Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2022 – 01:06 pm BST

PDB ID	:	7QIL
Title	:	Solution NMR structure of halophilic DnaE intein
Authors	:	Heikkinen, H.A.; Aranko, S.A.; Iwai, H.
Deposited on	:	2021-12-15

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

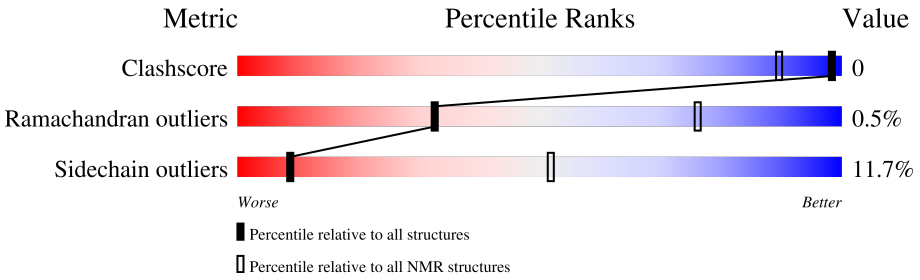
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 87%.

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	140	85% 9% 6%

2 Ensemble composition and analysis

This entry contains 20 models. Model 13 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:0-A:96, A:103-A:137 (132)	0.53	13

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 and 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2068 atoms, of which 984 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DnaE intein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	140	2068	667	984	166	246	5	0

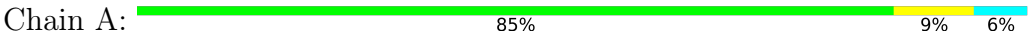
4 Residue-property plots (i)

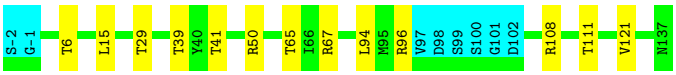
4.1 Average score per residue in the NMR ensemble

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: DnaE intein

Chain A:





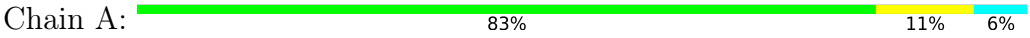
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: DnaE intein

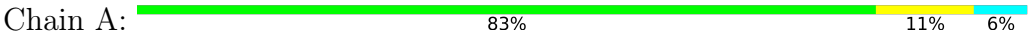
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: DnaE intein

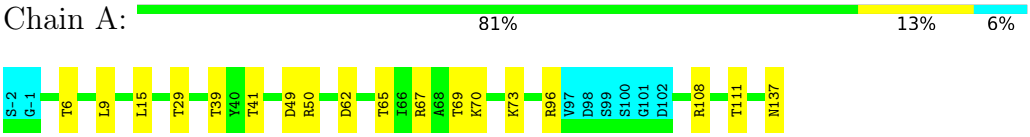
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: DnaE intein

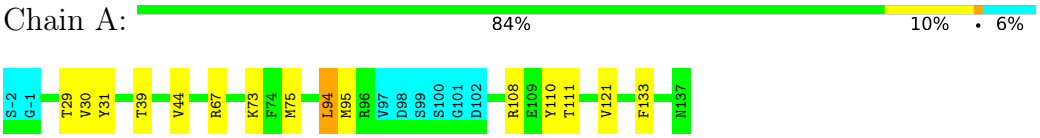
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: DnaE intein

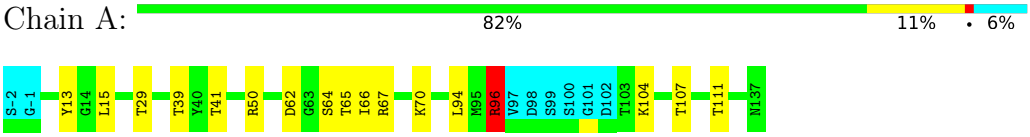
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: DnaE intein

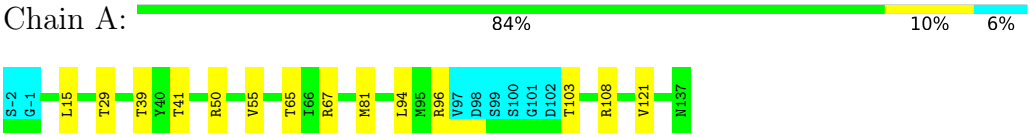
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: DnaE intein

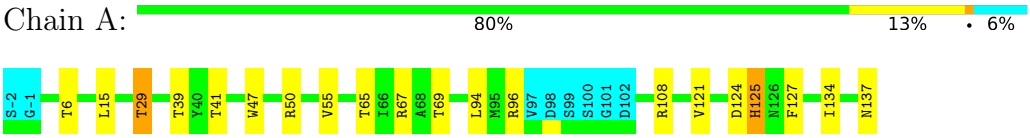
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: DnaE intein

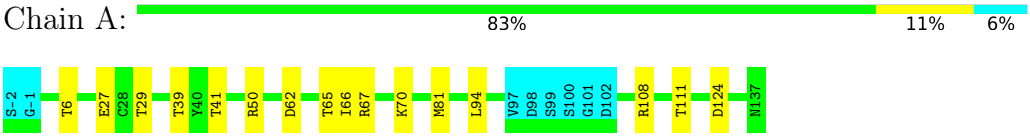
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: DnaE intein

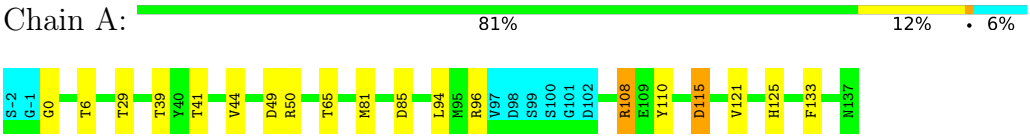
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: DnaE intein

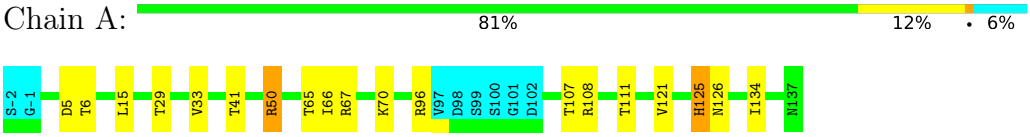
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: DnaE intein

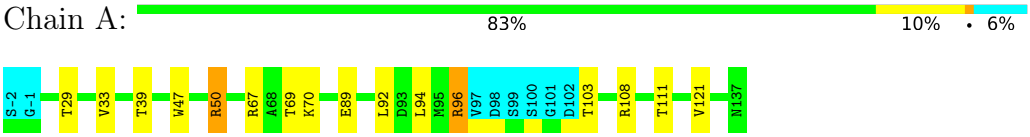
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: DnaE intein

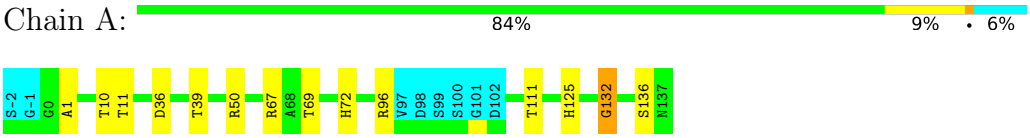
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: DnaE intein

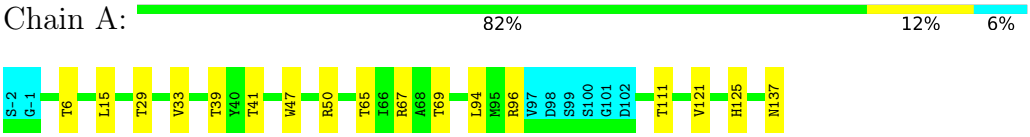
Chain A:



4.2.13 Score per residue for model 13 (medoid)

• Molecule 1: DnaE intein

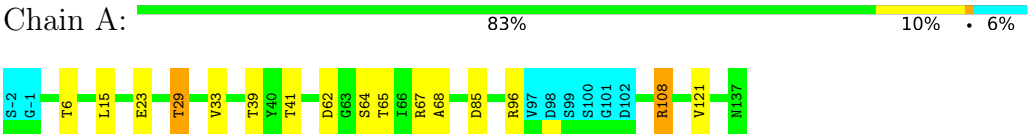
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: DnaE intein

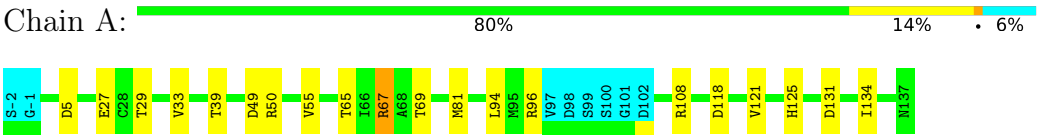
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: DnaE intein

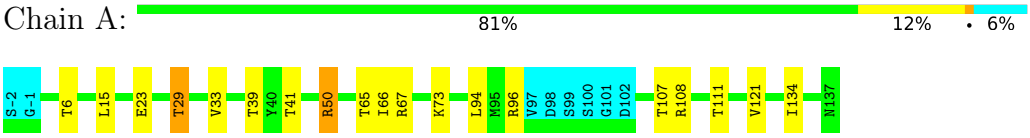
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: DnaE intein

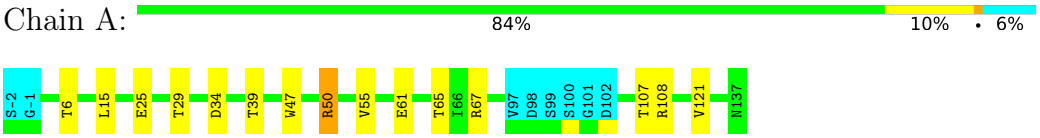
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: DnaE intein

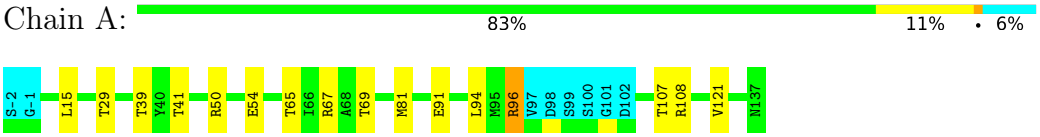
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: DnaE intein

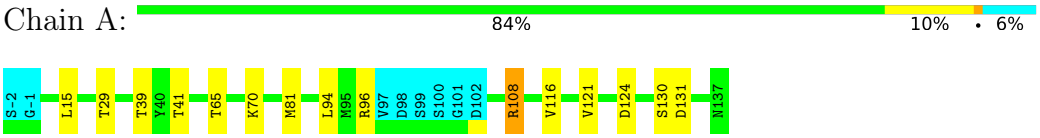
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: DnaE intein

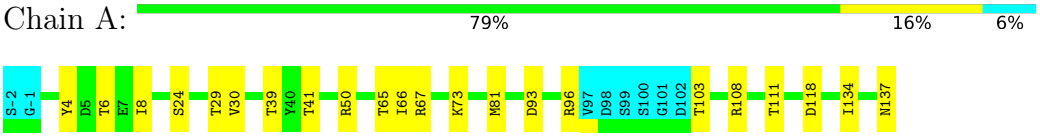
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: DnaE intein

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *DGSA-distance geometry simulated annealing.*

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *all calculated structures submitted.*

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	1551
Number of shifts mapped to atoms	1551
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

6 Model quality

6.1 Standard geometry

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	0.73±0.01	0±0/1053 (0.0± 0.0%)	1.13±0.02	4±1/1427 (0.2± 0.1%)
All	All	0.73	0/21060 (0.0%)	1.13	70/28540 (0.2%)

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
1	A	0.0±0.0	0.3±0.5
All	All	0	6

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	96	ARG	NE-CZ-NH1	9.39	124.99	120.30	3	16
1	A	108	ARG	NE-CZ-NH1	9.18	124.89	120.30	18	15
1	A	50	ARG	NE-CZ-NH1	9.12	124.86	120.30	16	14
1	A	67	ARG	NE-CZ-NH1	8.65	124.62	120.30	15	17
1	A	50	ARG	NE-CZ-NH2	-5.85	117.37	120.30	17	1
1	A	67	ARG	NE-CZ-NH2	-5.81	117.40	120.30	1	2
1	A	127	PHE	CB-CG-CD1	-5.57	116.90	120.80	7	1
1	A	96	ARG	NE-CZ-NH2	-5.48	117.56	120.30	6	2
1	A	115	ASP	CB-CG-OD2	5.35	123.12	118.30	9	1
1	A	108	ARG	NE-CZ-NH2	-5.17	117.71	120.30	20	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	96	ARG	Sidechain,Peptide	2
1	A	132	GLY	Peptide	1
1	A	68	ALA	Peptide	1
1	A	131	ASP	Peptide	1
1	A	4	TYR	Sidechain	1

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	1035	944	944	0±1
All	All	20700	18880	18880	6

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:134:ILE:HD12	1:A:134:ILE:N	0.45	2.26	20	4
1:A:94:LEU:HD22	1:A:133:PHE:CG	0.41	2.50	4	1
1:A:30:VAL:HG22	1:A:31:TYR:N	0.41	2.31	4	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	131/140 (94%)	124±3 (95±2%)	6±2 (5±2%)	1±1 (1±1%)	32	76
All	All	2620/2800 (94%)	2485 (95%)	121 (5%)	14 (1%)	32	76

All 9 unique Ramachandran outliers are listed below. They are sorted by the frequency of occur-

rence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	125	HIS	4
1	A	29	THR	3
1	A	13	TYR	1
1	A	0	GLY	1
1	A	103	THR	1
1	A	1	ALA	1
1	A	132	GLY	1
1	A	136	SER	1
1	A	131	ASP	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	115/121 (95%)	102±2 (88±2%)	13±2 (12±2%)	9 52
All	All	2300/2420 (95%)	2032 (88%)	268 (12%)	9 52

All 61 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	29	THR	19
1	A	39	THR	19
1	A	65	THR	17
1	A	121	VAL	14
1	A	41	THR	14
1	A	94	LEU	13
1	A	6	THR	12
1	A	111	THR	12
1	A	15	LEU	11
1	A	81	MET	9
1	A	33	VAL	8
1	A	69	THR	8
1	A	107	THR	6
1	A	70	LYS	6
1	A	50	ARG	6
1	A	55	VAL	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	73	LYS	5
1	A	108	ARG	5
1	A	137	ASN	5
1	A	66	ILE	5
1	A	62	ASP	4
1	A	47	TRP	4
1	A	125	HIS	4
1	A	103	THR	3
1	A	49	ASP	3
1	A	96	ARG	3
1	A	124	ASP	3
1	A	72	HIS	2
1	A	93	ASP	2
1	A	44	VAL	2
1	A	110	TYR	2
1	A	64	SER	2
1	A	27	GLU	2
1	A	85	ASP	2
1	A	5	ASP	2
1	A	67	ARG	2
1	A	23	GLU	2
1	A	118	ASP	2
1	A	9	LEU	1
1	A	75	MET	1
1	A	95	MET	1
1	A	104	LYS	1
1	A	115	ASP	1
1	A	133	PHE	1
1	A	126	ASN	1
1	A	89	GLU	1
1	A	92	LEU	1
1	A	10	THR	1
1	A	11	THR	1
1	A	36	ASP	1
1	A	134	ILE	1
1	A	25	GLU	1
1	A	34	ASP	1
1	A	61	GLU	1
1	A	54	GLU	1
1	A	91	GLU	1
1	A	116	VAL	1
1	A	130	SER	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	8	ILE	1
1	A	24	SER	1
1	A	30	VAL	1

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

There are no ligands in this entry.

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 87% for the well-defined parts and 86% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *NpuDnaE.str3*

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	1551
Number of shifts mapped to atoms	1551
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	138	-0.33 ± 0.21	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	125	-0.13 ± 0.19	None needed (< 0.5 ppm)
$^{13}C'$	135	-0.06 ± 0.12	None needed (< 0.5 ppm)
^{15}N	129	0.91 ± 0.38	Should be applied

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 87%, i.e. 1313 atoms were assigned a chemical shift out of a possible 1511. 16 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	640/656 (98%)	256/262 (98%)	259/264 (98%)	125/130 (96%)
Sidechain	614/721 (85%)	374/414 (90%)	236/288 (82%)	4/19 (21%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	59/134 (44%)	42/71 (59%)	17/59 (29%)	0/4 (0%)
Overall	1313/1511 (87%)	672/747 (90%)	512/611 (84%)	129/153 (84%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 86%, i.e. 1360 atoms were assigned a chemical shift out of a possible 1574. 17 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	669/696 (96%)	267/278 (96%)	273/280 (98%)	129/138 (93%)
Sidechain	632/744 (85%)	385/427 (90%)	243/298 (82%)	4/19 (21%)
Aromatic	59/134 (44%)	42/71 (59%)	17/59 (29%)	0/4 (0%)
Overall	1360/1574 (86%)	694/776 (89%)	533/637 (84%)	133/161 (83%)

7.1.4 Statistically unusual chemical shifts (i)

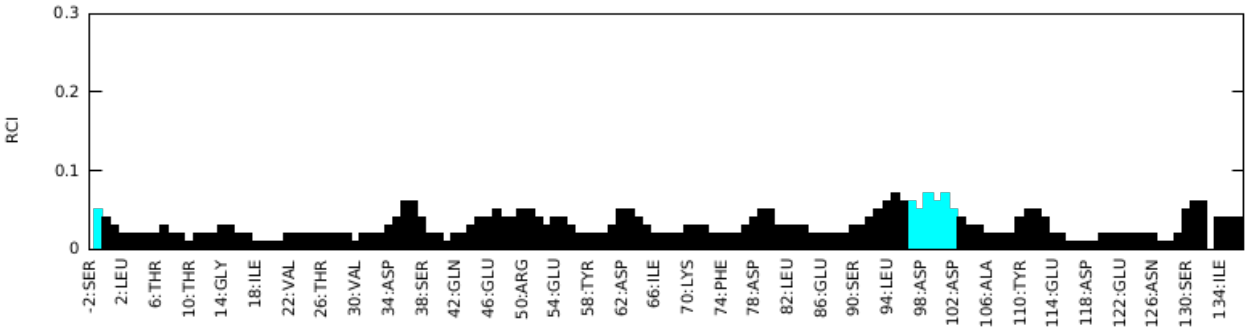
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	25	GLU	HB3	0.57	3.10 – 0.90	-6.5
1	A	19	GLY	HA3	1.53	5.80 – 2.00	-6.2
1	A	32	SER	HB3	2.11	5.25 – 2.45	-6.2
1	A	108	ARG	HB3	0.11	3.17 – 0.37	-5.9
1	A	19	GLY	HA2	1.85	5.87 – 2.07	-5.6

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:


RCI
0.3
0.2
0.1
0
-2:SER
2:LEU
6:THR
10:THR
14:GLY
18:ILE
22:VAL
26:THR
30:VAL
34:ASP
38:SER
42:GLN
46:GLU
50:ARG
54:GLU
58:TYR
62:ASP
66:ILE
70:LYS
74:PHE
78:ASP
82:LEU
86:GLU
90:SER
94:LEU
98:ASP
102:ASP
106:ALA
110:TYR
114:GLU
118:ASP
122:GLU
126:ASN
130:SER
134:ILE