wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 02:48 AM BST

PDB ID	:	2MDJ
Title	:	Solution structure of WW domain with polyproline stretch (PP2WW) of HYPB
Authors	:	Gao, Y.; Hu, H.
Deposited on	:	2013-09-11

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 (i) symbol.

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

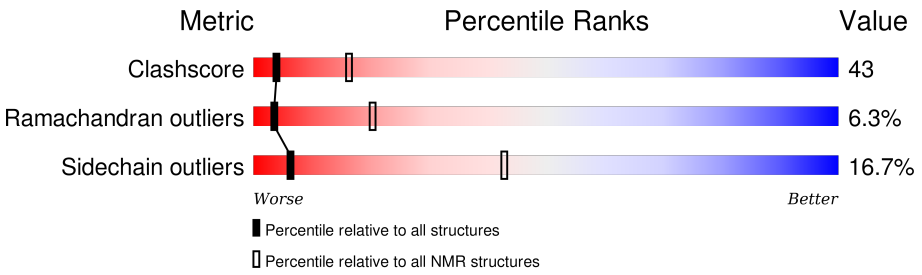
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 54%.

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	56	32% 32% 13% 23%

2 Ensemble composition and analysis (i)

This entry contains 15 models. Model 14 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:1-A:43 (43)	0.60	14

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 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 4, 5, 7, 9, 10, 11, 12, 13, 14
2	3, 8, 15
Single-model clusters	6

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 880 atoms, of which 430 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone-lysine N-methyltransferase SETD2.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	56	880	290	430	74	86	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP Q9BYW2
A	2	SER	-	EXPRESSION TAG	UNP Q9BYW2

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

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: Histone-lysine N-methyltransferase SETD2

Chain A: 32% 32% 13% 23%

G1 L4 K11 P12 K13 T14 I15 V16 L17 P18 P19 N20 W21 K22 T23 A24 R25 D26 P27 K30 I31 Y32 Y33 Y34 H35 V36 I37 Q40 T41 Q42 W43 D44 P45 P46 T47 W48 E49 S50 P51 G52 D53 D54 A55 S56

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 14. Colouring as in section 4.1 above.

- Molecule 1: Histone-lysine N-methyltransferase SETD2

Chain A: 43% 23% 11% 23%

G1 L4 K11 P12 K13 V16 L17 P18 P19 N20 W21 K22 T23 A24 R25 D26 P27 Y33 Y34 H35 V36 Q40 T41 Q42 W43 D44 P45 P46 T47 W48 E49 S50 P51 G52 D53 D54 A55 S56

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 15 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
CNS	structure solution	
ARIA	structure solution	
Molmol	structure solution	
ARIA	refinement	

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)	2mdj_cs.str
Number of chemical shift lists	1
Total number of shifts	441
Number of shifts mapped to atoms	441
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	54%

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)

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	0.61±0.16	1±1/370 (0.2±0.3%)	0.69±0.04	0±0/512 (0.0±0.0%)
All	All	0.63	10/5550 (0.2%)	0.69	0/7680 (0.0%)

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	0.1±0.2
All	All	0	1

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	34	TYR	CE2-CZ	-12.09	1.22	1.38	10	4
1	A	34	TYR	CE1-CZ	11.26	1.53	1.38	10	4
1	A	33	TYR	CE2-CZ	-5.56	1.31	1.38	10	1
1	A	33	TYR	CE1-CZ	5.27	1.45	1.38	10	1

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	34	TYR	Sidechain	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	353	356	355	31±6
All	All	5295	5340	5325	458

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 43.

5 of 227 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:21:TRP:HB3	1:A:35:HIS:HA	1.05	1.28	12	15
1:A:25:ARG:HG3	1:A:31:ILE:HA	1.01	1.32	7	1
1:A:34:TYR:HB2	1:A:41:THR:HG22	1.00	1.27	3	1
1:A:4:LEU:HB2	1:A:36:VAL:HG12	0.93	1.36	5	4
1:A:20:ASN:HD22	1:A:36:VAL:HG23	0.89	1.27	7	9

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	42/56 (75%)	34±1 (81±2%)	5±1 (13±3%)	3±1 (6±2%)	3 20
All	All	630/840 (75%)	508 (81%)	82 (13%)	40 (6%)	3 20

5 of 6 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	18	PRO	11
1	A	16	VAL	10
1	A	12	PRO	8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	15	ILE	7
1	A	11	LYS	2

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	40/51 (78%)	33±2 (83±4%)	7±2 (17±4%)	6 43
All	All	600/765 (78%)	500 (83%)	100 (17%)	6 43

5 of 18 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	42	GLN	15
1	A	36	VAL	14
1	A	20	ASN	12
1	A	21	TRP	10
1	A	34	TYR	9

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 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 (i)

The completeness of assignment taking into account all chemical shift lists is 54% for the well-defined parts and 57% for the entire structure.

7.1 Chemical shift list 1

File name: 2mdj_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 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	441
Number of shifts mapped to atoms	441
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	45	0.57 ± 0.41	None needed (imprecise)
$^{13}C_{\beta}$	43	-0.15 ± 0.28	None needed (< 0.5 ppm)
$^{13}C'$	38	0.59 ± 0.29	Should be applied
^{15}N	40	0.72 ± 0.75	None needed (imprecise)

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 54%, i.e. 305 atoms were assigned a chemical shift out of a possible 563. 1 out of 4 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	153/197 (78%)	61/77 (79%)	62/86 (72%)	30/34 (88%)
Sidechain	146/310 (47%)	91/188 (48%)	52/109 (48%)	3/13 (23%)

Continued on next page...

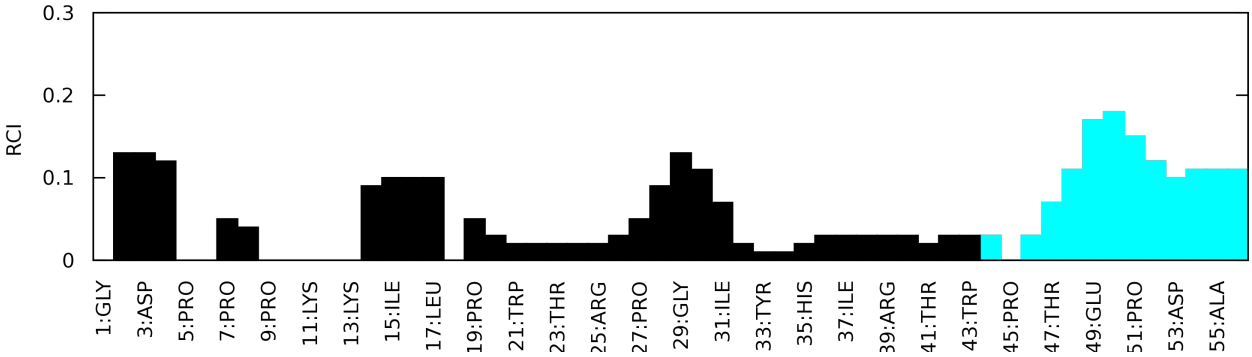
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	6/56 (11%)	4/28 (14%)	0/24 (0%)	2/4 (50%)
Overall	305/563 (54%)	156/293 (53%)	114/219 (52%)	35/51 (69%)

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: