wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 06:18 PM BST

PDB ID	:	1YDU
Title	:	Solution NMR structure of At5g01610, an Arabidopsis thaliana protein containing DUF538 domain
Authors	:	Zhao, Q.; Cornilescu, C.C.; Lee, M.S.; Markley, J.L.; Center for Eukaryotic Structural Genomics (CESG)
Deposited on	:	2004-12-26

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:

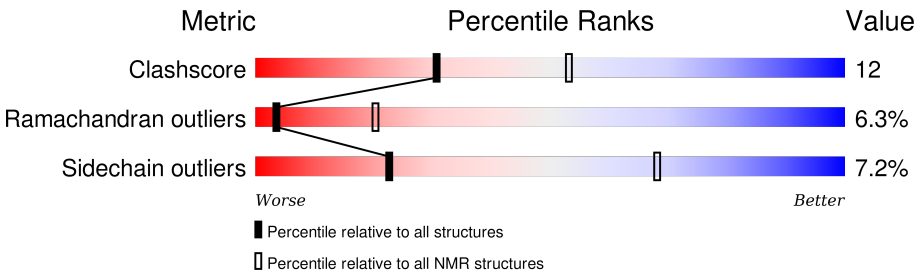
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 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	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	170	48% 18% • 34%

2 Ensemble composition and analysis (i)

This entry contains 20 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:54-A:95, A:100-A:170 (113)	0.80	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 3 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2703 atoms, of which 1368 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called At5g01610.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	170	2703	853	1368	221	257	4	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	CLONING ARTIFACT	UNP Q9M015
A	159	GLY	GLU	ENGINEERED	UNP Q9M015

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: At5g01610

Chain A: 48% 18% • 34%

S1 D2 Q3 I4 F5 N6 K7 V8 G9 S10 Y11 W12 L13 G14 Q15 K16 A17 N18 K19 Q20 F21 D22 S23 V24 G25 N26 D27 L28 N29 S30 V31 S32 T33 S34 I35 E36 G37 G38 T39 K40 W41 L42 V43 N44 K45 I46 K47 G48 K49 M50 Q51 K52 P53 L54 P55 L58 K59 L63

P64 I65 G66 I67 F68 A72 T73 N74 D78 T81 K82 K83 V86 L87 I88 P89 V94 G95 Y96 K97 D98 S99 T109 K116 D119 K126 V127 T134 S141 R155 Y158 G159 V160 Q161 R162 R166 F170

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: At5g01610

Chain A: 48% 15% • • 34%

S1 D2 Q3 I4 F5 N6 K7 V8 G9 S10 Y11 W12 L13 G14 Q15 K16 A17 N18 K19 Q20 F21 D22 S23 V24 G25 N26 D27 L28 N29 S30 V31 S32 T33 S34 I35 E36 G37 G38 T39 K40 W41 L42 V43 N44 K45 I46 K47 G48 K49 M50 Q51 K52 P53 L54 P55 K59 D62 L63

P64 I65 G66 I67 F68 A72 T73 E76 F77 D78 T81 K82 K83 L87 Y96 K97 D98 S99 T109 G110 K116 D119 K126 V127 M128 K132 V133 T134 S141 G148 M149 K150 R155 Y158 G159 V160 V167 F170

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *water refinement with XPLOR*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CYANA	structure solution	1.0.5
X-PLOR	refinement	2.9.3

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 6443
Number of chemical shift lists	1
Total number of shifts	1354
Number of shifts mapped to atoms	1354
Number of unparsed shifts	0
Number of shifts with mapping errors	0
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)

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	0.99±0.05	0±0/902 (0.0±0.0%)	0.79±0.03	0±0/1215 (0.0±0.0%)
All	All	0.99	1/18040 (0.0%)	0.79	1/24300 (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	0.1±0.2	0.2±0.4
All	All	1	4

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	120	VAL	N-CA	-5.39	1.35	1.46	5	1

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	77	PHE	N-CA-CB	5.93	121.27	110.60	19	1

All unique chiral outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	A	73	THR	CA	1

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	166	ARG	Sidechain	3
1	A	155	ARG	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	887	919	918	21±5
All	All	17740	18380	18360	421

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 12.

5 of 210 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:116:LYS:HE2	1:A:116:LYS:N	0.64	2.07	9	1
1:A:79:GLU:O	1:A:82:LYS:HB2	0.64	1.91	5	1
1:A:78:ASP:OD1	1:A:81:THR:HB	0.64	1.92	10	1
1:A:95:GLY:HA2	1:A:100:SER:O	0.64	1.91	16	3
1:A:67:ILE:O	1:A:160:VAL:HB	0.63	1.91	7	8

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	112/170 (66%)	87±3 (78±3%)	18±4 (16±4%)	7±2 (6±2%)	3	20
All	All	2240/3400 (66%)	1746 (78%)	353 (16%)	141 (6%)	3	20

5 of 31 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	68	PHE	20
1	A	141	SER	18
1	A	158	TYR	10
1	A	119	ASP	9
1	A	89	PRO	8

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	100/150 (67%)	93±2 (93±2%)	7±2 (7±2%)	23	68
All	All	2000/3000 (67%)	1856 (93%)	144 (7%)	23	68

5 of 38 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	65	ILE	20
1	A	116	LYS	17
1	A	126	LYS	11
1	A	155	ARG	8
1	A	74	ASN	8

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 57% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 6443

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	1354
Number of shifts mapped to atoms	1354
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	118	-0.70 ± 0.24	Should be applied
$^{13}C_{\beta}$	109	-0.36 ± 0.19	None needed (< 0.5 ppm)
$^{13}C'$	104	0.30 ± 0.20	None needed (< 0.5 ppm)
^{15}N	113	0.26 ± 0.32	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. 1141 atoms were assigned a chemical shift out of a possible 1387. 15 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	544/557 (98%)	222/222 (100%)	213/226 (94%)	109/109 (100%)
Sidechain	597/733 (81%)	374/426 (88%)	221/279 (79%)	2/28 (7%)

Continued on next page...

Continued from previous page...

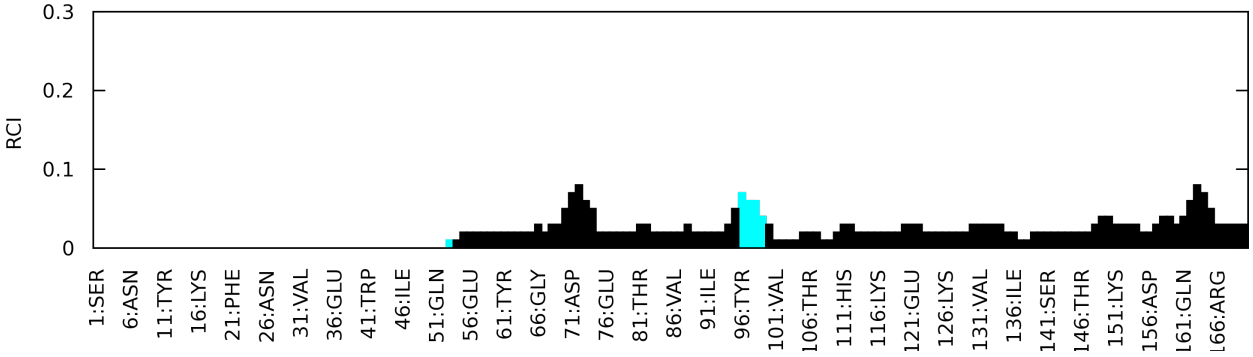
	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/97 (0%)	0/51 (0%)	0/43 (0%)	0/3 (0%)
Overall	1141/1387 (82%)	596/699 (85%)	434/548 (79%)	111/140 (79%)

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	153	ARG	NE	122.20	92.63 – 76.73	23.6
1	A	116	LYS	CE	35.60	46.00 – 37.80	-7.7
1	A	126	LYS	HB3	-0.22	3.10 – 0.40	-7.3
1	A	114	LYS	CE	36.80	46.00 – 37.80	-6.2
1	A	126	LYS	HG3	-0.25	2.76 – -0.04	-5.8
1	A	105	THR	HB	2.47	5.82 – 2.52	-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: