Full wwPDB NMR Structure Validation Report (i)

Feb 23, 2022 – 12:49 PM EST

PDB ID	:	1XYJ
Title	:	NMR Structure of the cat prion protein
Authors	:	Lysek, D.A.; Schorn, C.; Nivon, L.G.; Esteve-Moya, V.; Christen, B.; Calzolai, L.; von Schroetter, C.; Fiorito, F.; Herrmann, T.; Guntert, P.
Deposited on	:	2004-11-10

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
https://www.wwpdb.org/validation/2017/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:

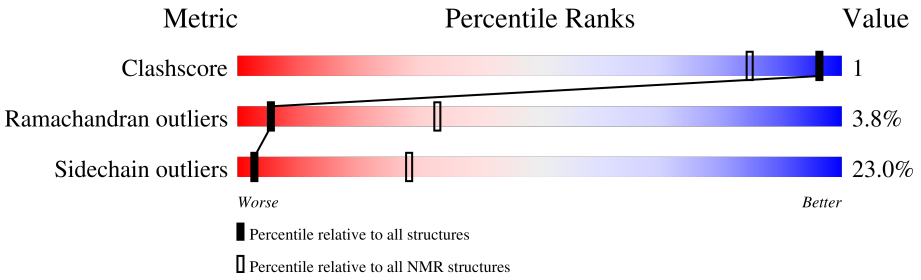
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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 was not calculated.

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	64% 16% 20%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 16 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:127-A:166, A:175-A:223 (89)	0.51	16

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 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1787 atoms, of which 864 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called prion protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	111	1787	573	864	163	179	8	0

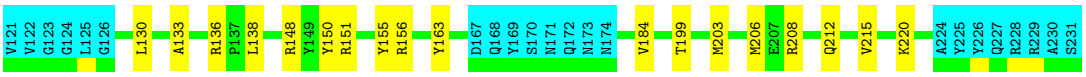
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: prion protein

Chain A: 64% 16% 20%



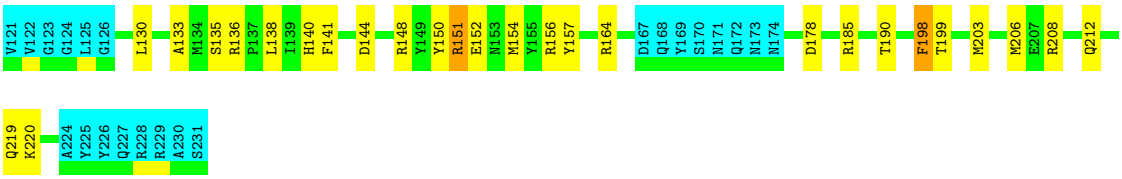
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: prion protein

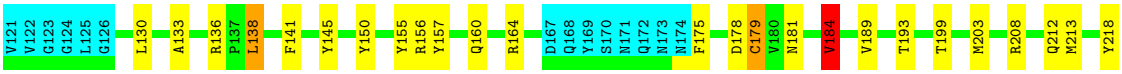
Chain A: 56% 23% • 20%

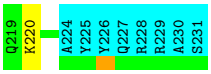


4.2.2 Score per residue for model 2

• Molecule 1: prion protein

Chain A: 57% 21% • • 20%





4.2.3 Score per residue for model 3

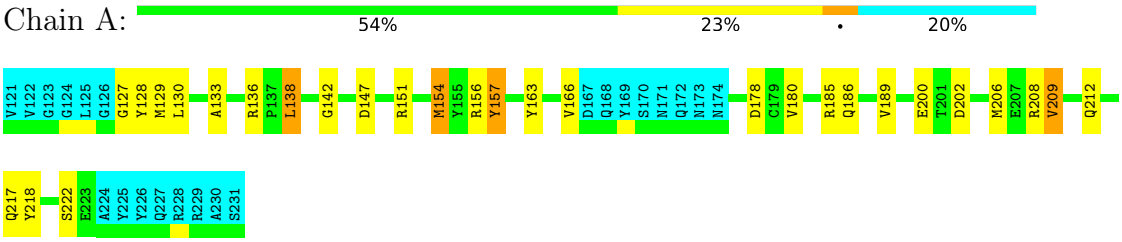
- Molecule 1: prion protein

Chain A:

4.2.4 Score per residue for model 4

- Molecule 1: prion protein

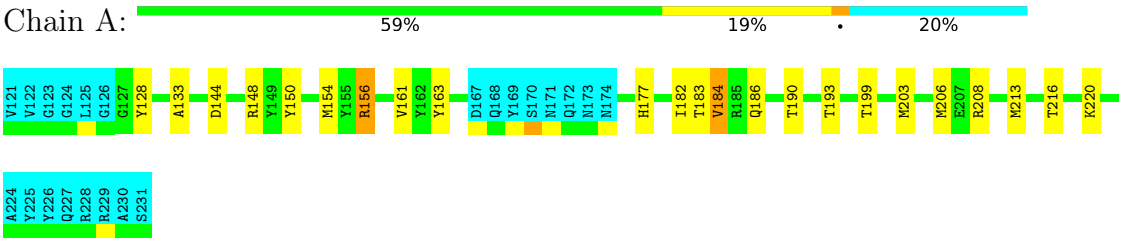
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: prion protein

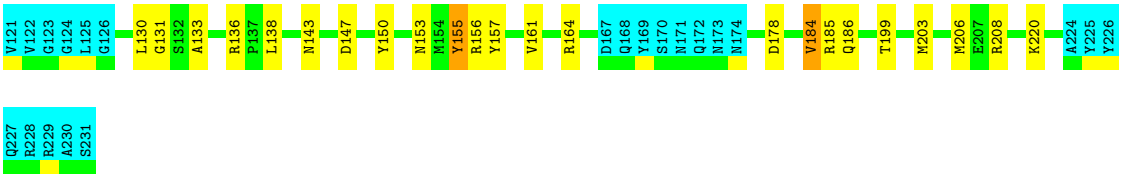
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: prion protein

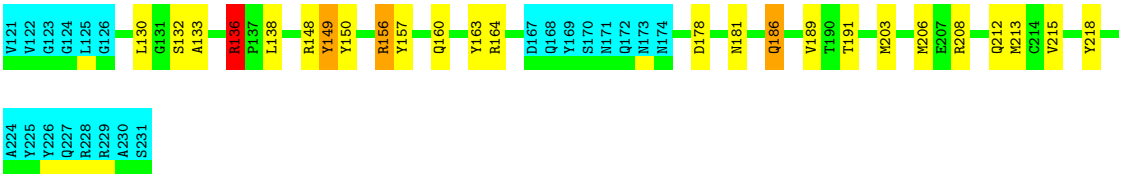
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: prion protein

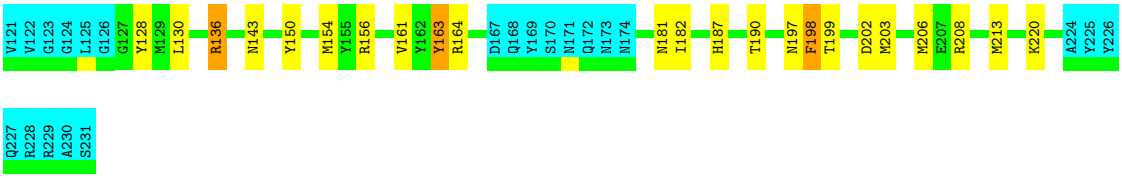
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: prion protein

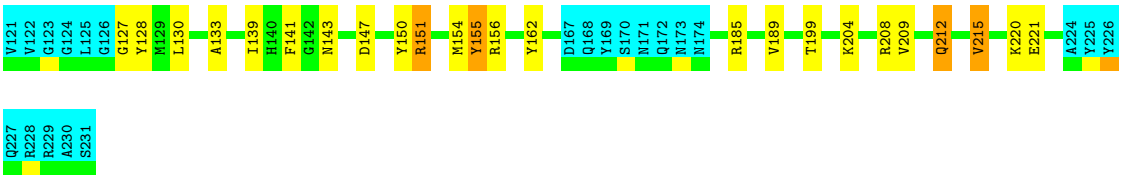
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: prion protein

Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: prion protein

Chain A: 56% 20% • • 20%

V121 V122 G123 G124 L125 G126 L130 A133 M134 S135 R136 P137 L138 F141 R148 Y149 Y150 M154 Y155 R156 Y157 Y163 D167 Q168 Y169 S170 N171 Q172 N173 N174 D178 V184 R185 Q186 H187 T188 D202 I205 M206 E207 R208 V209 Q212 M213 C214 V215 Y218

A224 Y225 Y226 Q227 R228 R229 A230 S231

4.2.11 Score per residue for model 11

- Molecule 1: prion protein

Chain A: 59% 16% • • 20%

V121 V122 G123 G124 L125 G126 L130 G131 S132 A133 R136 P137 L138 I139 Y150 Y155 R156 D167 Q168 Y169 S170 N171 Q172 N173 N174 C179 V180 N181 I182 T183 V184 F198 T199 M203 M206 E207 R208 V209 T216 Q219 K220 A224 Y225 Y226 Q227 R228 R229 A230

S231

4.2.12 Score per residue for model 12

- Molecule 1: prion protein

Chain A: 57% 18% 5% 20%

V121 V122 G123 G124 L125 G126 L130 A133 R136 P137 L138 I139 R148 Y149 Y150 R151 Y155 R156 V161 Y162 Y163 D167 Q168 Y169 S170 N171 Q172 N173 N174 N181 I182 T183 V184 R185 V189 T199 M203 R208 V209 Q212 K220 E221 A224 Y225 Y226

Q227 R228 R229 A230 S231

4.2.13 Score per residue for model 13

- Molecule 1: prion protein

Chain A: 52% 25% • • 20%

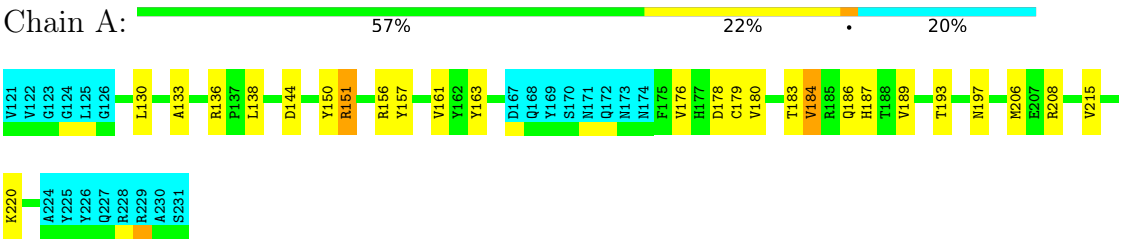
V121 V122 G123 G124 L125 G126 G127 Y128 M129 L130 A133 L138 F141 D147 R148 Y149 Y150 R151 M154 Y155 R156 Y162 D167 Q168 Y169 S170 N171 Q172 N173 N174 D178 C179 V180 N181 V184 R185 T190 F198 T199 M203 K204 I205 M206 E207 R208 Q212 M213

C214 V215 Y218 Q219 K220 A224 Y225 Y226 Q227 R228 R229 A230 S231

4.2.14 Score per residue for model 14

- Molecule 1: prion protein

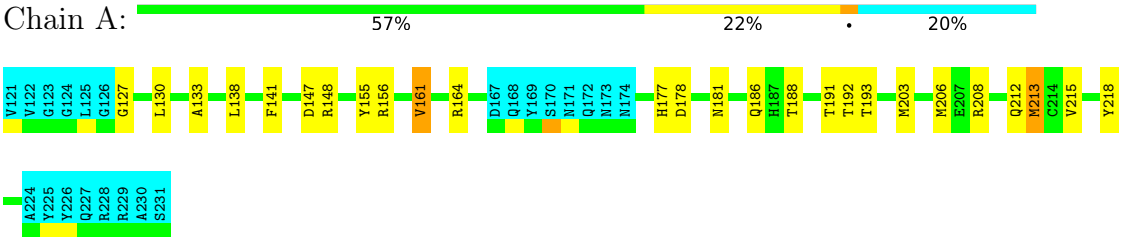
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: prion protein

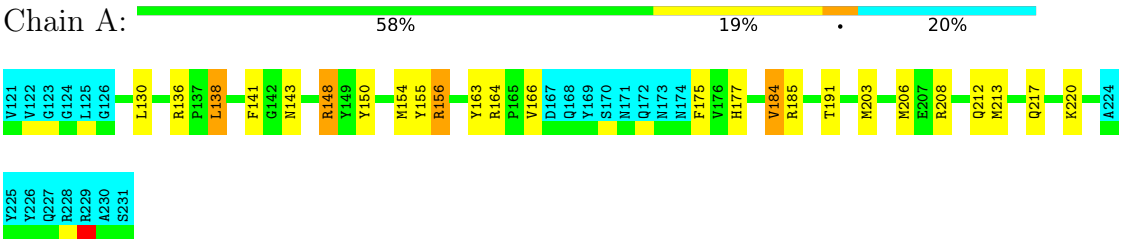
Chain A:



4.2.16 Score per residue for model 16 (medoid)

- Molecule 1: prion protein

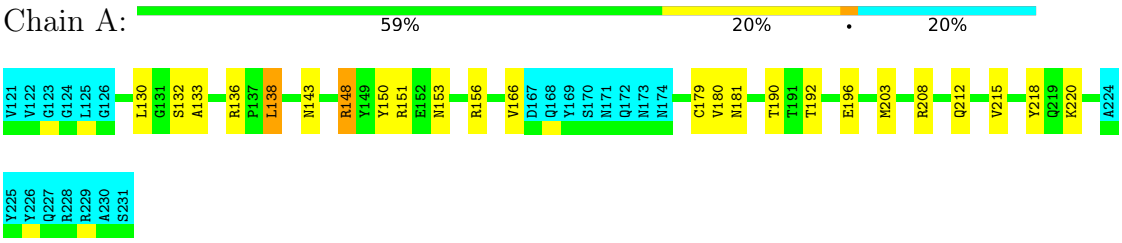
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: prion protein

Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: prion protein

Chain A: 59% 18% • 20%

V121 V122 G123 G124 L125 G126 L130 A133 M134 L138 Y150 R151 Y155 R156 V166 D167 Q168 Y169 S170 N171 Q172 N173 N174 F175 C179 I182 T183 V184* T199 D202 M203 K204 R208 Q212 M213 C214 V215 T216 Q217 Y218 Q219 K220 A224 Y225 Y226 Q227 R228

R229 A230 S231

4.2.19 Score per residue for model 19

- Molecule 1: prion protein

Chain A: 53% 25% • 20%

V121 V122 G123 G124 L125 G126 G127 Y128 M129 L130 G131 S132 A133 L138 F141 D144 R148 Y149 Y150 Y155 R156 Y157 Q160 Y163 R164 P165 V166 D167 Q168 Y169 S170 N171 Q172 N173 N174 N181 I182 Q186 V189 T192 T193 K194 T199 M203 M206 E207

R208 Q212 T216 K220 E221 S222 E223 A224 Y225 Y226 Q227 R228 R229 A230 S231

4.2.20 Score per residue for model 20

- Molecule 1: prion protein

Chain A: 58% 21% • 20%

V121 V122 G123 G124 L125 G126 L130 G131 S132 A133 R136 P137 L138 F141 D147 R148 Y149 Y150 R151 R156 M159 D167 Q168 Y169 S170 N171 Q172 N173 N174 F175 V176 H177 D178 C179 V180 V184 R185 M203 M206 E207 R208 M213 C214 Q219 K220 A224 Y225

Y226 Q227 R228 R229 A230 S231

5 Refinement protocol and experimental data overview (i)

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

Of the 20 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

No chemical shift data was provided.

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	0.68±0.01	0±0/768 (0.0± 0.0%)	1.15±0.03	2±1/1039 (0.2± 0.1%)
All	All	0.68	0/15360 (0.0%)	1.15	50/20780 (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
1	A	0.0±0.0	2.2±1.3
All	All	0	44

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	157	TYR	CB-CG-CD1	-9.11	115.54	121.00	7	1
1	A	184	VAL	CA-CB-CG2	8.52	123.69	110.90	18	9
1	A	180	VAL	CA-CB-CG2	7.51	122.16	110.90	14	5
1	A	148	ARG	CD-NE-CZ	7.18	133.66	123.60	16	1
1	A	209	VAL	CA-CB-CG2	7.06	121.49	110.90	10	2
1	A	185	ARG	NE-CZ-NH2	-6.82	116.89	120.30	1	1
1	A	151	ARG	NE-CZ-NH2	-6.47	117.06	120.30	1	3
1	A	164	ARG	NE-CZ-NH2	-6.41	117.09	120.30	1	4
1	A	141	PHE	CB-CG-CD2	-6.20	116.46	120.80	10	8
1	A	156	ARG	NE-CZ-NH2	-6.12	117.24	120.30	7	1
1	A	184	VAL	CA-CB-CG1	6.01	119.92	110.90	2	1
1	A	136	ARG	NE-CZ-NH2	-5.98	117.31	120.30	7	1
1	A	208	ARG	NE-CZ-NH2	-5.96	117.32	120.30	12	1
1	A	130	LEU	CB-CG-CD1	-5.95	100.88	111.00	18	1
1	A	157	TYR	CB-CA-C	5.91	122.22	110.40	7	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	208	ARG	CD-NE-CZ	5.82	131.75	123.60	12	1
1	A	148	ARG	NE-CZ-NH1	5.68	123.14	120.30	16	1
1	A	215	VAL	CA-CB-CG1	5.56	119.23	110.90	10	2
1	A	151	ARG	NE-CZ-NH1	5.47	123.03	120.30	18	1
1	A	164	ARG	NE-CZ-NH1	5.46	123.03	120.30	2	1
1	A	155	TYR	CB-CG-CD2	-5.25	117.85	121.00	6	1
1	A	163	TYR	CB-CG-CD2	-5.20	117.88	121.00	12	1
1	A	148	ARG	NE-CZ-NH2	-5.11	117.75	120.30	16	1
1	A	130	LEU	CB-CA-C	5.02	119.74	110.20	10	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	157	TYR	Sidechain	7
1	A	151	ARG	Sidechain	5
1	A	163	TYR	Sidechain,Peptide	4
1	A	136	ARG	Sidechain	4
1	A	128	TYR	Sidechain	3
1	A	148	ARG	Sidechain	2
1	A	177	HIS	Peptide	2
1	A	218	TYR	Sidechain	2
1	A	156	ARG	Sidechain	2
1	A	149	TYR	Sidechain	2
1	A	164	ARG	Sidechain	2
1	A	155	TYR	Sidechain	2
1	A	162	TYR	Sidechain	2
1	A	145	TYR	Sidechain	1
1	A	142	GLY	Peptide	1
1	A	150	TYR	Sidechain	1
1	A	196	GLU	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	750	707	709	1±1
All	All	15000	14140	14180	23

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:161:VAL:HG12	1:A:183:THR:HG21	0.57	1.76	12	2
1:A:181:ASN:HA	1:A:184:VAL:HG13	0.56	1.76	2	3
1:A:139:ILE:HD11	1:A:209:VAL:HA	0.54	1.79	11	1
1:A:161:VAL:HG11	1:A:213:MET:HG3	0.53	1.79	15	2
1:A:166:VAL:HG13	1:A:218:TYR:CE1	0.50	2.42	18	1
1:A:139:ILE:HD12	1:A:141:PHE:CE1	0.49	2.42	9	1
1:A:212:GLN:HA	1:A:215:VAL:HG22	0.48	1.84	9	1
1:A:212:GLN:O	1:A:216:THR:HG22	0.48	2.08	18	1
1:A:139:ILE:HD11	1:A:209:VAL:HG23	0.47	1.87	9	2
1:A:139:ILE:CD1	1:A:209:VAL:HG23	0.46	2.39	12	1
1:A:180:VAL:O	1:A:184:VAL:HG22	0.46	2.11	11	1
1:A:212:GLN:HA	1:A:215:VAL:CG2	0.45	2.42	9	1
1:A:198:PHE:CD2	1:A:206:MET:SD	0.44	3.10	13	1
1:A:150:TYR:CG	1:A:205:ILE:HG21	0.43	2.47	10	1
1:A:205:ILE:O	1:A:209:VAL:HG22	0.43	2.13	10	1
1:A:130:LEU:C	1:A:130:LEU:HD12	0.43	2.33	7	1
1:A:128:TYR:CE2	1:A:182:ILE:HG13	0.41	2.50	8	1
1:A:130:LEU:HD13	1:A:160:GLN:HB3	0.41	1.91	7	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	89/111 (80%)	73±3 (82±4%)	13±3 (14±4%)	3±1 (4±1%)	5	33
All	All	1780/2220 (80%)	1455 (82%)	257 (14%)	68 (4%)	5	33

All 21 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	133	ALA	18
1	A	138	LEU	12
1	A	198	PHE	4
1	A	178	ASP	4
1	A	166	VAL	4
1	A	179	CYS	3
1	A	127	GLY	3
1	A	156	ARG	3
1	A	131	GLY	2
1	A	186	GLN	2
1	A	155	TYR	2
1	A	176	VAL	2
1	A	140	HIS	1
1	A	154	MET	1
1	A	153	ASN	1
1	A	143	ASN	1
1	A	147	ASP	1
1	A	184	VAL	1
1	A	161	VAL	1
1	A	136	ARG	1
1	A	194	LYS	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	84/101 (83%)	65±2 (77±2%)	19±2 (23±2%)	3 28
All	All	1680/2020 (83%)	1294 (77%)	386 (23%)	3 28

All 67 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	156	ARG	20
1	A	208	ARG	20
1	A	150	TYR	18
1	A	130	LEU	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	203	MET	16
1	A	220	LYS	16
1	A	206	MET	14
1	A	212	GLN	13
1	A	184	VAL	12
1	A	136	ARG	11
1	A	199	THR	11
1	A	138	LEU	10
1	A	148	ARG	10
1	A	154	MET	9
1	A	185	ARG	9
1	A	213	MET	8
1	A	163	TYR	8
1	A	215	VAL	8
1	A	189	VAL	7
1	A	186	GLN	7
1	A	181	ASN	7
1	A	178	ASP	6
1	A	155	TYR	6
1	A	179	CYS	6
1	A	151	ARG	6
1	A	190	THR	5
1	A	218	TYR	5
1	A	143	ASN	5
1	A	147	ASP	5
1	A	144	ASP	4
1	A	193	THR	4
1	A	202	ASP	4
1	A	182	ILE	4
1	A	132	SER	4
1	A	214	CYS	4
1	A	198	PHE	3
1	A	219	GLN	3
1	A	175	PHE	3
1	A	134	MET	3
1	A	217	GLN	3
1	A	177	HIS	3
1	A	216	THR	3
1	A	161	VAL	3
1	A	164	ARG	3
1	A	191	THR	3
1	A	187	HIS	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	204	LYS	3
1	A	192	THR	3
1	A	160	GLN	2
1	A	129	MET	2
1	A	209	VAL	2
1	A	222	SER	2
1	A	197	ASN	2
1	A	221	GLU	2
1	A	188	THR	2
1	A	183	THR	2
1	A	128	TYR	2
1	A	135	SER	1
1	A	152	GLU	1
1	A	157	TYR	1
1	A	200	GLU	1
1	A	149	TYR	1
1	A	139	ILE	1
1	A	162	TYR	1
1	A	207	GLU	1
1	A	153	ASN	1
1	A	159	ASN	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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 monosaccharides 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)

No chemical shift data were provided