Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:47 pm BST

PDB ID	:	2L04
Title	:	The Solution Structure of the C-terminal Ig-like Domain of the Bacteriophage Lambda Tail Tube Protein
Authors	:	Pell, L.G.; Gasmi-Seabrook, G.M.C.; Donaldson, L.W.; Howell, P.; Davidson, A.R.; Maxwell, K.L.
Deposited on	:	2010-06-30

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 (1)) were used in the production of this report:

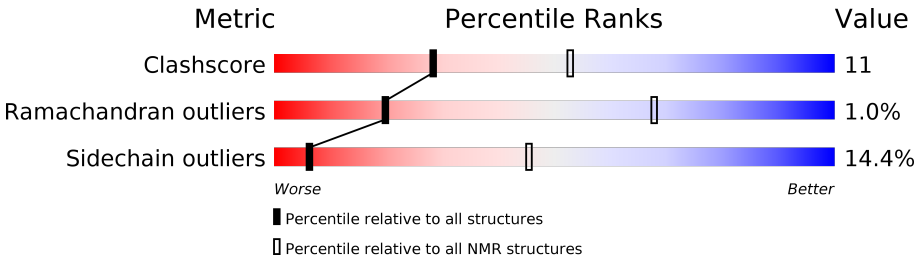
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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	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	91	63% 27% 10%

2 Ensemble composition and analysis

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:161-A:187, A:192-A:246 (82)	0.16	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, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19
2	2, 6, 10, 20

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1256 atoms, of which 637 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Major tail protein V.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	91	1256	383	637	104	129	3	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	156	GLY	-	expression tag	UNP P03733
A	157	ALA	-	expression tag	UNP P03733
A	158	MET	-	expression tag	UNP P03733
A	159	ALA	-	expression tag	UNP P03733

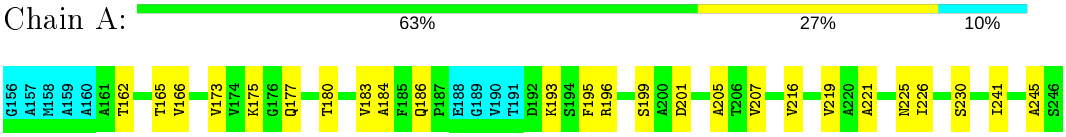
4 Residue-property plots

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: Major tail protein V

Chain A:



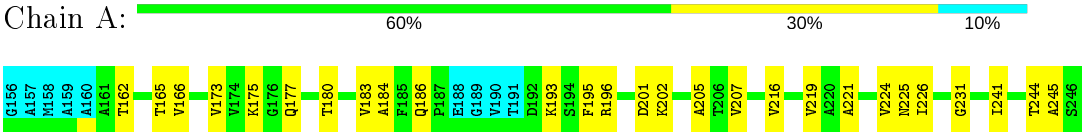
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Major tail protein V

Chain A:



4.2.2 Score per residue for model 2

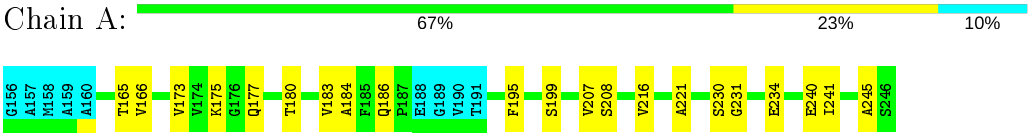
- Molecule 1: Major tail protein V

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Major tail protein V

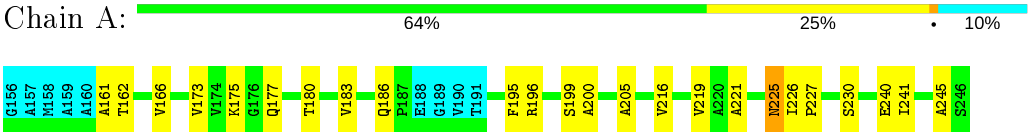
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Major tail protein V

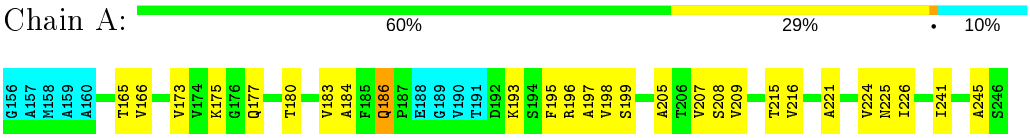
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Major tail protein V

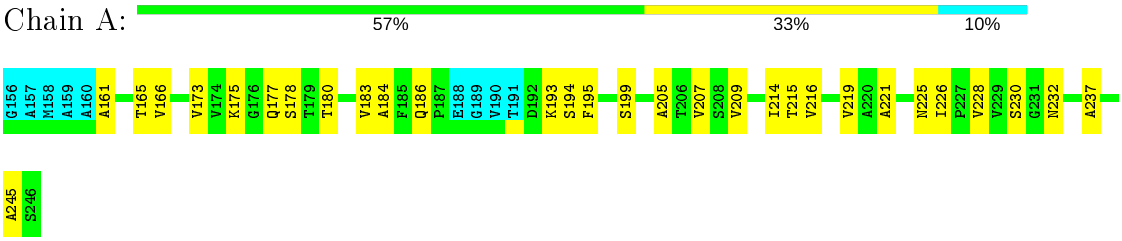
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Major tail protein V

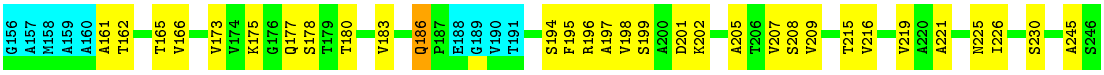
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Major tail protein V

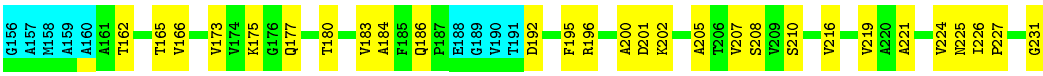
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Major tail protein V

Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Major tail protein V

Chain A:

4.2.10 Score per residue for model 10

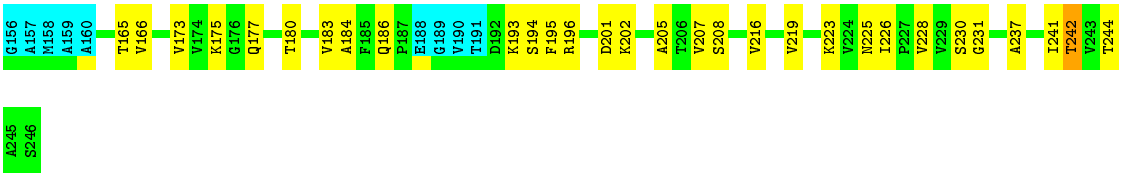
- Molecule 1: Major tail protein V

Chain A:

4.2.11 Score per residue for model 11

- Molecule 1: Major tail protein V

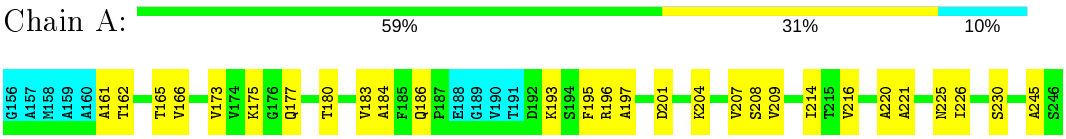
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Major tail protein V

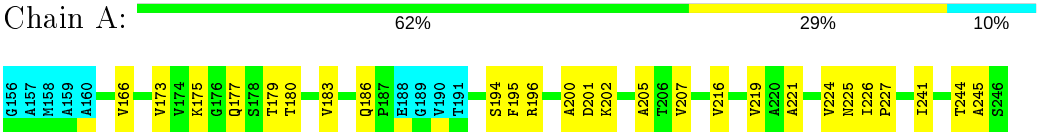
Chain A:



4.2.13 Score per residue for model 13

- Molecule 1: Major tail protein V

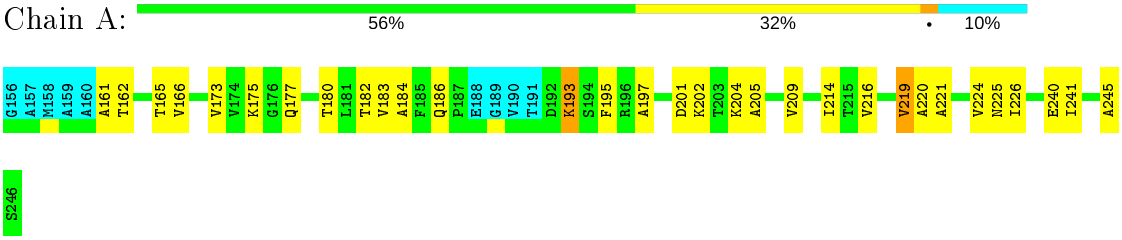
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Major tail protein V

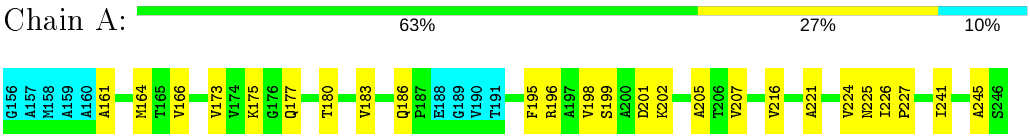
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Major tail protein V

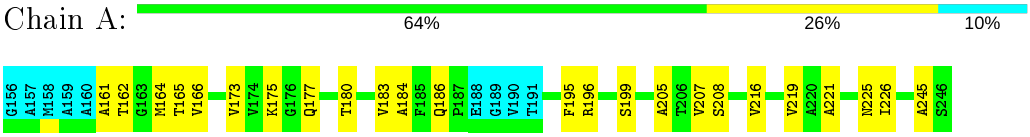
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Major tail protein V

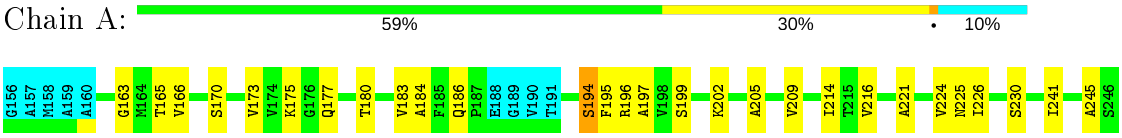
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Major tail protein V

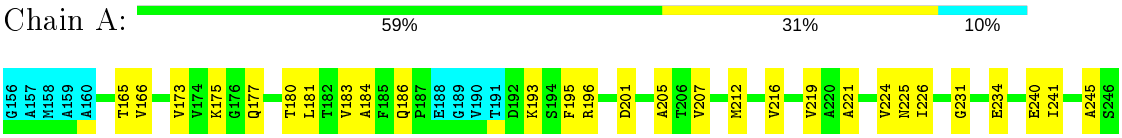
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Major tail protein V

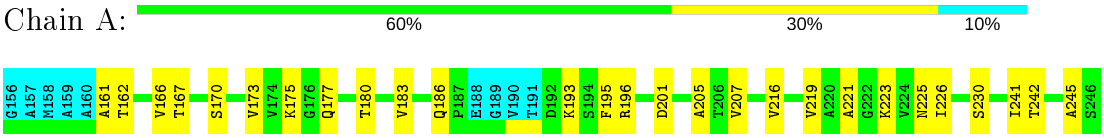
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: Major tail protein V

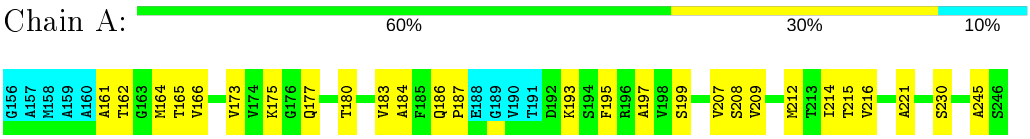
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Major tail protein V

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 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
CYANA	structure solution	
CYANA	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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1000
Number of shifts mapped to atoms	1000
Number of unparsed shifts	0
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

6.1 Standard geometry

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

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	565	585	585	12±2
All	All	11300	11700	11700	242

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 11.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:166:VAL:HG22	1:A:183:VAL:HG13	0.96	1.37	4	20
1:A:207:VAL:HG12	1:A:216:VAL:HG22	0.82	1.51	3	17
1:A:166:VAL:HG13	1:A:183:VAL:HG22	0.76	1.57	14	19
1:A:183:VAL:HG11	1:A:195:PHE:CE2	0.69	2.22	17	20
1:A:173:VAL:HG22	1:A:177:GLN:HB2	0.61	1.72	20	20
1:A:197:ALA:HB3	1:A:209:VAL:HB	0.58	1.75	2	5
1:A:223:LYS:HG3	1:A:242:THR:HG23	0.58	1.75	11	3
1:A:197:ALA:HB3	1:A:209:VAL:CG2	0.58	2.29	14	2
1:A:165:THR:OG1	1:A:184:ALA:HB3	0.56	2.01	18	15
1:A:225:ASN:C	1:A:226:ILE:HD12	0.55	2.21	13	18
1:A:204:LYS:HA	1:A:220:ALA:HB3	0.54	1.79	14	1
1:A:205:ALA:HA	1:A:219:VAL:HG22	0.53	1.80	9	2
1:A:162:THR:HG22	1:A:187:PRO:HA	0.53	1.80	20	1
1:A:205:ALA:CB	1:A:224:VAL:HG11	0.53	2.34	14	9

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:198:VAL:O	1:A:226:ILE:HG23	0.52	2.05	2	3
1:A:209:VAL:HG23	1:A:214:ILE:HG13	0.52	1.79	10	5
1:A:205:ALA:HA	1:A:219:VAL:HG23	0.52	1.81	11	12
1:A:221:ALA:HB2	1:A:245:ALA:HA	0.52	1.81	1	19
1:A:228:VAL:CG2	1:A:237:ALA:HB3	0.51	2.36	6	2
1:A:201:ASP:O	1:A:203:THR:HG23	0.50	2.07	9	1
1:A:164:MET:SD	1:A:166:VAL:HG23	0.49	2.46	16	1
1:A:240:GLU:O	1:A:241:ILE:HD13	0.49	2.07	10	6
1:A:216:VAL:HG11	1:A:241:ILE:HG13	0.48	1.84	19	11
1:A:200:ALA:HB2	1:A:227:PRO:HD3	0.46	1.88	9	4
1:A:183:VAL:HG11	1:A:195:PHE:CZ	0.46	2.45	13	1
1:A:205:ALA:HB3	1:A:224:VAL:HG11	0.45	1.88	9	1
1:A:166:VAL:CG2	1:A:183:VAL:HG13	0.45	2.25	16	1
1:A:197:ALA:HB3	1:A:209:VAL:HG21	0.45	1.89	14	1
1:A:221:ALA:HB2	1:A:245:ALA:CA	0.44	2.42	7	9
1:A:165:THR:HG23	1:A:186:GLN:HE22	0.44	1.72	5	2
1:A:208:SER:C	1:A:214:ILE:HG23	0.43	2.34	20	1
1:A:173:VAL:O	1:A:244:THR:HG23	0.42	2.15	1	5
1:A:198:VAL:HG23	1:A:227:PRO:HG2	0.41	1.91	15	1
1:A:204:LYS:CB	1:A:220:ALA:HB3	0.41	2.45	12	1
1:A:179:THR:HG21	1:A:241:ILE:HD12	0.40	1.93	13	1
1:A:209:VAL:HG23	1:A:214:ILE:CG1	0.40	2.47	10	1
1:A:181:LEU:HD12	1:A:181:LEU:N	0.40	2.32	18	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	81/91 (89%)	69±1 (85±2%)	12±1 (14±2%)	1±1 (1±1%)	20	68
All	All	1620/1820 (89%)	1372 (85%)	231 (14%)	17 (1%)	20	68

All 5 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	161	ALA	9
1	A	193	LYS	3
1	A	170	SER	2
1	A	234	GLU	2
1	A	163	GLY	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	63/67 (94%)	54±2 (86±3%)	9±2 (14±3%)	6 45
All	All	1260/1340 (94%)	1079 (86%)	181 (14%)	6 45

All 26 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	175	LYS	20
1	A	186	GLN	20
1	A	180	THR	20
1	A	196	ARG	16
1	A	230	SER	12
1	A	201	ASP	12
1	A	162	THR	11
1	A	199	SER	11
1	A	193	LYS	10
1	A	202	LYS	10
1	A	208	SER	7
1	A	215	THR	5
1	A	194	SER	5
1	A	212	MET	4
1	A	178	SER	3
1	A	164	MET	2
1	A	219	VAL	2
1	A	214	ILE	2
1	A	182	THR	2
1	A	232	ASN	1
1	A	192	ASP	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	242	THR	1
1	A	210	SER	1
1	A	225	ASN	1
1	A	167	THR	1
1	A	213	THR	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 carbohydrates 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 91% for the well-defined parts and 90% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

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	1000
Number of shifts mapped to atoms	1000
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	88	-0.02 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	81	-0.07 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	80	0.28 ± 0.21	None needed (< 0.5 ppm)
^{15}N	86	0.19 ± 0.55	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 91%, i.e. 772 atoms were assigned a chemical shift out of a possible 852. 1 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	391/404 (97%)	159/161 (99%)	153/164 (93%)	79/79 (100%)
Sidechain	363/421 (86%)	215/239 (90%)	143/169 (85%)	5/13 (38%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	18/27 (67%)	10/15 (67%)	8/12 (67%)	0/0 (—%)
Overall	772/852 (91%)	384/415 (93%)	304/345 (88%)	84/92 (91%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 90%, i.e. 834 atoms were assigned a chemical shift out of a possible 928. 1 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	428/449 (95%)	174/179 (97%)	168/182 (92%)	86/88 (98%)
Sidechain	388/452 (86%)	229/256 (89%)	154/183 (84%)	5/13 (38%)
Aromatic	18/27 (67%)	10/15 (67%)	8/12 (67%)	0/0 (—%)
Overall	834/928 (90%)	413/450 (92%)	330/377 (88%)	91/101 (90%)

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: