wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 11:32 PM BST

PDB ID	:	2KH9
Title	:	Solution structure of yeast Prp24-RRM2 bound to a fragment of U6 RNA
Authors	:	Martin-Tumasz, S.A.; Butcher, S.E.
Deposited on	:	2009-03-27

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:

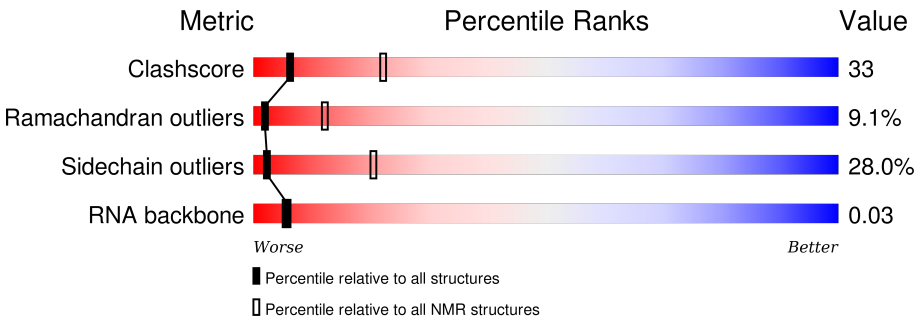
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 82%.

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
RNA backbone	3027	600

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	92	28% 32% 15% 18% 7%
2	B	6	33% 67%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 3 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:117-A:150, A:159-A:193 (69)	0.25	3

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, 3, 4, 5, 7, 9, 10
2	6, 8
Single-model clusters	2

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1612 atoms, of which 782 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called U4/U6 snRNA-associated-splicing factor PRP24.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	86	1416	445	715	120	132	4	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	114	MET	-	EXPRESSION TAG	UNP P49960
A	198	LEU	-	EXPRESSION TAG	UNP P49960
A	199	GLU	-	EXPRESSION TAG	UNP P49960
A	200	HIS	-	EXPRESSION TAG	UNP P49960
A	201	HIS	-	EXPRESSION TAG	UNP P49960
A	202	HIS	-	EXPRESSION TAG	UNP P49960
A	203	HIS	-	EXPRESSION TAG	UNP P49960
A	204	HIS	-	EXPRESSION TAG	UNP P49960
A	205	HIS	-	EXPRESSION TAG	UNP P49960

- Molecule 2 is a RNA chain called 5'-R(*AP*GP*AP*GP*AP*U)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	6	196	59	67	27	38	5	0

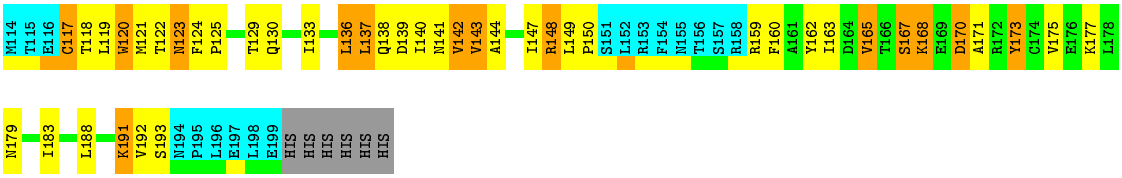
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA 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: U4/U6 snRNA-associated-splicing factor PRP24

Chain A: 28% 32% 15% 18% 7%



- Molecule 2: 5'-R(*AP*GP*AP*GP*AP*U)-3'

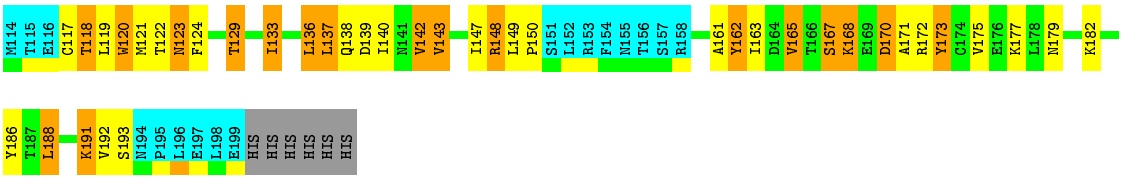
Chain B: 33% 67%

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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: U4/U6 snRNA-associated-splicing factor PRP24

Chain A: 32% 24% 20% 18% 7%



- Molecule 2: 5'-R(*AP*GP*AP*GP*AP*U)-3'

Chain B: 17% 33% 50%

A49
G50
A51
G52
A53
U54

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, torsion angle dynamics.*

Of the 100 calculated structures, 10 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
AMBER	refinement	
AtnosCandid	structure solution	
CNS	structure solution	
HADDOCK	refinement	2
HADDOCK	structure solution	2

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 16243, BMRB entry 16230, BMRB entry 162
Number of chemical shift lists	
Total number of shifts	
Number of shifts mapped to atoms	
Number of unparsed shifts	
Number of shifts with mapping errors	
Number of shifts with mapping warnings	
Assignment completeness (well-defined parts)	

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)

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	0.57±0.03	0±0/571 (0.0±0.0%)	0.79±0.04	1±1/777 (0.1±0.1%)
2	B	0.54±0.08	0±0/145 (0.0±0.0%)	1.16±0.11	1±1/225 (0.5±0.6%)
All	All	0.57	0/7160 (0.0%)	0.89	17/10020 (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
2	B	0.0±0.0	0.4±0.7
All	All	0	4

There are no bond-length outliers.

5 of 11 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
2	B	52	G	C3'-C2'-C1'	-9.59	93.83	101.50	3	1
2	B	51	A	C3'-C2'-C1'	-7.02	95.88	101.50	9	3
1	A	162	TYR	CB-CG-CD2	-6.77	116.94	121.00	8	4
2	B	51	A	O4'-C1'-N9	6.21	113.17	108.20	7	2
2	B	52	G	C5'-C4'-C3'	-5.75	106.80	116.00	8	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)
2	B	53	A	Sidechain	3
2	B	54	U	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	560	575	575	36±5
2	B	129	67	67	12±5
All	All	6890	6420	6420	442

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 33.

5 of 206 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:52:G:H3'	2:B:53:A:O4'	0.97	1.59	3	1
2:B:53:A:O3'	2:B:54:U:H2'	0.96	1.59	3	1
2:B:52:G:H5'	2:B:53:A:N7	0.91	1.79	8	1
1:A:138:GLN:HB2	1:A:143:VAL:HA	0.85	1.46	6	10
1:A:137:LEU:HD13	1:A:163:ILE:HG21	0.84	1.48	7	5

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	69/92 (75%)	57±1 (82±2%)	6±1 (9±2%)	6±1 (9±2%)	2 12
All	All	690/920 (75%)	565 (82%)	62 (9%)	63 (9%)	2 12

5 of 10 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	117	CYS	10
1	A	143	VAL	10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	123	ASN	10
1	A	168	LYS	10
1	A	142	VAL	9

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	64/87 (74%)	46±2 (72±3%)	18±2 (28±3%)	2 20
All	All	640/870 (74%)	461 (72%)	179 (28%)	2 20

5 of 37 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	136	LEU	10
1	A	191	LYS	10
1	A	120	TRP	10
1	A	147	ILE	10
1	A	173	TYR	9

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	5/6 (83%)	5±1 (90±10%)	2±1 (34±13%)	0.03±0.05
All	All	51/60 (85%)	45 (88%)	17 (33%)	0.03

The overall RNA backbone suiteness is 0.03.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	53	A	10
2	B	50	G	10
2	B	51	A	10
2	B	52	G	10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	54	U	5

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	50	G	10
2	B	52	G	4
2	B	53	A	2
2	B	49	A	1

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 82% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 16230

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

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	1100
Number of shifts mapped to atoms	1100
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	84	-0.81 ± 0.08	Should be applied
$^{13}C_{\beta}$	83	-0.01 ± 0.18	None needed (< 0.5 ppm)
$^{13}C'$	79	-0.29 ± 0.13	None needed (< 0.5 ppm)
^{15}N	83	0.66 ± 0.23	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 80%, i.e. 802 atoms were assigned a chemical shift out of a possible 999. 1 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	329/339 (97%)	133/135 (99%)	130/138 (94%)	66/66 (100%)
Sidechain	382/484 (79%)	234/282 (83%)	142/177 (80%)	6/25 (24%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	27/62 (44%)	27/32 (84%)	0/29 (0%)	0/1 (0%)
Overall	802/999 (80%)	436/515 (85%)	294/384 (77%)	72/100 (72%)

7.1.4 Statistically unusual chemical shifts (i)

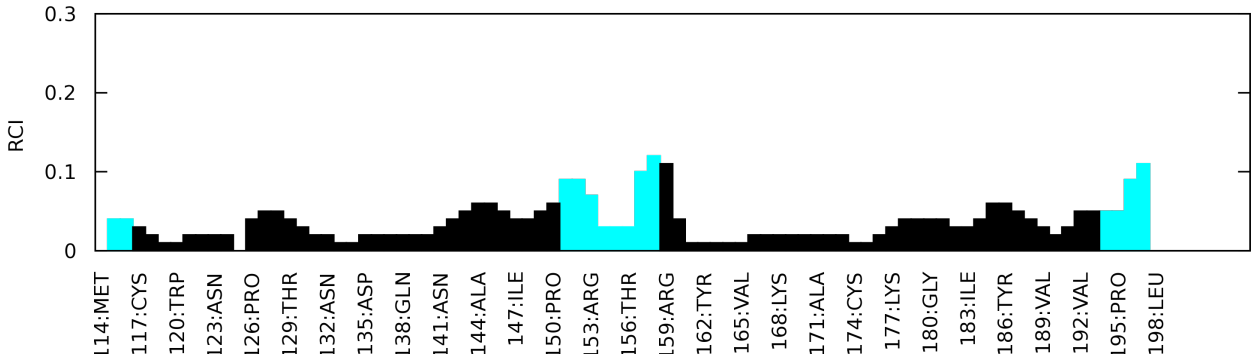
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	198	LEU	HD21	23.45	2.14 – -0.66	81.1
1	A	198	LEU	HD23	23.45	2.14 – -0.66	81.1
1	A	198	LEU	HD22	23.45	2.14 – -0.66	81.1

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: BMRB entry 16243

Chemical shift list name: *assigned_chem_shift_list_1*

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	316
Number of shifts mapped to atoms	148
Number of unparsed shifts	2
Number of shifts with mapping errors	166
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- Chemical shift has been reported more than once. All 2 occurences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
315	A	177	LYS	H	8.314	0.02	1
316	A	177	LYS	N	123.82	0.4	1

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Residue not found in structure. First 5 (of 166) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	92	ALA	N	124.707	0.4	1
A	153	ALA	H	6.514	0.02	1
A	121	PHE	N	116.963	0.4	1
A	102	ARG	H	9.215	0.02	1
A	133	GLN	N	118.252	0.4	1

7.2.2 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	—	—
$^{13}C_{\beta}$	0	—	—
$^{13}C'$	0	—	—
^{15}N	157	0.52 ± 0.36	None needed (imprecise)

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 12%, i.e. 122 atoms were assigned a chemical shift out of a possible 999. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	122/339 (36%)	61/135 (45%)	0/138 (0%)	61/66 (92%)
Sidechain	0/484 (0%)	0/282 (0%)	0/177 (0%)	0/25 (0%)
Aromatic	0/62 (0%)	0/32 (0%)	0/29 (0%)	0/1 (0%)
Overall	122/999 (12%)	61/515 (12%)	0/384 (0%)	61/100 (61%)

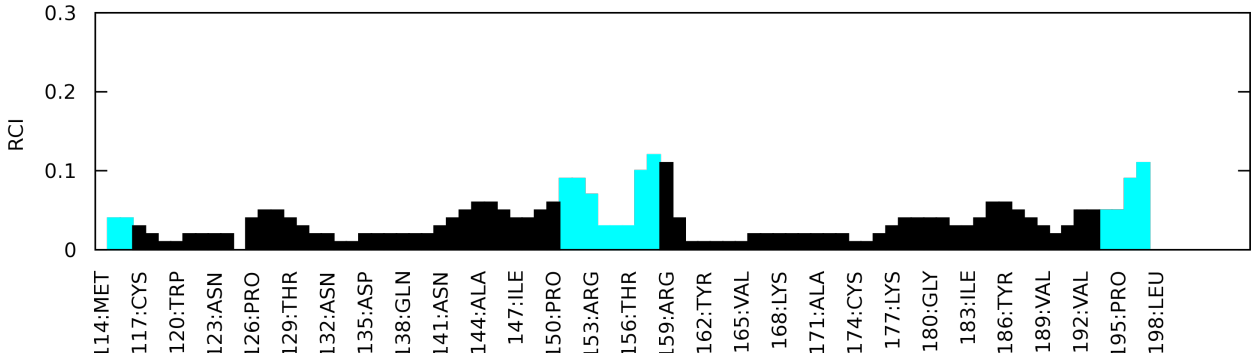
7.2.4 Statistically unusual chemical shifts (i)

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: BMRB entry 16244

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

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	262
Number of shifts mapped to atoms	0
Number of unparsed shifts	0
Number of shifts with mapping errors	262
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Chain not found in structure. First 5 (of 262) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	148	GLU	H	8.349	0.02	1
UNMAPPED	36	LEU	H	8.674	0.02	1
UNMAPPED	65	LEU	H	8.389	0.02	1
UNMAPPED	66	ASN	N	116.316	0.4	1
UNMAPPED	122	GLY	H	7.559	0.02	1

7.3.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	—
$^{13}C_{\beta}$	0	—	—
$^{13}C'$	0	—	—
^{15}N	131	0.00 ± 0.00	None needed (< 0.5 ppm)

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 0%, i.e. 0 atoms were assigned a chemical shift out of a possible 999. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	0/339 (0%)	0/135 (0%)	0/138 (0%)	0/66 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Sidechain	0/484 (0%)	0/282 (0%)	0/177 (0%)	0/25 (0%)
Aromatic	0/62 (0%)	0/32 (0%)	0/29 (0%)	0/1 (0%)
Overall	0/999 (0%)	0/515 (0%)	0/384 (0%)	0/100 (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)

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned_chem_shift_list_1). RCI is only applicable to proteins.

7.4 Chemical shift list 4

File name: BMRB entry 16246

Chemical shift list name: *assigned_chem_shift_list_1*

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	861
Number of shifts mapped to atoms	861
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.4.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	82	-0.55 ± 0.24	Should be applied
$^{13}C_{\beta}$	80	0.15 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	77	0.45 ± 0.55	None needed (< 0.5 ppm)

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 64%, i.e. 637 atoms were assigned a chemical shift out of a possible 999. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

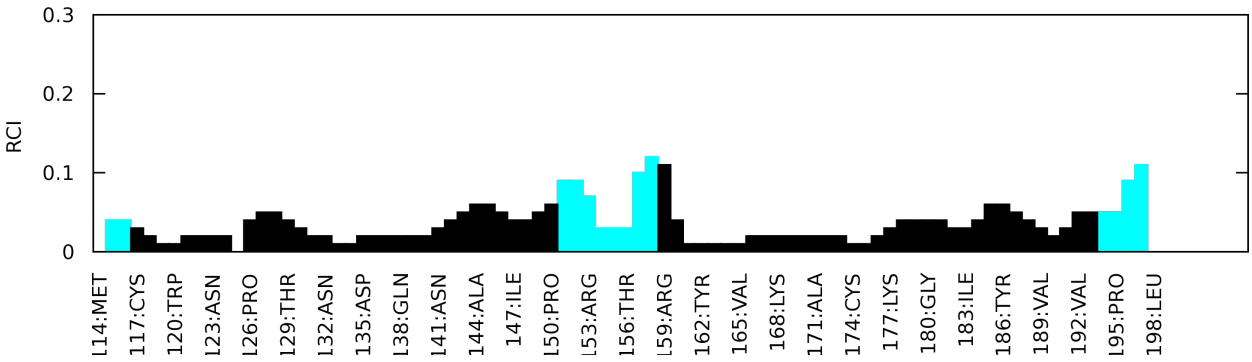
	Total	^{1}H	^{13}C	^{15}N
Backbone	264/339 (78%)	130/135 (96%)	68/138 (49%)	66/66 (100%)
Sidechain	372/484 (77%)	225/282 (80%)	141/177 (80%)	6/25 (24%)
Aromatic	1/62 (2%)	1/32 (3%)	0/29 (0%)	0/1 (0%)
Overall	637/999 (64%)	356/515 (69%)	209/384 (54%)	72/100 (72%)

7.4.4 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	153	ARG	CD	36.16	47.57 – 38.77	-8.0

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: