wwPDB NMR Structure Validation Summary Report

Sep 15, 2021 – 10:08 AM EDT

PDB ID	:	6UD0
Title	:	Solution-state NMR structural ensemble of human Tsg101 UEV in complex with K63-linked diubiquitin
Authors	:	Strickland, M.; Watanabe, S.; Bonn, S.M.; Camara, C.M.; Fushman, D.; Carter, C.A.; Tjandra, N.
Deposited on	:	2019-09-18

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 ⓘ symbol.

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

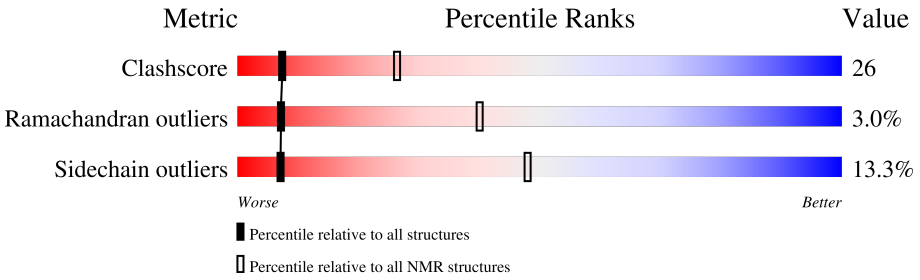
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.1
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.1

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 14%.

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	76	63% 33% .
2	B	77	64% 31% 5%
3	C	145	52% 39% 6% . .

2 Ensemble composition and analysis

This entry contains 20 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *target function*.

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:1-A:73, C:3-C:143 (214)	0.04	4
2	B:1-B:73 (73)	0.05	6

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. No single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4766 atoms, of which 2415 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ubiquitin.

Mol	Chain	Residues	Atoms						Trace
1	A	76	Total	C	H	N	O	S	0
			1232	378	629	107	117	1	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	63	ARG	LYS	engineered mutation	UNP P0CG47

- Molecule 2 is a protein called Ubiquitin.

Mol	Chain	Residues	Atoms						Trace
2	B	77	Total	C	H	N	O	S	0
			1241	382	631	106	121	1	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	77	ASP	MET	engineered mutation	UNP P0CG47

- Molecule 3 is a protein called Tumor susceptibility gene 101 protein.

Mol	Chain	Residues	Atoms						Trace
3	C	141	Total	C	H	N	O	S	0
			2293	743	1155	182	207	6	

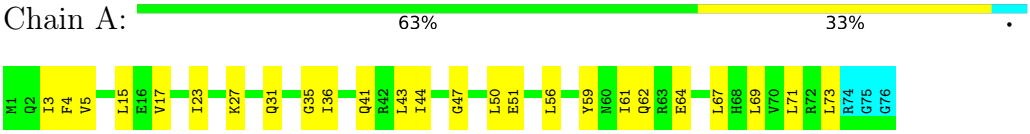
4 Residue-property plots

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.

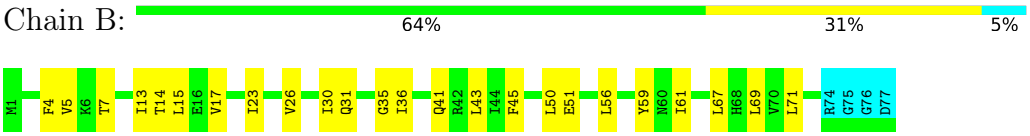
• Molecule 1: Ubiquitin

Chain A:



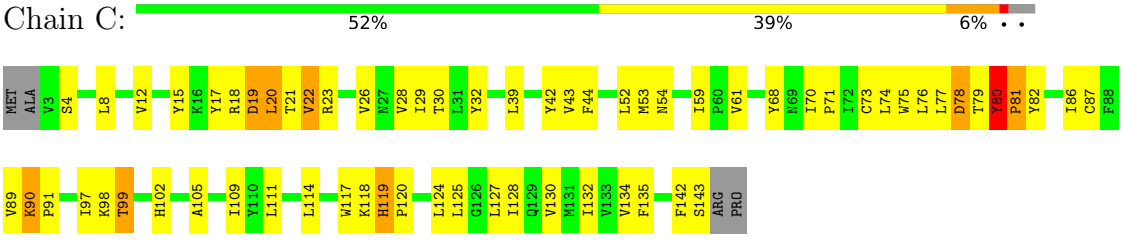
• Molecule 2: Ubiquitin

Chain B:



• Molecule 3: Tumor susceptibility gene 101 protein

Chain C:

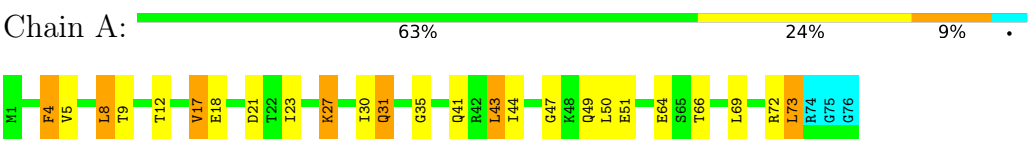


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

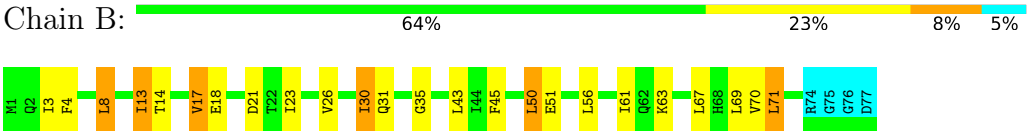
• Molecule 1: Ubiquitin

Chain A:



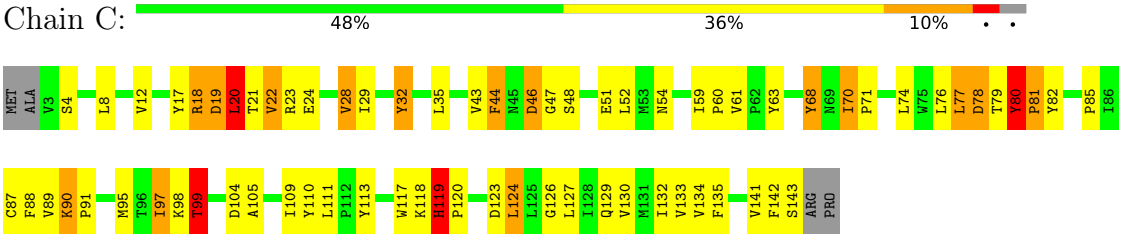
• Molecule 2: Ubiquitin

Chain B:



• Molecule 3: Tumor susceptibility gene 101 protein

Chain C:



5 Refinement protocol and experimental data overview

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

Of the 100 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
X-PLOR NIH	refinement	2.51
X-PLOR NIH	structure calculation	2.51

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	542
Number of shifts mapped to atoms	541
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	1
Assignment completeness (well-defined parts)	14%

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	0.66±0.03	0±0/591 (0.0± 0.0%)	0.75±0.04	0±1/798 (0.0± 0.1%)
2	B	0.66±0.02	0±0/589 (0.0± 0.0%)	0.76±0.03	0±0/795 (0.0± 0.1%)
3	C	0.68±0.05	1±1/1170 (0.1± 0.1%)	0.92±0.05	4±2/1596 (0.3± 0.1%)
All	All	0.67	15/47000 (0.0%)	0.84	101/63780 (0.2%)

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
3	C	0.0±0.0	0.2±0.4
All	All	0	4

5 of 10 unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
3	C	63	TYR	CB-CG	-10.07	1.36	1.51	17	1
3	C	32	TYR	CB-CG	-9.08	1.38	1.51	12	3
3	C	80	TYR	CG-CD2	-7.63	1.29	1.39	14	1
1	A	63	ARG	CD-NE	7.11	1.58	1.46	6	1
2	B	41	GLN	CA-CB	-6.46	1.39	1.53	2	1

5 of 44 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
3	C	80	TYR	CB-CG-CD2	-13.92	112.65	121.00	14	2
3	C	80	TYR	CB-CG-CD1	9.13	126.48	121.00	14	2
3	C	12	VAL	CB-CA-C	8.99	128.47	111.40	2	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
3	C	63	TYR	N-CA-CB	8.84	126.51	110.60	17	1
3	C	32	TYR	CA-CB-CG	-8.74	96.80	113.40	12	3

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)
3	C	32	TYR	Sidechain	2
3	C	80	TYR	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	584	610	610	24±10
2	B	582	608	608	24±10
3	C	1138	1155	1155	78±15
All	All	46080	47460	47457	2452

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 26.

5 of 1187 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:C:132:ILE:HA	3:C:135:PHE:CE1	1.12	1.79	3	2
3:C:80:TYR:CE2	3:C:85:PRO:HA	1.08	1.84	6	3
1:A:59:TYR:CD1	1:A:59:TYR:O	1.06	2.09	13	1
3:C:32:TYR:CE2	3:C:58:THR:O	1.06	2.08	9	3
3:C:89:VAL:HG21	3:C:97:ILE:HD11	1.03	1.30	1	2

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	72/76 (95%)	68±0 (95±1%)	3±0 (4±1%)	0±0 (1±1%)	26 73
2	B	72/77 (94%)	70±0 (97±1%)	2±0 (3±1%)	0±0 (0±0%)	100 100
3	C	139/145 (96%)	114±0 (82±0%)	18±0 (13±0%)	8±0 (6±0%)	3 21
All	All	5660/5960 (95%)	5028 (89%)	462 (8%)	170 (3%)	7 40

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	C	4	SER	20
3	C	19	ASP	20
3	C	22	VAL	20
3	C	78	ASP	20
3	C	80	TYR	20

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	67/68 (99%)	58±3 (87±4%)	9±3 (13±4%)	7 48
2	B	67/69 (97%)	58±2 (86±4%)	9±2 (14±4%)	6 46
3	C	133/136 (98%)	116±4 (87±3%)	17±4 (13±3%)	7 48
All	All	5340/5460 (98%)	4628 (87%)	712 (13%)	7 48

5 of 180 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	C	80	TYR	17
2	B	71	LEU	11
1	A	73	LEU	10
3	C	75	TRP	10
3	C	109	ILE	10

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

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

6.5 Carbohydrates

There are no monosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 14% for the well-defined parts and 15% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_ output*

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	542
Number of shifts mapped to atoms	541
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	1
Number of shift outliers (ShiftChecker)	0

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

- No matching atoms found in structure. The only occurence is reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
C	3	VAL	HT1	8.12385	-1.0	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	None (insufficient data)
$^{13}C_{\beta}$	0	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	271	0.01 ± 0.13	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 14%, i.e. 527 atoms were assigned a chemical shift out of a possible 3641. 0 out of 55 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	527/1399 (38%)	263/556 (47%)	0/574 (0%)	264/269 (98%)
Sidechain	0/2015 (0%)	0/1178 (0%)	0/754 (0%)	0/83 (0%)
Aromatic	0/227 (0%)	0/120 (0%)	0/100 (0%)	0/7 (0%)
Overall	527/3641 (14%)	263/1854 (14%)	0/1428 (0%)	264/359 (74%)

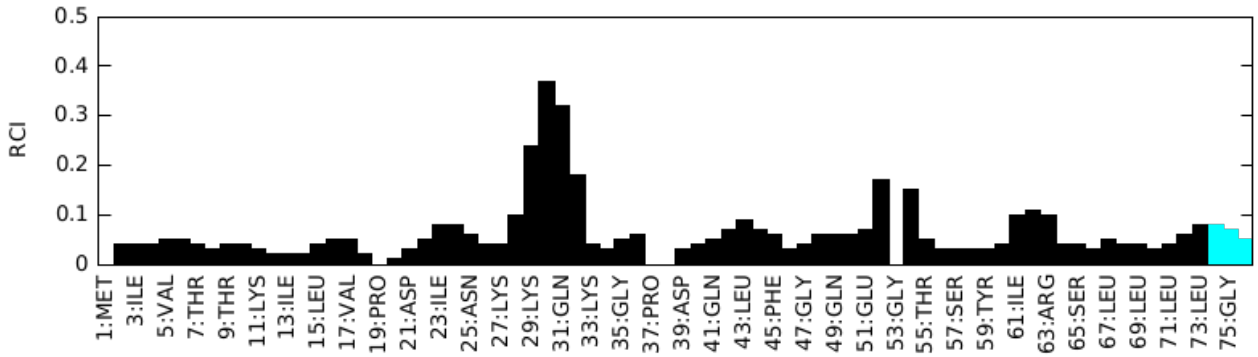
7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

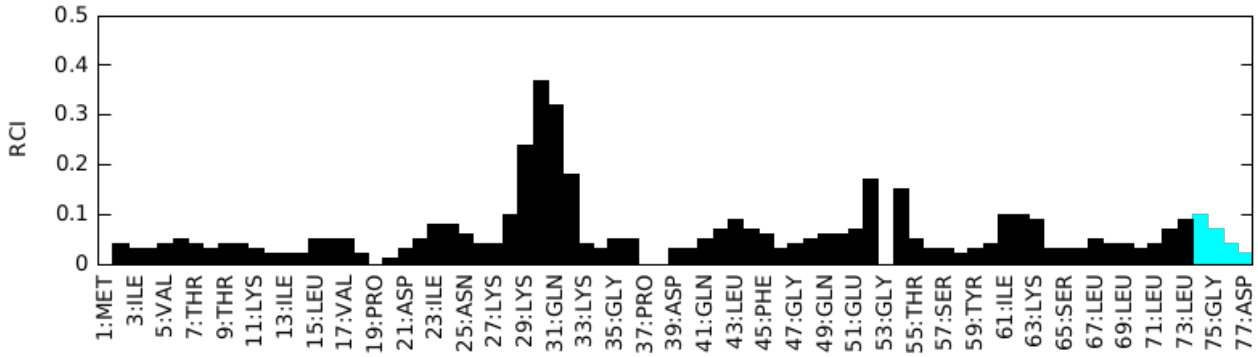
7.1.5 Random Coil Index (RCI) plots

The images below report *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:



Random coil index (RCI) for chain B:



Random coil index (RCI) for chain C: