Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 10:49 PM BST

PDB ID	:	2KFO
Title	:	Mouse Prion Protein (121-231) with Mutation V166A
Authors	:	Christen, B.; Hornemann, S.; Damberger, F.F.; Wuthrich, K.
Deposited on	:	2009-02-24

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
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:

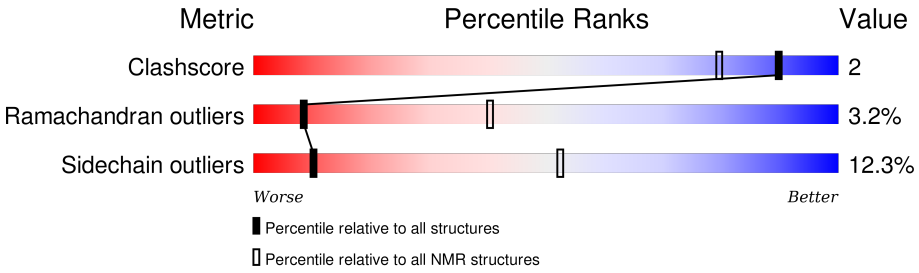
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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

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

The overall completeness of chemical shifts assignment is 64%.

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	114	82% 10% • 7%

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: *fewest violations*.

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:120-A:225 (106)	0.29	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 2 clusters. No single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1803 atoms, of which 868 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Major prion protein.

Mol	Chain	Residues	Atoms						Trace
1	A	114	Total	C	H	N	O	S	0
			1803	578	868	165	183	9	

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	119	GLY	-	EXPRESSION TAG	UNP P04925
A	120	SER	-	EXPRESSION TAG	UNP P04925
A	166	ALA	VAL	ENGINEERED	UNP P04925

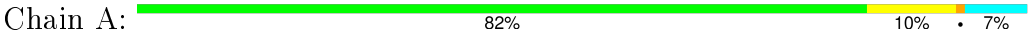
4 Residue-property plots (i)

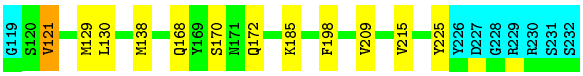
4.1 Average score per residue in the NMR ensemble

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: Major prion protein

Chain A:





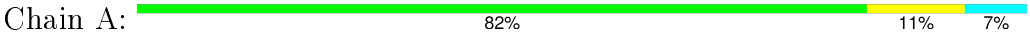
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Major prion protein

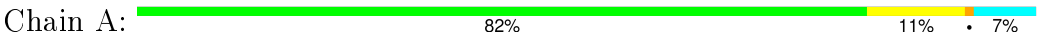
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Major prion protein

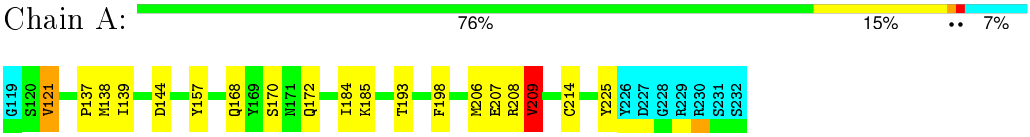
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Major prion protein

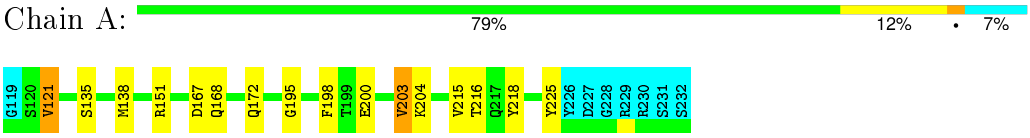
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Major prion protein

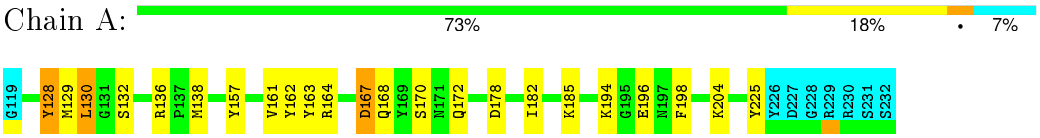
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Major prion protein

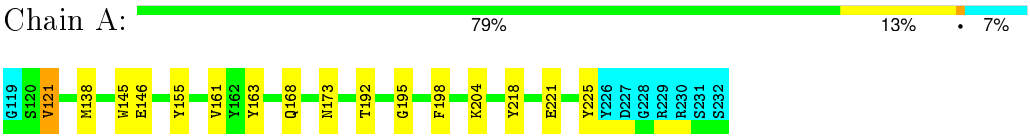
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Major prion protein

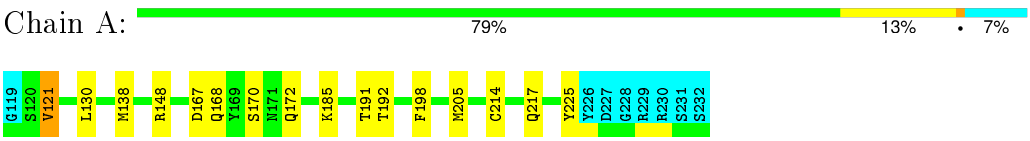
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Major prion protein

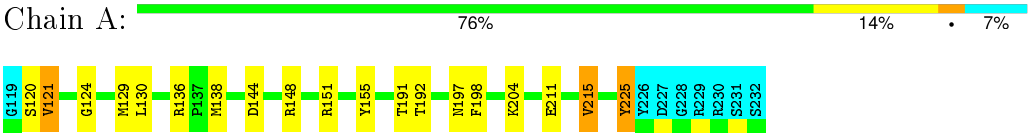
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Major prion protein

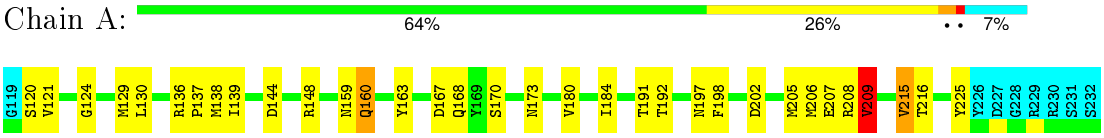
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Major prion protein

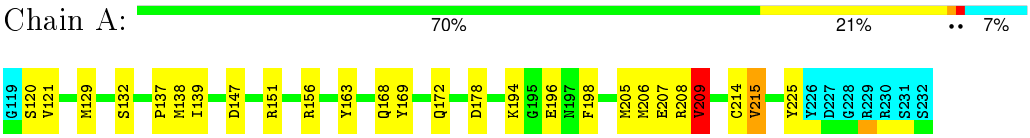
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Major prion protein

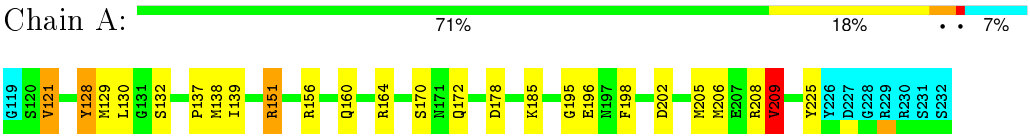
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Major prion protein

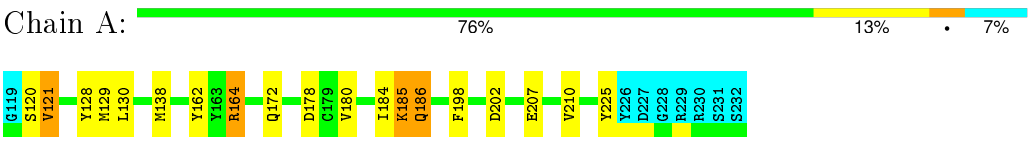
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Major prion protein

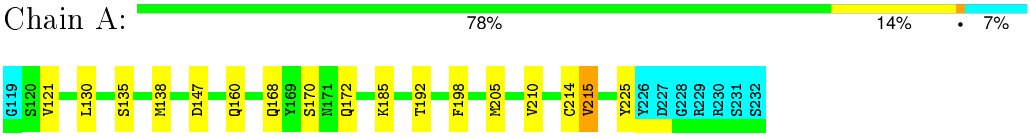
Chain A:



4.2.13 Score per residue for model 13 (medoid)

• Molecule 1: Major prion protein

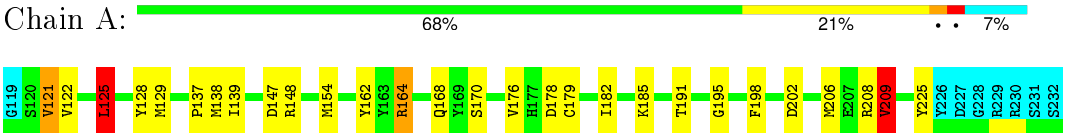
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Major prion protein

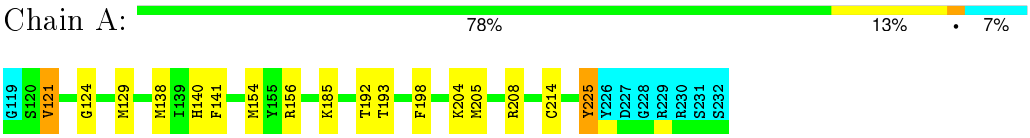
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Major prion protein

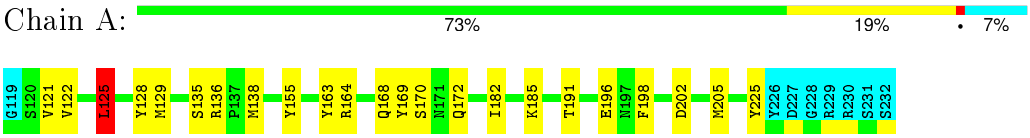
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Major prion protein

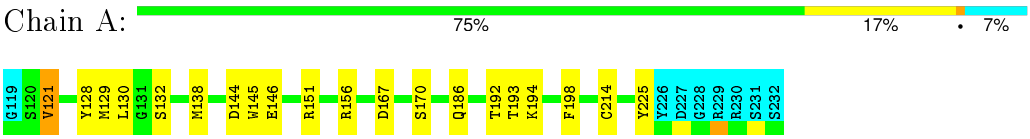
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Major prion protein

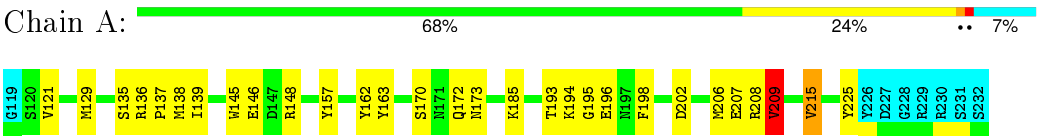
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Major prion protein

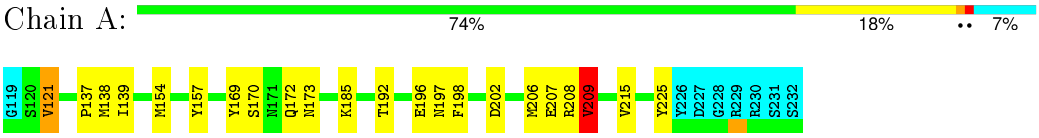
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Major prion protein

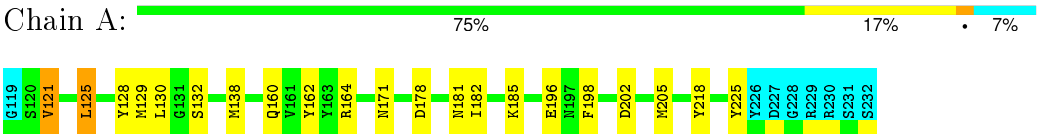
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Major prion protein

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *TORSION ANGLE DYNAMICS*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *TARGET FUNCTION*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
DYANA	structure solution	1.0.3
OPALP	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)	BMRB entry 16185
Number of chemical shift lists	1
Total number of shifts	924
Number of shifts mapped to atoms	924
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	64%

No validations of the models with respect to experimental NMR restraints is performed at this time.

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	0.67±0.01	0±0/894 (0.0±0.0%)	1.09±0.02	1±1/1210 (0.1±0.1%)
All	All	0.67	0/17880 (0.0%)	1.09	28/24200 (0.1%)

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	1.6±1.0
All	All	0	32

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	209	VAL	CA-CB-CG1	9.62	125.33	110.90	3	6
1	A	203	VAL	CG1-CB-CG2	8.08	123.83	110.90	4	1
1	A	125	LEU	CB-CG-CD1	-7.13	98.89	111.00	20	3
1	A	161	VAL	CA-CB-CG1	6.98	121.38	110.90	5	1
1	A	203	VAL	CA-CB-CG2	6.65	120.88	110.90	4	1
1	A	215	VAL	CA-CB-CG1	6.58	120.78	110.90	2	6
1	A	209	VAL	CA-CB-CG2	6.09	120.03	110.90	18	3
1	A	156	ARG	NE-CZ-NH2	-6.07	117.26	120.30	2	2
1	A	148	ARG	NE-CZ-NH2	-6.01	117.29	120.30	18	1
1	A	218	TYR	CB-CG-CD2	-5.77	117.54	121.00	4	2
1	A	151	ARG	NE-CZ-NH2	-5.23	117.68	120.30	11	1
1	A	148	ARG	NE-CZ-NH1	5.05	122.83	120.30	18	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	151	ARG	Sidechain	5
1	A	148	ARG	Sidechain	4
1	A	157	TYR	Sidechain	4
1	A	128	TYR	Sidechain	3
1	A	169	TYR	Sidechain	3
1	A	156	ARG	Sidechain	3
1	A	155	TYR	Sidechain	3
1	A	163	TYR	Sidechain	2
1	A	164	ARG	Sidechain	2
1	A	136	ARG	Sidechain	1
1	A	162	TYR	Sidechain	1
1	A	218	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	873	813	813	4±3
All	All	17460	16260	16260	74

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:122:VAL:HB	1:A:125:LEU:HD11	0.64	1.69	16	1
1:A:139:ILE:HD13	1:A:208:ARG:HB2	0.58	1.75	9	7
1:A:137:PRO:HD2	1:A:209:VAL:HG23	0.58	1.74	9	7
1:A:206:MET:HA	1:A:209:VAL:CG1	0.56	2.30	18	7
1:A:125:LEU:HD12	1:A:125:LEU:N	0.56	2.14	16	2
1:A:130:LEU:HD11	1:A:160:GLN:NE2	0.56	2.16	20	2
1:A:180:VAL:O	1:A:184:ILE:HD12	0.54	2.02	9	1
1:A:137:PRO:CD	1:A:209:VAL:HG23	0.51	2.36	14	3
1:A:125:LEU:H	1:A:125:LEU:HD12	0.51	1.64	16	1
1:A:122:VAL:HG21	1:A:162:TYR:CE1	0.51	2.40	14	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:139:ILE:HD11	1:A:209:VAL:HB	0.50	1.82	10	5
1:A:129:MET:HE3	1:A:163:TYR:CE2	0.50	2.40	10	2
1:A:125:LEU:HD12	1:A:125:LEU:H	0.48	1.67	14	1
1:A:128:TYR:CD2	1:A:182:ILE:HG13	0.48	2.42	16	2
1:A:128:TYR:CE2	1:A:164:ARG:HG3	0.47	2.44	20	5
1:A:200:GLU:HA	1:A:203:VAL:CG2	0.46	2.41	4	1
1:A:121:VAL:HG13	1:A:129:MET:HA	0.46	1.88	14	1
1:A:191:THR:HG21	1:A:197:ASN:HA	0.45	1.88	9	2
1:A:176:VAL:HA	1:A:179:CYS:SG	0.45	2.52	14	1
1:A:145:TRP:CE2	1:A:146:GLU:HG3	0.45	2.47	6	3
1:A:130:LEU:HD11	1:A:160:GLN:HE21	0.44	1.73	9	1
1:A:185:LYS:HG3	1:A:186:GLN:N	0.44	2.28	12	1
1:A:129:MET:CE	1:A:163:TYR:CE2	0.43	3.02	5	2
1:A:184:ILE:HD11	1:A:207:GLU:HA	0.43	1.89	9	3
1:A:128:TYR:CE2	1:A:182:ILE:HG13	0.43	2.49	16	2
1:A:130:LEU:HD23	1:A:162:TYR:CE1	0.43	2.49	20	2
1:A:180:VAL:HA	1:A:210:VAL:HG23	0.43	1.89	12	1
1:A:128:TYR:CZ	1:A:182:ILE:HD11	0.42	2.49	14	1
1:A:225:TYR:C	1:A:225:TYR:CD1	0.42	2.92	15	1
1:A:128:TYR:CE1	1:A:164:ARG:HG3	0.42	2.49	11	1
1:A:225:TYR:CD1	1:A:225:TYR:C	0.42	2.93	8	1
1:A:139:ILE:CD1	1:A:209:VAL:HB	0.42	2.45	10	1
1:A:130:LEU:HG	1:A:162:TYR:CE1	0.40	2.52	5	1
1:A:200:GLU:HA	1:A:203:VAL:CB	0.40	2.46	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	106/114 (93%)	89±2 (84±2%)	13±2 (13±2%)	3±1 (3±1%)	8	40
All	All	2120/2280 (93%)	1786 (84%)	267 (13%)	67 (3%)	8	40

All 13 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	198	PHE	20
1	A	121	VAL	19
1	A	170	SER	7
1	A	195	GLY	5
1	A	194	LYS	3
1	A	132	SER	3
1	A	124	GLY	3
1	A	120	SER	2
1	A	141	PHE	1
1	A	191	THR	1
1	A	171	ASN	1
1	A	136	ARG	1
1	A	167	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	96/102 (94%)	84±2 (88±2%)	12±2 (12±2%)	10	52
All	All	1920/2040 (94%)	1684 (88%)	236 (12%)	10	52

All 46 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	225	TYR	20
1	A	138	MET	20
1	A	172	GLN	13
1	A	185	LYS	12
1	A	121	VAL	11
1	A	168	GLN	11
1	A	192	THR	9
1	A	202	ASP	9
1	A	205	MET	8
1	A	215	VAL	8
1	A	209	VAL	7
1	A	196	GLU	7
1	A	129	MET	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	170	SER	6
1	A	130	LEU	6
1	A	214	CYS	6
1	A	178	ASP	6
1	A	204	LYS	5
1	A	144	ASP	5
1	A	167	ASP	5
1	A	173	ASN	5
1	A	193	THR	4
1	A	135	SER	4
1	A	191	THR	3
1	A	136	ARG	3
1	A	125	LEU	3
1	A	154	MET	3
1	A	207	GLU	3
1	A	147	ASP	3
1	A	160	GLN	2
1	A	186	GLN	2
1	A	216	THR	2
1	A	120	SER	2
1	A	211	GLU	2
1	A	132	SER	2
1	A	194	LYS	2
1	A	181	ASN	1
1	A	140	HIS	1
1	A	197	ASN	1
1	A	164	ARG	1
1	A	217	GLN	1
1	A	221	GLU	1
1	A	161	VAL	1
1	A	210	VAL	1
1	A	208	ARG	1
1	A	159	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 64% for the well-defined parts and 63% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 16185

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	924
Number of shifts mapped to atoms	924
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
$^{13}C_{\alpha}$	114	-0.46 ± 0.28	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	105	0.34 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	110	0.00 ± 0.26	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 64%, i.e. 857 atoms were assigned a chemical shift out of a possible 1331. 12 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	307/524 (59%)	98/209 (47%)	106/212 (50%)	103/103 (100%)
Sidechain	498/667 (75%)	283/393 (72%)	193/236 (82%)	22/38 (58%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	52/140 (37%)	14/73 (19%)	37/63 (59%)	1/4 (25%)
Overall	857/1331 (64%)	395/675 (59%)	336/511 (66%)	126/145 (87%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 63%, i.e. 904 atoms were assigned a chemical shift out of a possible 1428. 12 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	328/564 (58%)	104/225 (46%)	114/228 (50%)	110/111 (99%)
Sidechain	522/716 (73%)	295/423 (70%)	203/249 (82%)	24/44 (55%)
Aromatic	54/148 (36%)	14/77 (18%)	39/67 (58%)	1/4 (25%)
Overall	904/1428 (63%)	413/725 (57%)	356/544 (65%)	135/159 (85%)

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: