Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 07:36 pm BST

PDB ID	:	1HJM
Title	:	HUMAN PRION PROTEIN AT PH 7.0
Authors	:	Calzolai, L.; Zahn, R.
Deposited on	:	2003-02-27

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 (1)) were used in the production of this report:

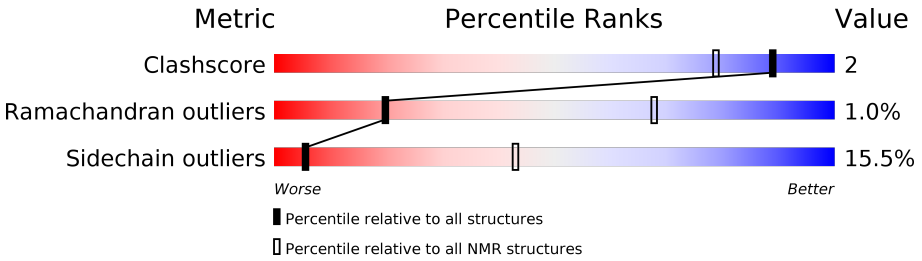
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 79%.

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	104	80% 17% • •

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1691 atoms, of which 814 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called MAJOR PRION PROTEIN PRECURSOR.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	104	1691	544	814	153	171	9	0

4 Residue-property plots (i)

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 PRECURSOR

Chain A:

5 Refinement protocol and experimental data overview (i)

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

Of the 20 calculated structures, 1 were deposited, based on the following criterion: *LOWER TARGET FUNCTION*.

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

Software name	Classification	Version
OPALp	refinement	
DYANA	structure solution	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1240
Number of shifts mapped to atoms	1180
Number of unparsed shifts	0
Number of shifts with mapping errors	60
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	79%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	877	814	813	3
All	All	877	814	813	3

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 clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
1:A:161:VAL:HA	1:A:183:THR:HG21	0.48	1.84
1:A:176:VAL:HG22	1:A:214:CYS:HB3	0.45	1.89
1:A:139:ILE:CG1	1:A:209:VAL:HG13	0.43	2.44

5.2 Torsion angles (i)

5.2.1 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	102/104 (98%)	88 (86%)	13 (13%)	1 (1%)	20 68
All	All	102/104 (98%)	88 (86%)	13 (13%)	1 (1%)	20 68

All 1 Ramachandran outliers are listed below.

Mol	Chain	Res	Type
1	A	135	SER

5.2.2 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	97/97 (100%)	82 (85%)	15 (15%)	5 43
All	All	97/97 (100%)	82 (85%)	15 (15%)	5 43

All 15 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	181	ASN
1	A	179	CYS
1	A	188	THR
1	A	172	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	187	HIS
1	A	209	VAL
1	A	220	ARG
1	A	153	ASN
1	A	190	THR
1	A	197	ASN
1	A	132	SER
1	A	159	ASN
1	A	148	ARG
1	A	163	TYR
1	A	183	THR

5.2.3 RNA (i)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 79% for the well-defined parts and 79% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 Bookkeeping (i)

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	1240
Number of shifts mapped to atoms	1180
Number of unparsed shifts	0
Number of shifts with mapping errors	60
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Residue not found in structure. All 60 occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	3	VAL	HG21	0.89	-1.0	2
A	3	VAL	CG1	21.0	-1.0	1
A	111	GLY	HA2	3.95	-1.0	1
A	4	VAL	HG13	0.88	-1.0	2
A	112	SER	HB3	3.89	-1.0	1
A	3	VAL	HA	4.18	-1.0	1
A	4	VAL	N	124.7	-1.0	1
A	1	GLY	HA2	3.88	-1.0	1
A	4	VAL	CB	32.4	-1.0	1
A	3	VAL	H	8.37	-1.0	1
A	4	VAL	HA	4.08	-1.0	1
A	1	GLY	CA	43.0	-1.0	1
A	4	VAL	CG2	20.5	-1.0	1
A	4	VAL	HG12	0.88	-1.0	2

Continued on next page...

Continued from previous page...

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	4	VAL	H	8.32	-1.0	1
A	112	SER	HB2	3.89	-1.0	1
A	3	VAL	CB	32.4	-1.0	1
A	111	GLY	CA	45.0	-1.0	1
A	4	VAL	HG22	0.9	-1.0	2
A	6	GLY	HA2	3.94	-1.0	1
A	3	VAL	HG23	0.89	-1.0	2
A	2	SER	N	115.7	-1.0	1
A	2	SER	CB	63.7	-1.0	1
A	2	SER	HB3	3.82	-1.0	1
A	112	SER	HA	4.45	-1.0	1
A	6	GLY	N	108.6	-1.0	1
A	111	GLY	HA3	3.95	-1.0	1
A	3	VAL	CA	61.9	-1.0	1
A	4	VAL	HG23	0.9	-1.0	2
A	6	GLY	HA3	3.94	-1.0	1
A	5	GLY	CA	45.1	-1.0	1
A	2	SER	H	8.64	-1.0	1
A	3	VAL	HG22	0.89	-1.0	2
A	112	SER	H	7.84	-1.0	1
A	111	GLY	N	109.9	-1.0	1
A	3	VAL	CG2	20.3	-1.0	1
A	2	SER	HB2	3.82	-1.0	1
A	6	GLY	H	8.3	-1.0	1
A	5	GLY	HA3	3.93	-1.0	1
A	3	VAL	HG11	0.86	-1.0	2
A	3	VAL	HB	2.03	-1.0	1
A	112	SER	CB	63.6	-1.0	1
A	4	VAL	HG11	0.88	-1.0	2
A	4	VAL	HB	1.99	-1.0	1
A	2	SER	HA	4.58	-1.0	1
A	4	VAL	CG1	20.9	-1.0	1
A	5	GLY	HA2	3.93	-1.0	1
A	3	VAL	HG12	0.86	-1.0	2
A	4	VAL	HG21	0.9	-1.0	2
A	5	GLY	H	8.58	-1.0	1
A	2	SER	CA	57.5	-1.0	1
A	111	GLY	H	8.3	-1.0	1
A	3	VAL	HG13	0.86	-1.0	2
A	6	GLY	CA	45.0	-1.0	1
A	3	VAL	N	122.0	-1.0	1

Continued on next page...

Continued from previous page...

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	112	SER	CA	58.3	-1.0	1
A	112	SER	N	120.0	-1.0	1
A	1	GLY	HA3	3.88	-1.0	1
A	4	VAL	CA	61.9	-1.0	1
A	5	GLY	N	113.4	-1.0	1

6.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	110	0.09 ± 0.11	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	100	0.45 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	104	0.18 ± 0.17	None needed (< 0.5 ppm)

6.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 79%, i.e. 1063 atoms were assigned a chemical shift out of a possible 1345. 9 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	399/514 (78%)	200/205 (98%)	102/208 (49%)	97/101 (96%)
Sidechain	571/688 (83%)	358/407 (88%)	191/240 (80%)	22/41 (54%)
Aromatic	93/143 (65%)	61/75 (81%)	32/64 (50%)	0/4 (0%)
Overall	1063/1345 (79%)	619/687 (90%)	325/512 (63%)	119/146 (82%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 79%, i.e. 1063 atoms were assigned a chemical shift out of a possible 1345. 9 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

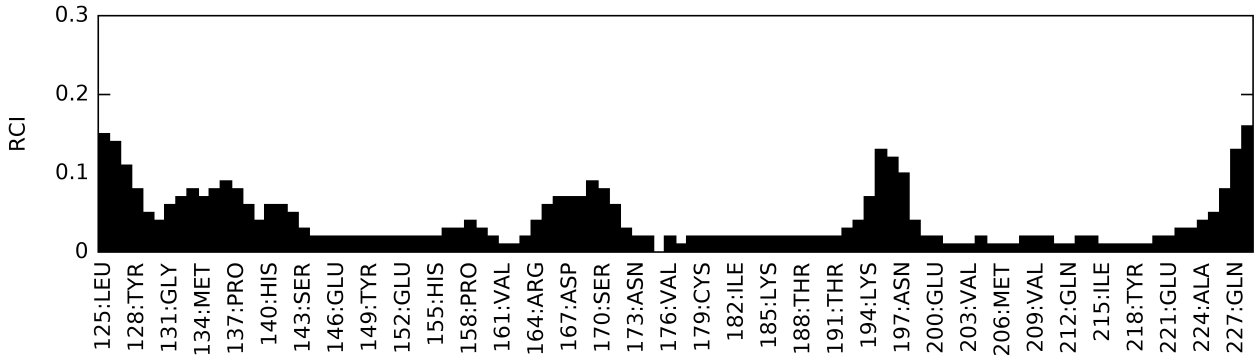
	Total	^{1}H	^{13}C	^{15}N
Backbone	399/514 (78%)	200/205 (98%)	102/208 (49%)	97/101 (96%)
Sidechain	571/688 (83%)	358/407 (88%)	191/240 (80%)	22/41 (54%)
Aromatic	93/143 (65%)	61/75 (81%)	32/64 (50%)	0/4 (0%)
Overall	1063/1345 (79%)	619/687 (90%)	325/512 (63%)	119/146 (82%)

6.1.4 Statistically unusual chemical shifts (i)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	128	TYR	HB2	23.28	4.76 – 1.06	55.1
1	A	128	TYR	HB3	23.28	4.75 – 0.95	53.8
1	A	168	GLU	HB2	3.32	3.08 – 0.98	6.1
1	A	168	GLU	HB3	3.24	3.10 – 0.90	5.6

6.1.5 Random Coil Index (RCI) plots (i)

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: