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 (i)

Nov 7, 2023 – 09:25 AM EST

PDB ID	:	2NDG
BMRB ID	:	26060
Title	:	Solution NMR structures of AF9 yeats domain in complex with histone H3 crotonylation at K18
Authors	:	Zeng, L.; Zhou, M.
Deposited on	:	2016-05-19

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:

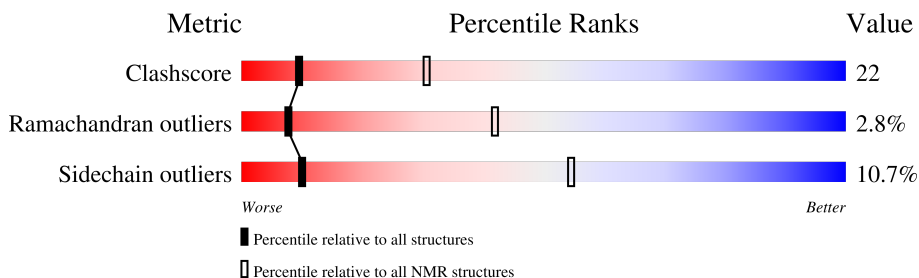|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 85%.

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		22
Ramachandran outliers		2.8%
Sidechain outliers		10.7%

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	A	141	50% 34% 7% 9%
2	B	13	15% 8% 77%

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: *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:9-A:111, A:117-A:141, B:215-B:217 (131)	0.38	4

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 5 clusters and 3 single-model clusters were found.

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

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 2554 atoms, of which 1283 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein AF-9.

Mol	Chain	Residues	Atoms						Trace
1	A	141	Total	C	H	N	O	S	0
			2345	755	1172	213	200	5	

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P42568
A	2	SER	-	expression tag	UNP P42568
A	3	HIS	-	expression tag	UNP P42568

- Molecule 2 is a protein called Histone H3 peptide.

Mol	Chain	Residues	Atoms					Trace
2	B	13	Total	C	H	N	O	0
			209	61	111	20	17	

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.
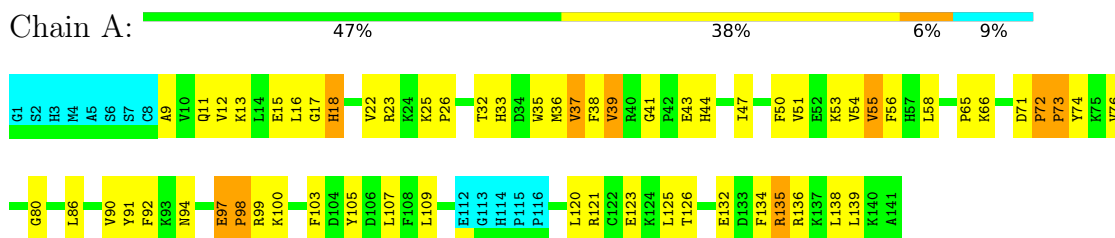
- Molecule 1: Protein AF-9
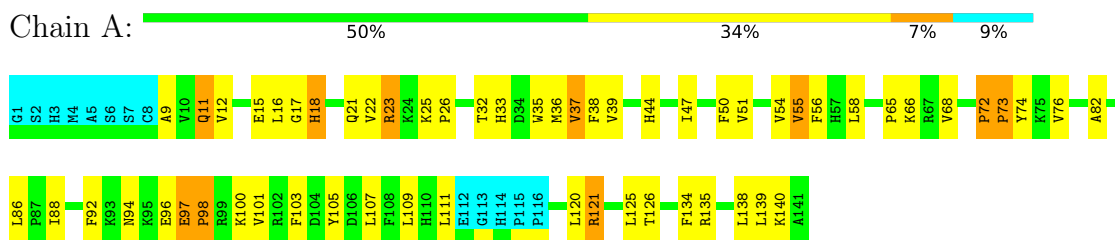
- Molecule 2: Histone H3 peptide

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: Protein AF-9

- Molecule 2: Histone H3 peptide

5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *distance geometry, torsion angle 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
ARIA	refinement	2.3
TALOS	refinement	talosn, talosplus
CNS	structure solution	1.2
CNS	geometry optimization	1.2
CNS	refinement	1.2

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	1795
Number of shifts mapped to atoms	1795
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

6   Model quality ⓘ

6.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: KCR

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

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	0.1±0.2
All	All	0	1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	121	ARG	Sidechain	1

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	1085	1094	1086	49±7
2	B	23	25	25	1±1
All	All	22160	22380	22220	996

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 22.

5 of 207 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:97:GLU:HB2	1:A:98:PRO:HD2	0.92	1.41	15	8
1:A:39:VAL:HG23	1:A:74:TYR:HB2	0.84	1.47	15	20
1:A:15:GLU:HA	1:A:121:ARG:NH2	0.81	1.90	20	1
1:A:121:ARG:HH21	1:A:122:CYS:HA	0.81	1.32	20	1
1:A:97:GLU:HB3	1:A:98:PRO:HD2	0.81	1.53	11	12

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	127/141 (90%)	112±2 (88±2%)	11±2 (9±2%)	4±0 (3±0%)	7	41
2	B	3/13 (23%)	3±0 (95±12%)	0±0 (5±12%)	0±0 (0±0%)	100	100
All	All	2600/3080 (84%)	2303 (89%)	224 (9%)	73 (3%)	8	42

All 4 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	72	PRO	20
1	A	73	PRO	20
1	A	97	GLU	20
1	A	98	PRO	13

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	119/129 (92%)	106±2 (89±2%)	13±2 (11±2%)	10	54
2	B	2/7 (29%)	2±0 (100±0%)	0±0 (0±0%)	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2420/2720 (89%)	2160 (89%)	260 (11%)	10 54

5 of 37 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	18	HIS	20
1	A	23	ARG	20
1	A	37	VAL	20
1	A	55	VAL	20
1	A	121	ARG	20

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	KCR	B	218	2	12,13,14	0.75±0.01	0±0 (0±1%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	KCR	B	218	2	9,14,16	1.33±0.03	<mark>1±0 (13±4%)</mark>

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	KCR	B	218	2	-	0±0,12,13,15	-

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	B	218	KCR	CX-CY	2.01	1.38	1.30	3	1

All 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(°)	Ideal(°)	Models	
								Worst	Total
2	B	218	KCR	CH3-CY-CX	3.07	119.28	125.34	12	20
2	B	218	KCR	CX-CH-NZ	2.16	119.04	114.97	13	4

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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.

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

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 85% 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: *assigned_chem_shift_list_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	1795
Number of shifts mapped to atoms	1795
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	12

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	143	0.09 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_\beta$	140	-0.14 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	110	0.60 ± 0.16	Should be applied
^{15}N	115	0.14 ± 0.39	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 85%, i.e. 1652 atoms were assigned a chemical shift out of a possible 1951. 0 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	595/640 (93%)	248/257 (96%)	235/262 (90%)	112/121 (93%)
Sidechain	890/1089 (82%)	587/701 (84%)	296/334 (89%)	7/54 (13%)

Continued on next page...

Continued from previous page...

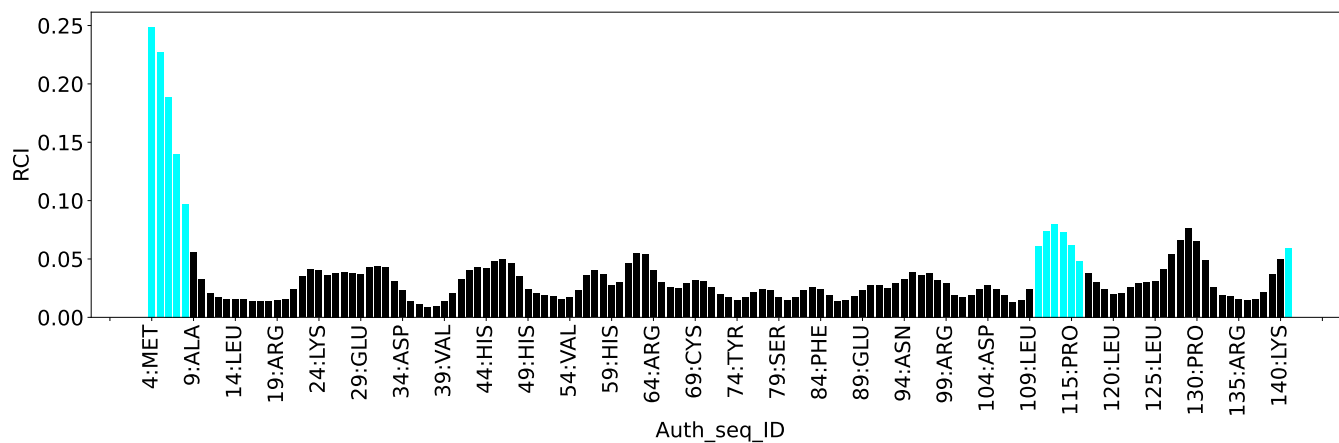|	**Total**	**^{1}H**	**^{13}C**	**^{15}N**
Aromatic	167/222 (75%)	84/109 (77%)	82/96 (85%)	1/17 (6%)
Overall	1652/1951 (85%)	919/1067 (86%)	613/692 (89%)	120/192 (62%)

7.1.4 Statistically unusual chemical shifts ⓘ

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	27	THR	HG1	5.57	$0.08 - 2.19$	21.0
1	A	78	GLU	HB2	-0.68	$1.00 - 3.05$	-13.2
1	A	29	GLU	HB2	0.23	$1.00 - 3.05$	-8.7
1	A	53	LYS	HE2	1.54	$1.95 - 3.88$	-7.1
1	A	53	LYS	HE3	1.62	$1.92 - 3.89$	-6.5
1	A	107	LEU	HD21	-0.88	$-0.65 - 2.13$	-5.8
1	A	107	LEU	HD22	-0.88	$-0.65 - 2.13$	-5.8
1	A	107	LEU	HD23	-0.88	$-0.65 - 2.13$	-5.8
1	A	107	LEU	HD11	-0.64	$-0.61 - 2.12$	-5.1
1	A	107	LEU	HD12	-0.64	$-0.61 - 2.12$	-5.1
1	A	107	LEU	HD13	-0.64	$-0.61 - 2.12$	-5.1
1	A	107	LEU	HG	-0.14	$-0.13 - 3.16$	-5.0

7.1.5 Random Coil Index (RCI) plots ⓘ

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:

Random coil index (RCI) for chain B: