wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 10:46 PM BST

PDB ID	:	2K6M
Title	:	Solution Structure of Human Supervillin Headpiece
Authors	:	Brown, J.W.; Didem, V.; McKnight, C.
Deposited on	:	2008-07-11

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:

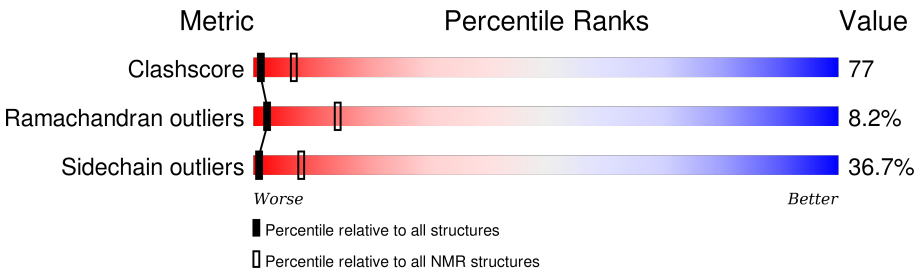
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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:
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	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	S	67	19% 43% 25% 12%

2 Ensemble composition and analysis (i)

This entry contains 22 models. Model 22 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	S:18-S:76 (59)	0.36	22

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 6 clusters and 2 single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 5, 6, 10, 14
2	11, 12, 16, 22
3	9, 15, 21
4	8, 13
5	1, 17
6	19, 20
Single-model clusters	7; 18

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1102 atoms, of which 560 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Supervillin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	S	67	1102	356	560	84	99	3	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
S	10	MET	-	EXPRESSION TAG	UNP O95425

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: Supervillin

Chain S: 19% 43% 25% 12%

M10 L11 A12 K13 L14 C15 K16 T17 I18 Y19 P20 L21 A22 D23 L24 L25 A26 R27 P28 L29 P30 V33 D34 P35 L36 K37 L38 E39 I40 Y41 L42 T43 D44 E45 D46 F47 E48 F49 A50 L51 D52 M53 T54 R55 D56 E57 Y58 N59 A60 L61 P62 A63 W64 K65 Q66 V67 N68 L69 K70

K73 G74 L75 F76

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

The representative model is number 22. Colouring as in section 4.1 above.

- Molecule 1: Supervillin

Chain S: 22% 37% 27% • 12%

M10 L11 A12 K13 L14 C15 K16 T17 I18 Y19 P20 L21 A22 D23 L24 L25 A26 R27 P28 L29 P30 E31 G32 V33 D34 P35 L36 K37 L38 E39 I40 Y41 L42 T43 F47 L51 D52 M53 T54 R55 D56 E57 Y58 L61 P62 A63 W64 K65 Q66 V67 N68 L69 K70 K71 K72 K73 G74

L75 F76

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-DISTANCE GEOMETRY SIMULATED ANNEALING.*

Of the 100 calculated structures, 22 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
CNS	refinement	1.1
CNS	structure solution	1.1
NMRPIPE	structure solution	
MOLMOL 2K.2	structure solution	2K.2
XWINNMR	structure solution	3.1
TALOS	structure solution	
PREDITOR	structure solution	
NMRVIEW	structure solution	7.0.16

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)	BMRB entry 15873
Number of chemical shift lists	1
Total number of shifts	677
Number of shifts mapped to atoms	677
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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)

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	S	482	484	484	74±4
All	All	10604	10648	10648	1636

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 77.

5 of 453 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:S:29:LEU:HD22	1:S:33:VAL:HG21	1.04	1.28	3	3
1:S:25:LEU:HD12	1:S:75:LEU:HD21	0.99	1.29	7	3
1:S:25:LEU:HD13	1:S:75:LEU:HD23	0.93	1.37	13	9
1:S:38:LEU:HD12	1:S:75:LEU:HD23	0.91	1.36	7	1
1:S:38:LEU:HD22	1:S:41:TYR:OH	0.91	1.65	17	18

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	S	58/67 (87%)	47±2 (80±3%)	7±2 (12±3%)	5±1 (8±2%)	2 14
All	All	1276/1474 (87%)	1023 (80%)	149 (12%)	104 (8%)	2 14

5 of 13 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	S	28	PRO	22
1	S	26	ALA	22
1	S	30	PRO	18
1	S	43	THR	12
1	S	75	LEU	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	S	51/58 (88%)	32±3 (63±5%)	19±3 (37±5%)	1 8
All	All	1122/1276 (88%)	710 (63%)	412 (37%)	1 8

5 of 38 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	S	42	LEU	22
1	S	36	LEU	22
1	S	24	LEU	22
1	S	47	PHE	22
1	S	65	LYS	22

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)

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 80% for the well-defined parts and 73% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 15873

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	677
Number of shifts mapped to atoms	677
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	121

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	55	0.00 ± 0.00	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	59	-0.11 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	55	1.41 ± 0.52	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 80%, i.e. 606 atoms were assigned a chemical shift out of a possible 755. 11 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	219/285 (77%)	112/113 (99%)	53/118 (45%)	54/54 (100%)
Sidechain	353/407 (87%)	227/238 (95%)	122/155 (79%)	4/14 (29%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	34/63 (54%)	33/33 (100%)	0/29 (0%)	1/1 (100%)
Overall	606/755 (80%)	372/384 (97%)	175/302 (58%)	59/69 (86%)

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
???	S	17	THR	CB	31.89	78.10 – 61.30	-22.5
???	S	54	THR	CB	33.93	78.10 – 61.30	-21.3
???	S	43	THR	CB	34.27	78.10 – 61.30	-21.1
???	S	62	PRO	CD	30.98	55.31 – 45.41	-19.6
???	S	30	PRO	CD	31.23	55.31 – 45.41	-19.3
???	S	55	ARG	NE	114.94	92.63 – 76.73	19.0
???	S	28	PRO	CD	31.56	55.31 – 45.41	-19.0
???	S	70	LYS	CG	46.20	30.67 – 19.17	18.5
???	S	35	PRO	CD	32.29	55.31 – 45.41	-18.3
???	S	20	PRO	CD	32.31	55.31 – 45.41	-18.2
???	S	43	THR	CG2	41.96	27.15 – 15.95	18.2
???	S	62	PRO	CG	46.69	32.66 – 21.76	17.9
???	S	30	PRO	CG	46.46	32.66 – 21.76	17.7
???	S	71	LYS	CG	45.21	30.67 – 19.17	17.6
???	S	17	THR	CG2	41.30	27.15 – 15.95	17.6
???	S	35	PRO	CG	46.06	32.66 – 21.76	17.3
???	S	21	LEU	CG	46.64	32.55 – 21.05	17.3
???	S	54	THR	CG2	40.78	27.15 – 15.95	17.2
???	S	51	LEU	CG	46.20	32.55 – 21.05	16.9
???	S	65	LYS	CG	44.00	30.67 – 19.17	16.6
???	S	69	LEU	CG	45.88	32.55 – 21.05	16.6
???	S	55	ARG	CG	46.56	33.23 – 21.23	16.1
???	S	38	LEU	CG	45.23	32.55 – 21.05	16.0
???	S	37	LYS	CG	43.26	30.67 – 19.17	15.9
???	S	61	LEU	CG	45.01	32.55 – 21.05	15.8
???	S	75	LEU	CG	44.98	32.55 – 21.05	15.8
???	S	42	LEU	CG	44.86	32.55 – 21.05	15.7
???	S	28	PRO	CG	44.24	32.66 – 21.76	15.6
???	S	27	ARG	CG	45.25	33.23 – 21.23	15.0
???	S	67	VAL	CG1	41.91	28.40 – 14.60	14.8
???	S	18	ILE	CG2	37.12	24.63 – 10.43	13.8
???	S	33	VAL	CG1	40.50	28.40 – 14.60	13.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
???	S	24	LEU	CD1	46.38	32.77 – 16.57	13.4
???	S	40	ILE	CG2	36.08	24.63 – 10.43	13.1
???	S	62	PRO	CA	43.18	71.13 – 55.53	-12.9
???	S	42	LEU	CD1	45.42	32.77 – 16.57	12.8
???	S	29	LEU	CD1	45.30	32.77 – 16.57	12.7
???	S	20	PRO	CA	43.58	71.13 – 55.53	-12.7
???	S	75	LEU	CD1	45.18	32.77 – 16.57	12.7
???	S	21	LEU	CD2	45.57	32.60 – 15.60	12.6
???	S	30	PRO	CA	43.65	71.13 – 55.53	-12.6
???	S	69	LEU	CD2	45.48	32.60 – 15.60	12.6
???	S	28	PRO	CA	43.72	71.13 – 55.53	-12.6
???	S	61	LEU	CD1	44.89	32.77 – 16.57	12.5
???	S	33	VAL	CG2	40.91	29.20 – 13.40	12.4
???	S	36	LEU	CD1	44.51	32.77 – 16.57	12.2
???	S	25	LEU	CD2	44.87	32.60 – 15.60	12.2
???	S	35	PRO	CA	44.88	71.13 – 55.53	-11.8
???	S	40	ILE	CD1	33.40	21.91 – 5.01	11.8
???	S	67	VAL	CG2	39.62	29.20 – 13.40	11.6
???	S	42	LEU	CD2	43.79	32.60 – 15.60	11.6
???	S	29	LEU	CD2	43.78	32.60 – 15.60	11.6
???	S	38	LEU	CD2	43.56	32.60 – 15.60	11.4
???	S	67	VAL	CA	29.56	76.93 – 48.03	-11.4
???	S	26	ALA	CB	39.30	28.03 – 9.93	11.2
???	S	51	LEU	CD2	43.18	32.60 – 15.60	11.2
???	S	21	LEU	CD1	42.75	32.77 – 16.57	11.2
???	S	50	ALA	CB	39.09	28.03 – 9.93	11.1
???	S	51	LEU	CD1	42.63	32.77 – 16.57	11.1
???	S	75	LEU	CD2	42.59	32.60 – 15.60	10.9
???	S	27	ARG	CA	32.44	68.35 – 45.25	-10.5
???	S	60	ALA	CB	38.06	28.03 – 9.93	10.5
???	S	25	LEU	CD1	41.53	32.77 – 16.57	10.4
???	S	36	LEU	CD2	41.73	32.60 – 15.60	10.4
???	S	22	ALA	CB	37.46	28.03 – 9.93	10.2
???	S	61	LEU	CD2	41.43	32.60 – 15.60	10.2
???	S	73	LYS	CA	34.63	67.97 – 45.97	-10.2
???	S	60	ALA	CA	33.07	63.07 – 43.27	-10.2
???	S	61	LEU	CA	33.94	66.36 – 44.96	-10.1
???	S	42	LEU	CA	33.95	66.36 – 44.96	-10.1
???	S	18	ILE	CG1	45.59	36.54 – 18.94	10.1
???	S	72	ALA	CB	37.24	28.03 – 9.93	10.1
???	S	75	LEU	CA	34.33	66.36 – 44.96	-10.0
???	S	24	LEU	CD2	40.92	32.60 – 15.60	9.9

Continued on next page...

Continued from previous page...

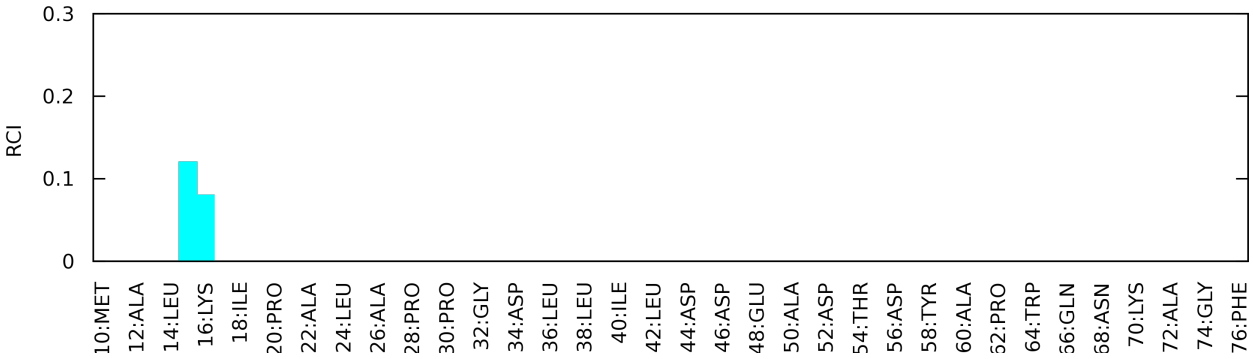
Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
???	S	18	ILE	CD1	30.17	21.91 – 5.01	9.9
???	S	63	ALA	CB	36.77	28.03 – 9.93	9.8
???	S	37	LYS	CA	35.41	67.97 – 45.97	-9.8
???	S	29	LEU	CA	34.89	66.36 – 44.96	-9.7
???	S	59	ASN	CA	35.73	63.05 – 44.05	-9.4
???	S	24	LEU	CA	35.76	66.36 – 44.96	-9.3
???	S	66	GLN	CB	46.15	38.36 – 19.96	9.2
???	S	51	LEU	CA	35.95	66.36 – 44.96	-9.2
???	S	39	GLU	CB	45.84	38.65 – 21.35	9.2
???	S	25	LEU	CA	36.37	66.36 – 44.96	-9.0
???	S	50	ALA	CA	35.45	63.07 – 43.27	-8.9
???	S	53	MET	CA	36.03	67.38 – 44.88	-8.9
???	S	40	ILE	CG1	43.43	36.54 – 18.94	8.9
???	S	52	ASP	CA	36.57	64.88 – 44.48	-8.9
???	S	72	ALA	CA	35.71	63.07 – 43.27	-8.8
???	S	36	LEU	CA	36.95	66.36 – 44.96	-8.7
???	S	16	LYS	CA	37.77	67.97 – 45.97	-8.7
???	S	31	GLU	CA	39.04	67.86 – 46.86	-8.7
???	S	68	ASN	CA	37.31	63.05 – 44.05	-8.5
???	S	39	GLU	CA	39.63	67.86 – 46.86	-8.4
???	S	18	ILE	CA	38.96	75.08 – 48.18	-8.4
???	S	22	ALA	CA	36.72	63.07 – 43.27	-8.3
???	S	63	ALA	CA	36.78	63.07 – 43.27	-8.3
???	S	48	GLU	CA	40.14	67.86 – 46.86	-8.2
???	S	46	ASP	CA	38.05	64.88 – 44.48	-8.1
???	S	57	GLU	CA	40.42	67.86 – 46.86	-8.1
???	S	64	TRP	CB	45.79	40.02 – 19.92	7.9
???	S	69	LEU	CA	39.08	66.36 – 44.96	-7.7
???	S	65	LYS	CA	39.98	67.97 – 45.97	-7.7
???	S	54	THR	CA	41.95	75.37 – 49.07	-7.7
???	S	38	LEU	CA	39.32	66.36 – 44.96	-7.6
???	S	45	GLU	CA	41.39	67.86 – 46.86	-7.6
???	S	21	LEU	CA	39.62	66.36 – 44.96	-7.5
???	S	43	THR	CA	42.62	75.37 – 49.07	-7.5
???	S	44	ASP	CA	39.55	64.88 – 44.48	-7.4
???	S	66	GLN	CA	40.77	67.31 – 45.91	-7.4
???	S	71	LYS	CA	41.33	67.97 – 45.97	-7.1
???	S	64	TRP	CA	39.81	70.54 – 44.84	-7.0
???	S	17	THR	CA	44.02	75.37 – 49.07	-6.9
???	S	33	VAL	CA	42.63	76.93 – 48.03	-6.9
???	S	55	ARG	CA	41.15	68.35 – 45.25	-6.8
???	S	40	ILE	CA	44.41	75.08 – 48.18	-6.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
???	S	49	PHE	CA	41.71	70.99 – 45.29	-6.4
???	S	76	PHE	CA	42.03	70.99 – 45.29	-6.3
???	S	58	TYR	CA	42.35	70.88 – 45.38	-6.2
???	S	70	LYS	CA	43.38	67.97 – 45.97	-6.2
???	S	47	PHE	CA	43.22	70.99 – 45.29	-5.8

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.

Random coil index (RCI) for chain S: