W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB NMR Structure Validation Summary Report ⓘ

May 29, 2020 – 09:01 am BST

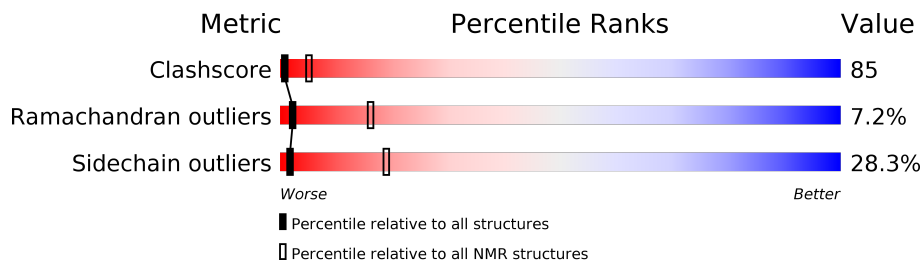|		
PDB ID	:	5YDX
Title	:	NMR structure of YAP1-2 WW1 domain with LATS1 PPxY motif complex
Authors	:	Fan, J.S.; Sivaraman, J.
Deposited on	:	2017-09-15

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 82%.

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	Percentile Ranks	Value
Clashscore		85
Ramachandran outliers		7.2%
Sidechain outliers		28.3%

Worse   *Better*

■ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	59	7% 32% 8% • 51%

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:16-A:25, A:29-A:42, A:58-A:62 (29)	0.16	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 3 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 898 atoms, of which 433 are hydrogens and 0 are deuteriums.

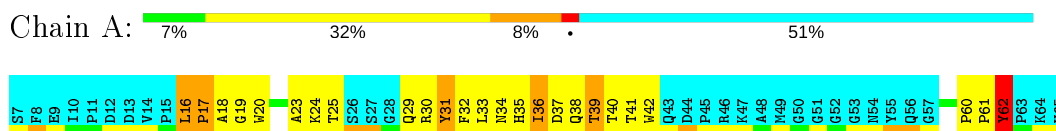- Molecule 1 is a protein called WW domain with PPxY motif.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	59	898	295	433	80	88	2	0

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: WW domain with PPxY motif

Chain A: 7%   32%   8%   •   51%

S7 F8 E9 I10 P11 D12 D13 V14 P15 L16 P17 A18 G19 W20 A23 K24 T25 S26 S27 G28 Q29 R30 Y31 F32 L33 N34 H35 I36 D37 Q38 T39 T40 T41 W42 Q43 D44 P45 R46 K47 A48 M49 G50 G51 G52 G53 N54 Y55 Q56 G57 P60 P61 Y62 P63 K64 H65

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: WW domain with PPxY motif

Chain A: 10%   27%   10%   •   51%

S7 F8 E9 I10 P11 D12 D13 V14 P15 L16 P17 A18 G19 W20 A23 K24 T25 S26 S27 G28 Q29 R30 Y31 F32 L33 N34 H35 I36 D37 Q38 T39 T40 T41 W42 Q43 D44 P45 R46 K47 A48 M49 G50 G51 G52 G53 N54 Y55 Q56 G57 P61 Y62 P63 K64 H65
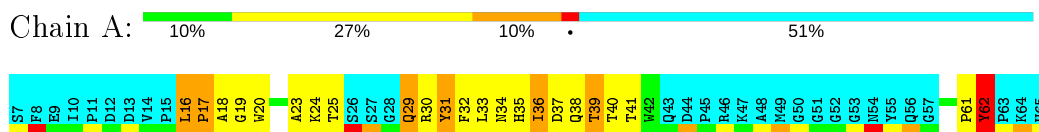
5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 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
X-PLOR NIH	structure calculation	

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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	606
Number of shifts mapped to atoms	582
Number of unparsed shifts	0
Number of shifts with mapping errors	24
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

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	245	231	230	40±4
All	All	4900	4620	4600	809

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 85.

5 of 135 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:16:LEU:HD23	1:A:16:LEU:N	1.05	1.65	12	9
1:A:16:LEU:N	1:A:16:LEU:HD23	1.01	1.67	1	11
1:A:37:ASP:O	1:A:39:THR:HG22	0.98	1.57	4	19
1:A:16:LEU:H	1:A:16:LEU:HD23	0.89	1.25	10	4
1:A:16:LEU:N	1:A:16:LEU:CD2	0.83	2.41	4	17

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	29/59 (49%)	25±1 (84±4%)	2±1 (8±4%)	2±0 (7±2%)	2	16
All	All	580/1180 (49%)	490 (84%)	48 (8%)	42 (7%)	2	16

All 4 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	17	PRO	20
1	A	62	TYR	19
1	A	60	PRO	2
1	A	16	LEU	1

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	26/49 (53%)	19±1 (72±4%)	7±1 (28±4%)	2	19
All	All	520/980 (53%)	373 (72%)	147 (28%)	2	19

5 of 13 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	16	LEU	20
1	A	39	THR	20
1	A	41	THR	20
1	A	31	TYR	19
1	A	36	ILE	18

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 82% for the well-defined parts and 79% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *ww1.str*

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	606
Number of shifts mapped to atoms	582
Number of unparsed shifts	0
Number of shifts with mapping errors	24
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

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

- Residue not found in structure. First 5 (of 24) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	3	LEU	CD1	23.551	0.01	2
A	2	PRO	HB3	1.897	0.0	2
A	3	LEU	HD21	0.884	0.01	2
A	2	PRO	CB	32.296	0.01	1
A	2	PRO	HG3	1.959	0.0	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	41	0.84 ± 0.38	Should be applied
$^{13}C_\beta$	51	-0.05 ± 0.33	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{15}N	49	-0.27 ± 0.27	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 82%, i.e. 312 atoms were assigned a chemical shift out of a possible 379. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	100/135 (74%)	53/53 (100%)	23/58 (40%)	24/24 (100%)
Sidechain	164/187 (88%)	104/113 (92%)	57/67 (85%)	3/7 (43%)
Aromatic	48/57 (84%)	26/29 (90%)	20/24 (83%)	2/4 (50%)
Overall	312/379 (82%)	183/195 (94%)	100/149 (67%)	29/35 (83%)

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.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	46	ARG	NE	107.94	92.63 − 76.73	14.6
1	A	34	ASN	HB2	0.36	4.36 − 1.26	-7.9
1	A	44	ASP	HA	2.56	6.15 − 3.05	-6.6
1	A	60	PRO	HD3	1.18	5.52 − 1.72	-6.4
1	A	59	PRO	HB3	-0.10	3.81 − 0.21	-5.9
1	A	63	PRO	HD2	1.77	5.45 − 1.85	-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.

Random coil index (RCI) for chain A: