wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 01:02 AM BST

PDB ID	:	2LEG
Title	:	Membrane protein complex DsbB-DsbA structure by joint calculations with solid-state NMR and X-ray experimental data
Authors	:	Tang, M.; Sperling, L.J.; Berthold, D.A.; Schwieters, C.D.; Nesbitt, A.E.; Nieuwkoop, A.J.; Gennis, R.B.; Rienstra, C.M.
Deposited on	:	2011-06-15

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
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

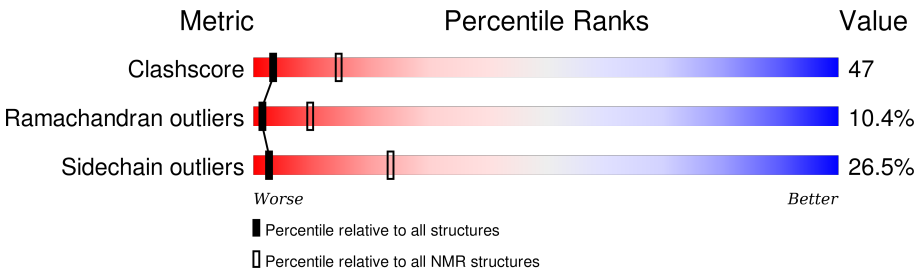
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLID-STATE NMR

The overall completeness of chemical shifts assignment is 36%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	189	26% 65% 8% ••
2	B	176	10% 46% 16% • 24%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 9 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:2-A:188, B:14-B:36, B:40-B:115, B:120-B:126, B:142-B:162 (314)	0.78	9

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 2 clusters and 5 single-model clusters were found.

Cluster number	Models
1	1, 5, 9
2	3, 8
Single-model clusters	2; 4; 6; 7; 10

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5154 atoms, of which 2579 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Thiol:disulfide interchange protein DsbA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	188	2930	946	1453	242	282	7	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	33	ALA	CYS	ENGINEERED MUTATION	UNP P0AEG4

- Molecule 2 is a protein called Disulfide bond formation protein B.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	134	2205	741	1126	164	165	9	0

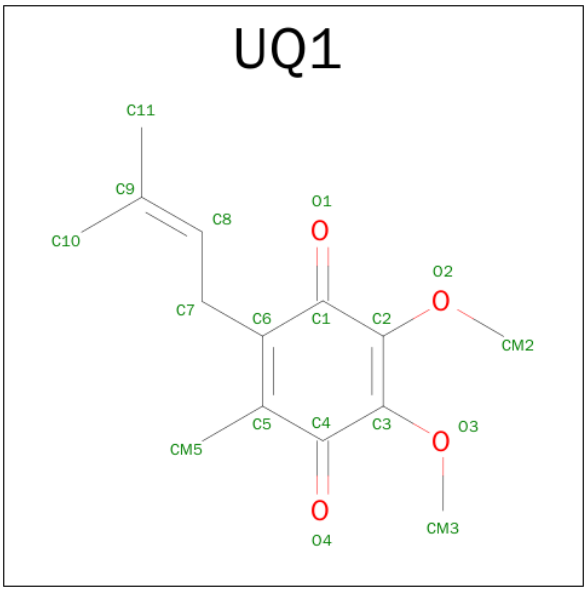
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	8	ALA	CYS	ENGINEERED MUTATION	UNP P0A6M2
B	49	VAL	CYS	ENGINEERED MUTATION	UNP P0A6M2
B	130	SER	CYS	ENGINEERED MUTATION	UNP P0A6M2

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
3	A	1	1	1

- Molecule 4 is UBIQUINONE-1 (three-letter code: UQ1) (formula: $C_{14}H_{18}O_4$).



Mol	Chain	Residues	Atoms		
			Total	C	O
4	B	1	18	14	4

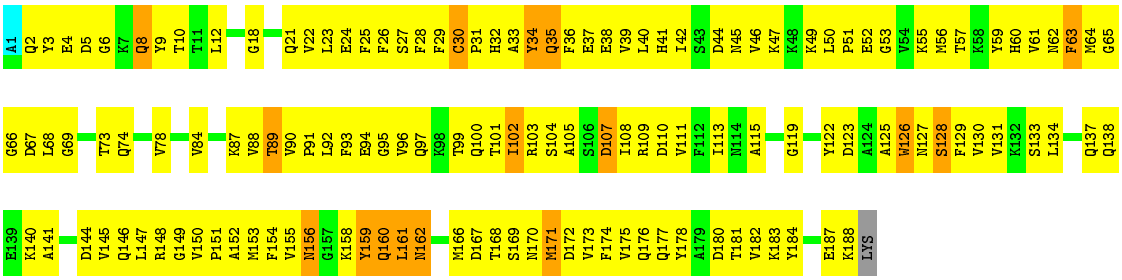
4 Residue-property plots (i)

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.

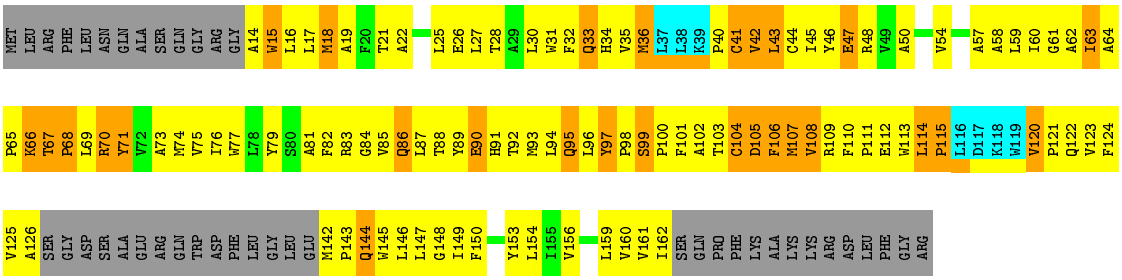
- Molecule 1: Thiol:disulfide interchange protein DsbA

Chain A: 26% 65% 8% ••



- Molecule 2: Disulfide bond formation protein B

Chain B: 10% 46% 16% • 24%



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

The representative model is number 9. Colouring as in section 4.1 above.

- Molecule 1: Thiol:disulfide interchange protein DsbA

Chain A: 42% 37% 17% • ••

A1 Q2 Y3 E4 D5 G6 K7 Q8 Y9 T10 T11 L12 P20 Q21 V22 L23 E24 F25 F26 F29 C30 P31 H32 A33 Y34 Q35 F36 E37 E38 V39 L40 H41 I42 S43 D44 N45 V46 K47 K48 K49 L50 P51 E52 K55 Y59 H60 V61 N62 F63 M64 G65 G66 D67 L68 G69 K70

D71 L72 W76 A81 L82 G83 V84 E85 D86 K87 V88 T89 V90 P91 L92 F93 E94 G95 V96 Q97 K98 I102 R103 S104 D107 D110 V111 A115 Y122 A125 W126 N127 S128 V131 Q137 K140 D144 V145 Q146 L147 R148 G149 V150 P151 A152 M153 F154 V155

N156 G157 K158 Y159 Q160 L161 N162 M166 D167 N170 M171 F174 V175 Q176 Q177 Y178 A179 Y184 K188 LYS

- Molecule 2: Disulfide bond formation protein B

Chain B: 23% 28% 19% • • 24%

MET LEU ARG PHE LEU ASN GLN ALA SER GLN GLY ARG GLY A14 W15 L16 L17 M18 T21 L25 E26 L27 T28 A29 L30 W31 F32 Q33 H34 V35 M36 L37 L38 K39 P40 C41 V42 L43 C44 I45 Y46 E47 A50 V54 A57 A58 L59 I60 G61 A62 I63 A64 P65 K66 T67

P68 L69 R70 Y71 M74 L78 V85 Q86 I87 T88 Y89 E90 H91 T92 N93 I94 Q95 L96 Y97 P98 S99 P100 F101 A102 T103 C104 D105 F106 M107 V108 P111 E112 W113 L114 P115 L116 D117 K118 W119 V120 P121 Q122 V123 F124 V125 A126 SER GLY ASP SER ALA GLU ARG GLN TRP ASP

PHE LEU GLY LEU GLU M142 P143 Q144 W145 L146 L147 G148 Y153 L154 I155 V156 A157 V158 L159 V160 V161 I162 SER GLN PRO PHE LYS LYS ALA LYS LYS ARG ASP LEU PHE GLY ARG

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 10 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
X-PLOR NIH	refinement	
X-PLOR NIH	structure solution	

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)	2leg_cs.str
Number of chemical shift lists	1
Total number of shifts	1555
Number of shifts mapped to atoms	1555
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	36%

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

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

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: UQ1, ZN

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	0.75±0.02	0±0/1504 (0.0±0.0%)	0.77±0.01	1±0/2034 (0.0±0.0%)
2	B	0.75±0.01	0±0/1049 (0.0±0.0%)	0.83±0.02	0±0/1439 (0.0±0.0%)
All	All	0.75	0/25530 (0.0%)	0.80	8/34730 (0.0%)

There are no bond-length outliers.

All 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	59	TYR	CB-CG-CD2	-6.15	117.31	121.00	9	6
2	B	89	TYR	CB-CG-CD2	-6.13	117.32	121.00	6	1
2	B	46	TYR	CB-CG-CD2	-5.16	117.90	121.00	9	1

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	1472	1446	1443	121±11
2	B	1015	1053	1049	121±18
4	B	18	0	18	9±4
All	All	25060	24990	25100	2340

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 1686 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:41:CYS:SG	4:B:501:UQ1:HM52	1.05	1.92	2	1
2:B:147:LEU:HD13	2:B:148:GLY:N	0.95	1.76	7	3
2:B:87:LEU:HD13	2:B:87:LEU:H	0.94	1.22	4	1
1:A:102:ILE:HD13	1:A:102:ILE:H	0.93	1.22	9	2
2:B:41:CYS:O	2:B:43:LEU:N	0.90	2.04	9	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	186/189 (98%)	150±4 (81±2%)	23±5 (13±3%)	13±2 (7±1%)	3 18
2	B	123/176 (70%)	85±3 (69±2%)	18±3 (15±3%)	20±2 (16±2%)	1 4
All	All	3090/3650 (85%)	2353 (76%)	415 (13%)	322 (10%)	1 10

5 of 102 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	120	VAL	10
2	B	108	VAL	9
1	A	30	CYS	8
1	A	89	THR	8
2	B	115	PRO	8

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	157/158 (99%)	118±7 (75±4%)	39±7 (25±4%)	3	26
2	B	106/147 (72%)	75±2 (71±2%)	31±2 (29±2%)	2	19
All	All	2630/3050 (86%)	1933 (73%)	697 (27%)	3	23

5 of 215 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)
2	B	71	TYR	9
1	A	159	TYR	9
2	B	18	MET	8
2	B	66	LYS	8
1	A	171	MET	8

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 carbohydrates in this entry.

6.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
4	UQ1	B	501	-	18,18,18	1.78±0.00	1±0 (5±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
4	UQ1	B	501	-	24,25,25	1.21±0.00	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	UQ1	B	501	-	-	0±0,9,33,33	0±0,1,1,1

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
4	B	501	UQ1	C6-C5	6.22	1.49	1.35	7	10

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 36% for the well-defined parts and 36% for the entire structure.

7.1 Chemical shift list 1

File name: 2leg_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

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	1555
Number of shifts mapped to atoms	1555
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	306	-0.68 ± 0.12	Should be applied
$^{13}C_{\beta}$	267	0.36 ± 0.10	None needed (< 0.5 ppm)
$^{13}C'$	277	-0.38 ± 0.13	None needed (< 0.5 ppm)
^{15}N	286	0.67 ± 0.27	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 36%, i.e. 1411 atoms were assigned a chemical shift out of a possible 3871. 5 out of 68 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	862/1538 (56%)	54/612 (9%)	550/628 (88%)	258/298 (87%)
Sidechain	517/1917 (27%)	0/1117 (0%)	517/736 (70%)	0/64 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	32/416 (8%)	0/218 (0%)	30/181 (17%)	2/17 (12%)
Overall	1411/3871 (36%)	54/1947 (3%)	1097/1545 (71%)	260/379 (69%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

7.1.4 Statistically unusual chemical shifts (i)

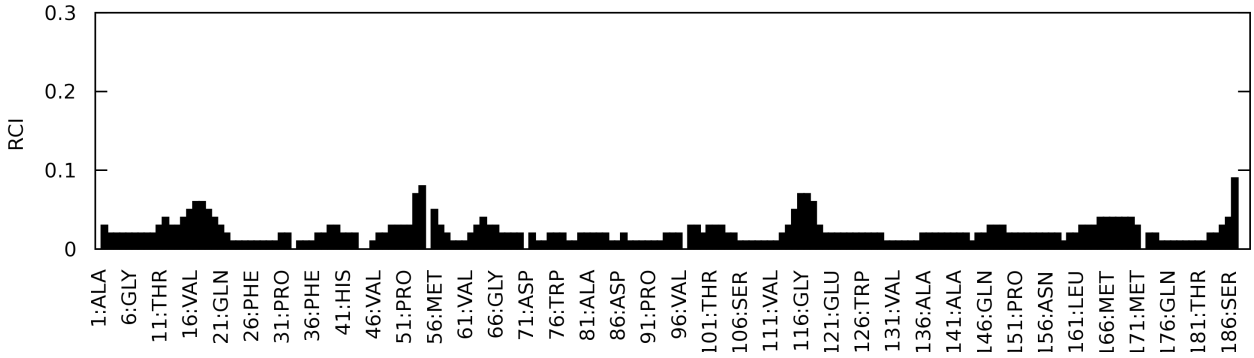
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
2	B	130	SER	CB	42.88	71.24 – 56.34	-14.0
1	A	24	GLU	CB	40.08	38.65 – 21.35	5.8
2	B	92	THR	CG2	27.48	27.15 – 15.95	5.3
1	A	5	ASP	CB	32.43	49.06 – 32.66	-5.1

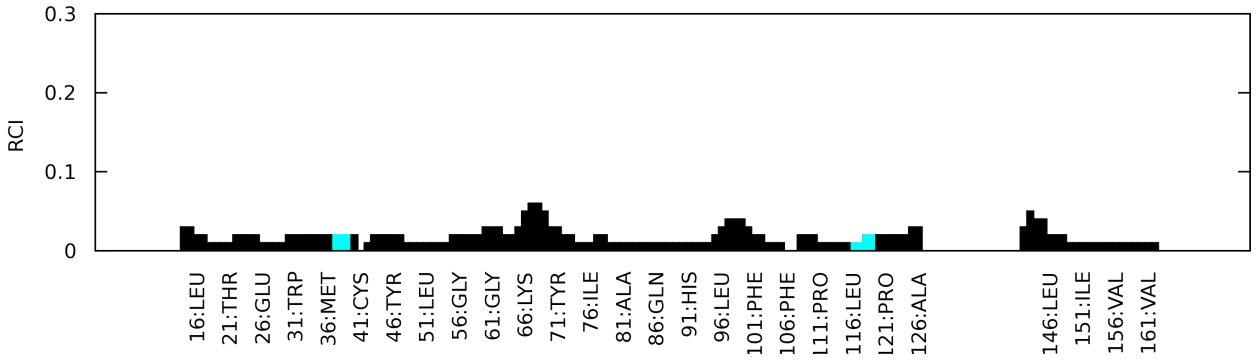
7.1.5 Random Coil Index (RCI) plots (i)

The images below report *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:



Random coil index (RCI) for chain B:


RCI
0.3
0.2
0.1
0
16:LEU
21:THR
26:GLU
31:TRP
36:MET
41:CYS
46:TYR
51:LEU
56:GLY
61:GLY
66:LYS
71:TYR
76:ILE
81:ALA
86:GLN
91:HIS
96:LEU
101:PHE
106:PHE
111:PRO
116:LEU
121:PRO
126:ALA
146:LEU
151:ILE
156:VAL
161:VAL