wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 02:08 AM EDT

PDB ID	:	2L5A
Title	:	Structural basis for recognition of centromere specific histone H3 variant by nonhistone Scm3
Authors	:	Zhou, Z.; Feng, H.; Zhou, B.; Ghirlando, R.; Hu, K.; Zwolak, A.; Jenkins, L.; Xiao, H.; Tjandra, N.; Wu, C.; Bai, Y.
Deposited on	:	2010-10-28

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:

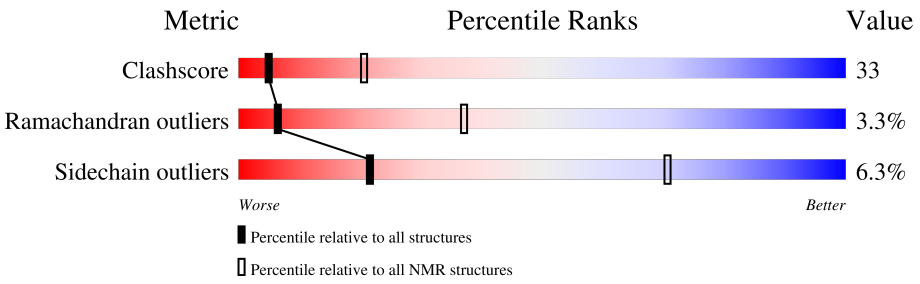
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 was not calculated.

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	235	29% 23% • 41% •

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 10 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:7-A:48, A:89-A:125, A:136-A:140, A:176-A:220 (129)	0.59	10

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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3657 atoms, of which 1854 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone H3-like centromeric protein CSE4, Protein SCM3, Histone H4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	225	3657	1131	1854	321	345	6	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-9	MET	-	expression tag	UNP P36012
A	-8	HIS	-	expression tag	UNP P36012
A	-7	HIS	-	expression tag	UNP P36012
A	-6	HIS	-	expression tag	UNP P36012
A	-5	HIS	-	expression tag	UNP P36012
A	-4	HIS	-	expression tag	UNP P36012
A	-3	HIS	-	expression tag	UNP P36012
A	-2	LYS	-	expression tag	UNP P36012
A	-1	LYS	-	expression tag	UNP P36012
A	78	LEU	-	linker	UNP P36012
A	79	VAL	-	linker	UNP P36012
A	80	PRO	-	linker	UNP P36012
A	81	ARG	-	linker	UNP P36012
A	82	GLY	-	linker	UNP P36012
A	83	SER	-	linker	UNP P36012

4 Residue-property plots (i)

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 H3-like centromeric protein CSE4, Protein SCM3, Histone H4

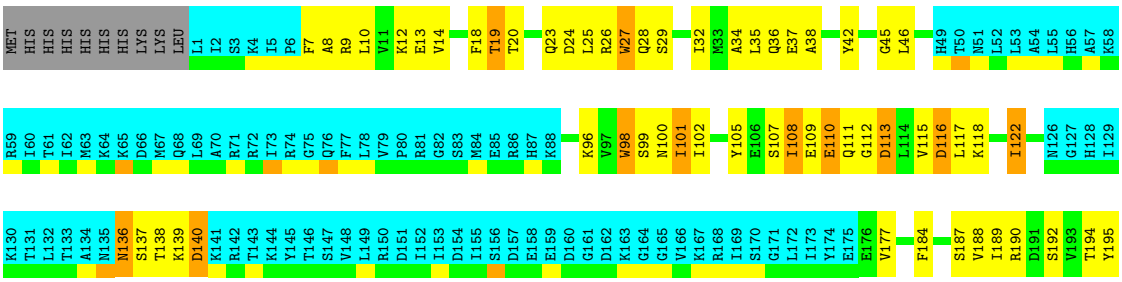
Chain A: 29% 23% • 41% •

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

The representative model is number 10. Colouring as in section 4.1 above.

• Molecule 1: Histone H3-like centromeric protein CSE4, Protein SCM3, Histone H4

Chain A: 25% 26% 5% 41% •



T196
E197
H198
R201
K202
T203
V204
T205
S206
L207
D208
V209
V210
Y211
A212
L213
Q216
Y221
G222
F223
G224
G225

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing, molecular dynamics.*

Of the 50 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	structure solution	2.25
X-PLOR NIH	refinement	2.25

No chemical shift data was provided.

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	0.71±0.01	0±0/1050 (0.0± 0.0%)	0.75±0.01	0±0/1417 (0.0± 0.0%)
All	All	0.71	0/21000 (0.0%)	0.75	3/28340 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	105	TYR	CB-CG-CD2	-5.24	117.86	121.00	18	3

There are no chirality outliers.

There are no planarity outliers.

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	1038	1047	1047	68±5
All	All	20760	20940	20940	1370

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 33.

5 of 295 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:105:TYR:CD2	1:A:108:ILE:HD11	0.88	2.04	19	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:20:THR:O	1:A:23:GLN:OE1	0.87	1.93	9	15
1:A:185:LEU:O	1:A:189:ILE:HG22	0.86	1.70	1	11
1:A:192:SER:O	1:A:196:THR:OG1	0.86	1.92	9	20
1:A:118:LYS:NZ	1:A:206:SER:H	0.85	1.70	1	5

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	129/235 (55%)	109±1 (84±1%)	16±2 (12±1%)	4±1 (3±1%)	6	37
All	All	2580/4700 (55%)	2176 (84%)	318 (12%)	86 (3%)	6	37

5 of 14 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	110	GLU	16
1	A	113	ASP	15
1	A	136	ASN	14
1	A	19	THR	14
1	A	140	ASP	8

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	116/208 (56%)	109±2 (94±2%)	7±2 (6±2%)	21	70
All	All	2320/4160 (56%)	2174 (94%)	146 (6%)	21	70

5 of 21 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	27	TRP	20
1	A	98	TRP	20
1	A	101	ILE	20
1	A	122	ILE	14
1	A	138	THR	14

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 (i)

No chemical shift data were provided