W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 10:11 pm BST

PDB ID	:	2JPA
Title	:	Structure of the Wilms Tumor Suppressor Protein Zinc Finger Domain Bound to DNA
Authors	:	Stoll, R.; Lee, B.M.; Debler, E.W.; Laity, J.H.; Wilson, I.A.; Dyson, H.J.; Wright, P.E.
Deposited on	:	2007-05-01

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
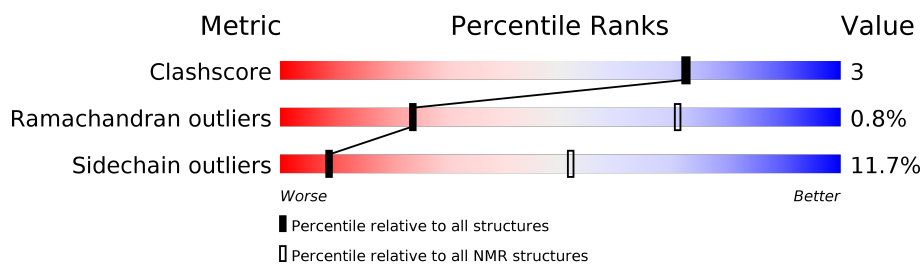with specific help available everywhere you see the (i) symbol.

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 11%.

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	Percentile Ranks	Value
Clashscore		3
Ramachandran outliers		0.8%
Sidechain outliers		11.7%

Worse   *Better*

■ Percentile relative to all structures

▯ Percentile relative to all NMR structures

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	B	14	14% / 79% / 7%
2	C	14	36% / 50% / 14%
3	A	119	72% / 20% / • 5%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:7-A:119 (113)	0.70	11

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 1 single-model cluster was found.

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

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2880 atoms, of which 1291 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(P*DCP*DGP*DCP*DGP*DGP*DGP*DGP *DGP*DCP*DGP*DTP*DCP*DTP*DG)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	14	448	136	156	56	86	14	0

- Molecule 2 is a DNA chain called DNA (5'-D(P*DCP*DAP*DGP*DAP*DCP*DGP*DCP* DCP*DCP*DCP*DCP*DGP*DCP*DG)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	14	436	132	154	54	82	14	0

- Molecule 3 is a protein called Wilms tumor 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	A	119	1992	622	981	210	168	11	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	ALA	-	expression tag	UNP Q4VXV4

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

Mol	Chain	Residues	Atoms	
			Total	Zn
4	A	4	4	4

4 Residue-property plots ⓘ

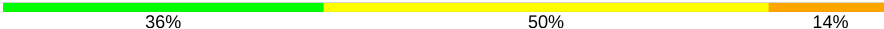## 4.1 Average score per residue in the NMR ensemble

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: DNA (5'-D(P*DCP*DGP*DCP*DGP*DGP*DGP*DGP*DGP*DCP*DGP*DTP*DCP*DTP*DG)-3')
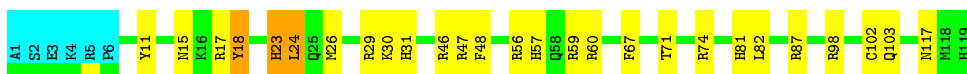
Chain B:

• Molecule 2: DNA (5'-D(P*DCP*DAP*DGP*DAP*DCP*DGP*DCP*DCP*DCP*DCP*DCP*DGP*DCP*DG)-3')

Chain C:

• Molecule 3: Wilms tumor 1

Chain A:

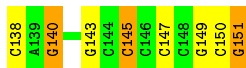## 4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 11. Colouring as in section 4.1 above.

• Molecule 1: DNA (5'-D(P*DCP*DGP*DCP*DGP*DGP*DGP*DGP*DGP*DCP*DGP*DTP*DCP*DTP*DG)-3')

Chain B:

C124 G125 C126 G127 G128 G129 G130 G133 T134 C135 T136 G137

● Molecule 2: DNA (5'-D(P*DCP*DAP*DGP*DAP*DCP*DGP*DCP*DCP*DCP*DCP*DCP* DGP*DCP*DG)-3')

Chain C:

43%   36%   21%

C138 A139 G140 G143 C144 C145 C146 C147 C148 G149 C150 G151

● Molecule 3: Wilms tumor 1

Chain A:

74%   16%   · · 5%

A1 S2 E3 K4 R5 P6 R17 Y18 H23 L24 Q25 R29 K30 Y37 K42 R46 R47 R50 L54 K55 R56 R59 R60 T71 H81 R87 R98 C102 Q103 K104 R114 N117 M118 H119

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

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
Amber	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 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	236
Number of shifts mapped to atoms	236
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	11%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	B	292	156	157	2±1
2	C	282	154	153	1±0
3	A	964	932	932	8±2
All	All	30840	24840	24835	188

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 3.

5 of 41 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:A:18:TYR:CD2	3:A:24:LEU:HD23	0.74	2.18	3	15
3:A:18:TYR:CD1	3:A:24:LEU:HD23	0.65	2.25	13	5
1:B:128:DG:C8	3:A:81:HIS:CE1	0.64	2.86	5	7
3:A:67:PHE:HB3	3:A:82:LEU:HD13	0.61	1.73	19	11
3:A:37:TYR:HB3	3:A:54:LEU:HD22	0.56	1.77	20	6

5.2 Torsion angles (i)

5.2.1 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	
3	A	112/119 (94%)	104±2 (93±2%)	7±2 (6±2%)	1±1 (1±1%)	24	71
All	All	2240/2380 (94%)	2083 (93%)	139 (6%)	18 (1%)	24	71

5 of 9 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	A	101	SER	5
3	A	102	CYS	4
3	A	73	GLN	2
3	A	103	GLN	2
3	A	99	TRP	1

5.2.2 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	
3	A	107/112 (96%)	94±2 (88±2%)	13±2 (12±2%)	9	52

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2140/2240 (96%)	1889 (88%)	251 (12%)	9 52

5 of 46 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)
3	A	23	HIS	20
3	A	24	LEU	20
3	A	71	THR	19
3	A	117	ASN	13
3	A	26	MET	12

5.2.3 RNA ⓘ

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.4 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.5 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

5.6 Other polymers ⓘ

There are no such molecules in this entry.

5.7 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 11% for the well-defined parts and 11% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

6.1.1 Bookkeeping ⓘ

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	236
Number of shifts mapped to atoms	236
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

6.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	0	—	None (insufficient data)
$^{13}C_\beta$	0	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	108	-0.12 ± 0.39	None needed (< 0.5 ppm)

6.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 11%, i.e. 230 atoms were assigned a chemical shift out of a possible 2139. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	210/555 (38%)	105/221 (48%)	0/226 (0%)	105/108 (97%)
Sidechain	18/831 (2%)	12/506 (2%)	0/261 (0%)	6/64 (9%)

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	2/185 (1%)	1/102 (1%)	0/71 (0%)	1/12 (8%)
Overall	230/2139 (11%)	118/1173 (10%)	0/742 (0%)	112/224 (50%)

6.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

6.1.5 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: