Full wwPDB NMR Structure Validation Report (i)

Jun 17, 2024 – 06:44 AM EDT

PDB ID	:	5IEJ
BMRB ID	:	30025
Title	:	Solution structure of the BeF3-activated conformation of SdrG from Pseudomonas melonis Fr1
Authors	:	Campagne, S.; Vorholt, J.A.; Allain, F.H.-T.
Deposited on	:	2016-02-25

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 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:

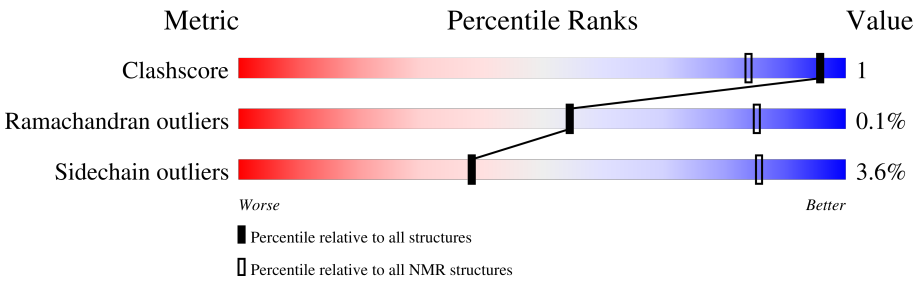
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 93%.

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	130	82% • 8% 9%

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: *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:4-A:58, A:66-A:118 (108)	1.13	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 3 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1759 atoms, of which 889 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called SdrG.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	118	1759	550	889	143	173	4	0

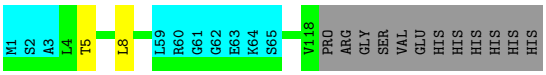
4 Residue-property plots (i)

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: SdrG

Chain A:



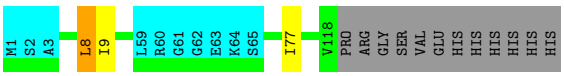
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: SdrG

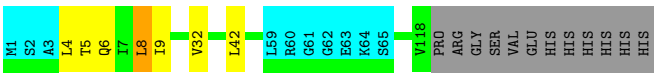
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: SdrG

Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: SdrG

Chain A: 80% • 8% 9%

M1 S2 A3 L4 T5 L8 K29 L59 R60 G61 G62 E63 K64 S65 K102 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.4 Score per residue for model 4

- Molecule 1: SdrG

Chain A: 79% • 8% 9%

M1 S2 A3 L4 T5 Q6 I7 L8 K29 V32 L59 R60 G61 G62 E63 K64 S65 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.5 Score per residue for model 5

- Molecule 1: SdrG

Chain A: 77% 6% 8% 9%

M1 S2 A3 L4 T5 Q6 I7 L8 L20 V32 L59 R60 G61 G62 E63 K64 S65 T66 P67 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.6 Score per residue for model 6 (medoid)

- Molecule 1: SdrG

Chain A: 82% • 8% 9%

M1 S2 A3 L8 K29 L59 R60 G61 G62 E63 K64 S65 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.7 Score per residue for model 7

- Molecule 1: SdrG

Chain A: 79% • 8% 9%

M1 S2 A3 L8 L59 R60 G61 G62 E63 K64 S65 T66 P67 F81 F94 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.8 Score per residue for model 8

- Molecule 1: SdrG

Chain A: 78% 5% 8% 9%

M1 S2 A3 L4 T5 L8 L20 L59 R60 G61 G62 E63 K64 S65 T66 P67 F81 F94 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS HIS

4.2.9 Score per residue for model 9

- Molecule 1: SdrG

Chain A: 82% • 8% 9%

M1 S2 A3 L4 T5 L8 L59 R60 G61 G62 E63 K64 S65 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.10 Score per residue for model 10

- Molecule 1: SdrG

Chain A: 76% 7% 8% 9%

M1 S2 A3 L4 T5 Q6 I7 L8 L20 V32 L59 R60 G61 G62 E63 K64 S65 T66 P67 R93 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.11 Score per residue for model 11

- Molecule 1: SdrG

Chain A: 80% • 8% 9%

M1 S2 A3 L4 T5 Q6 I7 L8 V32 L59 R60 G61 G62 E63 K64 S65 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.12 Score per residue for model 12

- Molecule 1: SdrG

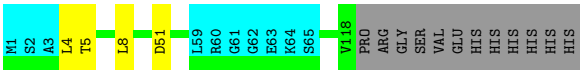
Chain A: 77% 5% • 8% 9%

M1 S2 A3 L8 I9 L20 K29 L59 R60 G61 G62 E63 K64 S65 T66 P67 F81 F94 V118 PRO ARG GLY SER VAL GLU HIS HIS HIS HIS HIS HIS

4.2.13 Score per residue for model 13

- Molecule 1: SdrG

Chain A:



4.2.14 Score per residue for model 14

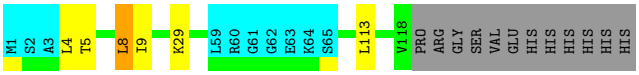
- Molecule 1: SdrG

Chain A:

4.2.15 Score per residue for model 15

- Molecule 1: SdrG

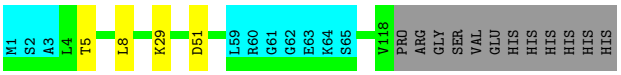
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: SdrG

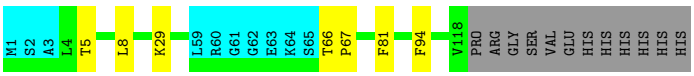
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: SdrG

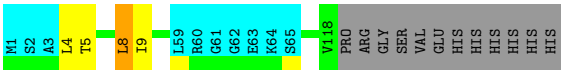
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: SdrG

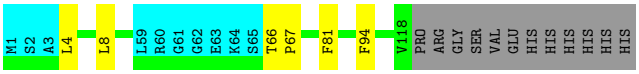
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: SdrG

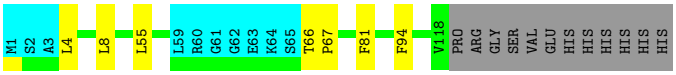
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: SdrG

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 30 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
CYANA	structure calculation	
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	1475
Number of shifts mapped to atoms	1410
Number of unparsed shifts	0
Number of shifts with mapping errors	65
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	93%

6 Model quality (i)

6.1 Standard geometry (i)

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 (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	800	814	814	1±1
All	All	16000	16280	16280	29

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:66:THR:N	1:A:67:PRO:CD	0.53	2.72	19	9
1:A:81:PHE:CD2	1:A:94:PHE:CD1	0.52	2.98	7	4
1:A:6:GLN:HB3	1:A:32:VAL:HG11	0.50	1.84	11	5
1:A:8:LEU:HD23	1:A:9:ILE:H	0.46	1.70	2	4
1:A:102:LYS:CE	1:A:102:LYS:HA	0.46	2.41	3	1
1:A:81:PHE:CD1	1:A:94:PHE:CD1	0.44	3.05	14	2
1:A:81:PHE:CD1	1:A:94:PHE:CD2	0.42	3.08	20	1
1:A:20:LEU:HD21	1:A:104:PHE:CZ	0.41	2.50	14	1
1:A:8:LEU:HD23	1:A:9:ILE:N	0.41	2.31	12	1
1:A:102:LYS:HA	1:A:102:LYS:HE3	0.40	1.94	3	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	107/130 (82%)	104±1 (98±1%)	2±1 (2±1%)	0±0 (0±0%)	54 85
All	All	2140/2600 (82%)	2088 (98%)	50 (2%)	2 (0%)	54 85

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	4	LEU	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	86/104 (83%)	83±1 (96±1%)	3±1 (4±1%)	38 86
All	All	1720/2080 (83%)	1658 (96%)	62 (4%)	38 86

All 11 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	8	LEU	20
1	A	5	THR	13
1	A	4	LEU	8
1	A	29	LYS	8
1	A	20	LEU	5
1	A	77	ILE	2
1	A	51	ASP	2
1	A	42	LEU	1
1	A	93	ARG	1
1	A	113	LEU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	55	LEU	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)

The completeness of assignment taking into account all chemical shift lists is 93% 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: *SdrG.str31*

7.1.1 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	1475
Number of shifts mapped to atoms	1410
Number of unparsed shifts	0
Number of shifts with mapping errors	65
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- No matching atom found in the structure. All 65 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	119	PRO	CA	63.08	0.3	1
1	A	119	PRO	CB	31.888	0.3	1
1	A	119	PRO	CG	27.417	0.3	1
1	A	119	PRO	CD	50.825	0.3	1
1	A	119	PRO	HA	4.259	0.020	1
1	A	119	PRO	HB2	2.173	0.020	2
1	A	119	PRO	HB3	1.733	0.020	2
1	A	119	PRO	HG2	1.84	0.020	1
1	A	119	PRO	HG3	1.84	0.020	1
1	A	119	PRO	HD2	3.822	0.020	2
1	A	119	PRO	HD3	3.512	0.020	2
1	A	120	ARG	C	176.966	0.3	1
1	A	120	ARG	CA	56.233	0.3	1
1	A	120	ARG	CB	30.579	0.3	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	120	ARG	CG	27.135	0.3	1
1	A	120	ARG	CD	43.037	0.3	1
1	A	120	ARG	H	8.457	0.020	1
1	A	120	ARG	HA	4.18	0.020	1
1	A	120	ARG	HB2	1.741	0.020	2
1	A	120	ARG	HB3	1.638	0.020	2
1	A	120	ARG	HG2	1.52	0.020	2
1	A	120	ARG	HG3	1.567	0.020	2
1	A	120	ARG	HD2	3.05	0.020	1
1	A	120	ARG	HD3	3.05	0.020	1
1	A	120	ARG	N	122.232	0.3	1
1	A	121	GLY	C	174.037	0.3	1
1	A	121	GLY	CA	45.098	0.3	1
1	A	121	GLY	H	8.387	0.020	1
1	A	121	GLY	HA2	3.846	0.020	1
1	A	121	GLY	HA3	3.846	0.020	1
1	A	121	GLY	N	109.779	0.3	1
1	A	122	SER	C	174.648	0.3	1
1	A	122	SER	CA	58.135	0.3	1
1	A	122	SER	CB	63.677	0.3	1
1	A	122	SER	H	8.105	0.020	1
1	A	122	SER	HA	4.344	0.020	1
1	A	122	SER	HB2	3.699	0.020	1
1	A	122	SER	HB3	3.699	0.020	1
1	A	122	SER	N	115.421	0.3	1
1	A	123	VAL	C	175.917	0.3	1
1	A	123	VAL	CA	62.252	0.3	1
1	A	123	VAL	CB	32.502	0.3	1
1	A	123	VAL	CG1	20.878	0.3	1
1	A	123	VAL	CG2	20.212	0.3	1
1	A	123	VAL	H	8.051	0.020	1
1	A	123	VAL	HA	3.932	0.020	1
1	A	123	VAL	HB	1.903	0.020	1
1	A	123	VAL	HG11	0.702	0.020	1
1	A	123	VAL	HG12	0.702	0.020	1
1	A	123	VAL	HG13	0.702	0.020	1
1	A	123	VAL	HG21	0.75	0.020	1
1	A	123	VAL	HG22	0.75	0.020	1
1	A	123	VAL	HG23	0.75	0.020	1
1	A	123	VAL	N	121.407	0.3	1
1	A	124	GLU	C	176.039	0.3	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	124	GLU	CA	56.305	0.3	1
1	A	124	GLU	CB	30.063	0.3	1
1	A	124	GLU	CG	35.912	0.3	1
1	A	124	GLU	H	8.277	0.020	1
1	A	124	GLU	HA	4.041	0.020	1
1	A	124	GLU	HB2	1.709	0.020	2
1	A	124	GLU	HB3	1.734	0.020	2
1	A	124	GLU	HG2	2.05	0.020	2
1	A	124	GLU	HG3	1.989	0.020	2
1	A	124	GLU	N	123.772	0.3	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
$^{13}C_{\alpha}$	122	0.32 ± 0.11	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	112	0.42 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	104	0.12 ± 0.10	None needed (< 0.5 ppm)
^{15}N	114	0.66 ± 0.30	Should be applied

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 93%, i.e. 1323 atoms were assigned a chemical shift out of a possible 1422. 0 out of 26 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	520/535 (97%)	217/217 (100%)	201/216 (93%)	102/102 (100%)
Sidechain	765/837 (91%)	527/553 (95%)	232/263 (88%)	6/21 (29%)
Aromatic	38/50 (76%)	24/25 (96%)	14/25 (56%)	0/0 (—%)
Overall	1323/1422 (93%)	768/795 (97%)	447/504 (89%)	108/123 (88%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 91%, i.e. 1409 atoms were assigned a chemical shift out of a possible 1545. 0 out of 27 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	558/587 (95%)	234/239 (98%)	215/236 (91%)	109/112 (97%)

Continued on next page...

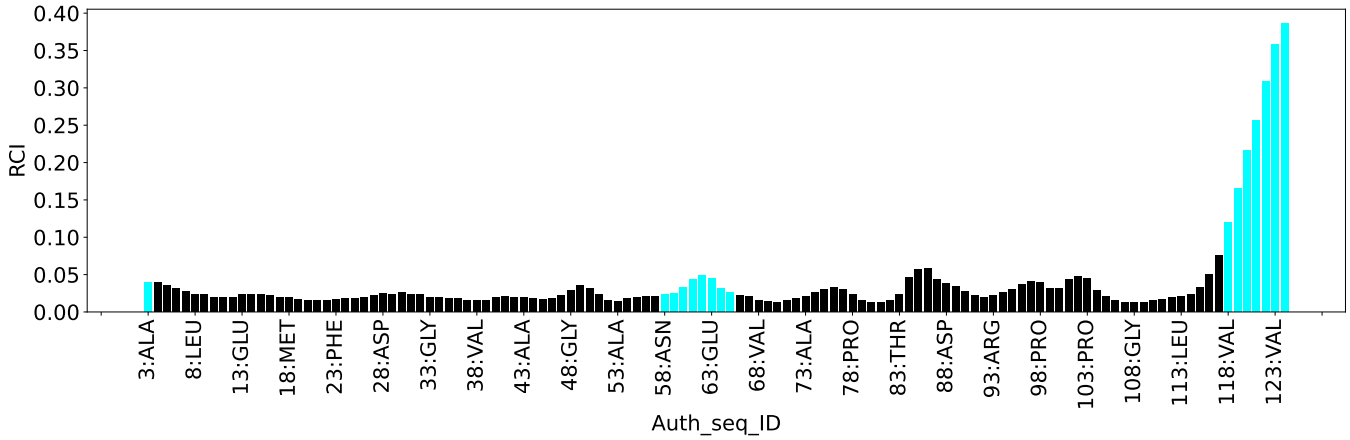
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Sidechain	813/908 (90%)	559/599 (93%)	247/284 (87%)	7/25 (28%)
Aromatic	38/50 (76%)	24/25 (96%)	14/25 (56%)	0/0 (—%)
Overall	1409/1545 (91%)	817/863 (95%)	476/545 (87%)	116/137 (85%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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: