wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 03:21 AM EDT

PDB ID	:	2LAW
BMRB ID	:	17538
Title	:	Structure of the second WW domain from human YAP in complex with a human Smad1 derived peptide
Authors	:	Macias, M.J.; Aragon, E.; Goerner, N.; Zaromytidou, A.; Xi, Q.; Escobedo, A.; Massague, J.
Deposited on	:	2011-03-22

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
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:

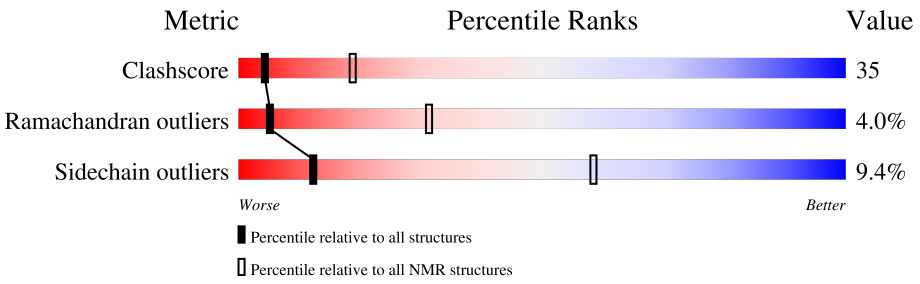
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 32%.

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	38	45% 26% 16% • 11%
2	B	12	42% 58%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 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:231-A:263, B:222-B:233 (45)	0.56	19

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 6 clusters and 4 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 721 atoms, of which 348 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Yorkie homolog.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	34	543	178	262	47	55	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	226	GLY	-	expression tag	UNP P46937
A	227	ALA	-	expression tag	UNP P46937
A	228	MET	-	expression tag	UNP P46937
A	229	GLU	-	expression tag	UNP P46937

- Molecule 2 is a protein called Mothers against decapentaplegic homolog 1.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	12	178	61	86	12	19	0

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: Yorkie homolog

Chain A: 45% 26% 16% • 11%

GLY ALA MET GLU G230 P231 L232 P233 W236 E237 Q238 I246 Y247 Y248 I249 N250 N253 T256 S257 W258 L259 D260 P261 R262 L263

• Molecule 2: Mothers against decapentaplegic homolog 1

Chain B: 42% 58%

T222 P223 P224 P225 L228 P229 P230 E231 D232 P233

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

The representative model is number 19. Colouring as in section 4.1 above.

• Molecule 1: Yorkie homolog

Chain A: 47% 29% 11% • 11%

GLY ALA MET GLU G230 P231 L232 W236 E237 Q238 Y247 Y248 I249 N250 N253 R254 T255 T256 S257 W258 L259 D260 P261 R262 L263

• Molecule 2: Mothers against decapentaplegic homolog 1

Chain B: 33% 58% 8%

T222 P223 P224 P225 A226 Y227 L228 P229 P230 E231 D232 P233

5 Refinement protocol and experimental data overview (i)

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

Of the 300 calculated structures, 20 were deposited, based on the following criterion: *structures with acceptable covalent geometry*.

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

Software name	Classification	Version
CNS	structure solution	1.3
CNS	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	194
Number of shifts mapped to atoms	194
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	32%

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	0.63±0.01	1±0/286 (0.3± 0.0%)	0.77±0.01	1±0/388 (0.3± 0.0%)
2	B	0.45±0.02	0±0/98 (0.0± 0.0%)	0.62±0.03	0±0/138 (0.0± 0.0%)
All	All	0.59	20/7680 (0.3%)	0.73	20/10520 (0.2%)

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	1.0±0.0
2	B	0.0±0.0	1.0±0.0
All	All	0	40

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	261	PRO	C-N	-5.76	1.20	1.34	5	20

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	261	PRO	CA-C-N	-5.79	104.47	117.20	14	20

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	236	TRP	Peptide	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
2	B	229	PRO	Peptide	20

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	277	259	257	24±3
2	B	92	86	83	6±1
All	All	7380	6900	6800	503

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 35.

5 of 93 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:236:TRP:HB3	1:A:250:ASN:HD22	0.96	1.20	5	2
1:A:250:ASN:HB3	1:A:253:ASN:HB2	0.92	1.39	4	20
1:A:236:TRP:CZ3	1:A:248:TYR:HB3	0.86	2.05	4	20
1:A:260:ASP:HB2	1:A:262:ARG:HG2	0.81	1.52	15	8
1:A:238:GLN:HA	1:A:247:TYR:O	0.78	1.78	2	18

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	32/38 (84%)	30±1 (92±3%)	2±1 (5±3%)	1±0 (3±0%)	7 39
2	B	10/12 (83%)	7±0 (74±5%)	2±1 (18±7%)	1±0 (7±5%)	2 17
All	All	840/1000 (84%)	739 (88%)	67 (8%)	34 (4%)	5 31

All 3 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	231	PRO	20
2	B	232	ASP	10
2	B	226	ALA	4

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	30/32 (94%)	26±1 (88±3%)	4±1 (12±3%)	8 51
2	B	11/11 (100%)	11±0 (97±4%)	0±0 (3±4%)	48 90
All	All	820/860 (95%)	743 (91%)	77 (9%)	12 58

5 of 14 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	232	LEU	20
1	A	238	GLN	12
1	A	247	TYR	12
1	A	262	ARG	11
1	A	255	THR	6

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)

The completeness of assignment taking into account all chemical shift lists is 32% for the well-defined parts and 32% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

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	194
Number of shifts mapped to atoms	194
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

7.1.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

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 32%, i.e. 192 atoms were assigned a chemical shift out of a possible 605. 0 out of 4 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

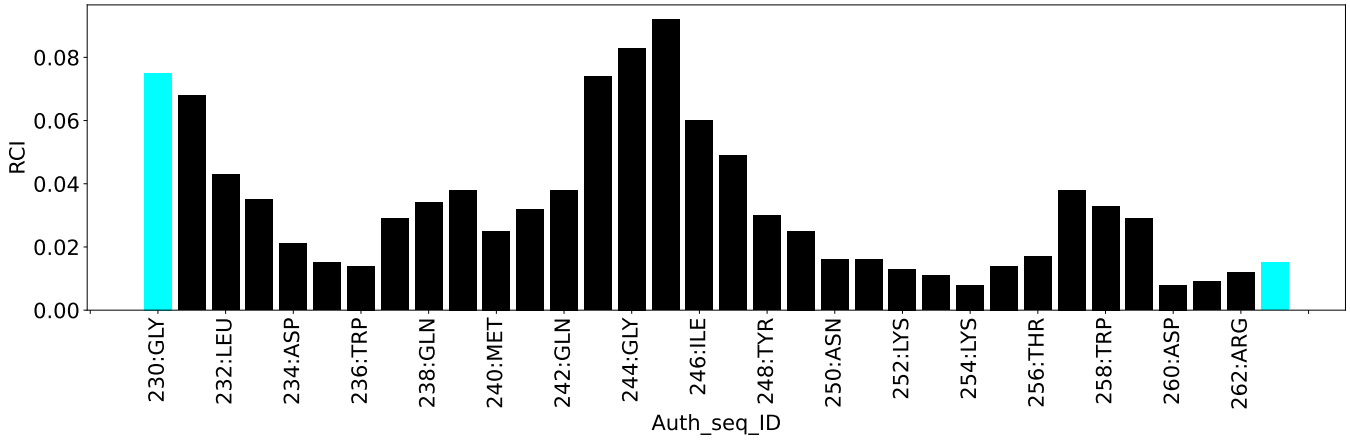
	Total	^{1}**H**	^{13}**C**	^{15}**N**
Backbone	63/209 (30%)	63/83 (76%)	0/90 (0%)	0/36 (0%)
Sidechain	111/337 (33%)	111/218 (51%)	0/110 (0%)	0/9 (0%)
Aromatic	18/59 (31%)	18/28 (64%)	0/27 (0%)	0/4 (0%)
Overall	192/605 (32%)	192/329 (58%)	0/227 (0%)	0/49 (0%)

7.1.4 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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	241	THR	HG1	6.78	0.08 – 2.19	26.7
1	A	250	ASN	HB3	-0.46	1.12 – 4.38	-9.8
1	A	260	ASP	HA	2.74	3.04 – 6.12	-6.0
1	A	261	PRO	HG3	0.10	0.33 – 3.48	-5.7
1	A	254	LYS	HD2	2.73	0.58 – 2.64	5.4
1	A	237	GLU	HB3	0.86	0.95 – 3.05	-5.4
1	A	261	PRO	HG2	0.34	0.41 – 3.45	-5.2

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: