wwPDB NMR Structure Validation Summary Report

Jun 24, 2024 – 03:06 PM EDT

PDB ID	:	6WLH
BMRB ID	:	50236
Title	:	RNA complex of WT1 zinc finger transcription factor
Authors	:	Wright, P.E.; Dyson, H.J.
Deposited on	:	2020-04-20

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.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

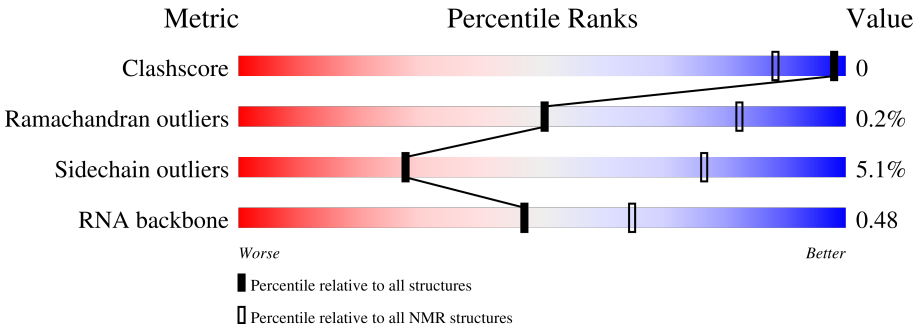
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 38%.

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428
RNA backbone	4643	676

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	118	85% 12% .
2	B	29	59% 28% 14%

2 Ensemble composition and analysis

This entry contains 20 models. Model 12 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:118 (114)	1.50	12

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 4 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2919 atoms, of which 1291 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Wilms tumor protein.

Mol	Chain	Residues	Atoms						Trace
1	A	118	Total	C	H	N	O	S	0
			1982	619	976	209	167	11	

- Molecule 2 is a RNA chain called RNA (29-MER).

Mol	Chain	Residues	Atoms						Trace
2	B	29	Total	C	H	N	O	P	0
			933	276	315	112	201	29	

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
3	A	4	Total	Zn
			4	4

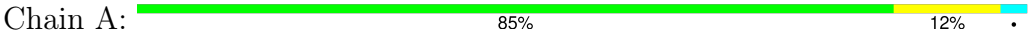
4 Residue-property plots

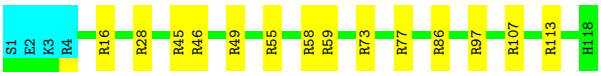
4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Wilms tumor protein

Chain A:





• Molecule 2: RNA (29-MER)

Chain B:

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

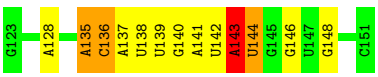
The representative model is number 12. Colouring as in section 4.1 above.

• Molecule 1: Wilms tumor protein

Chain A:

• Molecule 2: RNA (29-MER)

Chain B:



5 Refinement protocol and experimental data overview

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

Of the 800 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 calculation	
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)	working_cs.cif
Number of chemical shift lists	2
Total number of shifts	846
Number of shifts mapped to atoms	846
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	38%

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	0.82±0.01	0±0/1000 (0.0± 0.0%)	1.03±0.04	12±2/1331 (0.9± 0.2%)
2	B	1.30±0.02	0±0/690 (0.0± 0.0%)	1.31±0.04	4±2/1073 (0.3± 0.2%)
All	All	1.04	0/33800 (0.0%)	1.16	301/48080 (0.6%)

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	1.1±0.9
2	B	0.0±0.0	8.9±2.7
All	All	0	199

There are no bond-length outliers.

5 of 65 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	58	ARG	NE-CZ-NH2	11.42	126.01	120.30	15	10
2	B	142	U	C1'-O4'-C4'	-11.06	101.06	109.90	20	2
1	A	113	ARG	NE-CZ-NH1	10.81	125.71	120.30	7	9
1	A	55	ARG	NE-CZ-NH2	10.16	125.38	120.30	2	12
1	A	49	ARG	NE-CZ-NH1	10.02	125.31	120.30	8	4

There are no chirality outliers.

5 of 37 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	143	A	Sidechain	20
2	B	140	G	Sidechain	14
2	B	134	G	Sidechain	12
2	B	124	G	Sidechain	11
2	B	139	U	Sidechain	11

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	971	939	939	0±1
2	B	618	315	315	0±1
All	All	31860	25080	25080	19

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 0.

5 of 11 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:73:ARG:HH12	1:A:84:HIS:CE1	0.61	2.13	5	1
2:B:141:A:C5'	2:B:142:U:H1'	0.56	2.30	9	2
2:B:135:A:H2'	2:B:136:C:C5	0.51	2.39	9	1
2:B:135:A:H2'	2:B:136:C:C6	0.47	2.44	9	1
1:A:18:PHE:CD1	1:A:19:LYS:HE2	0.46	2.45	2	1

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	113/118 (96%)	108±1 (96±1%)	4±1 (4±1%)	0±1 (0±0%)	50 82

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	2260/2360 (96%)	2168 (96%)	88 (4%)	4 (0%)	50 82

All 2 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	35	PRO	3
1	A	5	PRO	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	108/112 (96%)	102±2 (95±2%)	6±2 (5±2%)	27 77
All	All	2160/2240 (96%)	2049 (95%)	111 (5%)	27 77

5 of 43 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	16	ARG	9
1	A	46	ARG	9
1	A	34	LYS	9
1	A	45	ARG	7
1	A	19	LYS	6

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	28/29 (97%)	6±2 (21±6%)	2±1 (5±4%)	0.48±0.04
All	All	560/580 (97%)	115 (21%)	30 (5%)	0.48

The overall RNA backbone suiteness is 0.48.

5 of 13 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	142	U	19
2	B	139	U	16
2	B	138	U	14
2	B	141	A	14
2	B	143	A	12

5 of 9 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	140	G	10
2	B	143	A	4
2	B	134	G	3
2	B	136	C	3
2	B	138	U	3

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 monosaccharides in this entry.

6.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 are 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

The completeness of assignment taking into account all chemical shift lists is 38% for the well-defined parts and 37% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chemical_shifts_1*

7.1.1 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	547
Number of shifts mapped to atoms	547
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	115	-0.03 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	105	0.95 ± 0.25	Should be checked
$^{13}C'$	109	-0.34 ± 0.11	None needed (< 0.5 ppm)
^{15}N	109	0.31 ± 0.36	None needed (< 0.5 ppm)

7.1.3 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 24%, i.e. 537 atoms were assigned a chemical shift out of a possible 2199. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	434/562 (77%)	107/226 (47%)	220/228 (96%)	107/108 (99%)
Sidechain	103/892 (12%)	0/564 (0%)	103/264 (39%)	0/64 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/196 (0%)	0/102 (0%)	0/82 (0%)	0/12 (0%)
Sugar	0/319 (0%)	0/174 (0%)	0/145 (0%)	0/0 (—%)
Base	0/230 (0%)	0/143 (0%)	0/50 (0%)	0/37 (0%)
Overall	537/2199 (24%)	107/1209 (9%)	323/769 (42%)	107/221 (48%)

7.1.4 Statistically unusual chemical shifts (i)

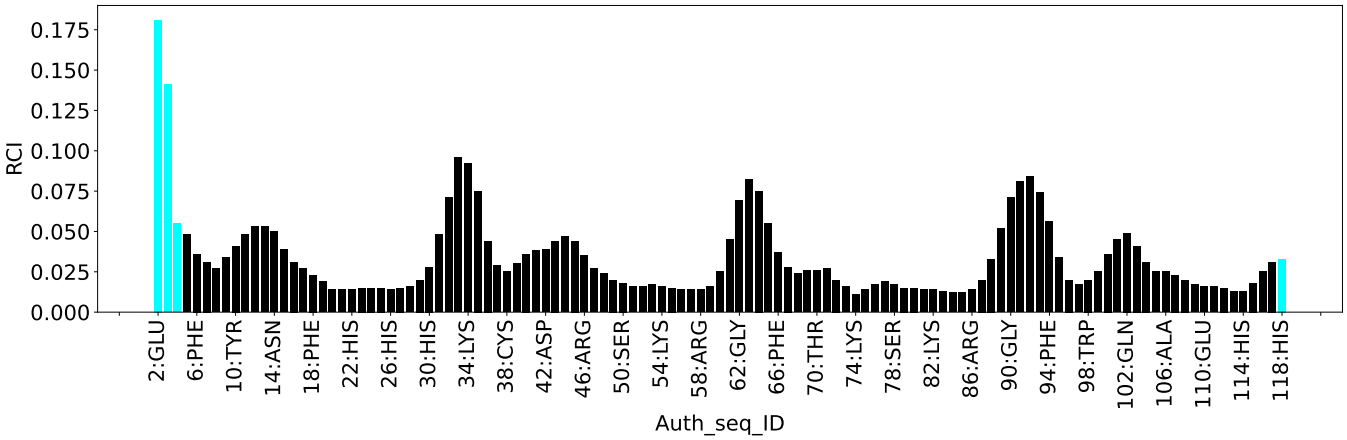
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	74	LYS	C	186.45	167.28 – 186.22	5.1

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:



7.2 Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *assigned_chemical_shifts_2*

7.2.1 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	299
Number of shifts mapped to atoms	299
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

7.2.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

7.2.3 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 13%, i.e. 289 atoms were assigned a chemical shift out of a possible 2199. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	0/562 (0%)	0/226 (0%)	0/228 (0%)	0/108 (0%)
Sidechain	0/892 (0%)	0/564 (0%)	0/264 (0%)	0/64 (0%)
Aromatic	0/196 (0%)	0/102 (0%)	0/82 (0%)	0/12 (0%)
Sugar	203/319 (64%)	107/174 (61%)	96/145 (66%)	0/0 (—%)
Base	86/230 (37%)	43/143 (30%)	43/50 (86%)	0/37 (0%)
Overall	289/2199 (13%)	150/1209 (12%)	139/769 (18%)	0/221 (0%)

7.2.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.2.5 Random Coil Index (RCI) plots

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