wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 11:53 pm BST

PDB ID	:	2MPT
Title	:	WW3 domain of Nedd4L in complex with its HECT domain PY motif
Authors	:	Escobedo, A.; Macias, M.J.; Gomes, T.; Aragon, E.; Martin-Malpartida, P.; Ruiz, L.
Deposited on	:	2014-06-02

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 following versions of software and data (see references (1)) were used in the production of this report:

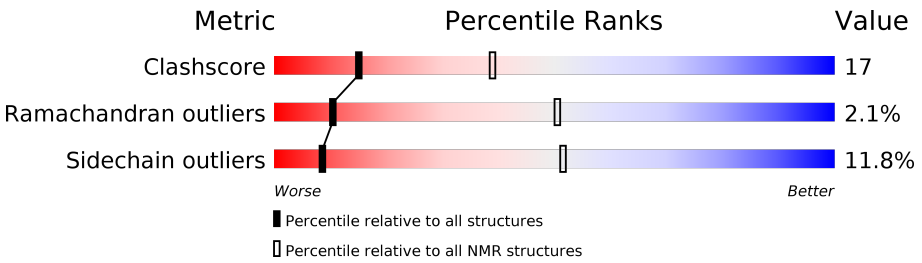
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 53%.

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	48	31% 29% 6% 33%
2	B	14	50% 21% 29%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 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:479-A:507, B:928-B:937 (39)	0.51	15

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 3 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 749 atoms, of which 364 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called E3 ubiquitin-protein ligase NEDD4-like.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	32	526	176	255	48	46	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	472	GLY	-	EXPRESSION TAG	UNP Q96PU5
A	473	ALA	-	EXPRESSION TAG	UNP Q96PU5
A	474	MET	-	EXPRESSION TAG	UNP Q96PU5
A	475	GLU	-	EXPRESSION TAG	UNP Q96PU5

- Molecule 2 is a protein called E3 ubiquitin-protein ligase NEDD4-like.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	14	223	74	109	17	23	1

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	938	NH2	-	AMIDATION	UNP Q96PU5

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: E3 ubiquitin-protein ligase NEDD4-like

Chain A: 31% 29% 6% 33%

- Molecule 2: E3 ubiquitin-protein ligase NEDD4-like

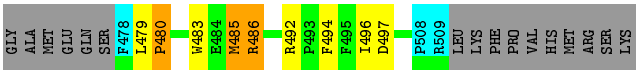
Chain B: 50% 21% 29%

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

The representative model is number 15. Colouring as in section 4.1 above.

- Molecule 1: E3 ubiquitin-protein ligase NEDD4-like

Chain A: 42% 13% 6% 6% 33%



- Molecule 2: E3 ubiquitin-protein ligase NEDD4-like

Chain B: 36% 36% 29%

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing.*

Of the 300 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
CNSSOLVE	refinement	
CNSSOLVE	structure solution	
TALOS	geometry optimization	

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	3
Total number of shifts	366
Number of shifts mapped to atoms	366
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	53%

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)

Bond lengths and bond angles in the following residue types are not validated in this section: NH2

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	242	226	225	9±3
2	B	86	77	77	4±2
All	All	6560	6060	6040	210

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 17.

5 of 106 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:488:ALA:HB3	1:A:492:ARG:HB3	0.90	1.42	19	3
1:A:503:THR:HB	2:B:928:LEU:HD23	0.89	1.43	6	1
1:A:492:ARG:HG2	1:A:505:TRP:CZ3	0.88	2.03	16	4
1:A:484:GLU:HB3	2:B:937:LEU:HD22	0.85	1.47	19	1
2:B:928:LEU:HD12	2:B:929:PRO:HD2	0.84	1.47	6	4

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	29/48 (60%)	26±1 (89±5%)	3±1 (9±5%)	1±1 (2±2%)	10	50
2	B	10/14 (71%)	9±1 (92±7%)	1±1 (6±6%)	0±0 (2±4%)	11	52
All	All	780/1240 (63%)	701 (90%)	63 (8%)	16 (2%)	10	50

All 5 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	480	PRO	7
1	A	489	PRO	3
2	B	935	GLU	3
1	A	481	PRO	2
2	B	937	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	26/43 (60%)	23±2 (87±8%)	4±2 (13±8%)	7	47
2	B	10/13 (77%)	9±1 (92±7%)	1±1 (8±7%)	17	65
All	All	720/1120 (64%)	635 (88%)	85 (12%)	8	51

5 of 22 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	497	ASP	15
1	A	484	GLU	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	492	ARG	7
1	A	500	THR	6
2	B	937	LEU	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 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 53% for the well-defined parts and 50% for the entire structure.

7.1 Chemical shift list 1

File name: input_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	195
Number of shifts mapped to atoms	195
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

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 29%, i.e. 148 atoms were assigned a chemical shift out of a possible 503. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}**H**	^{13}**C**	^{15}**N**
Backbone	53/183 (29%)	53/72 (74%)	0/78 (0%)	0/33 (0%)
Sidechain	77/253 (30%)	77/151 (51%)	0/93 (0%)	0/9 (0%)
Aromatic	18/67 (27%)	18/35 (51%)	0/28 (0%)	0/4 (0%)
Overall	148/503 (29%)	148/258 (57%)	0/199 (0%)	0/46 (0%)

7.1.4 Statistically unusual chemical shifts (i)

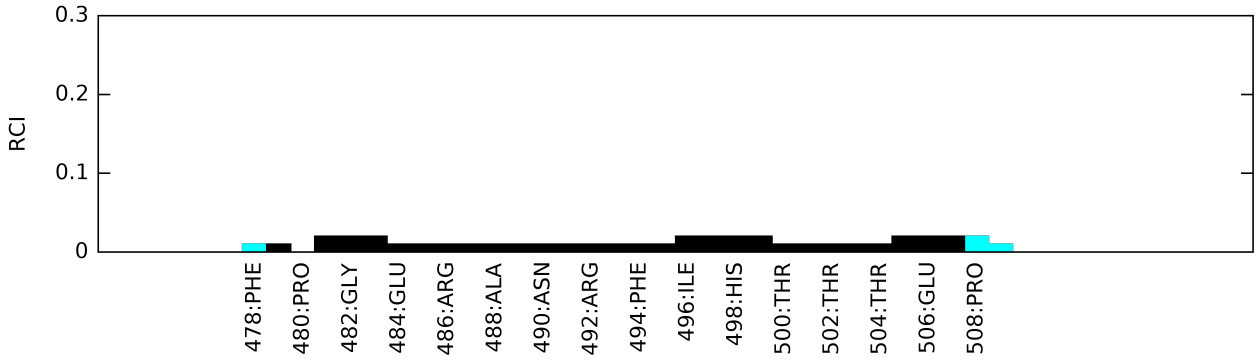
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	497	ASP	HB3	-0.04	4.07 – 1.27	-9.7
1	A	507	ASP	HA	2.52	6.15 – 3.05	-6.7
1	A	501	LYS	HE2	1.83	3.87 – 1.97	-5.7
1	A	508	PRO	HG3	0.16	3.56 – 0.26	-5.3

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:



7.2 Chemical shift list 2

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_2*

7.2.1 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	58

Number of shifts mapped to atoms	58
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

7.2.2 Chemical shift referencing (i)

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

7.2.3 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 8%, i.e. 38 atoms were assigned a chemical shift out of a possible 503. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	13/183 (7%)	13/72 (18%)	0/78 (0%)	0/33 (0%)
Sidechain	20/253 (8%)	20/151 (13%)	0/93 (0%)	0/9 (0%)
Aromatic	5/67 (7%)	5/35 (14%)	0/28 (0%)	0/4 (0%)
Overall	38/503 (8%)	38/258 (15%)	0/199 (0%)	0/46 (0%)

7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.2.5 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 B:

7.3 Chemical shift list 3

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_3*

7.3.1 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	113
Number of shifts mapped to atoms	113
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

7.3.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	31	2.49 ± 0.42	Should be applied
$^{13}C_{\beta}$	28	3.01 ± 0.51	Should be applied
$^{13}C'$	0	—	None (insufficient data)
^{15}N	27	0.49 ± 0.79	None needed (< 0.5 ppm)

7.3.3 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 21%, i.e. 104 atoms were assigned a chemical shift out of a possible 503. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned

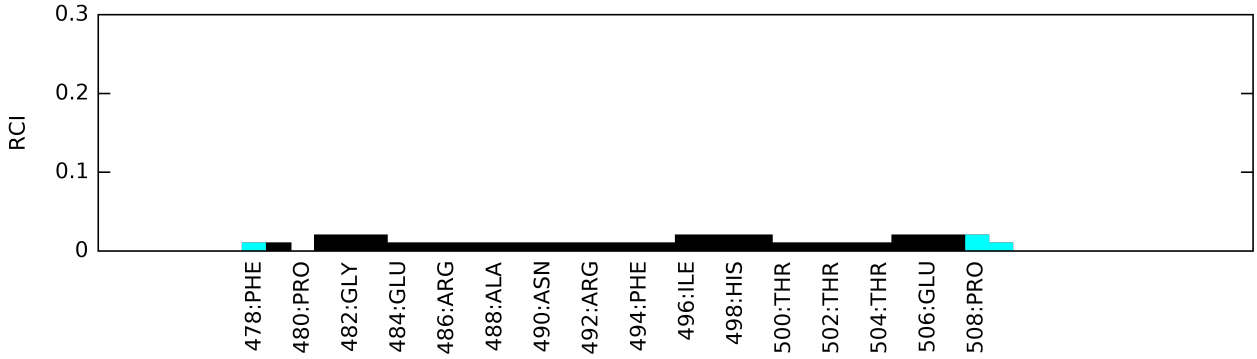
stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	78/183 (43%)	25/72 (35%)	28/78 (36%)	25/33 (76%)
Sidechain	26/253 (10%)	0/151 (0%)	26/93 (28%)	0/9 (0%)
Aromatic	0/67 (0%)	0/35 (0%)	0/28 (0%)	0/4 (0%)
Overall	104/503 (21%)	25/258 (10%)	54/199 (27%)	25/46 (54%)

7.3.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.3.5 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: