Full wwPDB NMR Structure Validation Report (i)

Mar 5, 2022 – 11:32 AM EST

PDB ID	:	2JV2
Title	:	Solution Structure of the N-terminal Domain of PH1500
Authors	:	Varnay, I.; Truffault, V.; Coles, M.
Deposited on	:	2007-09-11

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

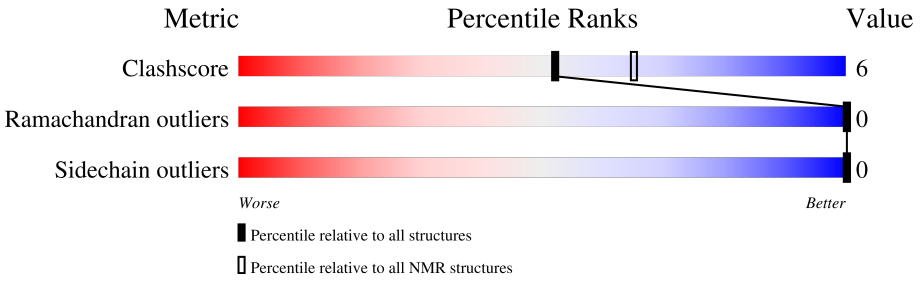
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	83	67% 14% 10% 8%

2 Ensemble composition and analysis (i)

This entry contains 22 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:13-A:80 (68)	0.18	1

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	1, 2, 4, 8, 10, 12, 19
2	6, 13, 15, 16, 17, 22
3	5, 14, 18
4	3, 11
Single-model clusters	7; 9; 20; 21

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1254 atoms, of which 657 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative uncharacterized protein PH1500.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	76	1254	389	657	101	106	1	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	HIS	-	expression tag	UNP O59169
A	2	HIS	-	expression tag	UNP O59169
A	3	HIS	-	expression tag	UNP O59169
A	4	HIS	-	expression tag	UNP O59169
A	5	HIS	-	expression tag	UNP O59169
A	6	HIS	-	expression tag	UNP O59169

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: Putative uncharacterized protein PH1500

Chain A: 67% 14% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L25 F29 I33 L37 I50 S51 I52 V57 K58 F59 A64 Y65 P66 S67 P68 L69 T81 H82 P83

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 (medoid)

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 70% 12% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L25 P26 P27 D28 F29 V30 I33 L37 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.2 Score per residue for model 2

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 70% 12% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 I33 L37 I50 S51 I52 V57 K58 F59 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.3 Score per residue for model 3

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 71% 11% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L20 I33 L37 I48 K58 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.4 Score per residue for model 4

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 67% 14% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L25 P26 P27 D28 F29 V30 I33 L37 I50 S51 I52 V57 K58 F59 P66 L69 T81 H82 P83

4.2.5 Score per residue for model 5

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 73% 8% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L25 F29 I33 L37 I48 P66 L69 T81 H82 P83

4.2.6 Score per residue for model 6

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 72% 10% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L25 F29 I33 L37 I50 F59 P66 L69 T81 H82 P83

4.2.7 Score per residue for model 7

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 69% 13% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L25 P26 P27 D28 F29 V30 I33 L37 R43 I50 F59 P66 L69 T81 H82 P83

4.2.8 Score per residue for model 8

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 70% 12% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 I33 L37 I50 S51 I52 V57 K58 F59 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.9 Score per residue for model 9

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 71% 11% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L25 P26 P27 D28 F29 V30 I33 L37 L53 P66 L69 T81 H82 P83

4.2.10 Score per residue for model 10

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 65% 17% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 S13 L20 L25 F29 I33 L37 I50 S51 I52 V57 K58 F59 P66 L69 T75 T81 H82 P83

4.2.11 Score per residue for model 11

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 61% 20% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L15 K16 L17 L20 L25 F29 I33 L37 V42 I50 S51 I52 V57 K58 F59 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.12 Score per residue for model 12

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 63% 19% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 S13 L20 L25 F29 I33 L37 I50 S51 I52 V57 K58 F59 A64 Y65 P66 S67 P68 L69 T75 T81 H82 P83

4.2.13 Score per residue for model 13

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 63% 19% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L20 L25 P26 P27 D28 F29 V30 I33 L37 I50 S51 I52 V57 K58 F59 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.14 Score per residue for model 14

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 71% 11% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 S13 L25 F29 I33 L37 I48 P66 L69 T75 T81 H82 P83

4.2.15 Score per residue for model 15

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 67% 14% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L25 F29 I33 L37 I50 S51 I52 V57 K58 F59 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.16 Score per residue for model 16

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 63% 19% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L15 L25 F29 I33 L37 V42 R43 T44 I50 S51 I52 V57 K58 F59 V62 Q63 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.17 Score per residue for model 17

• Molecule 1: Putative uncharacterized protein PH1500

Chain A: 63% 19% 10% 8%

HIS HIS HIS HIS HIS HIS MET E8 G9 V10 I11 M12 L15 L20 L25 F29 I33 L37 V42 I50 S51 I52 V57 K58 F59 A64 Y65 P66 S67 P68 L69 T81 H82 P83

4.2.18 Score per residue for model 18

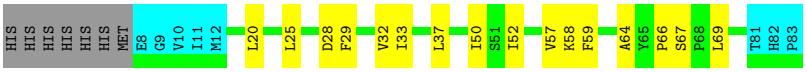
- Molecule 1: Putative uncharacterized protein PH1500

Chain A: 69% 13% 10% 8%

4.2.19 Score per residue for model 19

- Molecule 1: Putative uncharacterized protein PH1500

Chain A: 63% 19% 10% 8%



4.2.20 Score per residue for model 20

- Molecule 1: Putative uncharacterized protein PH1500

Chain A: 72% 10% 10% 8%

4.2.21 Score per residue for model 21

- Molecule 1: Putative uncharacterized protein PH1500

Chain A: 71% 11% 10% 8%

4.2.22 Score per residue for model 22

- Molecule 1: Putative uncharacterized protein PH1500

Chain A: 69% 13% 10% 8%

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 50 calculated structures, 22 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
X-PLOR NIH	structure solution	2.9.4A
X-PLOR NIH	refinement	2.9.4A

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	1.14±0.00	0±0/545 (0.0± 0.0%)	1.17±0.00	0±0/738 (0.0± 0.0%)
All	All	1.14	0/11990 (0.0%)	1.17	1/16236 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	43	ARG	NE-CZ-NH1	5.03	122.81	120.30	7	1

There are no chirality outliers.

There are no planarity outliers.

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	536	598	598	6±2
All	All	11792	13156	13156	143

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 6.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:66:PRO:O	1:A:69:LEU:HD23	0.62	1.94	10	22

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:33:ILE:HG22	1:A:37:LEU:HD12	0.59	1.75	17	21
1:A:53:LEU:O	1:A:53:LEU:HD23	0.56	2.00	9	2
1:A:15:LEU:HD23	1:A:17:LEU:HD11	0.55	1.77	11	1
1:A:64:ALA:HB3	1:A:67:SER:OG	0.53	2.03	1	12
1:A:25:LEU:HD22	1:A:29:PHE:CE2	0.52	2.40	17	17
1:A:15:LEU:HD13	1:A:42:VAL:HG21	0.52	1.81	21	5
1:A:52:ILE:HD12	1:A:57:VAL:HG11	0.51	1.83	12	12
1:A:50:ILE:HD12	1:A:59:PHE:CD2	0.50	2.41	22	15
1:A:25:LEU:HD22	1:A:29:PHE:CD2	0.47	2.44	20	6
1:A:77:ILE:HD12	1:A:77:ILE:N	0.47	2.24	22	1
1:A:13:SER:OG	1:A:75:THR:HG23	0.46	2.11	12	3
1:A:33:ILE:CG2	1:A:37:LEU:HD12	0.45	2.43	22	5
1:A:28:ASP:O	1:A:32:VAL:HG23	0.44	2.12	19	1
1:A:48:ILE:HD12	1:A:48:ILE:O	0.43	2.13	14	3
1:A:20:LEU:HD12	1:A:58:LYS:CB	0.43	2.43	3	4
1:A:27:PRO:O	1:A:30:VAL:HG12	0.42	2.13	4	6
1:A:20:LEU:HD12	1:A:58:LYS:HB3	0.42	1.90	11	4
1:A:44:THR:HG23	1:A:62:VAL:O	0.42	2.14	16	1
1:A:48:ILE:O	1:A:48:ILE:HD12	0.41	2.16	3	1
1:A:50:ILE:CG2	1:A:52:ILE:HD11	0.40	2.46	21	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	68/83 (82%)	67±0 (99±1%)	1±0 (1±1%)	0±0 (0±0%)	100	100
All	All	1496/1826 (82%)	1482 (99%)	14 (1%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	63/77 (82%)	63±0 (100±0%)	0±0 (0±0%)	100	100
All	All	1386/1694 (82%)	1386 (100%)	0 (0%)	100	100

There are no protein residues with a non-rotameric sidechain to report.

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