Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 12:58 AM BST

PDB ID	:	2LH9
Title	:	NMR structure of the first lotus domain of tudor domain-containing protein 7
Authors	:	Cui, G.; Botuyan, M.; Mer, G.
Deposited on	:	2011-08-06

This is a Full wwPDB NMR Structure Validation 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:

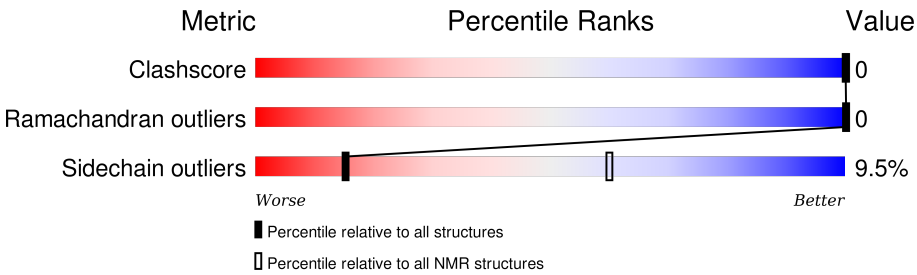
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 94%.

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	78	83% 5% • 10%

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:35-A:97, A:103-A:109 (70)	0.09	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 2 clusters and 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1240 atoms, of which 630 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Tudor domain-containing protein 7.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	78	1240	390	630	108	109	3	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP Q8K1H1
A	-1	HIS	-	EXPRESSION TAG	UNP Q8K1H1

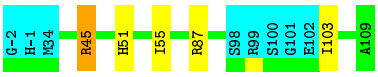
4 Residue-property plots (i)

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: Tudor domain-containing protein 7

Chain A:



4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Tudor domain-containing protein 7

Chain A:

4.2.2 Score per residue for model 2

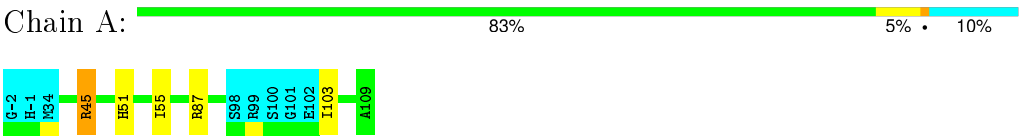
- Molecule 1: Tudor domain-containing protein 7

Chain A:

4.2.3 Score per residue for model 3

- Molecule 1: Tudor domain-containing protein 7

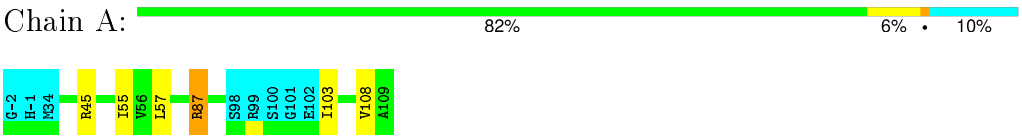
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Tudor domain-containing protein 7

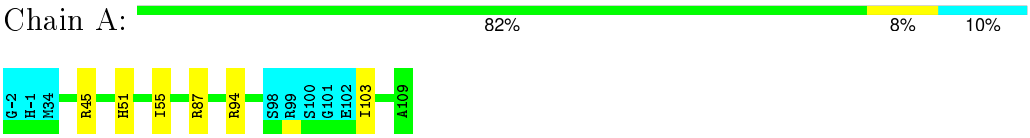
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Tudor domain-containing protein 7

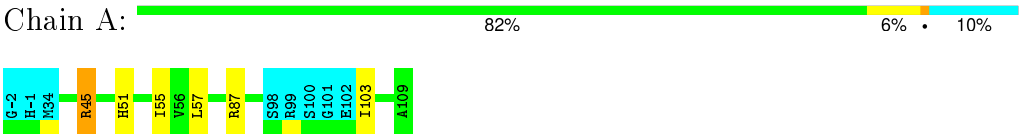
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Tudor domain-containing protein 7

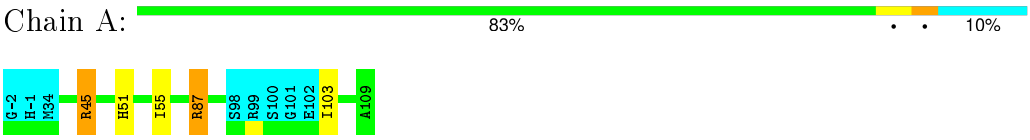
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Tudor domain-containing protein 7

Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Tudor domain-containing protein 7

Chain A: 82% 5% • 10%

G-2 H-1 M34 R45 H51 I55 V56 L57 R87 S98 R99 S100 G101 E102 I103 A109

4.2.9 Score per residue for model 9

- Molecule 1: Tudor domain-containing protein 7

Chain A: 81% 8% • 10%

G-2 H-1 M34 R45 I55 V56 L57 P58 R59 R87 S98 R99 S100 G101 E102 I103 V108 A109

4.2.10 Score per residue for model 10

- Molecule 1: Tudor domain-containing protein 7

Chain A: 83% 5% • 10%

G-2 H-1 M34 R45 H51 I55 R87 S98 R99 S100 G101 E102 I103 A109

4.2.11 Score per residue for model 11

- Molecule 1: Tudor domain-containing protein 7

Chain A: 82% 6% • 10%

G-2 H-1 M34 R45 H51 I55 R87 R94 S98 R99 S100 G101 E102 I103 A109

4.2.12 Score per residue for model 12

- Molecule 1: Tudor domain-containing protein 7

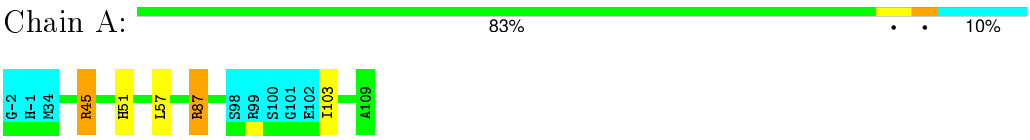
Chain A: 83% 5% • 10%

G-2 H-1 M34 R45 H51 I55 R87 S98 R99 S100 G101 E102 I103 A109

4.2.13 Score per residue for model 13

- Molecule 1: Tudor domain-containing protein 7

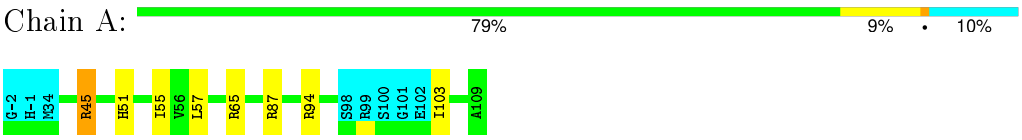
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Tudor domain-containing protein 7

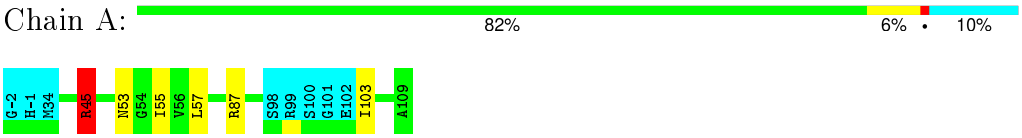
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Tudor domain-containing protein 7

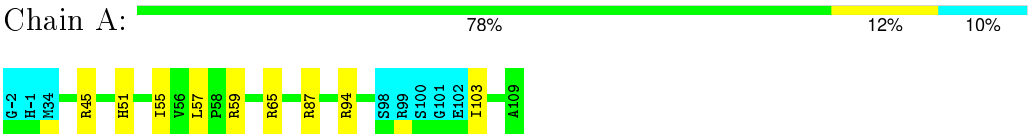
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Tudor domain-containing protein 7

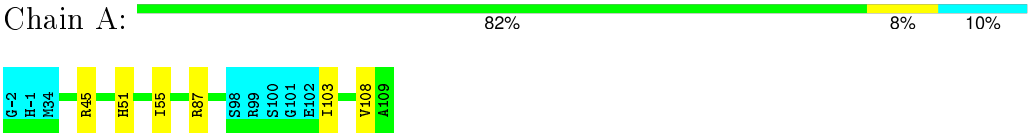
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Tudor domain-containing protein 7

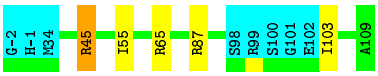
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Tudor domain-containing protein 7

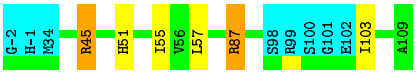
Chain A:



4.2.19 Score per residue for model 19 (medoid)

- Molecule 1: Tudor domain-containing protein 7

Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Tudor domain-containing protein 7

Chain A:

5 Refinement protocol and experimental data overview (i)

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

Of the 200 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
CYANA	structure solution	
AMBER	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)	2lh9_cs.str
Number of chemical shift lists	1
Total number of shifts	1018
Number of shifts mapped to atoms	1018
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	94%

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	0.65±0.01	0±0/563 (0.0±0.0%)	1.05±0.01	1±1/764 (0.2±0.1%)
All	All	0.65	0/11260 (0.0%)	1.05	24/15280 (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	0.3±0.5
All	All	0	6

There are no bond-length outliers.

All unique angle 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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	45	ARG	NE-CZ-NH2	-7.32	116.64	120.30	10	15
1	A	59	ARG	NE-CZ-NH2	-5.64	117.48	120.30	16	2
1	A	87	ARG	CD-NE-CZ	5.56	131.38	123.60	9	7

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	94	ARG	Sidechain	5
1	A	45	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
All	All	11040	11540	11540	-

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 -.

There are no clashes.

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	69/78 (88%)	64±1 (92±1%)	5±1 (8±1%)	0±0 (0±0%)	100	100
All	All	1380/1560 (88%)	1275 (92%)	105 (8%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	59/65 (91%)	53±1 (91±1%)	6±1 (9±1%)	15	60
All	All	1180/1300 (91%)	1068 (91%)	112 (9%)	15	60

All 9 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	103	ILE	20
1	A	87	ARG	20
1	A	45	ARG	20
1	A	55	ILE	18
1	A	51	HIS	16
1	A	57	LEU	10
1	A	65	ARG	4
1	A	108	VAL	3
1	A	53	ASN	1

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 94% for the well-defined parts and 93% for the entire structure.

7.1 Chemical shift list 1

File name: 2lh9_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	1018
Number of shifts mapped to atoms	1018
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	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
$^{13}C_{\alpha}$	78	-0.28 ± 0.18	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	72	0.33 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	78	-0.15 ± 0.19	None needed (< 0.5 ppm)
^{15}N	73	0.07 ± 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 94%, i.e. 824 atoms were assigned a chemical shift out of a possible 880. 18 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	342/342 (100%)	136/136 (100%)	140/140 (100%)	66/66 (100%)
Sidechain	426/478 (89%)	257/279 (92%)	160/177 (90%)	9/22 (41%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	56/60 (93%)	28/31 (90%)	27/27 (100%)	1/2 (50%)
Overall	824/880 (94%)	421/446 (94%)	327/344 (95%)	76/90 (84%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 93%, i.e. 898 atoms were assigned a chemical shift out of a possible 969. 18 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	380/382 (99%)	151/152 (99%)	156/156 (100%)	73/74 (99%)
Sidechain	460/520 (88%)	279/305 (91%)	171/190 (90%)	10/25 (40%)
Aromatic	58/67 (87%)	29/35 (83%)	28/29 (97%)	1/3 (33%)
Overall	898/969 (93%)	459/492 (93%)	355/375 (95%)	84/102 (82%)

7.1.4 Statistically unusual chemical shifts (i)

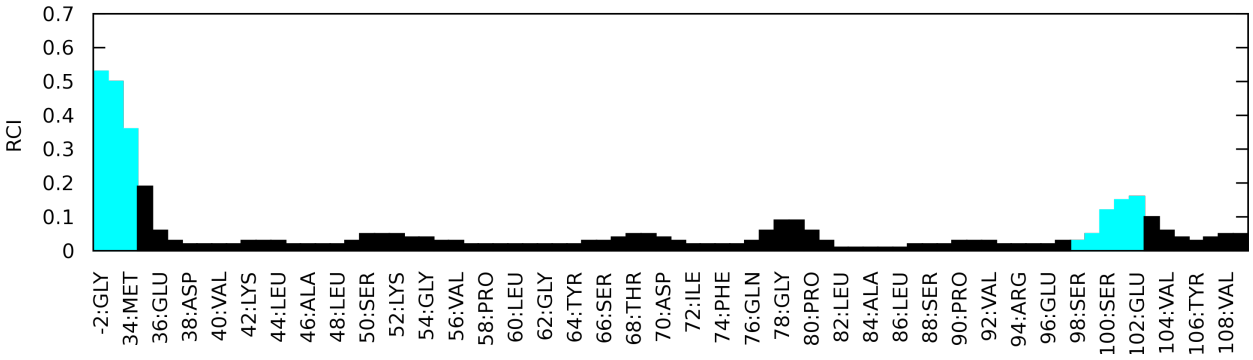
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	65	ARG	HD2	1.04	4.27 – 1.97	-9.0
1	A	65	ARG	HD3	1.72	4.36 – 1.86	-5.6

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:


RCI
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-2:GLY
34:MET
36:GLU
38:ASP
40:VAL
42:LYS
44:LEU
46:ALA
48:LEU
50:SER
52:LYS
54:GLY
56:VAL
58:PRO
60:LEU
62:GLY
64:TYR
66:SER
68:THR
70:ASP
72:ILE
74:PHE
76:GLN
78:GLY
80:PRO
82:LEU
84:ALA
86:LEU
88:SER
90:PRO
92:VAL
94:ARG
96:GLU
98:SER
100:SER
102:GLU
104:VAL
106:TYR
108:VAL