wwPDB NMR Structure Validation Summary Report

Jun 3, 2023 – 10:30 AM EDT

PDB ID	:	5W3G
BMRB ID	:	30303
Title	:	Solution Structure of ETS Transcription Factor PU.1
Authors	:	Lau, D.K.W.; Okon, M.; McIntosh, L.P.
Deposited on	:	2017-06-07

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:

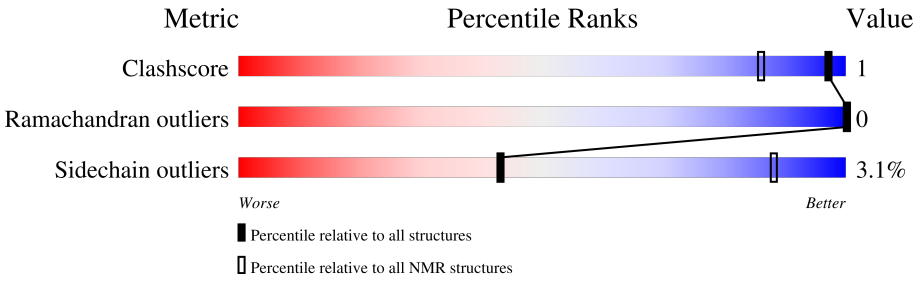
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

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 82%.

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

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	110	64% · 34%

2 Ensemble composition and analysis

This entry contains 10 models. Model 8 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:173-A:183, A:190-A:216, A:225-A:259 (73)	0.52	8

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 2 single-model clusters were found.

Cluster number	Models
1	3, 7, 8
2	6, 9, 10
3	1, 5
Single-model clusters	2; 4

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1868 atoms, of which 958 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transcription factor PU.1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	110	1868	581	958	175	150	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	163	HIS	-	expression tag	UNP P17433
A	164	ILE	-	expression tag	UNP P17433
A	165	HIS	-	expression tag	UNP P17433
A	166	MET	-	expression tag	UNP P17433

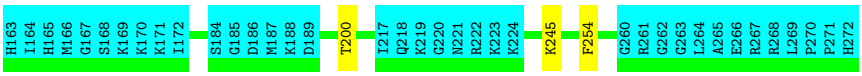
4 Residue-property plots

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: Transcription factor PU.1

Chain A:



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

The representative model is number 8. Colouring as in section 4.1 above.

- Molecule 1: Transcription factor PU.1

Chain A:

5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 10 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	refinement	
NMRe	refinement	
CYANA	structure calculation	

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	1
Total number of shifts	1244
Number of shifts mapped to atoms	1244
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

6 Model quality

6.1 Standard geometry

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	0.90±0.01	0±0/635 (0.0± 0.0%)	1.07±0.06	3±1/849 (0.4± 0.2%)
All	All	0.90	0/6350 (0.0%)	1.07	31/8490 (0.4%)

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	0.2±0.4
All	All	0	2

There are no bond-length outliers.

5 of 12 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	252	TYR	CB-CG-CD2	9.90	126.94	121.00	5	1
1	A	252	TYR	CB-CG-CD1	-8.99	115.61	121.00	5	1
1	A	175	TYR	CB-CG-CD2	-8.94	115.63	121.00	3	2
1	A	175	TYR	CB-CG-CD1	8.18	125.91	121.00	3	3
1	A	237	TYR	CB-CG-CD1	-7.23	116.66	121.00	4	2

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	252	TYR	Sidechain	1
1	A	175	TYR	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
1	A	619	641	639	1±1
All	All	6190	6410	6390	7

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 1.

5 of 7 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:178:LEU:HD11	1:A:211:LEU:HD21	0.55	1.78	6	1
1:A:181:LEU:CD2	1:A:191:ILE:HD13	0.55	2.32	9	1
1:A:181:LEU:HD23	1:A:191:ILE:HD13	0.53	1.79	9	1
1:A:245:LYS:HA	1:A:252:TYR:CD2	0.52	2.38	5	1
1:A:191:ILE:N	1:A:191:ILE:HD12	0.42	2.29	9	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	73/110 (66%)	73±0 (100±1%)	0±0 (0±1%)	0±0 (0±0%)	100	100
All	All	730/1100 (66%)	727 (100%)	3 (0%)	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	64/94 (68%)	62±1 (97±2%)	2±1 (3±2%)	43 88
All	All	640/940 (68%)	620 (97%)	20 (3%)	43 88

5 of 6 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	200	THR	7
1	A	245	LYS	5
1	A	248	LYS	3
1	A	253	GLN	2
1	A	239	LYS	2

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 monosaccharides 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

The completeness of assignment taking into account all chemical shift lists is 82% for the well-defined parts and 77% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *PU1_shifts.txt*

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	1244
Number of shifts mapped to atoms	1244
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
$^{13}C_{\alpha}$	107	-0.37 ± 0.08	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	96	0.19 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	106	-0.43 ± 0.10	None needed (< 0.5 ppm)
^{15}N	103	0.15 ± 0.31	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 82%, i.e. 895 atoms were assigned a chemical shift out of a possible 1092. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	366/370 (99%)	147/151 (97%)	146/146 (100%)	73/73 (100%)
Sidechain	457/594 (77%)	299/383 (78%)	153/180 (85%)	5/31 (16%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	72/128 (56%)	37/62 (60%)	32/59 (54%)	3/7 (43%)
Overall	895/1092 (82%)	483/596 (81%)	331/385 (86%)	81/111 (73%)

7.1.4 Statistically unusual chemical shifts (i)

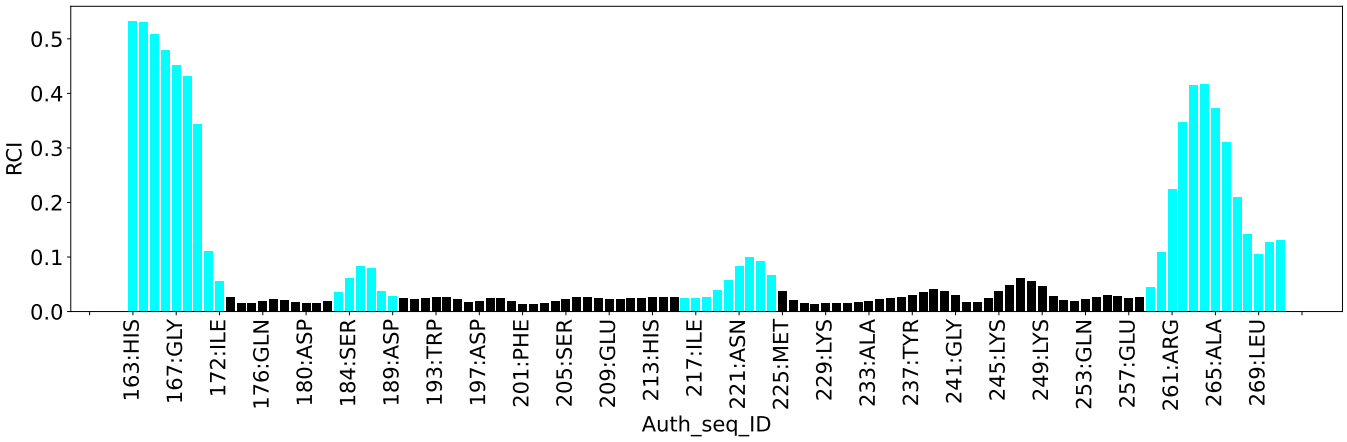
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	230	MET	CE	25.99	8.39 – 25.85	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:



8 NMR restraints analysis

8.1 Conformationally restricting restraints

The following table provides the summary of experimentally observed NMR restraints in different categories. Restraints are classified into different categories based on the sequence separation of the atoms involved.

Description	Value
Total distance restraints	1019
Intra-residue (|i-j|=0)	301
Sequential (|i-j|=1)	231
Medium range (|i-j|>1 and |i-j|<5)	183
Long range (|i-j|≥5)	304
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	9.3
Number of long range restraints per residue[1]	2.8

[1]Long range hydrogen bonds and disulfide bonds are counted as long range restraints while calculating the number of long range restraints per residue

8.2 Residual restraint violations

This section provides the overview of the restraint violations analysis. The violations are binned as small, medium and large violations based on its absolute value. Average number of violations per model is calculated by dividing the total number of violations in each bin by the size of the ensemble.

8.2.1 Average number of distance violations per model

Distance violations less than 0.1 Å are not included in the calculation.There are no distance violations

8.2.2 Average number of dihedral-angle violations per model

Dihedral-angle violations less than 1° are not included in the calculation.There are no dihedral-angle violations

9 Distance violation analysis (i)

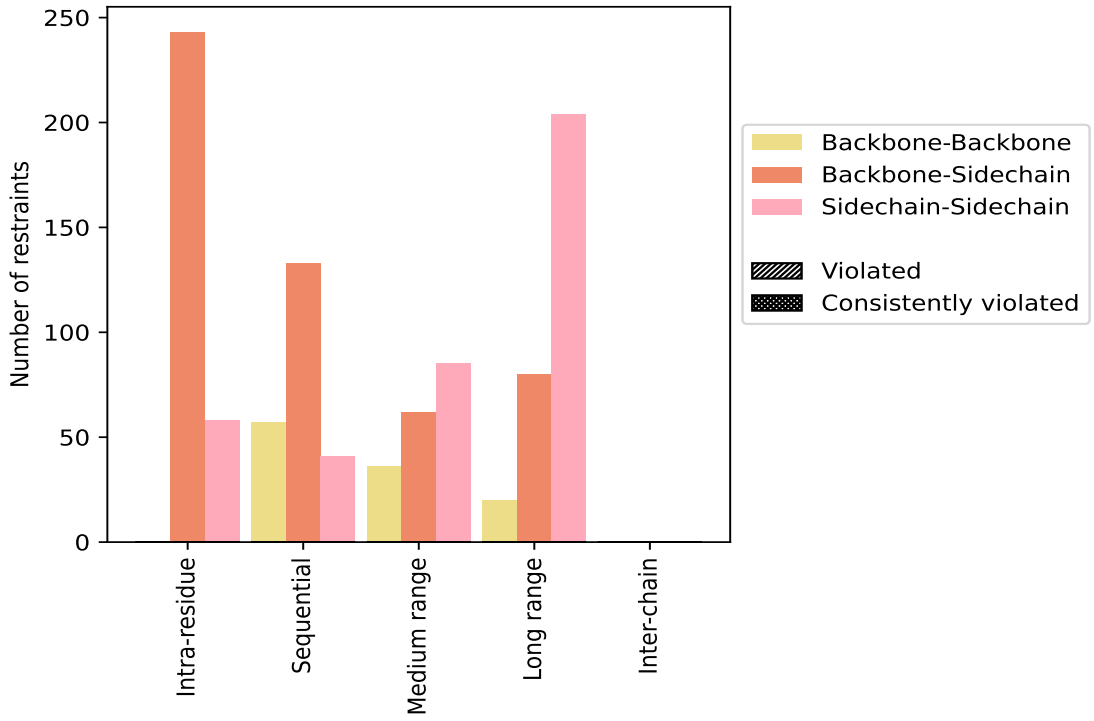
9.1 Summary of distance violations (i)

The following table shows the summary of distance violations in different restraint categories based on the sequence separation of the atoms involved. Each category is further sub-divided into three sub-categories based on the atoms involved. Violations less than 0.1 Å are not included in the statistics.

Restraints type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
Intra-residue (|i-j|=0)	301	29.5	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	243	23.8	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	58	5.7	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	231	22.7	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	57	5.6	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	133	13.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	41	4.0	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	183	18.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	36	3.5	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	62	6.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	85	8.3	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	304	29.8	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	20	2.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	80	7.9	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	204	20.0	0	0.0	0.0	0	0.0	0.0
Inter-chain	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	1019	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	113	11.1	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	518	50.8	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	388	38.1	0	0.0	0.0	0	0.0	0.0

[1] percentage calculated with respect to the total number of distance restraints, [2] percentage calculated with respect to the number of restraints in a particular restraint category, [3] violated in at least one model, [4] violated in all the models

9.1.1 Bar chart : Distribution of distance restraints and violations



Violated and consistently violated restraints are shown using different hatch patterns in their respective categories. The hydrogen bonds and disulfied bonds are counted in their appropriate category on the x-axis

9.2 Distance violation statistics for each model

No violations found

9.3 Distance violation statistics for the ensemble

No violations found

9.4 Most violated distance restraints in the ensemble

No violations found

9.5 All violated distance restraints

No violations found

10 Dihedral-angle violation analysis (i)

No dihedral-angle restraints found