



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 26, 2016 – 08:42 PM BST

PDB ID : 2GMO
Title : NMR-structure of an independently folded C-terminal domain of influenza polymerase subunit PB2
Authors : Boudet, J.; Tarendeau, F.; Guilligay, D.; Mas, P.; Bougault, C.M.; Cusack, S.; Simorre, J.-P.; Hart, D.J.
Deposited on : 2006-04-07

This is a wwPDB NMR Structure Validation Summary 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

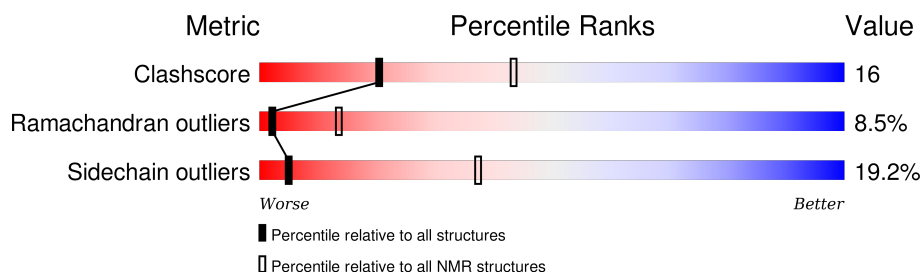
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 90%.

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	83	 55% 28% 6% • 10%

2 Ensemble composition and analysis

This entry contains 10 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:686-A:759 (74)	0.52	8

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 2 clusters and 4 single-model clusters were found.

Cluster number	Models
1	1, 2, 8, 10
2	3, 9
Single-model clusters	4; 5; 6; 7

3 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 1186 atoms, of which 613 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms						Trace
1	A	75	Total	C	H	N	O	S	0
			1186	355	613	108	108	2	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	677	GLY	-	CLONING ARTIFACT	UNP P31345

5 Refinement protocol and experimental data overview

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

Of the 1000 calculated structures, 10 were deposited, based on the following criterion: *10 structures of lowest energy*.

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

Software name	Classification	Version
CNS	structure solution	cns_solve_1.1
CNS	refinement	cns_solve_1.1

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 7056
Number of chemical shift lists	1
Total number of shifts	2043
Number of shifts mapped to atoms	1022
Number of unparsed shifts	1021
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

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

6 Model quality [i](#)

6.1 Standard geometry [i](#)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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	569	610	610	19±3
All	All	5690	6100	6100	187

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 16.

5 of 75 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:733:LEU:H	1:A:733:LEU:HD23	0.66	1.50	5	5
1:A:733:LEU:HD23	1:A:733:LEU:H	0.66	1.50	9	5
1:A:700:GLU:O	1:A:733:LEU:HD21	0.64	1.91	4	1
1:A:737:ARG:HA	1:A:744:LEU:HD11	0.62	1.69	5	4
1:A:733:LEU:HD23	1:A:733:LEU:N	0.61	2.10	4	2

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	73/83 (88%)	56±2 (76±3%)	11±2 (15±3%)	6±2 (8±3%)	2	13
All	All	730/830 (88%)	556 (76%)	112 (15%)	62 (8%)	2	13

5 of 24 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	706	PRO	10
1	A	739	ARG	6
1	A	727	GLY	4
1	A	692	ARG	4
1	A	707	ALA	4

6.3.2 Protein sidechains ⓘ

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	62/69 (90%)	50±2 (81±4%)	12±2 (19±4%)	5	37
All	All	620/690 (90%)	501 (81%)	119 (19%)	5	37

5 of 26 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	733	LEU	10
1	A	734	VAL	10
1	A	702	ARG	10
1	A	708	LEU	10
1	A	735	MET	10

6.3.3 RNA ⓘ

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 carbohydrates 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

The completeness of assignment taking into account all chemical shift lists is 90% for the well-defined parts and 90% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 7056

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	2043
Number of shifts mapped to atoms	1022
Number of unparsed shifts	1021
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following errors were found when reading this chemical shift list.

- Chemical shift has been reported more than once. First 5 (of 1021) occurrences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1023	A	677	GLY	HA2	3.695	0.01	2
1024	A	677	GLY	HA3	3.562	0.01	2
1025	A	677	GLY	CA	43.378	0.15	1
1026	A	678	ASP	HB2	2.623	0.01	2
1027	A	678	ASP	HB3	2.483	0.01	2

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	83	-0.29 ± 0.13	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	74	0.03 ± 0.22	None needed (< 0.5 ppm)
$^{13}\text{C}'$	81	-0.20 ± 0.15	None needed (< 0.5 ppm)

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction \pm precision, ppm	Suggested action
^{15}N	77	-0.49 \pm 0.29	None needed (< 0.5 ppm)

7.1.3 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 90%, i.e. 833 atoms were assigned a chemical shift out of a possible 925. 15 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	364/368 (99%)	145/147 (99%)	148/148 (100%)	71/73 (97%)
Sidechain	454/540 (84%)	274/314 (87%)	174/193 (90%)	6/33 (18%)
Aromatic	15/17 (88%)	9/9 (100%)	6/8 (75%)	0/0 (—%)
Overall	833/925 (90%)	428/470 (91%)	328/349 (94%)	77/106 (73%)

7.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

7.1.5 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:

