Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 01:34 AM BST

PDB ID	:	2LO7
Title	:	Ts16 NMR solution structure
Authors	:	del Rio-Portilla, F.; Saucedo, A.L.; Flores-Solis, D.
Deposited on	:	2012-01-17

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:

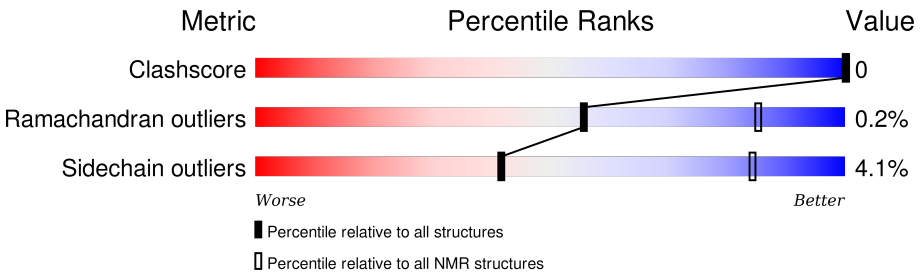
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 49%.

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	31	

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *no violations*.

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:4-A:29 (26)	0.26	2

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 8 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 464 atoms, of which 229 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Toxin Ts16.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	31	464	141	229	45	42	7	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP P86271
A	2	SER	-	EXPRESSION TAG	UNP P86271

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: Toxin Ts16

Chain A:

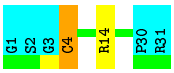
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Toxin Ts16

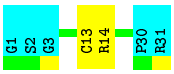
Chain A:



4.2.2 Score per residue for model 2 (medoid)

- Molecule 1: Toxin Ts16

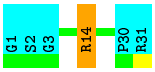
Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Toxin Ts16

Chain A:



4.2.4 Score per residue for model 4

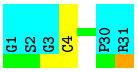
- Molecule 1: Toxin Ts16

Chain A:

4.2.5 Score per residue for model 5

- Molecule 1: Toxin Ts16

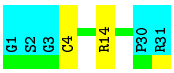
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Toxin Ts16

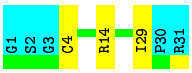
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Toxin Ts16

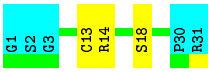
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Toxin Ts16

Chain A: 74% 10% 16%



4.2.9 Score per residue for model 9

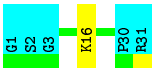
- Molecule 1: Toxin Ts16

Chain A: 77% 6% 16%

4.2.10 Score per residue for model 10

- Molecule 1: Toxin Ts16

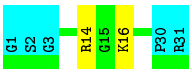
Chain A: 81% • 16%



4.2.11 Score per residue for model 11

- Molecule 1: Toxin Ts16

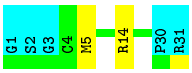
Chain A: 77% 6% 16%



4.2.12 Score per residue for model 12

- Molecule 1: Toxin Ts16

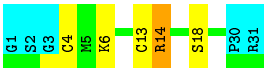
Chain A: 77% 6% 16%



4.2.13 Score per residue for model 13

- Molecule 1: Toxin Ts16

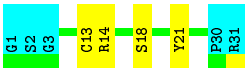
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Toxin Ts16

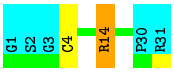
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Toxin Ts16

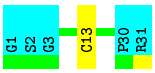
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Toxin Ts16

Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Toxin Ts16

Chain A:

4.2.18 Score per residue for model 18

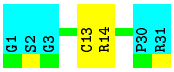
- Molecule 1: Toxin Ts16

Chain A:

4.2.19 Score per residue for model 19

- Molecule 1: Toxin Ts16

Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Toxin Ts16

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: *20 structures for lowest energy and no violations*.

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

Software name	Classification	Version
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)	2lo7_cs.str
Number of chemical shift lists	1
Total number of shifts	192
Number of shifts mapped to atoms	192
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	49%

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	0.68±0.01	0±0/205 (0.0±0.0%)	1.00±0.07	1±1/272 (0.3±0.2%)
All	All	0.69	0/4100 (0.0%)	1.00	16/5440 (0.3%)

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	0.5±0.7
All	All	0	9

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	14	ARG	NE-CZ-NH1	8.15	124.38	120.30	2	15
1	A	21	TYR	CB-CG-CD1	-5.41	117.76	121.00	14	1

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	14	ARG	Sidechain,Peptide	4
1	A	4	CYS	Peptide	4
1	A	29	ILE	Peptide	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	4040	3920	3920	-

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	26/31 (84%)	24±1 (94±3%)	2±1 (6±3%)	0±0 (0±1%)	56 85
All	All	520/620 (84%)	487 (94%)	32 (6%)	1 (0%)	56 85

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	16	LYS	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	23/26 (88%)	22±1 (96±3%)	1±1 (4±3%)	42 84

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	460/520 (88%)	441 (96%)	19 (4%)	42 84

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	13	CYS	6
1	A	4	CYS	4
1	A	18	SER	3
1	A	9	CYS	1
1	A	6	LYS	1
1	A	5	MET	1
1	A	16	LYS	1
1	A	14	ARG	1
1	A	7	GLU	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 49% for the well-defined parts and 48% for the entire structure.

7.1 Chemical shift list 1

File name: 2lo7_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	192
Number of shifts mapped to atoms	192
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

7.1.2 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 49%, i.e. 156 atoms were assigned a chemical shift out of a possible 318. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	52/130 (40%)	52/52 (100%)	0/52 (0%)	0/26 (0%)
Sidechain	94/165 (57%)	94/100 (94%)	0/56 (0%)	0/9 (0%)
Aromatic	10/23 (43%)	10/12 (83%)	0/10 (0%)	0/1 (0%)
Overall	156/318 (49%)	156/164 (95%)	0/118 (0%)	0/36 (0%)

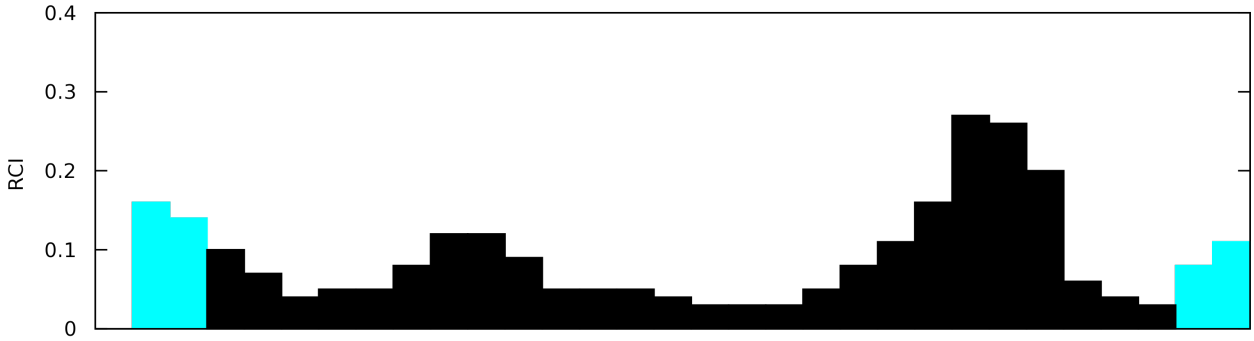
The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 48%, i.e. 177 atoms were assigned a chemical shift out of a possible 371. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	59/153 (39%)	59/61 (97%)	0/62 (0%)	0/30 (0%)
Sidechain	108/195 (55%)	108/119 (91%)	0/64 (0%)	0/12 (0%)
Aromatic	10/23 (43%)	10/12 (83%)	0/10 (0%)	0/1 (0%)
Overall	177/371 (48%)	177/192 (92%)	0/136 (0%)	0/43 (0%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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: