Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 09:58 PM BST

PDB ID	:	2K1D
Title	:	NMR Studies of a Pathogenic Mutant (D178N) of the Human Prion Protein
Authors	:	Mills, J.L.; Surewicz, K.; Surewicz, W.K.; Sonnichsen, F.D.
Deposited on	:	2008-02-29

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:

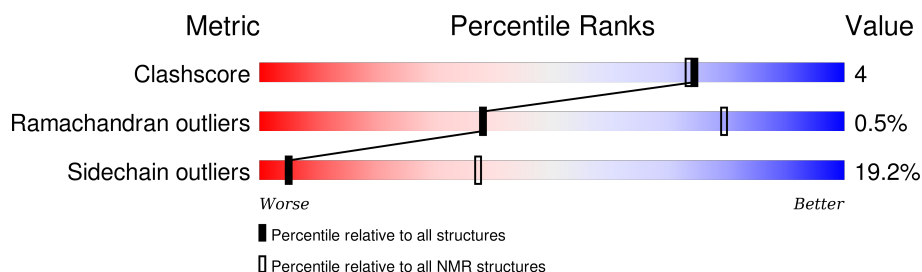|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457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

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

The overall completeness of chemical shifts assignment is 73%.

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		4
Ramachandran outliers		0.5%
Sidechain outliers		19.2%

■ Percentile relative to all structures

▯ Percentile relative to all NMR structures

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	146	40% 12% 22% 27%

2 Ensemble composition and analysis ⓘ

This entry contains 20 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: *closest to the average*.

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:133-A:162,　 A:174-A:218 (75)	0.57	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 3 clusters and 1 single-model cluster was found.

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

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 1721 atoms, of which 829 are hydrogens and 0 are deuteriums.
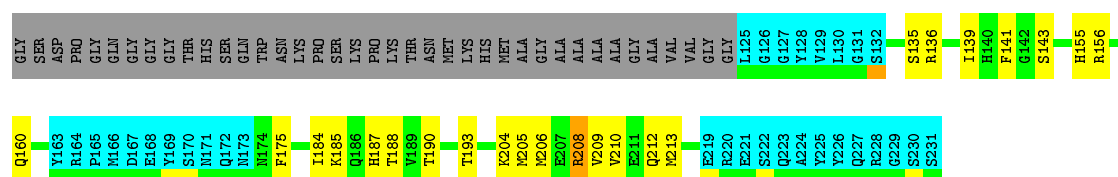
- Molecule 1 is a protein called Major prion protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	107	1721	552	829	157	175	8	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	86	GLY	-	EXPRESSION TAG	UNP P04156
A	87	SER	-	EXPRESSION TAG	UNP P04156
A	88	ASP	-	EXPRESSION TAG	UNP P04156
A	89	PRO	-	EXPRESSION TAG	UNP P04156
A	129	VAL	MET	VARIANT	UNP P04156
A	178	ASN	ASP	ENGINEERED	UNP P04156

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

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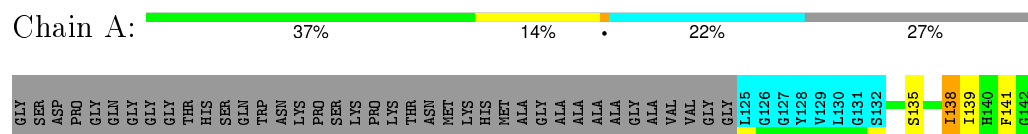
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Major prion protein

4.2.2 Score per residue for model 2

- Molecule 1: Major prion protein

4.2.3 Score per residue for model 3 (medoid)

• Molecule 1: Major prion protein
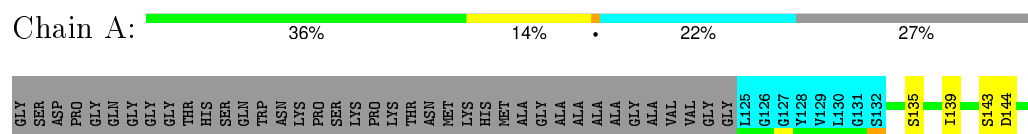
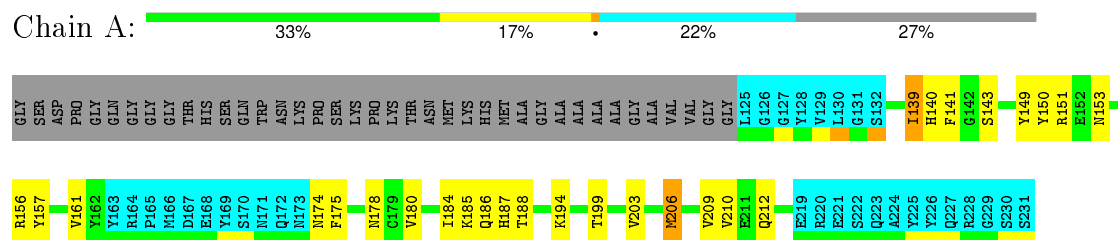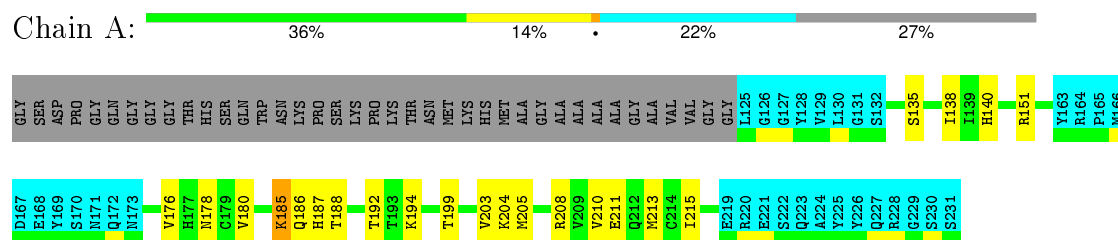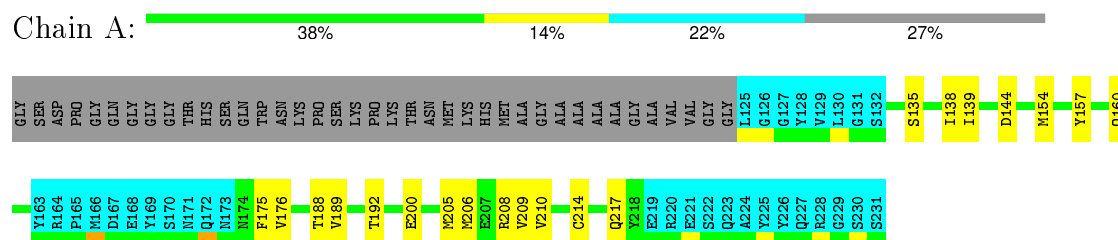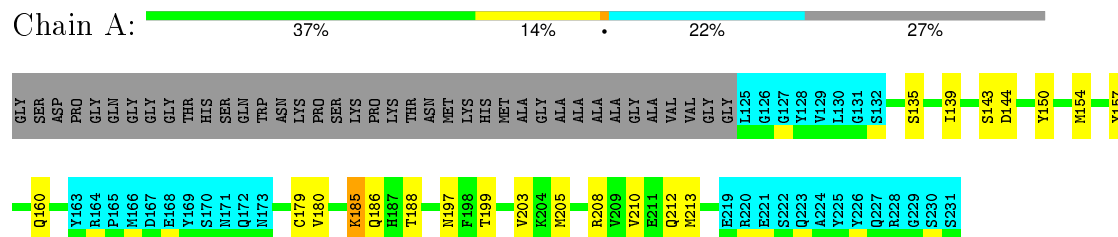### 4.2.4 Score per residue for model 4

• Molecule 1: Major prion protein
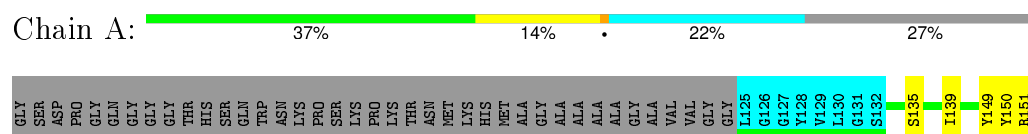
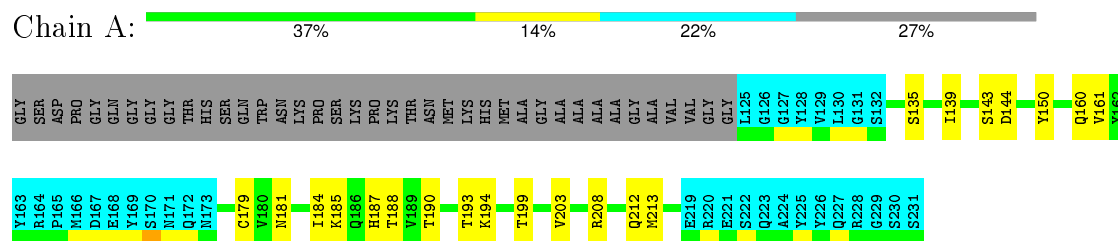### 4.2.5 Score per residue for model 5

• Molecule 1: Major prion protein

4.2.6 Score per residue for model 6

• Molecule 1: Major prion protein

Chain A: 33% 17% • 22% 27%

4.2.7 Score per residue for model 7

• Molecule 1: Major prion protein

Chain A: 36% 14% • 22% 27%

4.2.8 Score per residue for model 8

• Molecule 1: Major prion protein

Chain A: 38% 14% 22% 27%

4.2.9 Score per residue for model 9

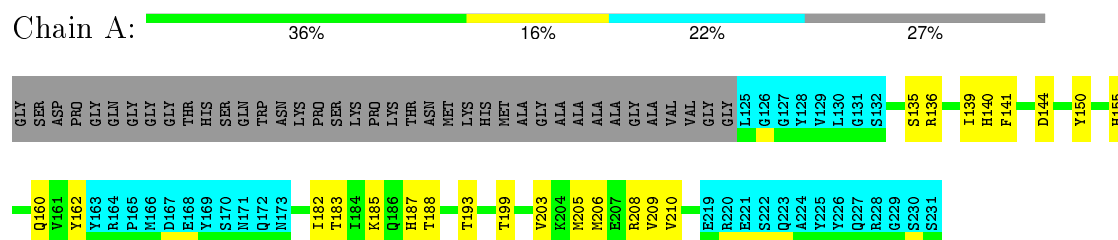• Molecule 1: Major prion protein

Chain A: 37% 14% • 22% 27%

4.2.10 Score per residue for model 10

- Molecule 1: Major prion protein
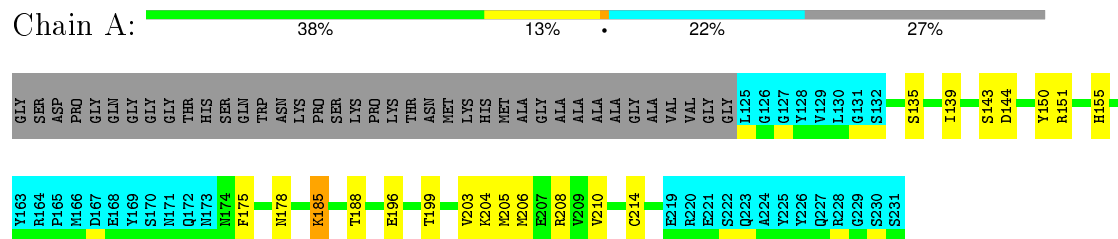

4.2.11 Score per residue for model 11

- Molecule 1: Major prion protein

4.2.12 Score per residue for model 12

- Molecule 1: Major prion protein

4.2.13 Score per residue for model 13

- Molecule 1: Major prion protein

4.2.14 Score per residue for model 14

- Molecule 1: Major prion protein

4.2.15 Score per residue for model 15

- Molecule 1: Major prion protein

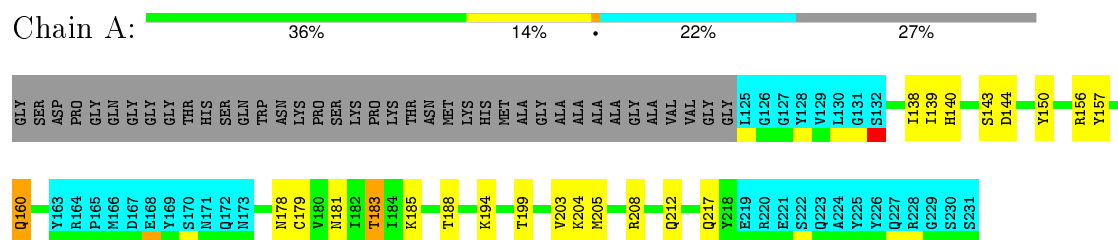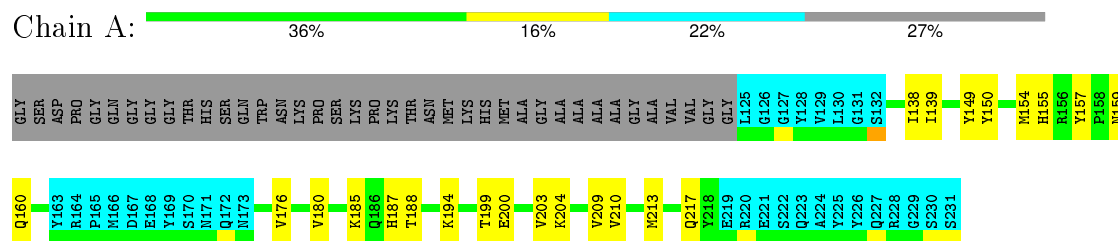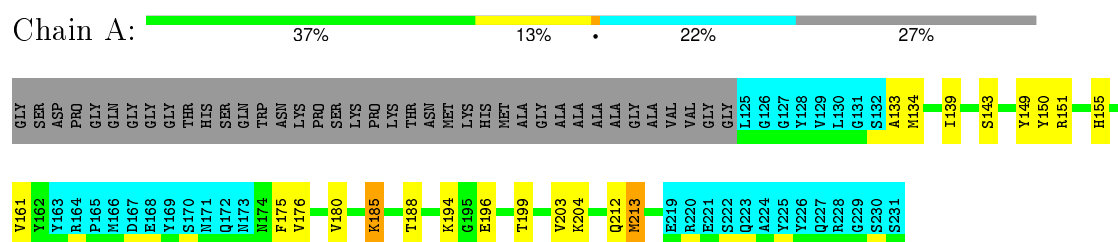## 4.2.16 Score per residue for model 16

- Molecule 1: Major prion protein

4.2.17 Score per residue for model 17

- Molecule 1: Major prion protein

4.2.18 Score per residue for model 18

• Molecule 1: Major prion protein

4.2.19 Score per residue for model 19

• Molecule 1: Major prion protein

4.2.20 Score per residue for model 20
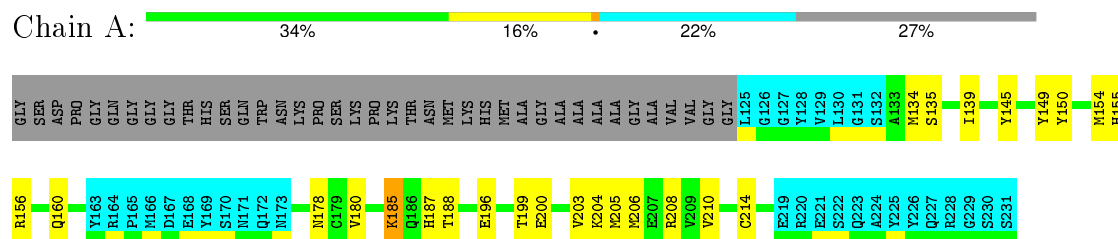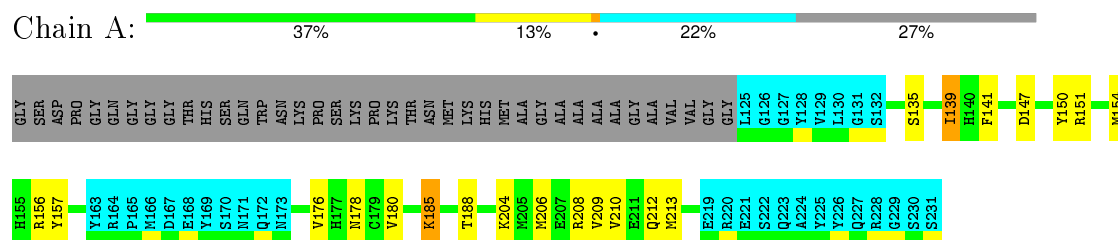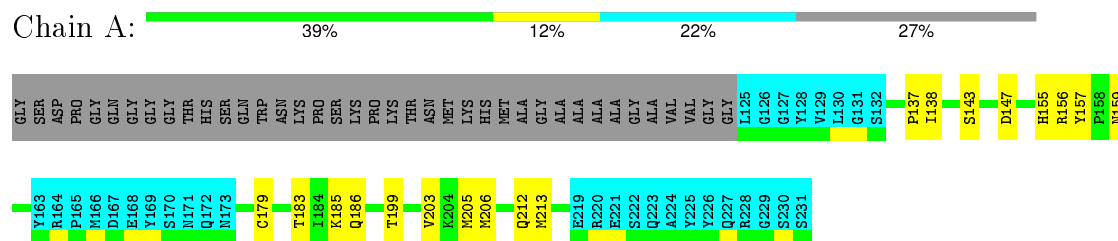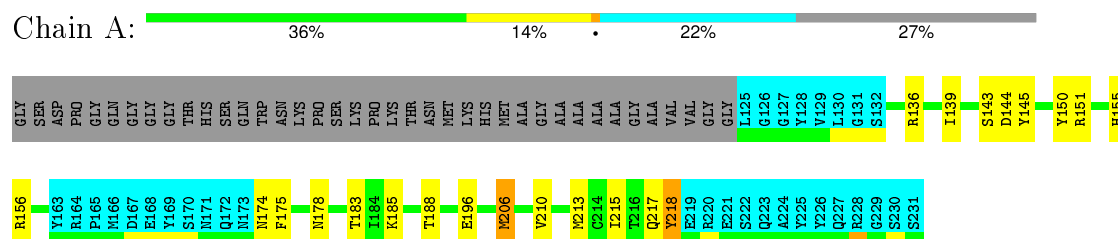
• Molecule 1: Major prion protein
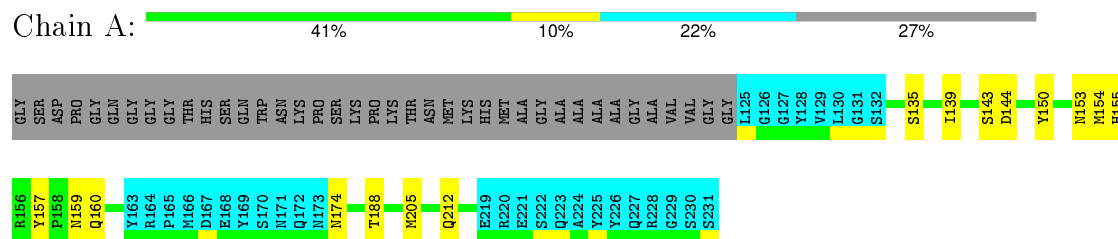
5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *STRUC-TURES WITH THE LEAST RESTRAINT VIOLATIONS*.

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

Software name	Classification	Version
CYANA	structure solution	2.1
CYANA	refinement	2.1

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 15676
Number of chemical shift lists	1
Total number of shifts	1053
Number of shifts mapped to atoms	1053
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	73%

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	632	598	600	5±2
All	All	12640	11960	12000	98

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 4.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:139:ILE:HD12	1:A:209:VAL:HA	0.64	1.68	4	4
1:A:176:VAL:O	1:A:180:VAL:HG23	0.63	1.94	17	4
1:A:139:ILE:HG22	1:A:150:TYR:OH	0.62	1.95	1	15
1:A:206:MET:O	1:A:210:VAL:HG23	0.62	1.94	6	10
1:A:139:ILE:HD12	1:A:209:VAL:CA	0.61	2.26	2	2
1:A:139:ILE:HG22	1:A:150:TYR:CZ	0.60	2.32	17	8
1:A:138:ILE:HD11	1:A:141:PHE:CZ	0.56	2.35	5	1
1:A:185:LYS:O	1:A:188:THR:HG22	0.56	2.01	19	12
1:A:180:VAL:HG13	1:A:210:VAL:HB	0.55	1.78	7	5
1:A:180:VAL:HG13	1:A:210:VAL:CG1	0.55	2.32	14	4
1:A:184:ILE:HD11	1:A:210:VAL:HG21	0.55	1.78	2	1
1:A:139:ILE:HD13	1:A:209:VAL:HA	0.55	1.78	6	3
1:A:176:VAL:HG13	1:A:214:CYS:SG	0.53	2.43	3	1
1:A:199:THR:O	1:A:203:VAL:HG12	0.50	2.07	12	14

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:161:VAL:HG13	1:A:213:MET:HG2	0.49	1.82	10	1
1:A:184:ILE:O	1:A:188:THR:HG23	0.48	2.09	10	2
1:A:211:GLU:O	1:A:215:ILE:HD12	0.48	2.08	7	1
1:A:137:PRO:HB2	1:A:138:ILE:HD12	0.47	1.86	18	1
1:A:179:CYS:HA	1:A:182:ILE:HD12	0.46	1.87	5	1
1:A:190:THR:O	1:A:193:THR:HG22	0.46	2.11	2	2
1:A:139:ILE:HG22	1:A:150:TYR:CE1	0.45	2.46	6	1
1:A:162:TYR:CD2	1:A:182:ILE:HG22	0.43	2.48	11	1
1:A:189:VAL:O	1:A:192:THR:HG22	0.42	2.13	8	1
1:A:161:VAL:HG11	1:A:213:MET:HB3	0.42	1.91	15	1
1:A:180:VAL:HG13	1:A:210:VAL:CB	0.42	2.44	6	1
1:A:139:ILE:HD11	1:A:208:ARG:HD2	0.42	1.90	2	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	75/146 (51%)	70±1 (93±1%)	5±1 (7±1%)	0±0 (0±1%)	38	79
All	All	1500/2920 (51%)	1395 (93%)	98 (7%)	7 (0%)	38	79

All 4 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	138	ILE	4
1	A	133	ALA	1
1	A	218	TYR	1
1	A	143	SER	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	72/123 (59%)	58±2 (81±3%)	14±2 (19±3%)	5	37
All	All	1440/2460 (59%)	1164 (81%)	276 (19%)	5	37

All 45 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	185	LYS	15
1	A	135	SER	14
1	A	212	GLN	13
1	A	143	SER	12
1	A	208	ARG	12
1	A	205	MET	11
1	A	155	HIS	11
1	A	204	LYS	11
1	A	160	GLN	10
1	A	144	ASP	10
1	A	213	MET	9
1	A	175	PHE	8
1	A	151	ARG	8
1	A	156	ARG	8
1	A	154	MET	8
1	A	178	ASN	8
1	A	187	HIS	8
1	A	157	TYR	8
1	A	149	TYR	7
1	A	194	LYS	7
1	A	196	GLU	6
1	A	217	GLN	5
1	A	214	CYS	5
1	A	188	THR	4
1	A	141	PHE	4
1	A	174	ASN	4
1	A	140	HIS	4
1	A	186	GLN	4
1	A	179	CYS	4
1	A	183	THR	4
1	A	206	MET	3
1	A	153	ASN	3
1	A	136	ARG	3
1	A	159	ASN	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	200	GLU	3
1	A	147	ASP	3
1	A	181	ASN	3
1	A	193	THR	2
1	A	134	MET	2
1	A	139	ILE	2
1	A	192	THR	2
1	A	145	TYR	2
1	A	138	ILE	1
1	A	197	ASN	1
1	A	218	TYR	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 73% for the well-defined parts and 69% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 15676

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	1053
Number of shifts mapped to atoms	1053
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	-0.22 ± 0.17	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	87	0.14 ± 0.09	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	94	0.24 ± 0.20	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 73%, i.e. 712 atoms were assigned a chemical shift out of a possible 973. 6 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	285/371 (77%)	138/148 (93%)	75/150 (50%)	72/73 (99%)
Sidechain	387/499 (78%)	236/293 (81%)	136/178 (76%)	15/28 (54%)

Continued on next page...

Continued from previous page...

	Total	**¹H**	**¹³C**	**¹⁵N**
Aromatic	40/103 (39%)	26/55 (47%)	14/44 (32%)	0/4 (0%)
Overall	712/973 (73%)	400/496 (81%)	225/372 (60%)	87/105 (83%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 69%, i.e. 942 atoms were assigned a chemical shift out of a possible 1367. 9 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**¹H**	**¹³C**	**¹⁵N**
Backbone	373/529 (71%)	181/211 (86%)	98/214 (46%)	94/104 (90%)
Sidechain	499/695 (72%)	306/411 (74%)	170/242 (70%)	23/42 (55%)
Aromatic	70/143 (49%)	46/75 (61%)	24/64 (38%)	0/4 (0%)
Overall	942/1367 (69%)	533/697 (76%)	292/520 (56%)	117/150 (78%)

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: