140	22230	22590	2111

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 47.

5 of 375 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:9:LEU:HD13	1:A:74:PHE:CE1	0.97	1.95	20	6
1:A:48:LEU:HD12	1:A:80:VAL:HG13	0.96	1.32	17	23
1:A:28:ALA:HB3	1:A:31:THR:HG21	0.95	1.39	29	1
1:A:9:LEU:HD13	1:A:74:PHE:CZ	0.91	2.01	30	2
1:A:59:LEU:HD12	1:A:79:LEU:HD11	0.90	1.41	12	9

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	96/99 (97%)	69±3 (72±3%)	22±3 (23±3%)	5±1 (5±1%)	4 24
All	All	2880/2970 (97%)	2076 (72%)	657 (23%)	147 (5%)	4 24

5 of 16 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	103	PHE	29
1	A	70	LYS	20
1	A	62	THR	16
1	A	88	GLU	15
1	A	13	SER	13

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	79/81 (98%)	56±3 (71±3%)	23±3 (29±3%)	1 17
All	All	2370/2430 (98%)	1679 (71%)	691 (29%)	1 17

5 of 57 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	14	SER	30
1	A	20	GLU	30
1	A	51	LEU	30
1	A	13	SER	29
1	A	26	VAL	28

6.3.3 RNA (i)

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 90% for the well-defined parts and 91% 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	1399
Number of shifts mapped to atoms	1206
Number of unparsed shifts	0
Number of shifts with mapping errors	193
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

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 193) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	1	MET	C	176.1	.	1
1	A	2	ALA	H	8.33	.	1
1	A	2	ALA	HA	4.24	.	1
1	A	2	ALA	HB1	1.38	.	1
1	A	2	ALA	HB2	1.38	.	1
1	A	2	ALA	HB3	1.38	.	1
1	A	2	ALA	C	177.8	.	1
1	A	2	ALA	CA	52.6	.	1
1	A	2	ALA	CB	18.7	.	1
1	A	2	ALA	N	125.6	.	1
1	A	3	ALA	H	8.22	.	1
1	A	3	ALA	HA	4.18	.	1
1	A	3	ALA	HB1	1.39	.	1
1	A	3	ALA	HB2	1.39	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	3	ALA	HB3	1.39	.	1
1	A	3	ALA	C	178.3	.	1
1	A	3	ALA	CA	52.6	.	1
1	A	3	ALA	CB	18.9	.	1
1	A	3	ALA	N	123.2	.	1
1	A	4	ALA	H	8.27	.	1
1	A	4	ALA	HA	4.18	.	1
1	A	4	ALA	HB1	1.39	.	1
1	A	4	ALA	HB2	1.39	.	1
1	A	4	ALA	HB3	1.39	.	1
1	A	4	ALA	C	178.0	.	1
1	A	4	ALA	CA	52.9	.	1
1	A	4	ALA	CB	18.6	.	1
1	A	4	ALA	N	122.2	.	1
1	A	5	ALA	H	7.98	.	1
1	A	5	ALA	HA	4.22	.	1
1	A	5	ALA	HB1	1.39	.	1
1	A	5	ALA	HB2	1.39	.	1
1	A	5	ALA	HB3	1.39	.	1
1	A	5	ALA	C	178.0	.	1
1	A	5	ALA	CA	52.8	.	1
1	A	5	ALA	CB	18.7	.	1
1	A	5	ALA	N	120.4	.	1
1	A	6	GLN	H	8.01	.	1
1	A	6	GLN	HA	4.26	.	1
1	A	6	GLN	HB2	1.95	.	2
1	A	6	GLN	HB3	2.12	.	2
1	A	6	GLN	HG2	2.34	.	2
1	A	6	GLN	HE21	6.81	.	2
1	A	6	GLN	HE22	7.47	.	2
1	A	6	GLN	C	175.9	.	1
1	A	6	GLN	CA	55.4	.	1
1	A	6	GLN	CB	28.8	.	1
1	A	6	GLN	CG	33.5	.	1
1	A	6	GLN	CD	180.5	.	1
1	A	6	GLN	N	117.6	.	1
1	A	6	GLN	NE2	112.4	.	1
1	A	106	SER	H	8.45	.	1
1	A	106	SER	HA	4.47	.	1
1	A	106	SER	HB2	3.89	.	2
1	A	106	SER	C	174.6	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	106	SER	CA	58.4	.	1
1	A	106	SER	CB	63.7	.	1
1	A	106	SER	N	116.3	.	1
1	A	107	SER	H	8.46	.	1
1	A	107	SER	HA	4.49	.	1
1	A	107	SER	HB2	3.9	.	2
1	A	107	SER	C	174.2	.	1
1	A	107	SER	CA	58.3	.	1
1	A	107	SER	CB	63.8	.	1
1	A	107	SER	N	117.6	.	1
1	A	108	SER	H	8.18	.	1
1	A	108	SER	HA	4.78	.	1
1	A	108	SER	HB2	3.85	.	2
1	A	108	SER	CA	56.3	.	1
1	A	108	SER	CB	63.2	.	1
1	A	108	SER	N	118.5	.	1
1	A	109	PRO	HA	4.42	.	1
1	A	109	PRO	HB2	1.89	.	2
1	A	109	PRO	HB3	2.27	.	2
1	A	109	PRO	HG2	1.99	.	2
1	A	109	PRO	HD2	3.72	.	2
1	A	109	PRO	HD3	3.8	.	2
1	A	109	PRO	C	176.8	.	1
1	A	109	PRO	CA	62.9	.	1
1	A	109	PRO	CB	31.7	.	1
1	A	109	PRO	CG	27.0	.	1
1	A	109	PRO	CD	50.5	.	1
1	A	110	LEU	H	8.22	.	1
1	A	110	LEU	HA	4.28	.	1
1	A	110	LEU	HB2	1.58	.	2
1	A	110	LEU	HG	1.58	.	1
1	A	110	LEU	HD11	0.88	.	1
1	A	110	LEU	HD12	0.88	.	1
1	A	110	LEU	HD13	0.88	.	1
1	A	110	LEU	HD21	0.86	.	1
1	A	110	LEU	HD22	0.86	.	1
1	A	110	LEU	HD23	0.86	.	1
1	A	110	LEU	C	177.3	.	1
1	A	110	LEU	CA	55.2	.	1
1	A	110	LEU	CB	42.0	.	1
1	A	110	LEU	CG	27.1	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	110	LEU	CD1	24.7	.	1
1	A	110	LEU	CD2	23.3	.	1
1	A	110	LEU	N	121.7	.	1
1	A	111	LEU	H	8.2	.	1
1	A	111	LEU	HA	4.43	.	1
1	A	111	LEU	HB2	1.6	.	2
1	A	111	LEU	HB3	1.66	.	2
1	A	111	LEU	HG	1.6	.	1
1	A	111	LEU	HD11	0.91	.	1
1	A	111	LEU	HD12	0.91	.	1
1	A	111	LEU	HD13	0.91	.	1
1	A	111	LEU	HD21	0.84	.	1
1	A	111	LEU	HD22	0.84	.	1
1	A	111	LEU	HD23	0.84	.	1
1	A	111	LEU	C	177.5	.	1
1	A	111	LEU	CA	54.7	.	1
1	A	111	LEU	CB	42.1	.	1
1	A	111	LEU	CG	26.8	.	1
1	A	111	LEU	CD1	24.7	.	1
1	A	111	LEU	CD2	23.2	.	1
1	A	111	LEU	N	123.0	.	1
1	A	112	THR	H	8.07	.	1
1	A	112	THR	HA	4.39	.	1
1	A	112	THR	HB	4.26	.	1
1	A	112	THR	HG21	1.17	.	1
1	A	112	THR	HG22	1.17	.	1
1	A	112	THR	HG23	1.17	.	1
1	A	112	THR	C	174.5	.	1
1	A	112	THR	CA	61.3	.	1
1	A	112	THR	CB	69.6	.	1
1	A	112	THR	CG2	21.1	.	1
1	A	112	THR	N	114.1	.	1
1	A	113	SER	H	8.26	.	1
1	A	113	SER	HA	4.51	.	1
1	A	113	SER	HB2	3.88	.	2
1	A	113	SER	C	174.5	.	1
1	A	113	SER	CA	58.0	.	1
1	A	113	SER	CB	63.8	.	1
1	A	113	SER	N	117.5	.	1
1	A	114	GLY	H	8.23	.	1
1	A	114	GLY	HA2	4.14	.	2

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	114	GLY	HA3	4.1	.	2
1	A	114	GLY	CA	44.1	.	1
1	A	114	GLY	N	110.6	.	1
1	A	115	PRO	HA	4.47	.	1
1	A	115	PRO	HB2	2.0	.	2
1	A	115	PRO	HB3	2.26	.	2
1	A	115	PRO	HG2	2.0	.	2
1	A	115	PRO	HD2	3.61	.	2
1	A	115	PRO	C	176.2	.	1
1	A	115	PRO	CA	63.0	.	1
1	A	115	PRO	CB	31.9	.	1
1	A	115	PRO	CG	26.7	.	1
1	A	115	PRO	CD	49.6	.	1
1	A	116	SER	H	8.45	.	1
1	A	116	SER	HA	4.48	.	1
1	A	116	SER	HB2	3.88	.	2
1	A	116	SER	C	176.6	.	1
1	A	116	SER	CA	58.1	.	1
1	A	116	SER	CB	63.7	.	1
1	A	116	SER	N	116.3	.	1
1	A	117	VAL	H	8.11	.	1
1	A	117	VAL	HA	4.16	.	1
1	A	117	VAL	HB	2.08	.	1
1	A	117	VAL	HG11	0.91	.	1
1	A	117	VAL	HG12	0.91	.	1
1	A	117	VAL	HG13	0.91	.	1
1	A	117	VAL	HG21	0.9	.	1
1	A	117	VAL	HG22	0.9	.	1
1	A	117	VAL	HG23	0.9	.	1
1	A	117	VAL	C	175.8	.	1
1	A	117	VAL	CA	61.8	.	1
1	A	117	VAL	CB	32.8	.	1
1	A	117	VAL	CG1	20.7	.	1
1	A	117	VAL	CG2	19.9	.	1
1	A	117	VAL	N	121.5	.	1
1	A	118	ALA	H	8.27	.	1
1	A	118	ALA	HA	4.32	.	1
1	A	118	ALA	HB1	1.37	.	1
1	A	118	ALA	HB2	1.37	.	1
1	A	118	ALA	HB3	1.37	.	1
1	A	118	ALA	C	178.2	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	118	ALA	CA	52.2	.	1
1	A	118	ALA	CB	18.8	.	1
1	A	118	ALA	N	127.2	.	1
1	A	119	GLU	H	8.25	.	1
1	A	119	GLU	HA	4.28	.	1
1	A	119	GLU	HB2	1.95	.	2
1	A	119	GLU	HB3	2.11	.	2
1	A	119	GLU	HG3	2.34	.	2
1	A	119	GLU	C	175.3	.	1
1	A	119	GLU	CA	55.6	.	1
1	A	119	GLU	N	120.6	.	1
1	A	120	LEU	H	7.85	.	1
1	A	120	LEU	HA	4.19	.	1
1	A	120	LEU	CA	56.1	.	1
1	A	120	LEU	N	128.8	.	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
$^{13}C_{\alpha}$	119	0.10 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	108	0.47 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	111	-0.26 ± 0.09	None needed (< 0.5 ppm)
^{15}N	111	0.31 ± 0.30	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 90%, i.e. 1179 atoms were assigned a chemical shift out of a possible 1303. 0 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	473/483 (98%)	194/197 (98%)	188/194 (97%)	91/92 (99%)
Sidechain	656/734 (89%)	435/482 (90%)	210/227 (93%)	11/25 (44%)
Aromatic	50/86 (58%)	25/42 (60%)	24/42 (57%)	1/2 (50%)
Overall	1179/1303 (90%)	654/721 (91%)	422/463 (91%)	103/119 (87%)

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	70	LYS	HG2	-0.45	0.13 – 2.61	-7.3
1	A	13	SER	HA	2.47	2.50 – 6.44	-5.1

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: