WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 06:33 AM EDT

PDB ID	:	2LRL
BMRB ID	:	18379
Title	:	Solution Structures of the IIA(Chitobiose)-HPr complex of the N,N'-Diacetyl chitobiose Branch of the Escherichia coli Phosphotransferase System
Authors	:	Jung, Y.; Cai, M.; Clore, M.
Deposited on	:	2012-04-06

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

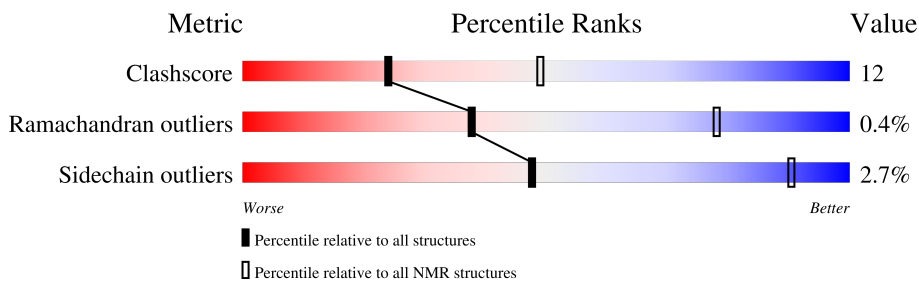|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & 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 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	Percentile Ranks	Value
Clashscore		12
Ramachandran outliers		0.4%
Sidechain outliers		2.7%

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	103	68% 17% 15%
1	B	103	66% 19% 15%
1	C	103	68% 18% 14%
2	D	85	73% 24% ·

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 19 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:60, A:71-A:101, B:4-B:60, B:71-B:101, C:4-C:60, C:72-C:103, D:301-D:385 (350)	0.43	19

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 2 single-model clusters were found.

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

3 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 6109 atoms, of which 3116 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component.

Mol	Chain	Residues	Atoms						Trace
1	A	103	Total	C	H	N	O	S	0
			1604	491	821	136	149	7	
1	B	103	Total	C	H	N	O	S	0
			1604	491	821	136	149	7	
1	C	103	Total	C	H	N	O	S	0
			1604	491	821	136	149	7	

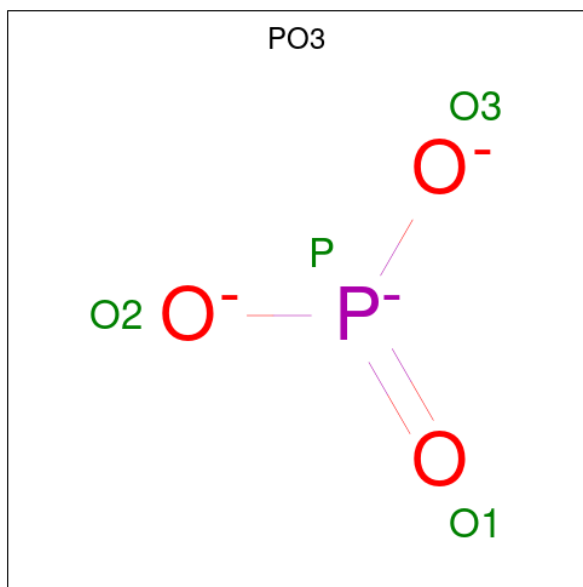There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	79	LEU	ASP	engineered mutation	UNP P69791
B	79	LEU	ASP	engineered mutation	UNP P69791
C	79	LEU	ASP	engineered mutation	UNP P69791

- Molecule 2 is a protein called Phosphocarrier protein HPr.

Mol	Chain	Residues	Atoms						Trace
2	D	85	Total	C	H	N	O	S	0
			1293	401	653	107	130	2	

- Molecule 3 is PHOSPHITE ION (three-letter code: PO3) (formula: O_3P).

PO3

O3

O⁻

P

O2 O⁻ — P⁻

O

O1

Mol	Chain	Residues	Atoms		
3	D	1	Total	O	P
			4	3	1

W O R L D W I D E

wwwPDB

PROTEIN DATA BANK

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.
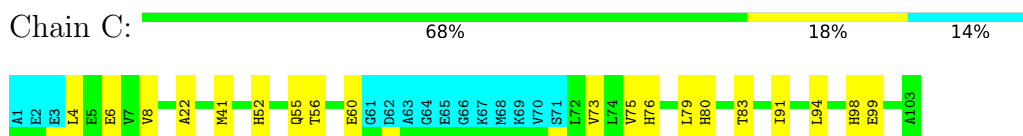
• Molecule 1: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component
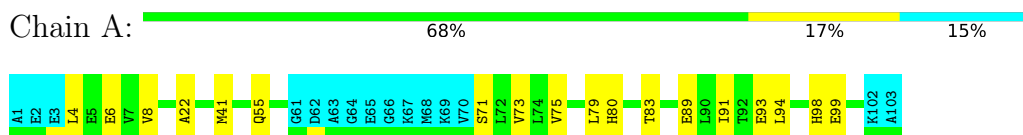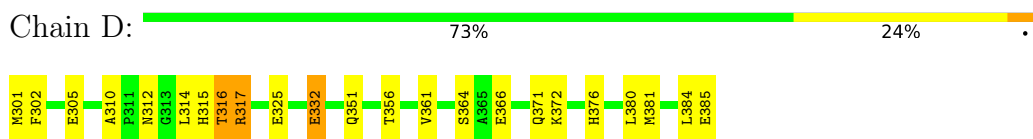
Chain A:

• Molecule 1: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain B:

• Molecule 1: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component
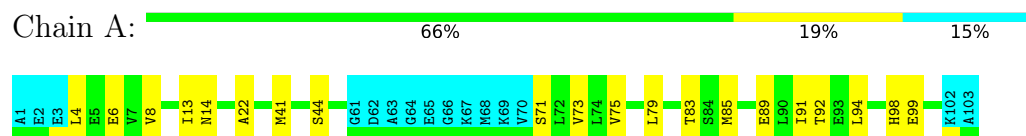
Chain C:

• Molecule 2: Phosphocarrier protein HPr
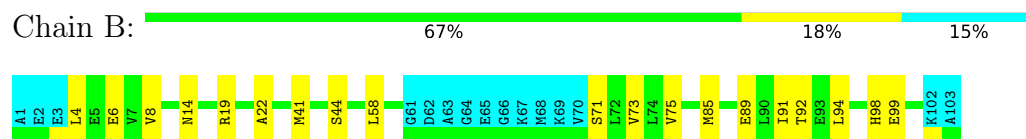
Chain D:
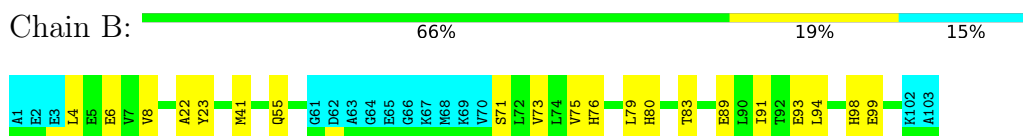
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

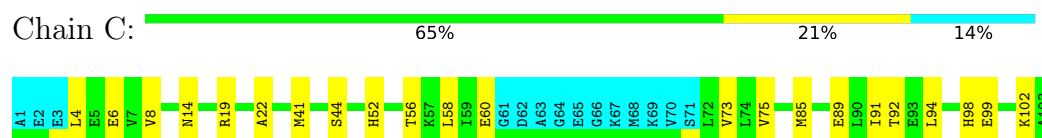The representative model is number 19. Colouring as in section 4.1 above.

• Molecule 1: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain A:

66% 19% 15%

A1 E2 E3 L4 E5 E6 V7 V8 I13 N14 A22 M41 S44 G61 D62 A63 G64 E65 G66 K67 M68 K69 V70 S71 L72 V73 L74 V75 L79 T83 S84 M85 E89 L90 I91 T92 E93 L94 H98 E99 K102 A103

• Molecule 1: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain B:

67% 18% 15%

A1 E2 E3 L4 E5 E6 V7 V8 N14 R19 A22 M41 S44 L58 G61 D62 A63 G64 E65 G66 K67 M68 K69 V70 S71 L72 V73 L74 V75 M85 E89 L90 I91 T92 E93 L94 H98 E99 K102 A103

• Molecule 1: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain C:

65% 21% 14%

A1 E2 E3 L4 E5 E6 V7 V8 N14 R19 A22 M41 S44 H52 T56 K57 L58 I59 E60 G61 D62 A63 G64 E65 G66 K67 M68 K69 V70 S71 L72 V73 L74 V75 M85 E89 L90 I91 T92 E93 L94 H98 E99 K102 A103

• Molecule 2: Phosphocarrier protein HPr

Chain D:

72% 24% ● ●

M301 F302 E305 A310 P311 N312 G313 L314 H315 T316 R317 K324 E325 E332 Q351 T356 V361 S364 A365 E366 Q371 K372 H376 L380 M381 L384 E385
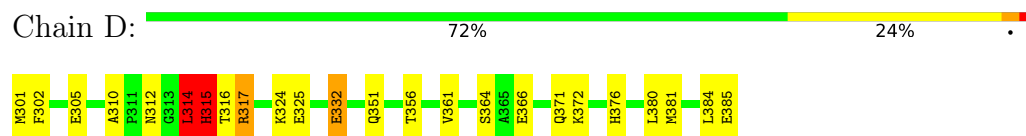
5   Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *torsion angle dynamics.*

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations.*

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

Software name	Classification	Version
X-PLOR NIH	geometry optimization	
X-PLOR NIH	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	1868
Number of shifts mapped to atoms	1868
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%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: PO3

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

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
2	D	0.0±0.0	1.0±0.0
All	All	0	20

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
2	D	317	ARG	Sidechain	20

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	678	716	716	18±3
1	B	678	716	716	18±4
1	C	687	729	729	21±4
2	D	640	653	650	22±3
3	D	4	0	0	4±2
All	All	53740	56280	56220	1315

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 12.

5 of 300 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:C:56:THR:HG21	2:D:356:THR:HG22	0.89	1.44	7	20
2:D:372:LYS:HZ2	2:D:376:HIS:CE1	0.79	1.94	10	20
2:D:310:ALA:CB	2:D:385:GLU:OXT	0.79	2.31	3	20
2:D:310:ALA:HB1	2:D:385:GLU:OXT	0.78	1.78	20	20
1:B:97:LEU:HD21	1:C:98:HIS:NE2	0.76	1.95	16	3

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	88/103 (85%)	87±0 (99±0%)	1±0 (1±0%)	0±0 (0±0%)	54	85
1	B	88/103 (85%)	87±0 (99±0%)	1±0 (1±0%)	0±0 (0±0%)	54	85
1	C	88/103 (85%)	87±0 (99±0%)	1±0 (1±1%)	0±0 (0±1%)	50	82
2	D	83/85 (98%)	79±1 (95±1%)	3±0 (4±1%)	1±0 (1±1%)	18	66
All	All	6940/7880 (88%)	6794 (98%)	121 (2%)	25 (0%)	38	78

5 of 7 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	D	316	THR	16
1	A	4	LEU	2
1	B	4	LEU	2
1	C	4	LEU	2
1	C	102	LYS	1

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	72/81 (89%)	72±0 (100±1%)	0±0 (0±1%)	92	98
1	B	72/81 (89%)	72±0 (100±1%)	0±0 (0±1%)	92	98
1	C	72/81 (89%)	72±0 (100±0%)	0±0 (0±0%)	93	98
2	D	70/70 (100%)	63±1 (90±1%)	7±1 (10±1%)	11	56
All	All	5720/6260 (91%)	5567 (97%)	153 (3%)	48	90

5 of 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)
2	D	305	GLU	20
2	D	332	GLU	20
2	D	364	SER	20
2	D	366	GLU	20
2	D	371	GLN	20

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 ⓘ

1 ligand is modelled in this entry.

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
3	PO3	D	200	-	0,3,3	0.00±0.00	-

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
3	PO3	D	200	-	0,3,3	0.00±0.00	-

There are no bond-length outliers.

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 ⓘ

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 36% for the well-defined parts and 35% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *IIAChb*

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	1868
Number of shifts mapped to atoms	1868
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	185	-0.26 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_\beta$	173	0.56 ± 0.07	Should be checked
$^{13}C'$	100	-0.49 ± 0.08	None needed (< 0.5 ppm)
^{15}N	181	-0.29 ± 0.17	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 36%, i.e. 1736 atoms were assigned a chemical shift out of a possible 4832. 0 out of 78 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	750/1761 (43%)	324/713 (45%)	258/700 (37%)	168/348 (48%)
Sidechain	947/2876 (33%)	592/1899 (31%)	355/892 (40%)	0/85 (0%)

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	39/195 (20%)	15/103 (15%)	24/78 (31%)	0/14 (0%)
Overall	1736/4832 (36%)	931/2715 (34%)	637/1670 (38%)	168/447 (38%)

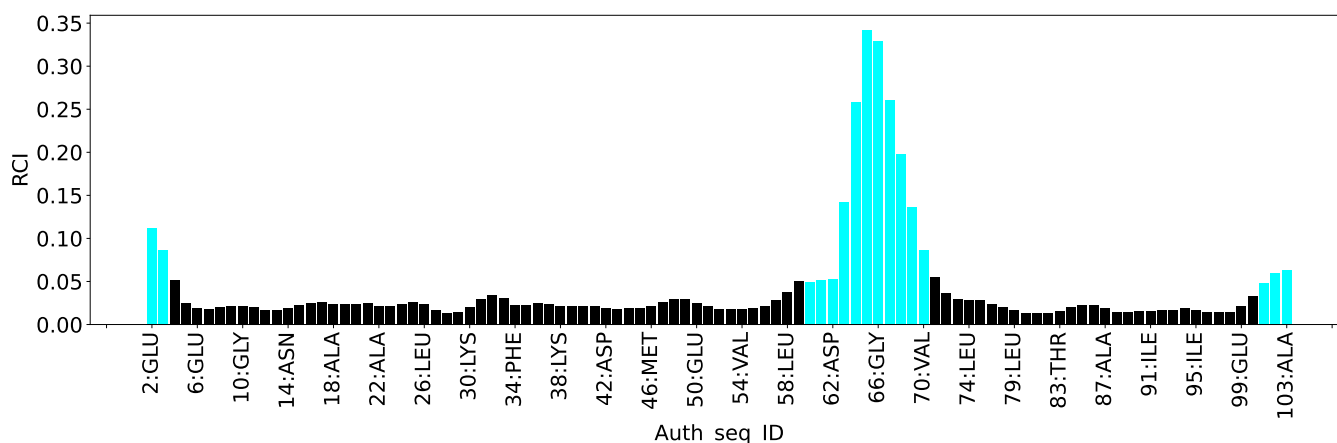### 7.1.4 Statistically unusual chemical shifts (i)

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	D	373	ALA	HA	1.91	2.13 − 6.34	-5.5

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:

Random coil index (RCI) for chain D: