Full wwPDB NMR Structure Validation Report (i)

Nov 9, 2024 – 10:35 am GMT

PDB ID	:	1E1U
BMRB ID	:	4620
Title	:	Human prion protein variant R220K
Authors	:	Calzolai, L.; Lysek, D.A.; Guntert, P.; Von Schroetter, C.; Zahn, R.; Riek, R.; Wuthrich, K.
Deposited on	:	2000-05-11

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

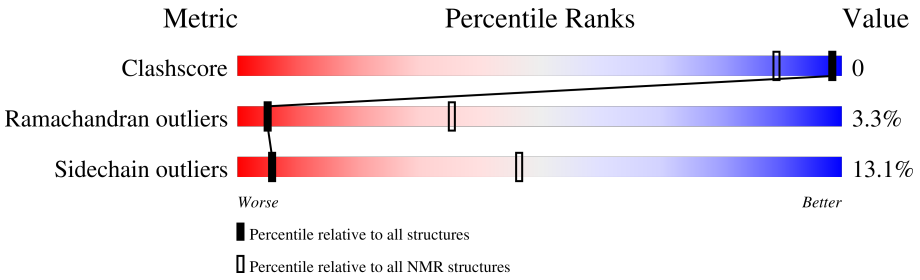
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	104	83% 6% 12%

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 7 as representative.

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:128-A:219 (92)	0.93	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 5 clusters and 3 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1689 atoms, of which 814 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	104	1689	544	814	151	171	9	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	220	LYS	ARG	engineered mutation	UNP P78446

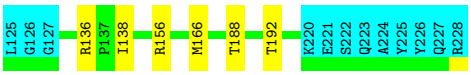
4 Residue-property plots

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:



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

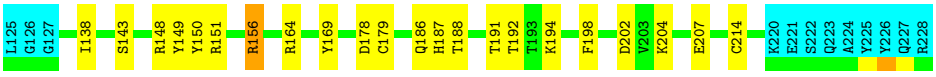
• Molecule 1: PRION PROTEIN

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: PRION PROTEIN

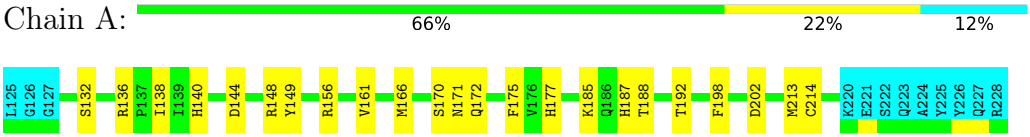
Chain A:



4.2.3 Score per residue for model 3 (medoid)

• Molecule 1: PRION PROTEIN

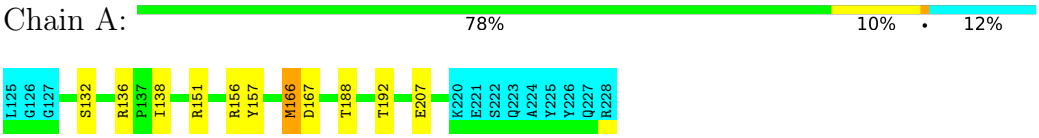
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: PRION PROTEIN

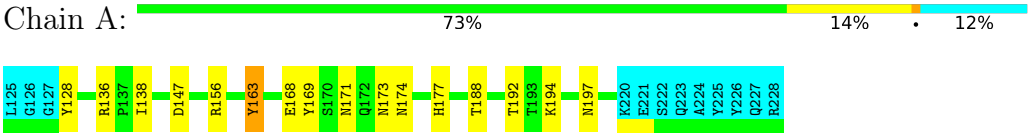
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: PRION PROTEIN

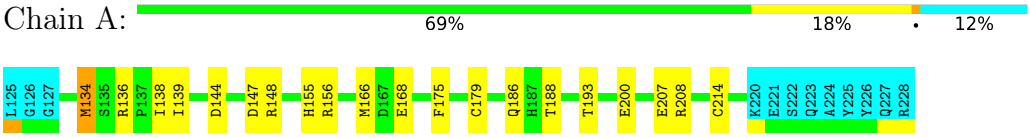
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: PRION PROTEIN

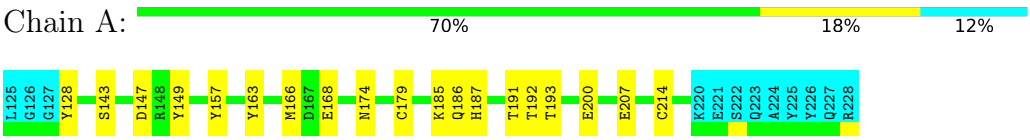
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: PRION PROTEIN

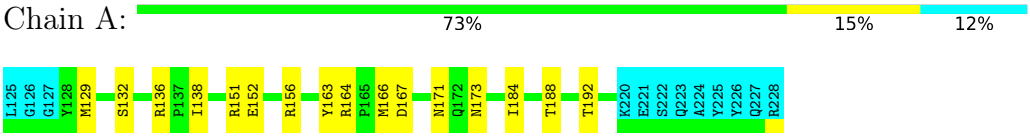
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: PRION PROTEIN

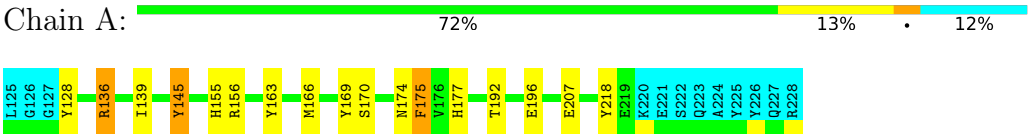
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: PRION PROTEIN

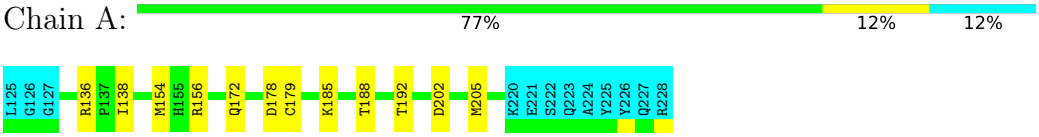
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: PRION PROTEIN

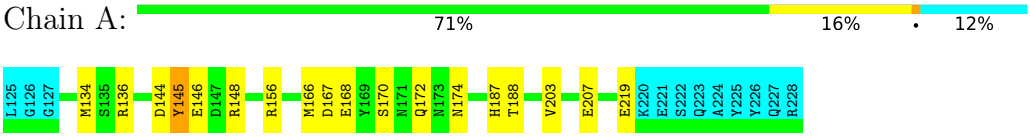
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: PRION PROTEIN

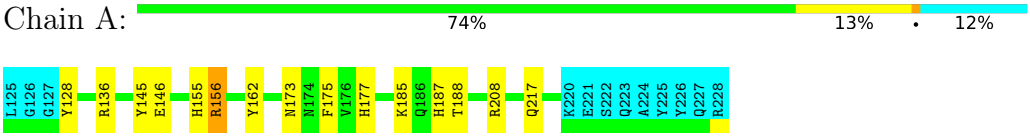
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: PRION PROTEIN

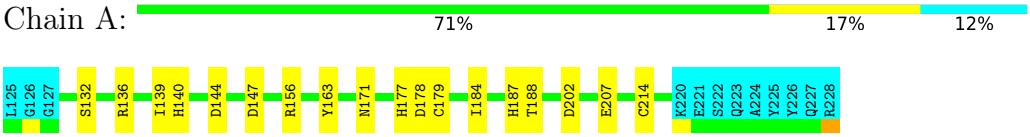
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: PRION PROTEIN

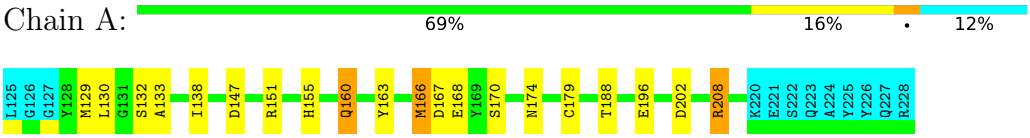
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: PRION PROTEIN

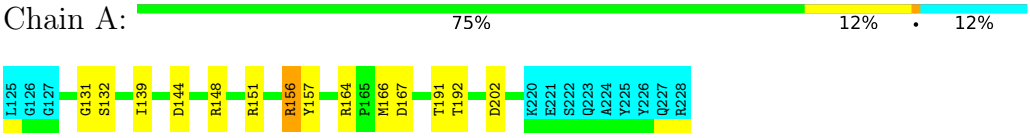
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: PRION PROTEIN

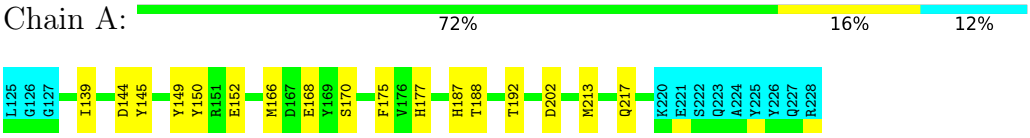
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: PRION PROTEIN

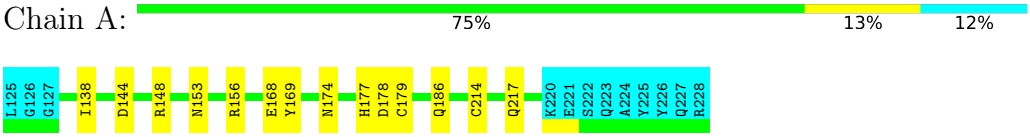
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: PRION PROTEIN

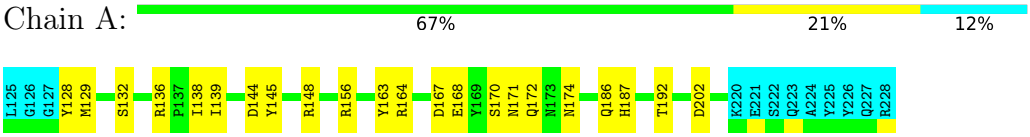
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: PRION PROTEIN

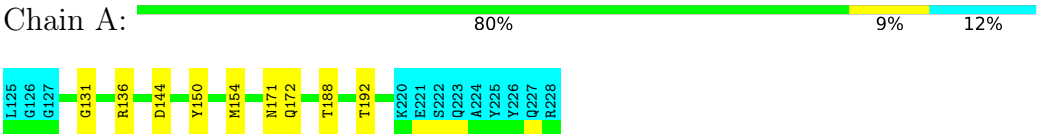
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: PRION PROTEIN

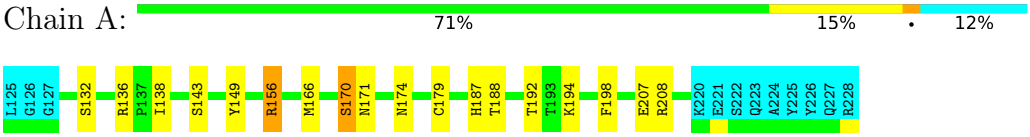
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: PRION PROTEIN

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
OPALp	refinement	
DYANA	structure solution	

No chemical shift data was provided.

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±0/796 (0.0± 0.0%)	1.03±0.03	1±1/1077 (0.1± 0.1%)
All	All	0.67	0/15920 (0.0%)	1.03	16/21540 (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	1.9±1.4
All	All	0	38

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	151	ARG	NE-CZ-NH2	-6.93	116.84	120.30	2	2
1	A	157	TYR	CB-CG-CD2	-6.19	117.29	121.00	7	3
1	A	163	TYR	CB-CG-CD2	-6.18	117.29	121.00	8	3
1	A	203	VAL	CA-CB-CG1	5.99	119.88	110.90	11	1
1	A	208	ARG	NE-CZ-NH2	-5.79	117.41	120.30	14	2
1	A	148	ARG	NE-CZ-NH2	-5.71	117.44	120.30	15	1
1	A	151	ARG	CD-NE-CZ	5.51	131.32	123.60	2	1
1	A	136	ARG	NE-CZ-NH2	-5.42	117.59	120.30	1	1
1	A	136	ARG	NE-CZ-NH1	5.16	122.88	120.30	1	1
1	A	164	ARG	NE-CZ-NH1	5.14	122.87	120.30	2	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	5
1	A	149	TYR	Sidechain	4
1	A	136	ARG	Sidechain	4
1	A	145	TYR	Sidechain	4
1	A	208	ARG	Sidechain	3
1	A	150	TYR	Sidechain	3
1	A	156	ARG	Sidechain	3
1	A	128	TYR	Sidechain	3
1	A	134	MET	Peptide	1
1	A	164	ARG	Sidechain	1
1	A	162	TYR	Sidechain	1
1	A	131	GLY	Peptide	1
1	A	151	ARG	Sidechain	1
1	A	153	ASN	Peptide	1
1	A	163	TYR	Sidechain	1
1	A	167	ASP	Peptide	1
1	A	170	SER	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	777	721	721	0±1
All	All	15540	14420	14420	6

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:175:PHE:CD2	1:A:218:TYR:CE1	0.50	2.99	9	1
1:A:175:PHE:CD2	1:A:218:TYR:CD1	0.49	3.00	9	1
1:A:145:TYR:CE2	1:A:146:GLU:HG3	0.42	2.50	11	1
1:A:161:VAL:CG1	1:A:213:MET:SD	0.41	3.08	3	1
1:A:145:TYR:CE1	1:A:146:GLU:HG3	0.41	2.50	12	1
1:A:133:ALA:HA	1:A:160:GLN:CG	0.40	2.47	14	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	92/104 (88%)	81±2 (88±2%)	8±2 (9±2%)	3±1 (3±1%)	5 35
All	All	1840/2080 (88%)	1612 (88%)	168 (9%)	60 (3%)	5 35

All 18 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	11
1	A	170	SER	6
1	A	139	ILE	6
1	A	166	MET	5
1	A	167	ASP	5
1	A	168	GLU	4
1	A	169	TYR	3
1	A	198	PHE	3
1	A	172	GLN	3
1	A	219	GLU	2
1	A	171	ASN	2
1	A	136	ARG	2
1	A	128	TYR	2
1	A	132	SER	2
1	A	134	MET	1
1	A	140	HIS	1
1	A	156	ARG	1
1	A	131	GLY	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	88/97 (91%)	76±3 (87±3%)	12±3 (13±3%)	6 46
All	All	1760/1940 (91%)	1529 (87%)	231 (13%)	6 46

All 52 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	16
1	A	188	THR	14
1	A	192	THR	13
1	A	187	HIS	9
1	A	144	ASP	9
1	A	179	CYS	8
1	A	202	ASP	8
1	A	207	GLU	8
1	A	166	MET	8
1	A	136	ARG	8
1	A	174	ASN	8
1	A	214	CYS	7
1	A	177	HIS	7
1	A	147	ASP	6
1	A	163	TYR	6
1	A	186	GLN	6
1	A	132	SER	6
1	A	178	ASP	5
1	A	175	PHE	5
1	A	171	ASN	5
1	A	191	THR	4
1	A	185	LYS	4
1	A	168	GLU	4
1	A	155	HIS	4
1	A	151	ARG	3
1	A	208	ARG	3
1	A	143	SER	3
1	A	194	LYS	3
1	A	173	ASN	3
1	A	129	MET	3
1	A	217	GLN	3
1	A	170	SER	2
1	A	213	MET	2
1	A	169	TYR	2
1	A	172	GLN	2
1	A	193	THR	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	200	GLU	2
1	A	152	GLU	2
1	A	184	ILE	2
1	A	196	GLU	2
1	A	154	MET	2
1	A	164	ARG	2
1	A	204	LYS	1
1	A	140	HIS	1
1	A	197	ASN	1
1	A	149	TYR	1
1	A	145	TYR	1
1	A	205	MET	1
1	A	134	MET	1
1	A	130	LEU	1
1	A	160	GLN	1
1	A	148	ARG	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 oligosaccharides 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

No chemical shift data were provided