WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

Feb 12, 2017 – 10:43 pm GMT

PDB ID : 2K5V

Title : SOLUTION NMR STRUCTURE OF the second OB-fold domain of replication protein A from Methanococcus maripaludis. NORTHEAST STRUCTURAL GENOMICS TARGET MrR110B.

Authors : Aramini, J.M.; Maglaqui, M.; Jiang, M.; Ciccosanti, C.; Xiao, R.; Nair, R.; Everett, J.K.; Swapna, G.VT.; Acton, T.B.; Rost, B.; Montelione, G.T.; Northeast Structural Genomics Consortium (NESG)
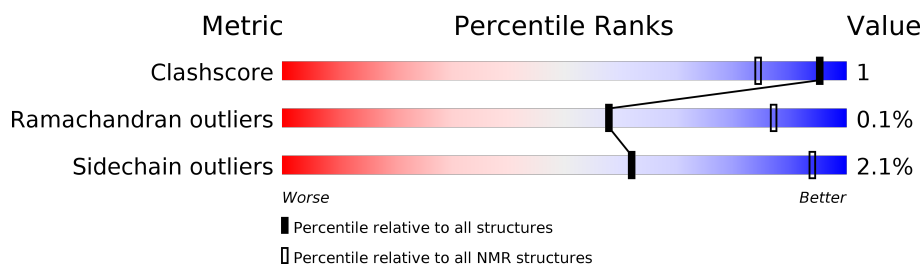
Deposited on : 2008-06-30

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 97%.

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	Percentile Ranks	Value
Clashscore		1
Ramachandran outliers		0.1%
Sidechain outliers		2.1%

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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	104	77% • 16% 6%

W O R L D W I D E

wwPDB

PROTEIN DATA BANK

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 15 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	A:173-A:200, A:209-A:249, A:254-A:265 (81)	0.37	15

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1527 atoms, of which 769 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Replication protein A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	98	1527	481	769	123	152	2	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	172	MET	-	EXPRESSION TAG	UNP Q6LYF9
A	193	ALA	THR	ENGINEERED	UNP Q6LYF9
A	268	LEU	-	EXPRESSION TAG	UNP Q6LYF9
A	269	GLU	-	EXPRESSION TAG	UNP Q6LYF9
A	270	HIS	-	EXPRESSION TAG	UNP Q6LYF9
A	271	HIS	-	EXPRESSION TAG	UNP Q6LYF9
A	272	HIS	-	EXPRESSION TAG	UNP Q6LYF9
A	273	HIS	-	EXPRESSION TAG	UNP Q6LYF9
A	274	HIS	-	EXPRESSION TAG	UNP Q6LYF9
A	275	HIS	-	EXPRESSION TAG	UNP Q6LYF9

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.

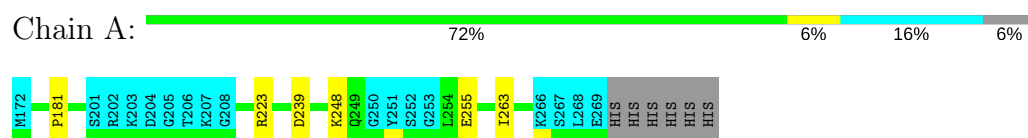• Molecule 1: Replication protein A

Chain A:
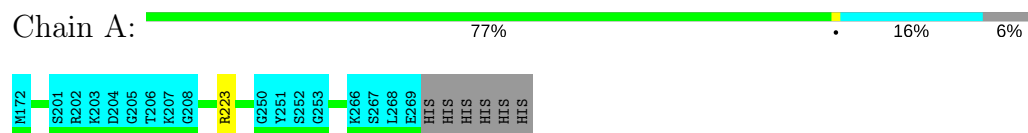
4.2 Scores per residue for each member of the ensemble
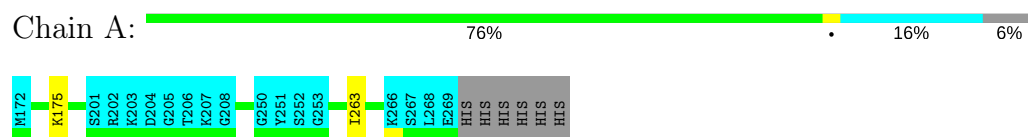
Colouring as in section 4.1 above.

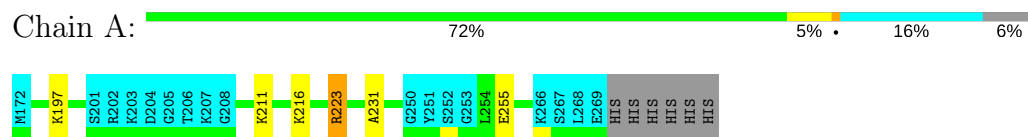### 4.2.1 Score per residue for model 1

• Molecule 1: Replication protein A

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: Replication protein A

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Replication protein A

Chain A: 72% 5% 16% 6%

M172 K197 S201 R202 K203 D204 G205 T206 K207 G208 K211 K216 R223 A231 G250 Y251 S252 G253 L254 E255 K266 S267 L268 E269 HIS HIS HIS HIS HIS HIS

4.2.4 Score per residue for model 4

• Molecule 1: Replication protein A

Chain A: 74% 16% 6%

M172 S201 R202 K203 D204 G205 T206 K207 G208 L215 E229 V243 G250 Y251 S252 G253 L254 E255 K266 S267 L268 E269 HIS HIS HIS HIS HIS HIS

4.2.5 Score per residue for model 5

• Molecule 1: Replication protein A

Chain A: 74% 16% 6%

M172 F200 S201 R202 K203 D204 G205 T206 K207 G208 R223 D239 G250 Y251 S252 G253 I263 K266 S267 L268 E269 HIS HIS HIS HIS HIS HIS

4.2.6 Score per residue for model 6

• Molecule 1: Replication protein A

Chain A: 75% 16% 6%

M172 S201 R202 K203 D204 G205 T206 K207 G208 K211 K248 Q249 G250 Y251 S252 G253 L254 E255 K266 S267 L268 E269 HIS HIS HIS HIS HIS HIS

4.2.7 Score per residue for model 7

• Molecule 1: Replication protein A

Chain A: 74% 16% 6%

M172 K175 K197 S201 R202 K203 D204 G205 T206 K207 G208 K216 D232 G250 Y251 S252 G253 K266 S267 L268 E269 HIS HIS HIS HIS HIS HIS

4.2.8 Score per residue for model 8

• Molecule 1: Replication protein A

Chain A:

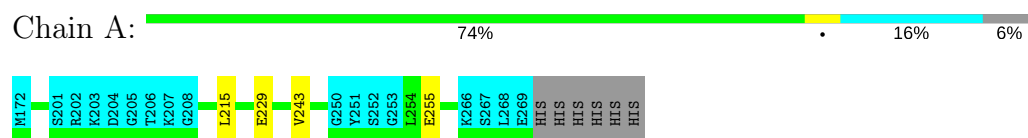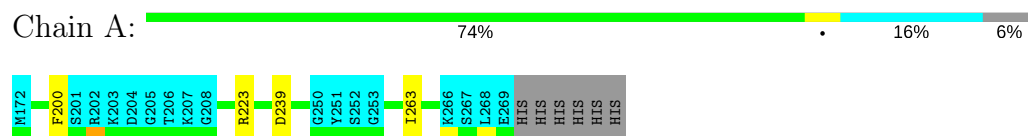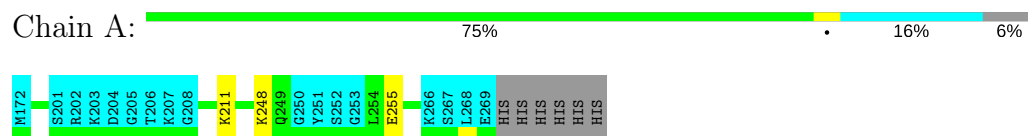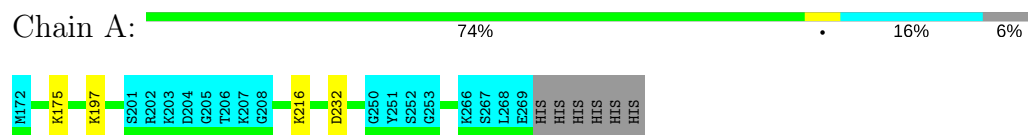#### 4.2.9 Score per residue for model 9

• Molecule 1: Replication protein A

Chain A:

4.2.10 Score per residue for model 10

• Molecule 1: Replication protein A

Chain A:

4.2.11 Score per residue for model 11

• Molecule 1: Replication protein A

Chain A:

4.2.12 Score per residue for model 12

• Molecule 1: Replication protein A

Chain A:

4.2.13 Score per residue for model 13
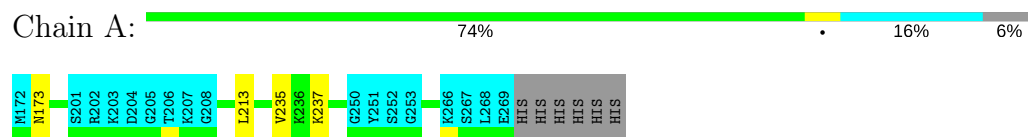
• Molecule 1: Replication protein A

Chain A:

4.2.14 Score per residue for model 14

• Molecule 1: Replication protein A

Chain A:
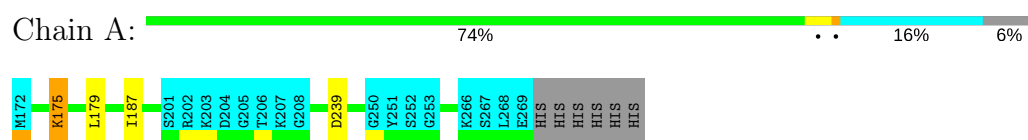
4.2.15 Score per residue for model 15 (medoid)

• Molecule 1: Replication protein A

Chain A:
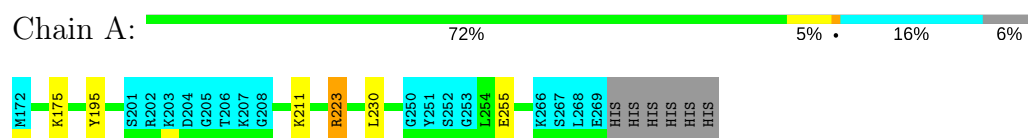
4.2.16 Score per residue for model 16

• Molecule 1: Replication protein A

Chain A:
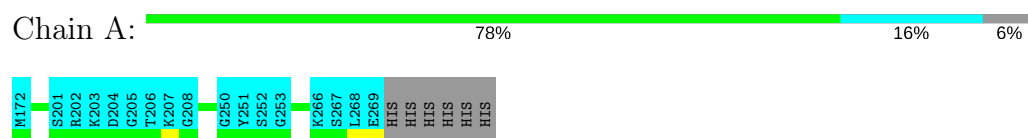
4.2.17 Score per residue for model 17
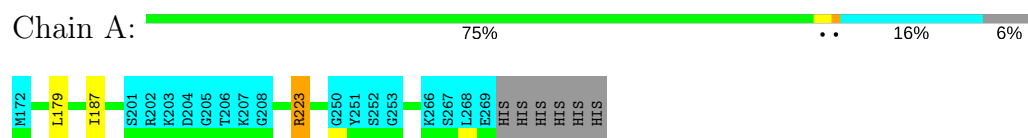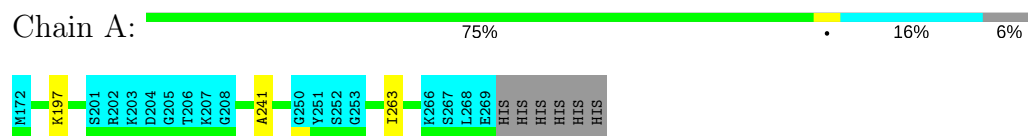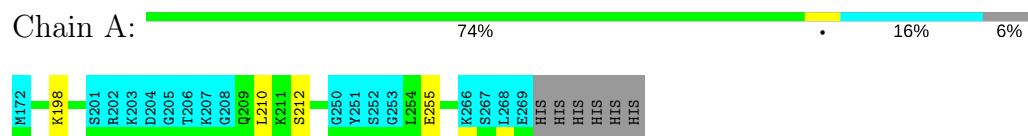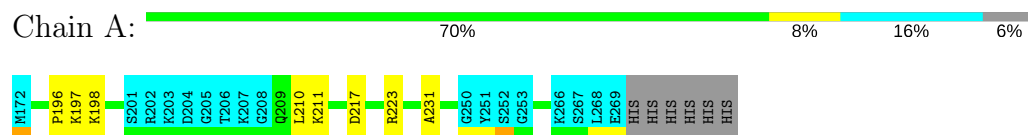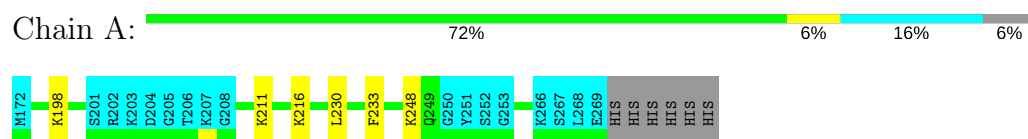
• Molecule 1: Replication protein A

Chain A:
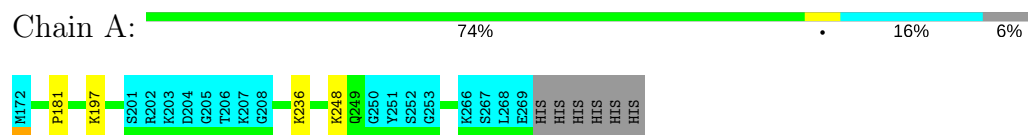
4.2.18 Score per residue for model 18

• Molecule 1: Replication protein A

Chain A:

76%   •   16%   6%

M172 K198 S201 R202 K203 D204 G205 T206 K207 G208 K216 G250 Y251 S252 G253 K266 S267 L268 E269 HIS HIS HIS HIS HIS HIS
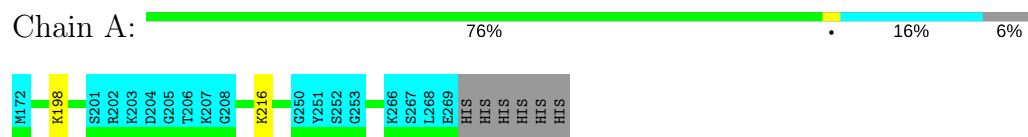
4.2.19 Score per residue for model 19

• Molecule 1: Replication protein A

Chain A:

75%   •   16%   6%

M172 S201 R202 K203 D204 G205 T206 K207 G208 E229 A241 G250 Y251 S252 G253 I263 K266 S267 L268 E269 HIS HIS HIS HIS HIS HIS
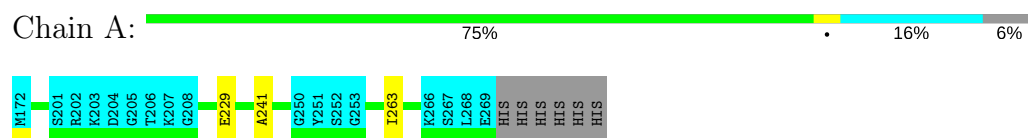
4.2.20 Score per residue for model 20

• Molecule 1: Replication protein A

Chain A:

72%   6%   16%   6%

M172 N173 T186 S201 R202 K203 D204 G205 T206 K207 G208 K211 R223 L226 E229 G250 Y251 S252 G253 K266 S267 L268 E269 HIS HIS HIS HIS HIS HIS
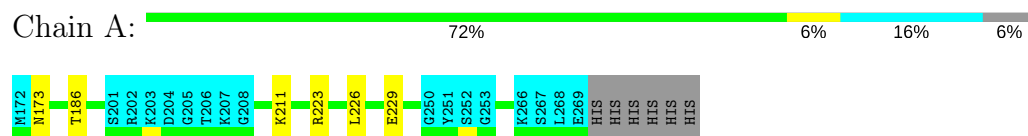
5 Refinement protocol and experimental data overview ⓘ

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
CNS	refinement	1.2
TOPSPIN	structure solution	2.1
AUTOASSIGN	structure solution	2.4.0
NMRPIPE	structure solution	2.3
CYANA	structure solution	2.1
CNS	structure solution	1.2
SPARKY	structure solution	3.113
AUTOSTRUCTURE	structure solution	2.2.1
PSVS	structure solution	1.3
PDBSTAT	structure solution	5.0

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 15849
Number of chemical shift lists	1
Total number of shifts	1263
Number of shifts mapped to atoms	1263
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	97%

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	634	642	642	2±1
All	All	12680	12840	12840	30

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	Models Total
1:A:223:ARG:HD2	1:A:223:ARG:H	0.54	1.63	15	1
1:A:179:LEU:HD11	1:A:187:ILE:HD13	0.52	1.82	14	2
1:A:241:ALA:HB2	1:A:263:ILE:HD13	0.51	1.82	19	2
1:A:198:LYS:HE2	1:A:210:LEU:HG	0.51	1.81	12	2
1:A:197:LYS:HB3	1:A:231:ALA:HB1	0.50	1.83	12	1
1:A:223:ARG:H	1:A:223:ARG:HD2	0.50	1.65	9	1
1:A:173:ASN:HA	1:A:186:THR:O	0.48	2.08	20	1
1:A:197:LYS:HD3	1:A:232:ASP:HB2	0.47	1.86	7	1
1:A:223:ARG:HH11	1:A:223:ARG:N	0.46	2.09	15	1
1:A:211:LYS:NZ	1:A:230:LEU:O	0.46	2.48	13	2
1:A:195:TYR:O	1:A:211:LYS:HG3	0.45	2.10	15	1
1:A:175:LYS:HD2	1:A:175:LYS:N	0.45	2.26	2	2
1:A:248:LYS:HB2	1:A:255:GLU:HG3	0.45	1.86	6	1
1:A:181:PRO:HA	1:A:248:LYS:HA	0.45	1.88	17	2

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:223:ARG:HD3	1:A:223:ARG:H	0.44	1.72	3	1
1:A:198:LYS:NZ	1:A:212:SER:OG	0.44	2.51	11	1
1:A:211:LYS:HB2	1:A:231:ALA:HA	0.44	1.90	3	1
1:A:215:LEU:HD11	1:A:243:VAL:HG21	0.44	1.90	4	1
1:A:223:ARG:HD2	1:A:223:ARG:N	0.44	2.28	1	1
1:A:211:LYS:HD2	1:A:231:ALA:HA	0.43	1.90	12	1
1:A:211:LYS:HB3	1:A:226:LEU:HB2	0.42	1.92	20	1
1:A:198:LYS:HD2	1:A:198:LYS:N	0.41	2.30	18	2
1:A:213:LEU:HD21	1:A:235:VAL:HG11	0.40	1.94	8	1

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	81/104 (78%)	78±1 (96±1%)	3±1 (4±1%)	0±0 (0±0%)	58	87
All	All	1620/2080 (78%)	1557 (96%)	61 (4%)	2 (0%)	58	87

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	233	PHE	1
1	A	196	PRO	1

6.3.2 Protein sidechains ⓘ

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	70/89 (79%)	69±1 (98±1%)	2±1 (2±1%)	62	94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1400/1780 (79%)	1370 (98%)	30 (2%)	62 94

All 13 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	255	GLU	4
1	A	216	LYS	4
1	A	223	ARG	4
1	A	175	LYS	3
1	A	229	GLU	3
1	A	239	ASP	3
1	A	197	LYS	3
1	A	248	LYS	1
1	A	236	LYS	1
1	A	237	LYS	1
1	A	217	ASP	1
1	A	211	LYS	1
1	A	173	ASN	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 97% for the well-defined parts and 96% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 15849

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	1263
Number of shifts mapped to atoms	1263
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
$^{13}C_\alpha$	98	-0.08 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_\beta$	88	0.02 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	98	0.38 ± 0.17	None needed (< 0.5 ppm)
^{15}N	95	-0.36 ± 0.44	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 97%, i.e. 950 atoms were assigned a chemical shift out of a possible 984. 14 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	401/401 (100%)	160/160 (100%)	162/162 (100%)	79/79 (100%)
Sidechain	486/520 (93%)	294/301 (98%)	184/201 (92%)	8/18 (44%)

Continued on next page...

Continued from previous page...

	Total	**¹H**	**¹³C**	**¹⁵N**
Aromatic	63/63 (100%)	33/33 (100%)	29/29 (100%)	1/1 (100%)
Overall	950/984 (97%)	487/494 (99%)	375/392 (96%)	88/98 (90%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 96%, i.e. 1128 atoms were assigned a chemical shift out of a possible 1178. 15 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**¹H**	**¹³C**	**¹⁵N**
Backbone	484/486 (100%)	193/194 (99%)	196/196 (100%)	95/96 (99%)
Sidechain	573/621 (92%)	350/362 (97%)	214/235 (91%)	9/24 (38%)
Aromatic	71/71 (100%)	37/37 (100%)	33/33 (100%)	1/1 (100%)
Overall	1128/1178 (96%)	580/593 (98%)	443/464 (95%)	105/121 (87%)

7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots ⓘ

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: