Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 06:14 PM BST

PDB ID	:	1XYW
Title	:	elk prion protein
Authors	:	Gossert, A.D.; Bonjour, S.; Lysek, D.A.; Fiorito, F.; Wuthrich, K.
Deposited on	:	2004-11-11

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:

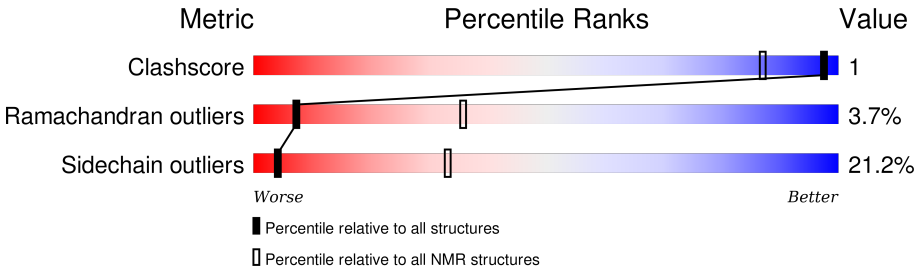
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 80%.

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	111	69% 23% • 7%

2 Ensemble composition and analysis

This entry contains 20 models. Model 4 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:124-A:226 (103)	0.41	4

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 4 clusters and 4 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1776 atoms, of which 858 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	111	1776	571	858	160	179	8	0

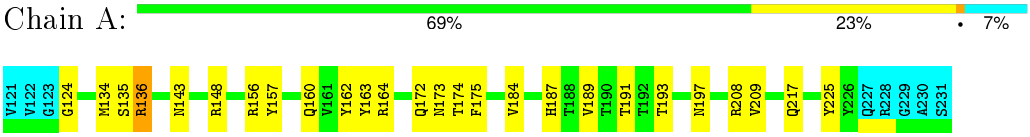
4 Residue-property plots (i)

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: Major prion protein

Chain A:



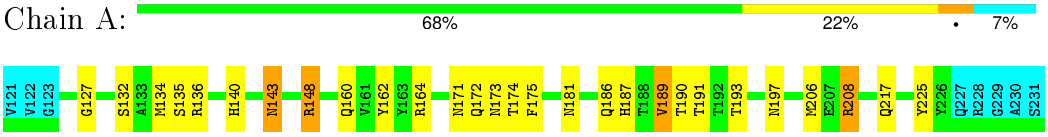
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Major prion protein

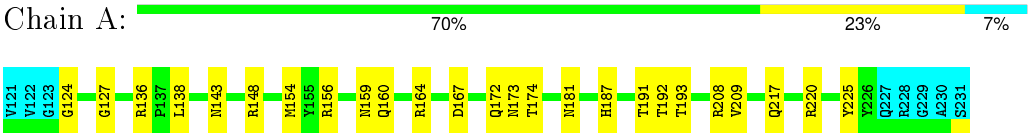
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Major prion protein

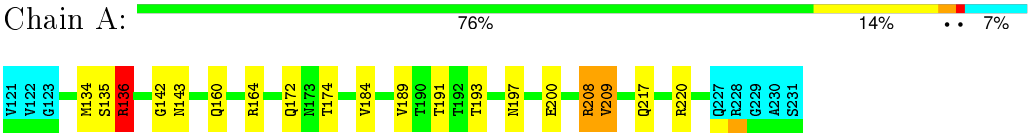
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Major prion protein

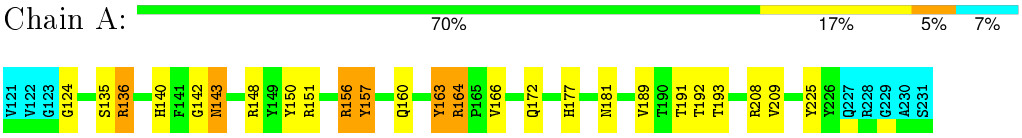
Chain A:



4.2.4 Score per residue for model 4 (medoid)

• Molecule 1: Major prion protein

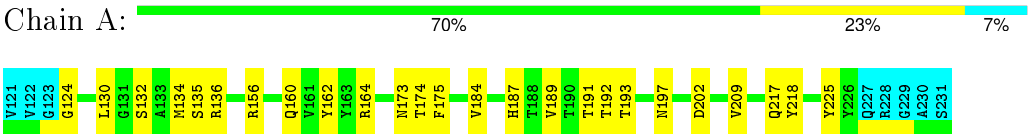
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Major prion protein

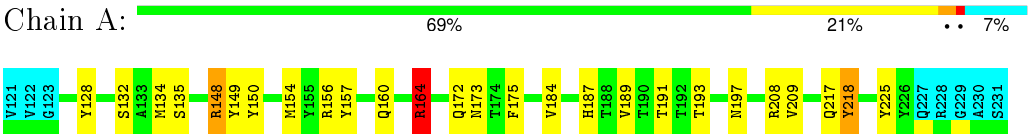
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Major prion protein

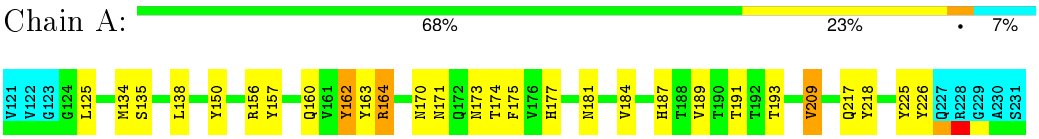
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Major prion protein

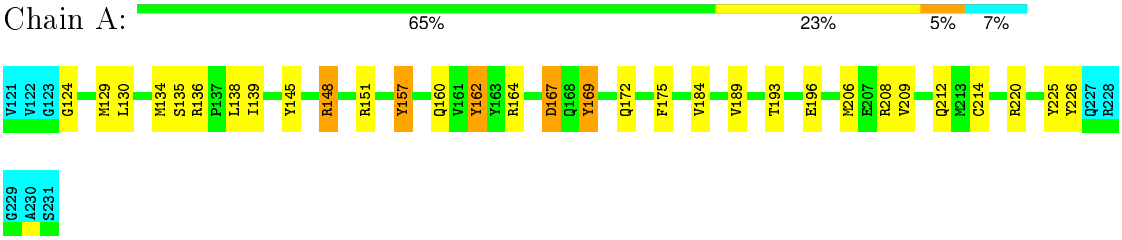
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Major prion protein

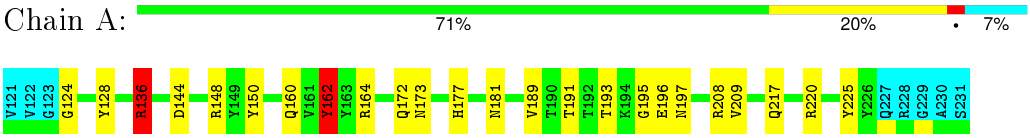
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Major prion protein

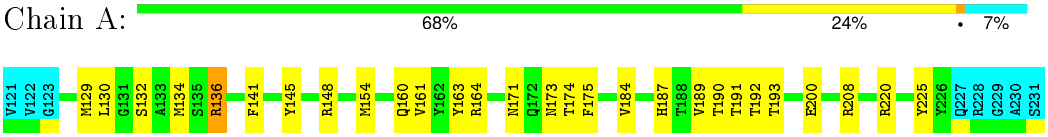
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: Major prion protein

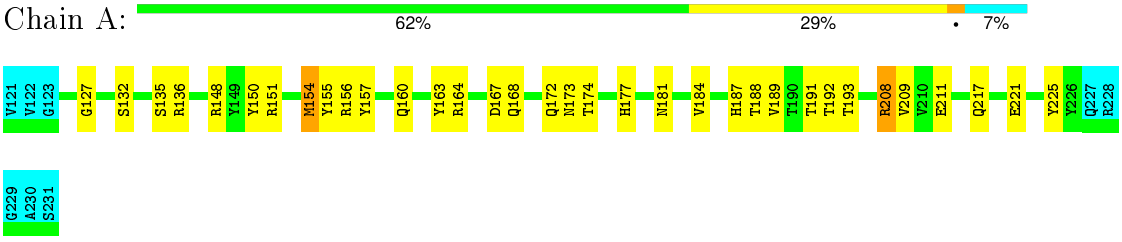
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: Major prion protein

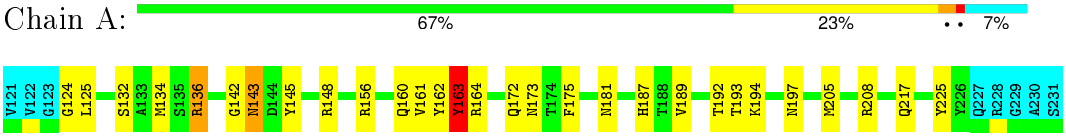
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Major prion protein

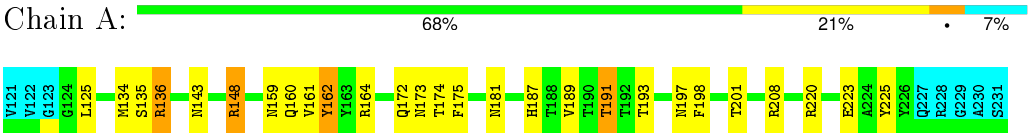
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Major prion protein

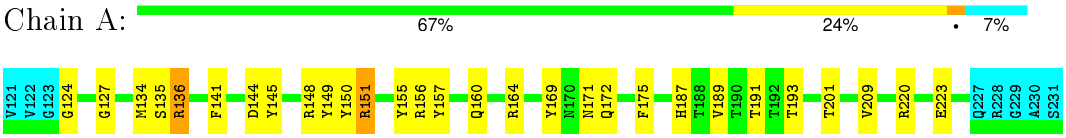
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Major prion protein

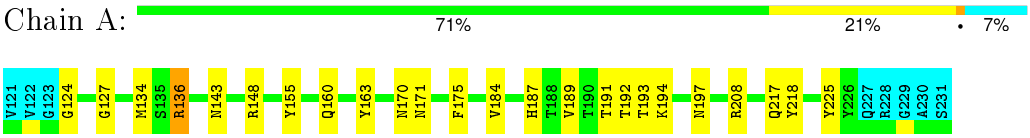
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Major prion protein

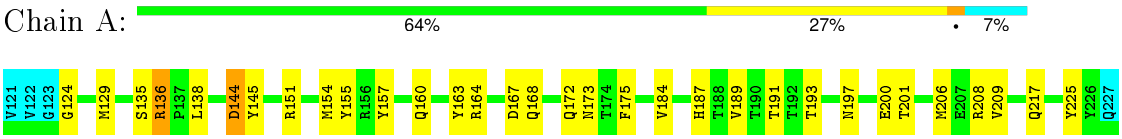
Chain A:



4.2.16 Score per residue for model 16

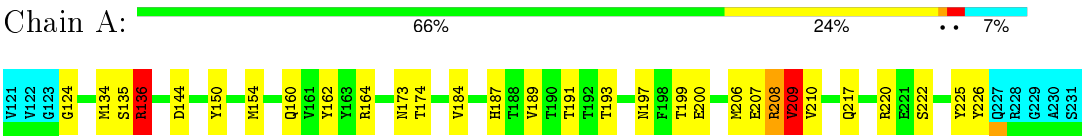
• Molecule 1: Major prion protein

Chain A:



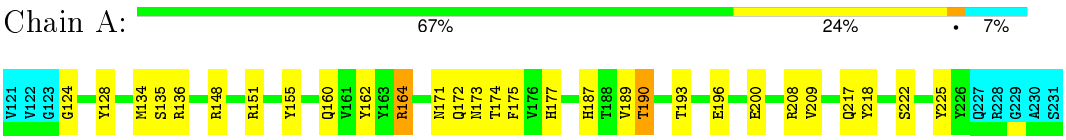
4.2.17 Score per residue for model 17

• Molecule 1: Major prion protein



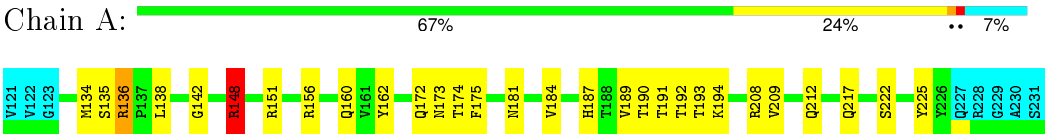
4.2.18 Score per residue for model 18

• Molecule 1: Major prion protein



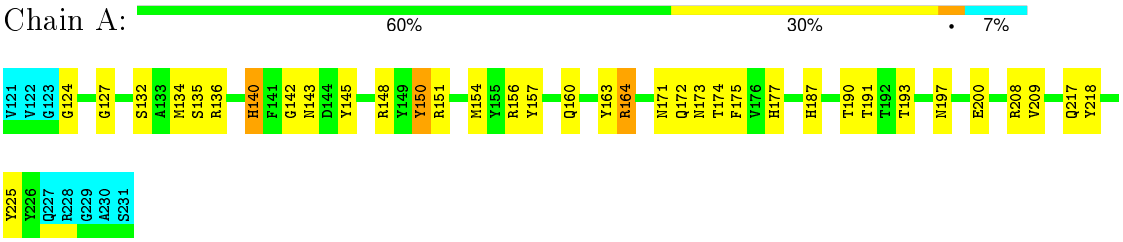
4.2.19 Score per residue for model 19

• 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

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	6.2
CANDID	refinement	1.0

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 6383
Number of chemical shift lists	1
Total number of shifts	1257
Number of shifts mapped to atoms	1247
Number of unparsed shifts	0
Number of shifts with mapping errors	10
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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	0.69±0.01	0±0/886 (0.0±0.0%)	1.11±0.03	2±1/1200 (0.1±0.1%)
All	All	0.69	0/17720 (0.0%)	1.11	31/24000 (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	3.1±1.6
All	All	0	63

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	184	VAL	CA-CB-CG2	10.57	126.76	110.90	11	10
1	A	169	TYR	CB-CG-CD2	-7.82	116.31	121.00	8	1
1	A	163	TYR	CB-CG-CD2	-7.16	116.70	121.00	12	2
1	A	162	TYR	CB-CG-CD2	-5.94	117.43	121.00	9	2
1	A	156	ARG	NE-CZ-NH2	-5.39	117.60	120.30	4	1
1	A	220	ARG	NE-CZ-NH2	5.38	122.99	120.30	13	1
1	A	209	VAL	CA-CB-CG1	5.35	118.93	110.90	7	2
1	A	184	VAL	N-CA-CB	-5.35	99.72	111.50	5	3
1	A	162	TYR	CB-CG-CD1	-5.18	117.89	121.00	13	1
1	A	148	ARG	NE-CZ-NH1	5.16	122.88	120.30	19	1
1	A	189	VAL	CA-CB-CG1	5.14	118.61	110.90	18	1
1	A	226	TYR	C-N-CA	5.10	134.45	121.70	7	1
1	A	209	VAL	CA-CB-CG2	5.09	118.54	110.90	17	1
1	A	135	SER	N-CA-CB	-5.07	102.90	110.50	11	2
1	A	208	ARG	NE-CZ-NH2	-5.06	117.77	120.30	16	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	189	VAL	CB-CA-C	5.04	120.98	111.40	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	148	ARG	Sidechain	8
1	A	136	ARG	Peptide,Sidechain	7
1	A	150	TYR	Sidechain	6
1	A	151	ARG	Sidechain	5
1	A	157	TYR	Sidechain	4
1	A	162	TYR	Sidechain	4
1	A	155	TYR	Sidechain	4
1	A	208	ARG	Sidechain	4
1	A	220	ARG	Sidechain	4
1	A	164	ARG	Sidechain	3
1	A	156	ARG	Sidechain	3
1	A	149	TYR	Sidechain	2
1	A	226	TYR	Sidechain	2
1	A	169	TYR	Sidechain	2
1	A	195	GLY	Peptide	1
1	A	175	PHE	Sidechain	1
1	A	218	TYR	Sidechain	1
1	A	163	TYR	Sidechain	1
1	A	142	GLY	Peptide	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	865	803	803	1±1
All	All	17300	16060	16060	26

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	Total
1:A:184:VAL:HG22	1:A:206:MET:SD	0.54	2.43	8	1
1:A:128:TYR:CE2	1:A:164:ARG:CZ	0.50	2.94	18	1
1:A:191:THR:HG21	1:A:198:PHE:CE2	0.48	2.44	13	1
1:A:154:MET:HA	1:A:157:TYR:CD2	0.47	2.44	6	2
1:A:175:PHE:CE2	1:A:218:TYR:HB2	0.46	2.46	18	6
1:A:206:MET:O	1:A:210:VAL:HG23	0.46	2.11	17	1
1:A:150:TYR:CE1	1:A:157:TYR:CD2	0.46	3.04	20	1
1:A:150:TYR:CE1	1:A:157:TYR:CE2	0.45	3.04	14	3
1:A:128:TYR:CE1	1:A:164:ARG:HD2	0.45	2.46	6	1
1:A:139:ILE:HD11	1:A:212:GLN:HG3	0.44	1.89	8	1
1:A:125:LEU:CD1	1:A:162:TYR:CE1	0.44	3.01	13	2
1:A:163:TYR:CD2	1:A:164:ARG:O	0.43	2.72	4	1
1:A:186:GLN:HA	1:A:189:VAL:HG22	0.42	1.92	1	1
1:A:163:TYR:CD2	1:A:175:PHE:CE1	0.42	3.08	12	1
1:A:206:MET:HA	1:A:209:VAL:CG2	0.41	2.46	17	1
1:A:140:HIS:CD2	1:A:143:ASN:HA	0.40	2.51	20	1
1:A:157:TYR:CD1	1:A:206:MET:HG3	0.40	2.51	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	103/111 (93%)	89±2 (87±2%)	10±2 (10±2%)	4±1 (4±1%)	7	36
All	All	2060/2220 (93%)	1784 (87%)	199 (10%)	77 (4%)	7	36

All 15 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	136	ARG	16
1	A	189	VAL	16
1	A	124	GLY	12
1	A	127	GLY	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	197	ASN	5
1	A	167	ASP	4
1	A	142	GLY	4
1	A	144	ASP	3
1	A	143	ASN	3
1	A	154	MET	2
1	A	196	GLU	2
1	A	125	LEU	1
1	A	141	PHE	1
1	A	145	TYR	1
1	A	190	THR	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	95/100 (95%)	75±3 (79±3%)	20±3 (21±3%)	4 33
All	All	1900/2000 (95%)	1497 (79%)	403 (21%)	4 33

All 63 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	160	GLN	20
1	A	193	THR	20
1	A	164	ARG	18
1	A	225	TYR	18
1	A	191	THR	17
1	A	208	ARG	16
1	A	187	HIS	16
1	A	217	GLN	15
1	A	172	GLN	15
1	A	173	ASN	15
1	A	209	VAL	15
1	A	134	MET	15
1	A	135	SER	13
1	A	148	ARG	12

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	174	THR	12
1	A	136	ARG	9
1	A	181	ASN	9
1	A	192	THR	8
1	A	156	ARG	7
1	A	132	SER	7
1	A	171	ASN	7
1	A	163	TYR	7
1	A	143	ASN	7
1	A	162	TYR	6
1	A	177	HIS	6
1	A	175	PHE	6
1	A	200	GLU	6
1	A	197	ASN	6
1	A	138	LEU	5
1	A	145	TYR	5
1	A	190	THR	5
1	A	151	ARG	4
1	A	154	MET	4
1	A	201	THR	3
1	A	194	LYS	3
1	A	140	HIS	3
1	A	220	ARG	3
1	A	222	SER	3
1	A	129	MET	3
1	A	130	LEU	3
1	A	161	VAL	3
1	A	206	MET	2
1	A	170	ASN	2
1	A	223	GLU	2
1	A	144	ASP	2
1	A	168	GLN	2
1	A	159	ASN	2
1	A	221	GLU	1
1	A	141	PHE	1
1	A	166	VAL	1
1	A	188	THR	1
1	A	167	ASP	1
1	A	155	TYR	1
1	A	214	CYS	1
1	A	205	MET	1
1	A	211	GLU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	207	GLU	1
1	A	189	VAL	1
1	A	202	ASP	1
1	A	199	THR	1
1	A	196	GLU	1
1	A	128	TYR	1
1	A	212	GLN	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 80% for the well-defined parts and 80% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 6383

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	1257
Number of shifts mapped to atoms	1247
Number of unparsed shifts	0
Number of shifts with mapping errors	10
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.

- Residue not found in structure. All 10 occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	1	GLY	HA2	3.89	0.005	1
A	1	GLY	CA	42.8	0.05	1
A	2	SER	N	115.7	0.05	1
A	2	SER	CB	63.5	0.05	1
A	2	SER	HB3	3.83	0.005	1
A	2	SER	H	8.64	0.005	1
A	2	SER	HB2	3.83	0.005	1
A	2	SER	HA	4.58	0.005	1
A	2	SER	CA	57.7	0.05	1
A	1	GLY	HA3	3.89	0.005	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	112	-0.08 ± 0.20	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	103	0.60 ± 0.13	Should be applied
$^{13}C'$	0	—	—
^{15}N	108	0.21 ± 0.14	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 80%, i.e. 1054 atoms were assigned a chemical shift out of a possible 1319. 11 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	406/509 (80%)	203/203 (100%)	103/206 (50%)	100/100 (100%)
Sidechain	580/666 (87%)	368/392 (94%)	190/235 (81%)	22/39 (56%)
Aromatic	68/144 (47%)	68/75 (91%)	0/66 (0%)	0/3 (0%)
Overall	1054/1319 (80%)	639/670 (95%)	293/507 (58%)	122/142 (86%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 80%, i.e. 1118 atoms were assigned a chemical shift out of a possible 1404. 13 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	434/549 (79%)	217/219 (99%)	110/222 (50%)	107/108 (99%)
Sidechain	616/711 (87%)	390/418 (93%)	202/250 (81%)	24/43 (56%)
Aromatic	68/144 (47%)	68/75 (91%)	0/66 (0%)	0/3 (0%)
Overall	1118/1404 (80%)	675/712 (95%)	312/538 (58%)	131/154 (85%)

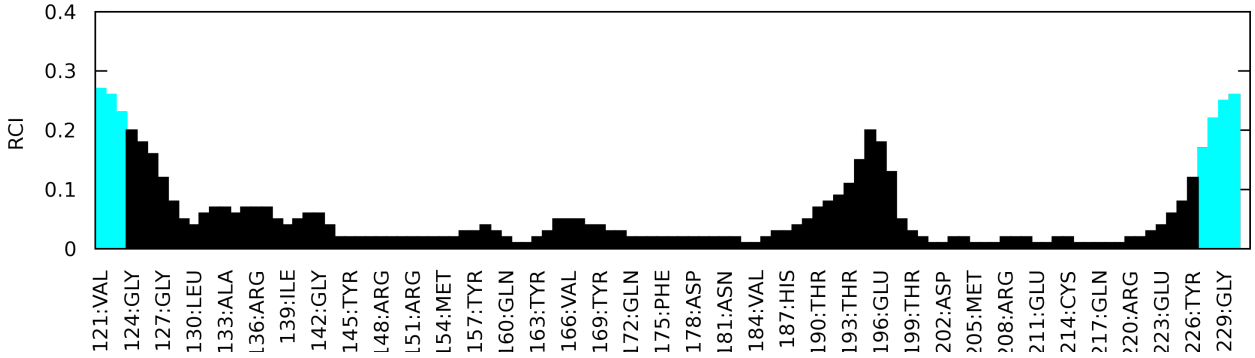
7.1.4 Statistically unusual chemical shifts

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:


RCI
0.4
0.3
0.2
0.1
0
121:VAL
124:GLY
127:GLY
130:LEU
133:ALA
136:ARG
139:ILE
142:GLY
145:TYR
148:ARG
151:ARG
154:MET
157:TYR
160:GLN
163:TYR
166:VAL
169:TYR
172:GLN
175:PHE
178:ASP
181:ASN
184:VAL
187:HIS
190:THR
193:THR
196:GLU
199:THR
202:ASP
205:MET
208:ARG
211:GLU
214:CYS
217:GLN
220:ARG
223:GLU
226:TYR
229:GLY