W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB NMR Structure Validation Summary Report ⓘ

Feb 16, 2022 – 12:51 PM EST

PDB ID	:	5VEY
Title	:	Solution NMR structure of histone H2A-H2B mono-ubiquitylated at H2A Lys15 in complex with RNF169 (653-708)
Authors	:	Hu, Q.; Botuyan, M.V.; Cui, G.; Mer, G.
Deposited on	:	2017-04-06

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:

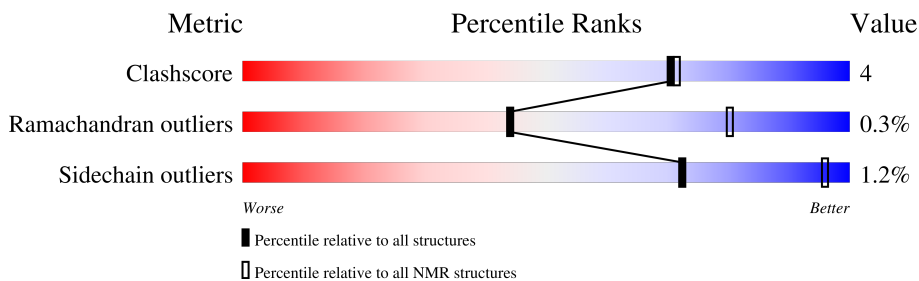|		
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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 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	Percentile Ranks	Value
Clashscore		4
Ramachandran outliers		0.3%
Sidechain outliers		1.2%

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	A	194	88% · · ·
2	B	76	88% 7% 5%
3	C	59	29% 5% 66%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 16 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:11-A:100, A:104-A:191, B:2-B:73, C:666-C:681, C:697-C:700 (270)	0.61	16

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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 5158 atoms, of which 2622 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone H2B type 1-J,Histone H2A type 1-B/E.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	186	2952	906	1505	273	266	2	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P06899
A	2	HIS	-	expression tag	UNP P06899
A	3	HIS	-	expression tag	UNP P06899
A	4	HIS	-	expression tag	UNP P06899
A	5	HIS	-	expression tag	UNP P06899
A	6	HIS	-	expression tag	UNP P06899
A	7	HIS	-	expression tag	UNP P06899
A	8	MET	-	expression tag	UNP P06899
A	100	SER	-	linker	UNP P06899
A	102	SER	LYS	linker	UNP P04908

- Molecule 2 is a protein called Polyubiquitin-B.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	76	1230	378	629	105	117	1	0

- Molecule 3 is a protein called E3 ubiquitin-protein ligase RNF169.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	C	59	976	292	488	97	95	4	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	650	GLY	-	expression tag	UNP Q8NCN4
C	651	HIS	-	expression tag	UNP Q8NCN4
C	652	MET	-	expression tag	UNP Q8NCN4

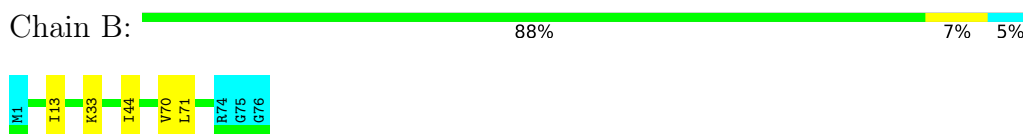# 4   Residue-property plots ⓘ

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.
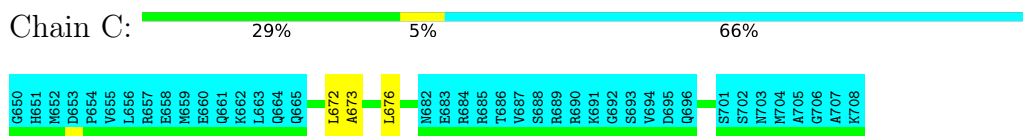
• Molecule 1: Histone H2B type 1-J,Histone H2A type 1-B/E
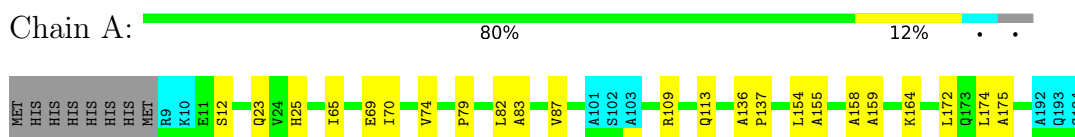
Chain A:
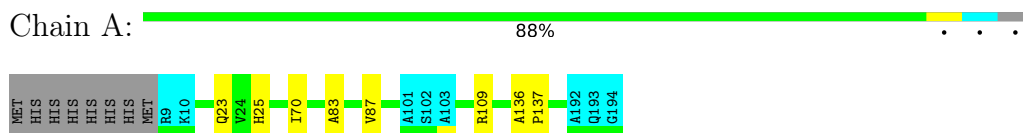
• Molecule 2: Polyubiquitin-B

Chain B:

• Molecule 3: E3 ubiquitin-protein ligase RNF169

Chain C:

4.2   Residue scores for the representative (medoid) model from the NMR ensemble
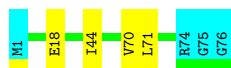
The representative model is number 16. Colouring as in section 4.1 above.

• Molecule 1: Histone H2B type 1-J,Histone H2A type 1-B/E

Chain A:

• Molecule 2: Polyubiquitin-B

Chain B:

89% 5% 5%

M1 E18 I44 V70 L71 R74 G75 G76

• Molecule 3: E3 ubiquitin-protein ligase RNF169

Chain C:

27% 7% 66%

G650 H651 M652 D653 P654 V655 L656 R657 E658 M659 E660 Q661 K662 L663 Q664 Q665 L672 A673 L676 N682 E683 R684 R685 T686 V687 S688 R689 R690 K691 G692 S693 V694 D695 Q696 R700 S701 S702 N703 M704 A705 G706 A707 K708

5 Refinement protocol and experimental data overview ⓘ

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

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
X-PLOR NIH	refinement	
TopSpin	structure solution	
NMRDraw	structure solution	
NMRView	structure solution	
NMRPipe	structure solution	
Sparky	structure solution	
TALOS	structure solution	
X-PLOR NIH	structure solution	

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	3731
Number of shifts mapped to atoms	3731
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 ⓘ

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts ⓘ

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	1392	1448	1448	9±3
2	B	574	599	599	6±2
3	C	179	176	176	4±1
All	All	42900	44460	44455	313

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 4.

5 of 99 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:116:VAL:HG12	1:A:141:ALA:CB	0.74	2.12	12	1
1:A:159:ALA:HB1	1:A:164:LYS:O	0.68	1.87	10	9
2:B:70:VAL:HG22	3:C:673:ALA:HB1	0.66	1.66	14	14
2:B:42:ARG:HH12	3:C:676:LEU:HD22	0.66	1.51	1	1
2:B:70:VAL:CG2	3:C:673:ALA:HB1	0.64	2.23	10	15

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	178/194 (92%)	175±1 (98±1%)	3±1 (2±1%)	0±0 (0±0%)	44	80
2	B	72/76 (95%)	71±1 (98±1%)	1±1 (2±1%)	0±0 (0±1%)	44	80
3	C	20/59 (34%)	20±0 (99±2%)	0±0 (0±2%)	0±0 (0±2%)	32	76
All	All	5400/6580 (82%)	5301 (98%)	84 (2%)	15 (0%)	44	80

All 5 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	79	PRO	7
2	B	64	GLU	4
3	C	699	LEU	2
1	A	100	SER	1
1	A	187	GLY	1

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	148/160 (92%)	146±1 (99±1%)	2±1 (1±1%)	70	96
2	B	66/68 (97%)	66±1 (99±1%)	0±1 (1±1%)	86	97
3	C	19/53 (36%)	19±1 (98±3%)	0±1 (2±3%)	51	92
All	All	4660/5620 (83%)	4606 (99%)	54 (1%)	72	96

5 of 19 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	23	GLN	14
1	A	25	HIS	11
3	C	669	ASP	5
3	C	697	TYR	3
1	A	113	GLN	3

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 82% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *Final_H2Aub-H2B-RNF169_BMRB.str*

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	3731
Number of shifts mapped to atoms	3731
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

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	314	-0.36 ± 0.11	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	290	0.20 ± 0.10	None needed (< 0.5 ppm)
$^{13}C'$	270	1.09 ± 0.19	Should be applied
^{15}N	295	0.03 ± 0.21	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. 2832 atoms were assigned a chemical shift out of a possible 3441. 44 out of 50 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	1263/1334 (95%)	517/532 (97%)	496/540 (92%)	250/262 (95%)
Sidechain	1469/1931 (76%)	909/1127 (81%)	540/696 (78%)	20/108 (19%)

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

|	Total	^{1}H	^{13}C	^{15}N
Aromatic	100/176 (57%)	69/94 (73%)	31/76 (41%)	0/6 (0%)
Overall	2832/3441 (82%)	1495/1753 (85%)	1067/1312 (81%)	270/376 (72%)

7.1.4 Statistically unusual chemical shifts ⓘ
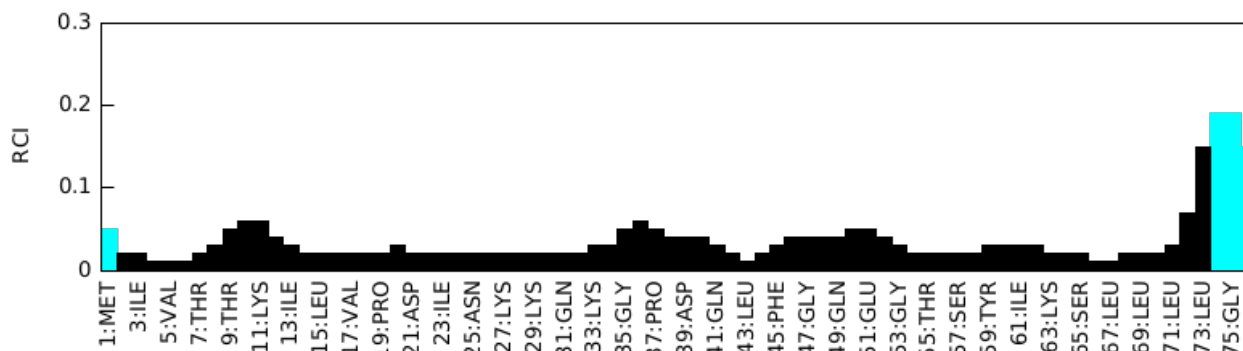
There are no statistically unusual chemical shifts.

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:
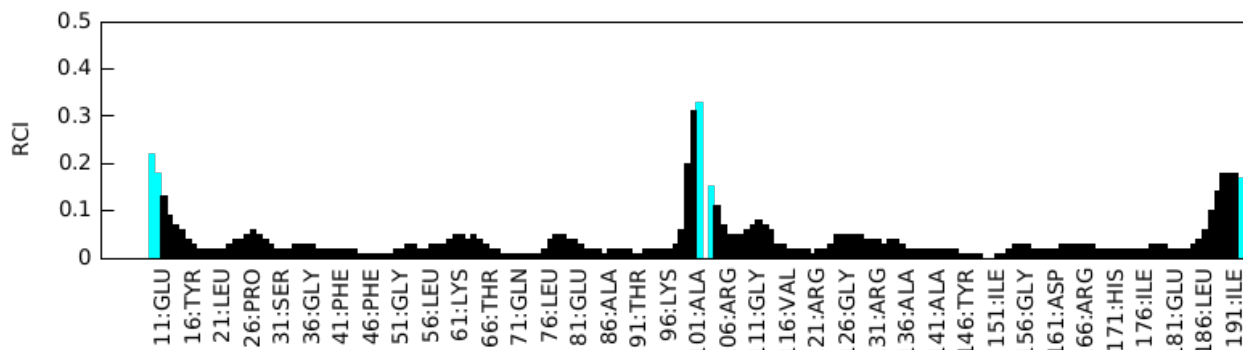
Random coil index (RCI) for chain B:

Random coil index (RCI) for chain C: