wwPDB NMR Structure Validation Summary Report (i)

Nov 14, 2016 – 06:59 PM EST

PDB ID	:	5M1G
Title	:	Structure of a Spumaretrovirus Gag central domain reveals an ancient retroviral capsid
Authors	:	Nicastro, G.; Ball, N.; Taylor, I.A.
Deposited on	:	2016-10-07

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:

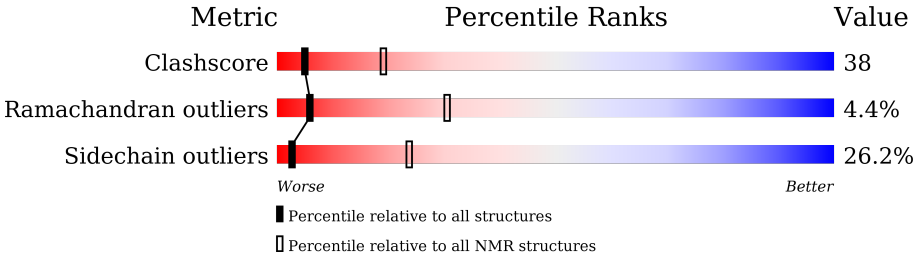
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20028320
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028320

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 39%.

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	99	34% 43% 14% 8%
1	B	99	35% 42% 14% 8%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 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:386-A:476, B:386-B:476 (182)	0.18	8

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3086 atoms, of which 1548 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Gag protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	99	1543	480	774	139	147	3	0
1	B	99	1543	480	774	139	147	3	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	449	ASP	ASN	conflict	UNP P14349
A	478	LEU	-	expression tag	UNP P14349
A	479	GLU	-	expression tag	UNP P14349
B	449	ASP	ASN	conflict	UNP P14349
B	478	LEU	-	expression tag	UNP P14349
B	479	GLU	-	expression tag	UNP P14349

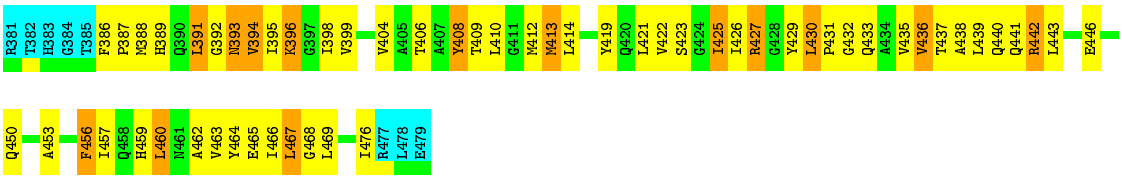
4 Residue-property plots (i)

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.

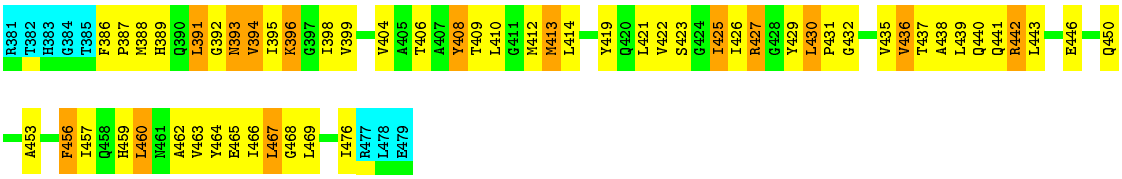
- Molecule 1: Gag protein

Chain A: 34% 43% 14% 8%



- Molecule 1: Gag protein

Chain B: 35% 42% 14% 8%

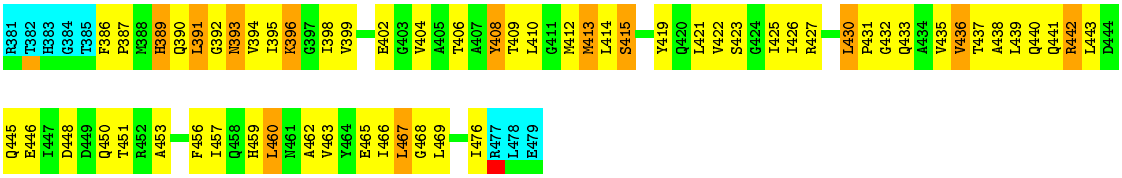


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 8. Colouring as in section 4.1 above.

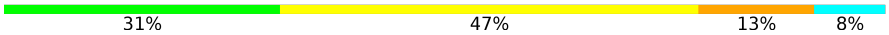
- Molecule 1: Gag protein

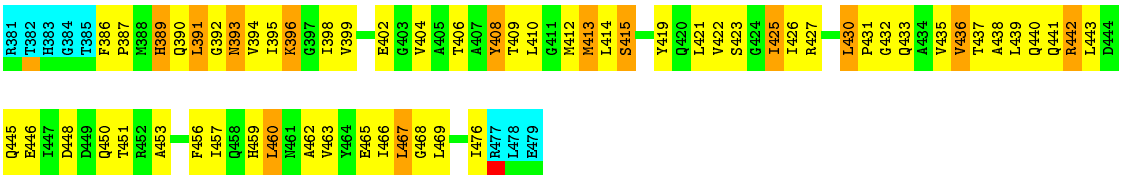
Chain A: 31% 48% 12% 8%



- Molecule 1: Gag protein

Chain B:





5 Refinement protocol and experimental data overview (i)

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

Of the 1000 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

The authors did not provide any information on software used for structure solution, optimization or 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)	5m1g_cs.cif
Number of chemical shift lists	4
Total number of shifts	2512
Number of shifts mapped to atoms	2512
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	39%

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.

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	0.1±0.3
1	B	0.0±0.0	0.1±0.3
All	All	0	4

There are no bond-length outliers.

There are no bond-angle outliers.

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	B	442	ARG	Sidechain	2
1	A	442	ARG	Sidechain	2

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	701	704	702	55±6
1	B	701	704	702	55±6
All	All	28040	28160	28080	2109

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 38.

5 of 304 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:413:MET:SD	1:A:414:LEU:HG	0.85	2.11	4	20
1:B:413:MET:SD	1:B:414:LEU:HG	0.85	2.11	4	20
1:A:398:ILE:HD13	1:A:410:LEU:HD12	0.81	1.52	19	20
1:B:439:LEU:HD23	1:B:442:ARG:NH2	0.81	1.89	19	2
1:A:439:LEU:HD23	1:A:442:ARG:NH2	0.81	1.89	19	2

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	91/99 (92%)	80±1 (88±2%)	7±1 (7±1%)	4±1 (4±1%)	6	30
1	B	91/99 (92%)	80±1 (88±2%)	7±1 (7±1%)	4±1 (4±1%)	6	30
All	All	3640/3960 (92%)	3214 (88%)	265 (7%)	161 (4%)	6	30

5 of 18 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	B	446	GLU	20
1	A	446	GLU	20
1	A	391	LEU	17
1	B	391	LEU	17
1	B	431	PRO	16

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	74/81 (91%)	55±1 (74±1%)	19±1 (26±1%)	3	24
1	B	74/81 (91%)	55±1 (74±1%)	19±1 (26±1%)	3	24

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2960/3240 (91%)	2183 (74%)	777 (26%)	3 24

5 of 56 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	B	393	ASN	20
1	A	465	GLU	20
1	A	422	VAL	20
1	A	460	LEU	20
1	B	421	LEU	20

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 39% for the well-defined parts and 38% for the entire structure.

7.1 Chemical shift list 1

File name: 5m1g_cs.cif

Chemical shift list name: *13C-ARO-HPFV-381-477-chemical-shifts*

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	151
Number of shifts mapped to atoms	151
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 (i)

No chemical shift referencing corrections were calculated (not enough data).

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 5%, i.e. 104 atoms were assigned a chemical shift out of a possible 2194. 0 out of 34 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	14/902 (2%)	14/360 (4%)	0/364 (0%)	0/178 (0%)
Sidechain	52/1160 (4%)	52/670 (8%)	0/434 (0%)	0/56 (0%)
Aromatic	38/132 (29%)	25/68 (37%)	13/56 (23%)	0/8 (0%)
Overall	104/2194 (5%)	91/1098 (8%)	13/854 (2%)	0/242 (0%)

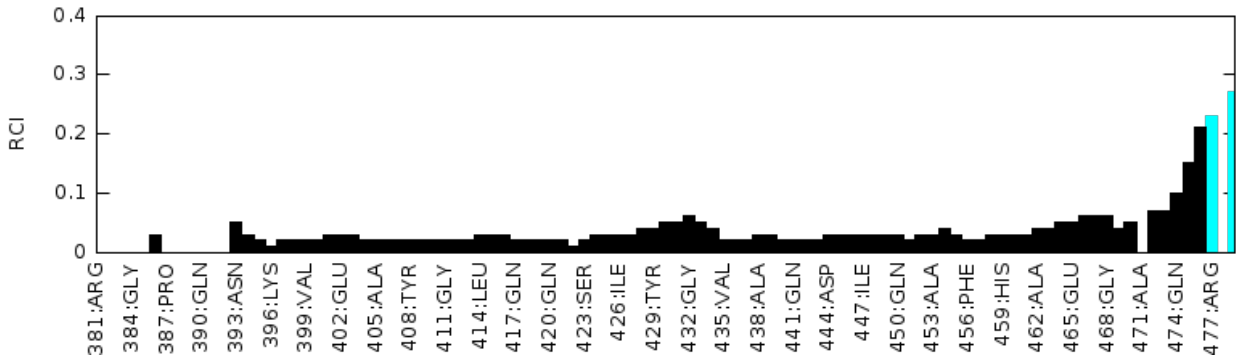
7.1.4 Statistically unusual chemical shifts (i)

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:



7.2 Chemical shift list 2

File name: 5m1g_cs.cif

Chemical shift list name: *13C-HPFV-381-477-27012015-chemical-shifts*

7.2.1 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	911
Number of shifts mapped to atoms	911
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

7.2.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	89	-0.57 ± 0.08	Should be applied
$^{13}C_{\beta}$	82	0.17 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	0	—	None (insufficient data)

7.2.3 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 33%, i.e. 719 atoms were assigned a chemical shift out of a possible 2194. 15 out of 34 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	244/902 (27%)	161/360 (45%)	83/364 (23%)	0/178 (0%)
Sidechain	451/1160 (39%)	277/670 (41%)	174/434 (40%)	0/56 (0%)
Aromatic	24/132 (18%)	24/68 (35%)	0/56 (0%)	0/8 (0%)
Overall	719/2194 (33%)	462/1098 (42%)	257/854 (30%)	0/242 (0%)

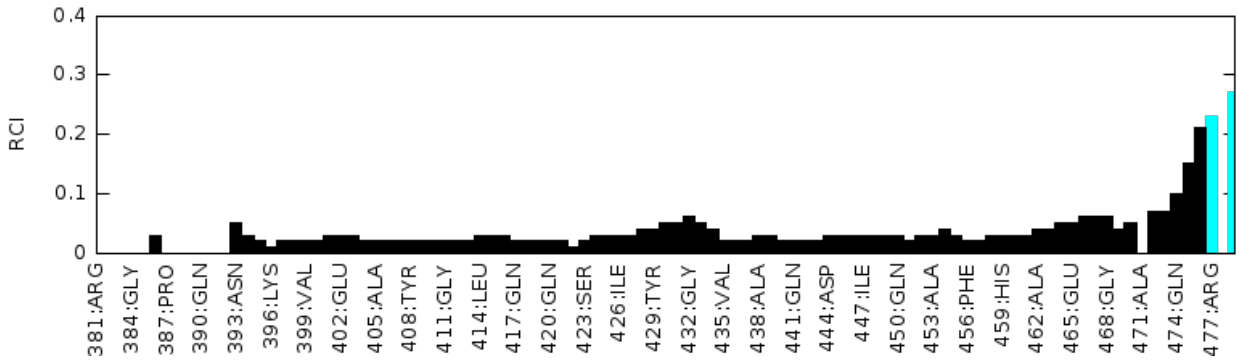
7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.2.5 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:



7.3 Chemical shift list 3

File name: 5m1g_cs.cif

Chemical shift list name: *13C-HPFV-381-477-D2O-chemical-shifts*

7.3.1 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	818
Number of shifts mapped to atoms	818
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

7.3.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	92	-0.49 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	80	0.14 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	0	—	None (insufficient data)

7.3.3 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 28%, i.e. 619 atoms were assigned a chemical shift out of a possible 2194. 14 out of 34 assigned methyl groups (LEU and VAL) were assigned

stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	171/902 (19%)	86/360 (24%)	85/364 (23%)	0/178 (0%)
Sidechain	419/1160 (36%)	252/670 (38%)	167/434 (38%)	0/56 (0%)
Aromatic	29/132 (22%)	29/68 (43%)	0/56 (0%)	0/8 (0%)
Overall	619/2194 (28%)	367/1098 (33%)	252/854 (30%)	0/242 (0%)

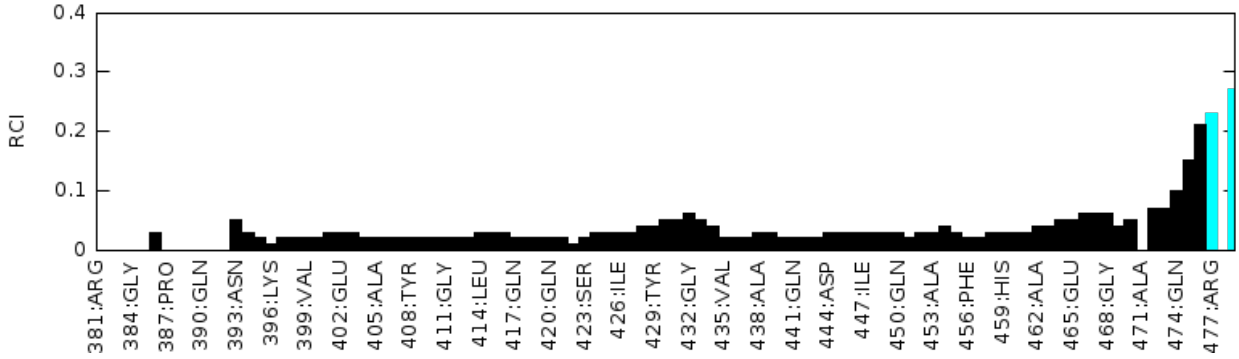
7.3.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.3.5 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:



7.4 Chemical shift list 4

File name: 5m1g_cs.cif

Chemical shift list name: *15N-HPFV-381-477-chemical-shifts*

7.4.1 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	632
Number of shifts mapped to atoms	632
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

7.4.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	None (insufficient data)
$^{13}C_{\beta}$	0	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	79	0.73 ± 0.31	Should be applied

7.4.3 Completeness of resonance assignments (i)

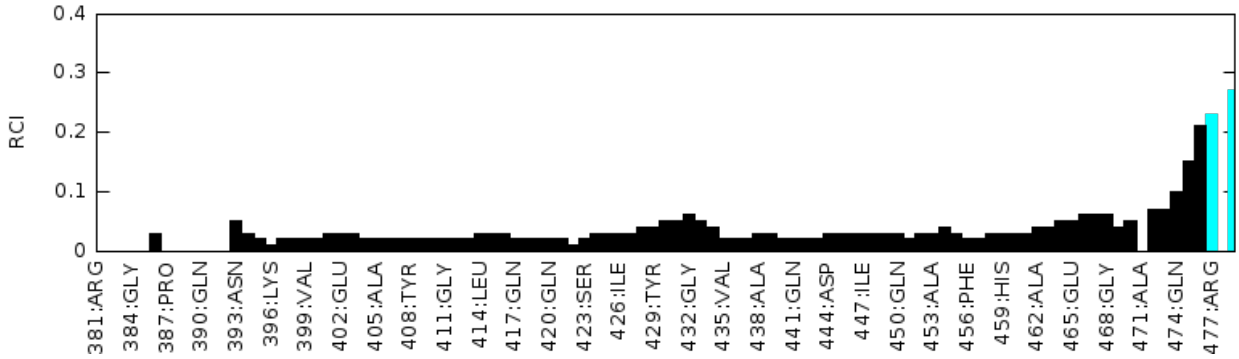
The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 23%, i.e. 509 atoms were assigned a chemical shift out of a possible 2194. 0 out of 34 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	235/902 (26%)	158/360 (44%)	0/364 (0%)	77/178 (43%)
Sidechain	265/1160 (23%)	251/670 (37%)	0/434 (0%)	14/56 (25%)
Aromatic	9/132 (7%)	9/68 (13%)	0/56 (0%)	0/8 (0%)
Overall	509/2194 (23%)	418/1098 (38%)	0/854 (0%)	91/242 (38%)

7.4.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.4.5 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: