Full wwPDB NMR Structure Validation Report (i)

Aug 8, 2016 – 07:25 PM EDT

PDB ID	:	5IEB
Title	:	Solution structure of SdrG from Sphingomonas 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
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:

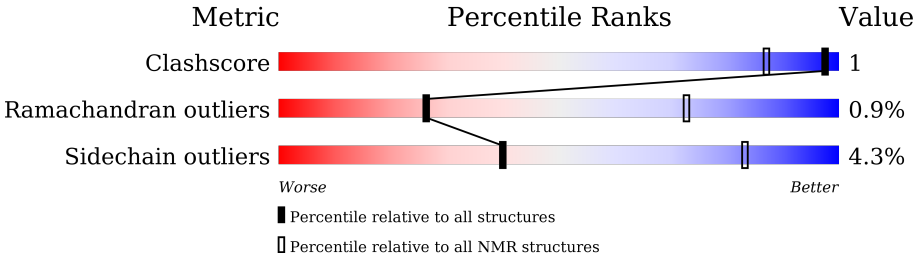
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027939
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 91%.

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	130	85% • • • 9%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:116 (113)	0.29	1

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. No single-model clusters were found.

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

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 Sensory transduction regulatory protein.

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

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	117	LEU	-	expression tag	UNP A0A0D1MA58
A	118	VAL	-	expression tag	UNP A0A0D1MA58
A	119	PRO	-	expression tag	UNP A0A0D1MA58
A	120	ARG	-	expression tag	UNP A0A0D1MA58
A	121	GLY	-	expression tag	UNP A0A0D1MA58
A	122	SER	-	expression tag	UNP A0A0D1MA58
A	123	VAL	-	expression tag	UNP A0A0D1MA58
A	124	GLU	-	expression tag	UNP A0A0D1MA58
A	125	HIS	-	expression tag	UNP A0A0D1MA58
A	126	HIS	-	expression tag	UNP A0A0D1MA58
A	127	HIS	-	expression tag	UNP A0A0D1MA58
A	128	HIS	-	expression tag	UNP A0A0D1MA58
A	129	HIS	-	expression tag	UNP A0A0D1MA58
A	130	HIS	-	expression tag	UNP A0A0D1MA58

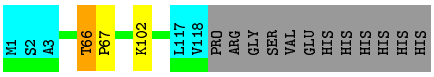
4 Residue-property plots (i)

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: Sensory transduction regulatory protein

Chain A:



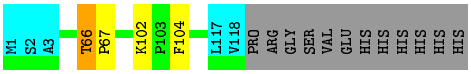
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Sensory transduction regulatory protein

Chain A:



4.2.2 Score per residue for model 2

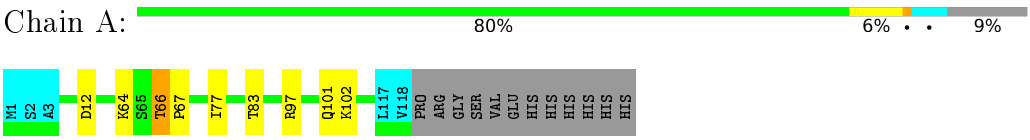
- Molecule 1: Sensory transduction regulatory protein

Chain A:

4.2.3 Score per residue for model 3

- Molecule 1: Sensory transduction regulatory protein

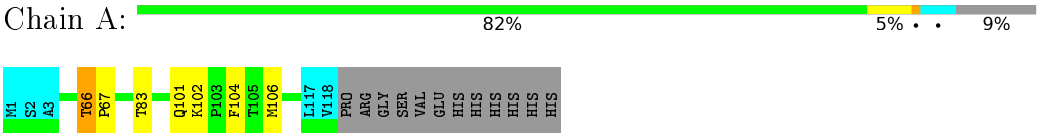
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Sensory transduction regulatory protein

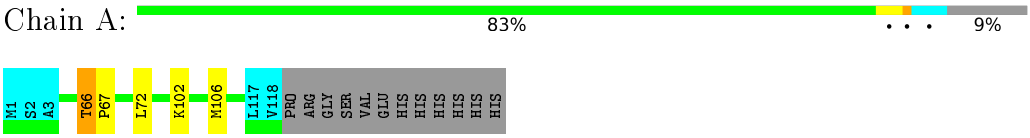
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Sensory transduction regulatory protein

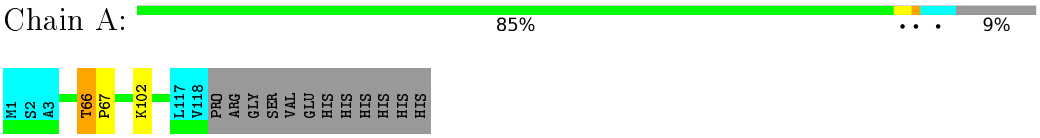
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Sensory transduction regulatory protein

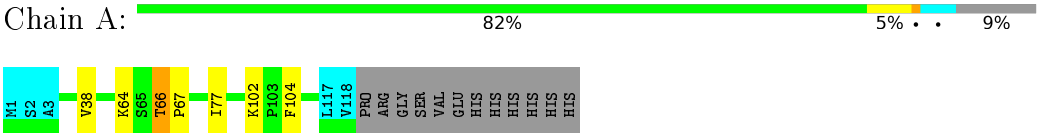
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Sensory transduction regulatory protein

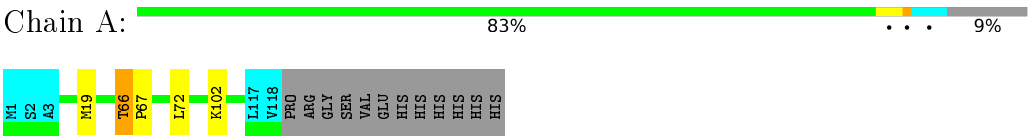
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Sensory transduction regulatory protein

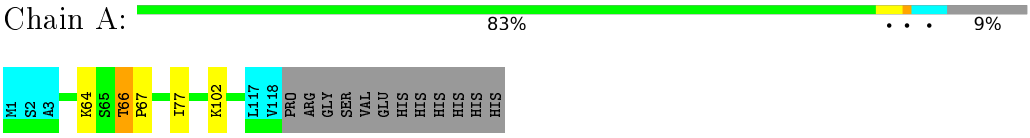
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Sensory transduction regulatory protein

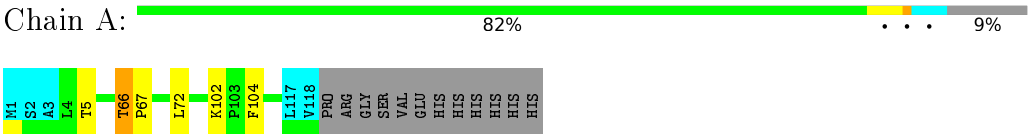
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: Sensory transduction regulatory protein

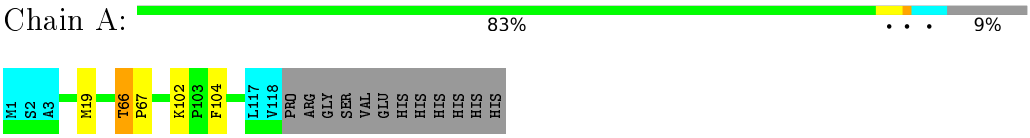
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: Sensory transduction regulatory protein

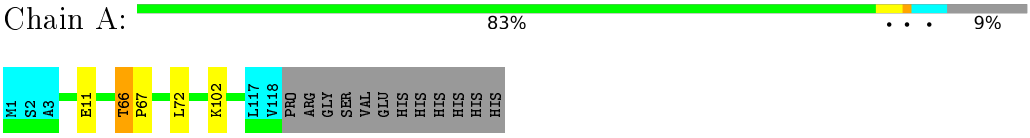
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Sensory transduction regulatory protein

Chain A:



4.2.13 Score per residue for model 13

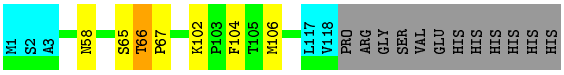
- Molecule 1: Sensory transduction regulatory protein

Chain A:

4.2.14 Score per residue for model 14

- Molecule 1: Sensory transduction regulatory protein

Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Sensory transduction regulatory protein

Chain A:

4.2.16 Score per residue for model 16

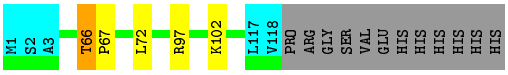
- Molecule 1: Sensory transduction regulatory protein

Chain A:

4.2.17 Score per residue for model 17

- Molecule 1: Sensory transduction regulatory protein

Chain A:



4.2.18 Score per residue for model 18

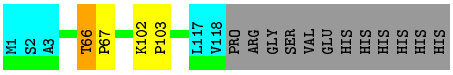
- Molecule 1: Sensory transduction regulatory protein

Chain A:

4.2.19 Score per residue for model 19

- Molecule 1: Sensory transduction regulatory protein

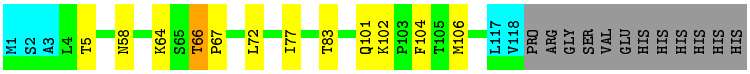
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Sensory transduction regulatory protein

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 least restraint violations*.

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

Software name	Classification	Version
AMBER	refinement	12

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)	5ieb_cs.cif
Number of chemical shift lists	1
Total number of shifts	1485
Number of shifts mapped to atoms	1407
Number of unparsed shifts	78
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

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	A	836	848	848	1±0
All	All	16720	16960	16960	27

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	0.64	2.61	7	20
1:A:65:SER:C	1:A:67:PRO:HD2	0.47	2.30	14	1
1:A:66:THR:H	1:A:67:PRO:CD	0.45	2.23	7	6

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	113/130 (87%)	104±1 (92±1%)	8±1 (7±1%)	1±0 (1±0%)	26 73
All	All	2260/2600 (87%)	2089 (92%)	150 (7%)	21 (1%)	26 73

All 2 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	66	THR	20
1	A	103	PRO	1

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	89/104 (86%)	85±2 (96±2%)	4±2 (4±2%)	40 83
All	All	1780/2080 (86%)	1703 (96%)	77 (4%)	40 83

All 18 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	102	LYS	20
1	A	104	PHE	8
1	A	72	LEU	8
1	A	77	ILE	6
1	A	64	LYS	5
1	A	106	MET	5
1	A	5	THR	5
1	A	101	GLN	3
1	A	19	MET	3
1	A	97	ARG	3
1	A	83	THR	3
1	A	58	ASN	2
1	A	38	VAL	1
1	A	95	ARG	1
1	A	93	ARG	1
1	A	4	LEU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	11	GLU	1
1	A	12	ASP	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 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 91% for the well-defined parts and 89% for the entire structure.

7.1 Chemical shift list 1

File name: 5ieb_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	1485
Number of shifts mapped to atoms	1407
Number of unparsed shifts	78
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following errors were found when reading this chemical shift list.

- Entity instance (chain) must be specified. All 78 occurences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1408	?	119	PRO	CA	63.179	0.3	1
1409	?	119	PRO	CB	31.999	0.3	1
1410	?	119	PRO	CG	27.417	0.3	1
1411	?	119	PRO	CD	51.021	0.3	1
1412	?	119	PRO	HA	4.335	0.020	1
1413	?	119	PRO	HB2	2.238	0.020	2
1414	?	119	PRO	HB3	1.809	0.020	2
1415	?	119	PRO	HG2	1.915	0.020	1
1416	?	119	PRO	HG3	1.915	0.020	1
1417	?	119	PRO	HD2	3.872	0.020	2
1418	?	119	PRO	HD3	3.584	0.020	2
1419	?	120	ARG	C	176.966	0.3	1
1420	?	120	ARG	CA	56.289	0.3	1
1421	?	120	ARG	CB	30.694	0.3	1
1422	?	120	ARG	CG	27.105	0.3	1

Continued on next page...

Continued from previous page...

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1423	?	120	ARG	CD	43.165	0.3	1
1424	?	120	ARG	H	8.452	0.020	1
1425	?	120	ARG	HA	4.250	0.020	1
1426	?	120	ARG	HB2	1.820	0.020	2
1427	?	120	ARG	HB3	1.713	0.020	2
1428	?	120	ARG	HG2	1.641	0.020	2
1429	?	120	ARG	HG3	1.598	0.020	2
1430	?	120	ARG	HD2	3.129	0.020	1
1431	?	120	ARG	HD3	3.129	0.020	1
1432	?	120	ARG	N	122.049	0.3	1
1433	?	121	GLY	C	174.037	0.3	1
1434	?	121	GLY	CA	45.164	0.3	1
1435	?	121	GLY	H	8.410	0.020	1
1436	?	121	GLY	HA2	3.918	0.020	1
1437	?	121	GLY	HA3	3.918	0.020	1
1438	?	121	GLY	N	109.689	0.3	1
1439	?	122	SER	C	174.648	0.3	1
1440	?	122	SER	CA	58.114	0.3	1
1441	?	122	SER	CB	63.799	0.3	1
1442	?	122	SER	H	8.128	0.020	1
1443	?	122	SER	HA	4.396	0.020	1
1444	?	122	SER	HB2	3.771	0.020	1
1445	?	122	SER	HB3	3.771	0.020	1
1446	?	122	SER	N	115.439	0.3	1
1447	?	123	VAL	C	175.917	0.3	1
1448	?	123	VAL	CA	62.260	0.3	1
1449	?	123	VAL	CB	32.610	0.3	1
1450	?	123	VAL	CG1	20.968	0.3	1
1451	?	123	VAL	CG2	20.222	0.3	1
1452	?	123	VAL	H	8.073	0.020	1
1453	?	123	VAL	HA	4.018	0.020	1
1454	?	123	VAL	HB	1.982	0.020	1
1455	?	123	VAL	HG11	0.780	0.020	1
1456	?	123	VAL	HG12	0.780	0.020	1
1457	?	123	VAL	HG13	0.780	0.020	1
1458	?	123	VAL	HG21	0.821	0.020	1
1459	?	123	VAL	HG22	0.821	0.020	1
1460	?	123	VAL	HG23	0.821	0.020	1
1461	?	123	VAL	N	121.284	0.3	1
1462	?	124	GLU	C	176.039	0.3	1
1463	?	124	GLU	CA	56.446	0.3	1

Continued on next page...

Continued from previous page...

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1464	?	124	GLU	CB	30.198	0.3	1
1465	?	124	GLU	CG	36.098	0.3	1
1466	?	124	GLU	H	8.316	0.020	1
1467	?	124	GLU	HA	4.120	0.020	1
1468	?	124	GLU	HB2	1.764	0.020	2
1469	?	124	GLU	HB3	1.813	0.020	2
1470	?	124	GLU	HG2	2.059	0.020	2
1471	?	124	GLU	HG3	2.118	0.020	2
1472	?	124	GLU	N	123.718	0.3	1
1473	?	125	HIS	C	173.842	0.3	1
1474	?	125	HIS	CA	56.450	0.3	1
1475	?	125	HIS	CB	30.018	0.3	1
1476	?	125	HIS	H	8.250	0.020	1
1477	?	125	HIS	HA	4.117	0.020	1
1478	?	125	HIS	N	120.281	0.3	1
1479	?	126	HIS	CA	57.117	0.3	1
1480	?	126	HIS	CB	30.065	0.3	1
1481	?	126	HIS	H	8.077	0.020	1
1482	?	126	HIS	HA	4.359	0.020	1
1483	?	126	HIS	HB2	3.136	0.020	2
1484	?	126	HIS	HB3	3.019	0.020	2
1485	?	126	HIS	N	125.180	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
$^{13}C_{\alpha}$	116	0.27 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	107	0.33 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	99	0.11 ± 0.18	None needed (< 0.5 ppm)
^{15}N	109	0.72 ± 0.32	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 91%, i.e. 1170 atoms were assigned a chemical shift out of a possible 1292. 25 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	538/553 (97%)	220/220 (100%)	211/226 (93%)	107/107 (100%)
Sidechain	599/694 (86%)	359/397 (90%)	233/272 (86%)	7/25 (28%)
Aromatic	33/45 (73%)	23/25 (92%)	10/20 (50%)	0/0 (—%)
Overall	1170/1292 (91%)	602/642 (94%)	454/518 (88%)	114/132 (86%)

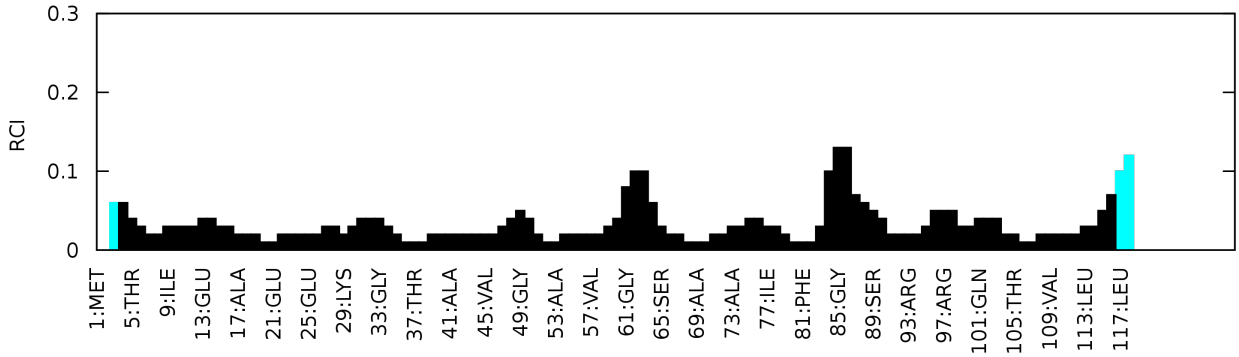
The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 89%, i.e. 1197 atoms were assigned a chemical shift out of a possible 1345. 26 out of 27 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	549/578 (95%)	225/230 (98%)	215/236 (91%)	109/112 (97%)
Sidechain	615/722 (85%)	368/413 (89%)	240/284 (85%)	7/25 (28%)
Aromatic	33/45 (73%)	23/25 (92%)	10/20 (50%)	0/0 (—%)
Overall	1197/1345 (89%)	616/668 (92%)	465/540 (86%)	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.

Random coil index (RCI) for chain A: