



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 26, 2016 – 11:31 PM BST

PDB ID : 2KO0
Title : Solution structure of the THAP zinc finger of THAP1 in complex with its DNA target
Authors : Campagne, S.; Gervais, V.; Saurel, O.; Milon, A.
Deposited on : 2009-09-08

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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

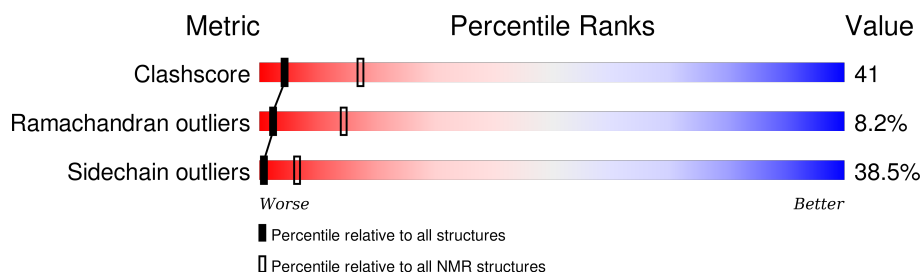
1 Overall quality at a glance

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	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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	87	
2	B	16	
3	C	16	

2 Ensemble composition and analysis

This entry contains 15 models. Model 9 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:5-A:82 (78)	0.15	9

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 3 single-model clusters were found.

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

3 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2447 atoms, of which 1078 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called THAP domain-containing protein 1.

Mol	Chain	Residues	Atoms						Trace
1	A	87	Total	C	H	N	O	S	0
			1434	455	719	127	128	5	

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	62	SER	CYS	ENGINEERED	UNP Q9NVV9
A	67	SER	CYS	ENGINEERED	UNP Q9NVV9
A	83	GLU	-	EXPRESSION TAG	UNP Q9NVV9
A	84	LEU	-	EXPRESSION TAG	UNP Q9NVV9
A	85	VAL	-	EXPRESSION TAG	UNP Q9NVV9
A	86	PRO	-	EXPRESSION TAG	UNP Q9NVV9
A	87	ARG	-	EXPRESSION TAG	UNP Q9NVV9

- Molecule 2 is a DNA chain called RRM1.

Mol	Chain	Residues	Atoms						Trace
2	B	16	Total	C	H	N	O	P	0
			512	157	181	62	97	15	

- Molecule 3 is a DNA chain called ZINC ION.

Mol	Chain	Residues	Atoms						Trace
3	C	16	Total	C	H	N	O	P	0
			500	152	178	61	93	16	

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

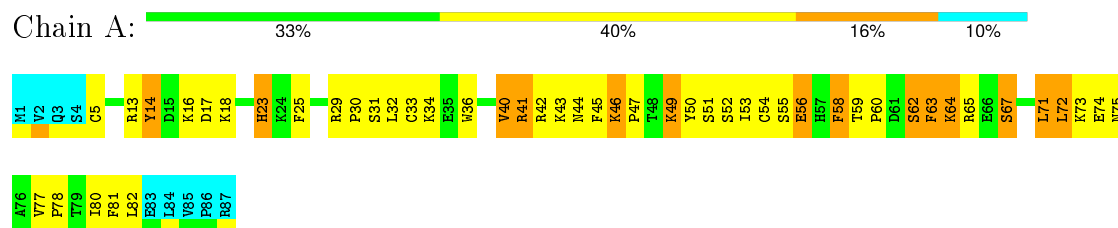
Mol	Chain	Residues	Atoms	
4	A	1	Total	Zn
			1	1

4 Residue-property plots

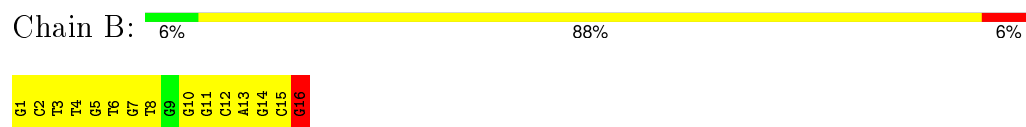
4.1 Average score per residue in the NMR ensemble

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: THAP domain-containing protein 1



- Molecule 2: RRM1



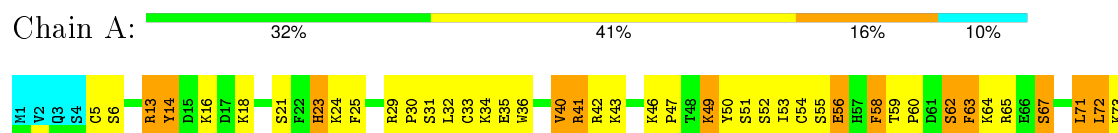
- Molecule 3: ZINC ION

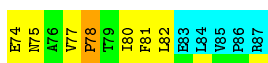


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

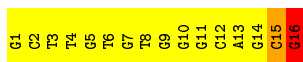
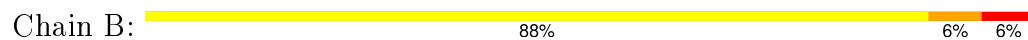
The representative model is number 9. Colouring as in section 4.1 above.

- Molecule 1: THAP domain-containing protein 1

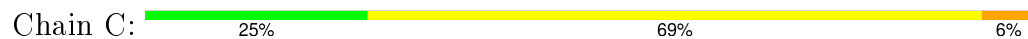




- Molecule 2: RRM1



- Molecule 3: ZINC ION



5 Refinement protocol and experimental data overview

The models were refined using the following method: *Rigid body docking, Semi flexible simulated annealing, Water refinement.*

Of the 200 calculated structures, 15 were deposited, based on the following criterion: *structures with the least restraint violations.*

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

Software name	Classification	Version
HADDOCK	refinement	2.0

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 16485
Number of chemical shift lists	2
Total number of shifts	1337
Number of shifts mapped to atoms	1042
Number of unparsed shifts	0
Number of shifts with mapping errors	295
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

6.1 Standard geometry

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

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	0.42±0.01	0±0/661 (0.0±0.0%)	0.59±0.01	0±0/890 (0.0±0.0%)
2	B	0.42±0.01	0±0/371 (0.0±0.0%)	0.83±0.01	2±0/573 (0.3±0.0%)
3	C	0.28±0.01	0±0/360 (0.0±0.0%)	0.68±0.01	0±0/551 (0.0±0.0%)
All	All	0.39	0/20880 (0.0%)	0.69	30/30210 (0.1%)

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
2	B	0.0±0.0	1.5±0.6
3	C	0.0±0.0	0.8±0.5
All	All	0	34

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(°)	Ideal(°)	Models	
								Worst	Total
2	B	16	DG	O4'-C1'-N9	6.43	112.50	108.00	8	15
2	B	16	DG	O4'-C4'-C3'	-6.41	101.94	104.50	2	15

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)
2	B	16	DG	Sidechain	15

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
3	C	22	DC	Sidechain	10
2	B	15	DC	Sidechain	5
2	B	10	DG	Sidechain	2
3	C	23	DC	Sidechain	2

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	642	640	639	31±3
2	B	331	181	182	46±3
3	C	322	178	178	23±2
All	All	19440	14985	14985	1410

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 41.

5 of 184 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:15:DC:H1'	2:B:16:DG:N7	1.14	1.58	5	15
2:B:14:DG:H2''	2:B:15:DC:C5	0.88	2.03	9	15
2:B:15:DC:C1'	2:B:16:DG:N7	0.88	2.37	4	15
2:B:7:DG:H2'	2:B:8:DT:H72	0.87	1.45	5	10
2:B:7:DG:C2'	2:B:8:DT:H71	0.87	1.99	9	5

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	78/87 (90%)	57±1 (73±1%)	15±1 (19±2%)	6±1 (8±1%)	2	14
All	All	1170/1305 (90%)	855 (73%)	219 (19%)	96 (8%)	2	14

5 of 10 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	67	SER	15
1	A	40	VAL	15
1	A	64	LYS	15
1	A	56	GLU	15
1	A	78	PRO	11

6.3.2 Protein sidechains ⓘ

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	73/82 (89%)	45±2 (61±2%)	28±2 (39±2%)	1	6
All	All	1095/1230 (89%)	673 (61%)	422 (39%)	1	6

5 of 45 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	14	TYR	15
1	A	71	LEU	15
1	A	58	PHE	15
1	A	41	ARG	15
1	A	49	LYS	15

6.3.3 RNA ⓘ

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](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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 55% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 16485

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	1042
Number of shifts mapped to atoms	1042
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	8

7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	86	-0.62 ± 0.26	Should be applied
$^{13}\text{C}_\beta$	84	-0.18 ± 0.10	None needed (< 0.5 ppm)
$^{13}\text{C}'$	73	-0.16 ± 0.08	None needed (< 0.5 ppm)
^{15}N	77	-0.49 ± 0.52	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 53%, i.e. 883 atoms were assigned a chemical shift out of a possible 1670. 8 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	365/378 (97%)	150/150 (100%)	145/156 (93%)	70/72 (97%)
Sidechain	446/546 (82%)	287/329 (87%)	151/186 (81%)	8/31 (26%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	72/105 (69%)	44/56 (79%)	27/45 (60%)	1/4 (25%)
Overall	883/1670 (53%)	481/920 (52%)	323/600 (54%)	79/150 (53%)

7.1.4 Statistically unusual chemical shifts [i](#)

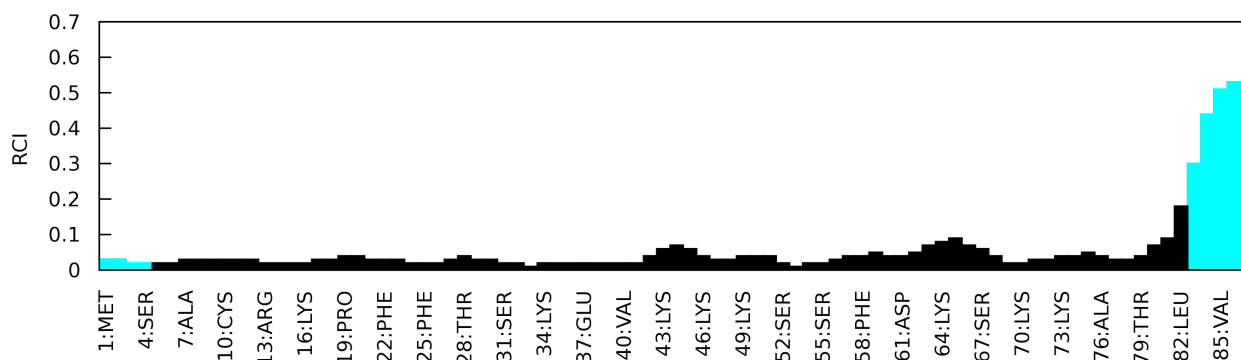
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	29	ARG	NH1	109.27	94.37 – 52.87	8.6
1	A	72	LEU	HD23	-1.28	2.14 – -0.66	-7.2
1	A	72	LEU	HD22	-1.28	2.14 – -0.66	-7.2
1	A	72	LEU	HD21	-1.28	2.14 – -0.66	-7.2
1	A	65	ARG	NH1	100.90	94.37 – 52.87	6.6
1	A	26	PRO	HD3	1.21	5.52 – 1.72	-6.3
1	A	78	PRO	HB3	-0.08	3.81 – 0.21	-5.8
1	A	78	PRO	HD3	1.48	5.52 – 1.72	-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:



7.2 Chemical shift list 2

File name: BMRB entry 16485

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	295
Number of shifts mapped to atoms	0
Number of unparsed shifts	0
Number of shifts with mapping errors	295
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- Chain not found in structure. First 5 (of 295) occurrences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	14	DG	H1'	5.706	0.04	1
UNMAPPED	26	DC	H6	7.224	0.04	1
UNMAPPED	13	DA	H2''	2.838	0.04	1
UNMAPPED	28	DC	H3'	4.952	0.04	1
UNMAPPED	8	DT	H2''	1.883	0.04	1

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 0%, i.e. 0 atoms were assigned a chemical shift out of a possible 1670. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	0/378 (0%)	0/150 (0%)	0/156 (0%)	0/72 (0%)

Continued on next page...

Continued from previous page...

	Total	¹H	¹³C	¹⁵N
Sidechain	0 / 546 (0%)	0 / 329 (0%)	0 / 186 (0%)	0 / 31 (0%)
Aromatic	0 / 105 (0%)	0 / 56 (0%)	0 / 45 (0%)	0 / 4 (0%)
Overall	0 / 1670 (0%)	0 / 920 (0%)	0 / 600 (0%)	0 / 150 (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](#)

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned_chem_shift_list_2). RCI is only applicable to proteins.