wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 04:59 AM BST

PDB ID	:	2RNW
Title	:	The Structural Basis for Site-Specific Lysine-Acetylated Histone Recognition by the Bromodomains of the Human Transcriptional Co-Activators PCAf and CBP
Authors	:	Zeng, L.; Zhang, Q.; Gerona-Navarro, G.; Zhou, M.M.
Deposited on	:	2008-02-03

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
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

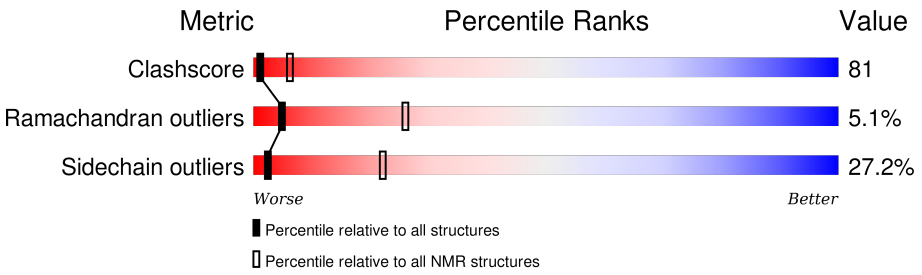
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	118	11% 49% 19% • 20%
2	B	15	100%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 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:724-A:749, A:764-A:831 (94)	0.17	3

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 4 clusters and 8 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2211 atoms, of which 1112 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone acetyltransferase PCAF.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	118	1976	636	989	164	180	7	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	715	GLY	-	EXPRESSION TAG	UNP Q92831
A	716	SER	-	EXPRESSION TAG	UNP Q92831
A	717	HIS	-	EXPRESSION TAG	UNP Q92831
A	718	MET	-	EXPRESSION TAG	UNP Q92831

- Molecule 2 is a protein called Histone H3.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	15	235	65	123	25	22	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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 acetyltransferase PCAF

Chain A: 11% 49% 19% • 20%

G715 S716 H717 M718 S719 K720 E721 P722 R723 D724 Q727 L728 Y729 S730 T731 L732 K733 S734 I735 L736 Q737 Q738 V739 K740 S741 H742 Q743 S744 A745 W746 P747 F748 M749 E750 P751 V752 K753 R754 T755 E756 A757 P758 G759 Y760 Y761 E762 V763 I764 R765 F766 P767 M768 D769 L770 K771 T772 M773 S774 E775

R776 L777 K778 N779 R780 Y781 Y782 V783 S784 K785 K786 L787 F788 M789 A790 D791 L792 Q793 R794 V795 F796 T797 N798 C799 K800 E801 Y802 N803 P804 P805 E806 S807 E808 Y809 Y810 K811 C812 A813 L816 E817 K818 F819 F820 F821 S822 K823 I824 K825 E826 A827 G828 L829 I830 D831 K832

- Molecule 2: Histone H3

Chain B: 100%

A1 R2 T3 K4 Q5 T6 A7 R8 K9 S10 T11 G12 G13 K14 A15

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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: Histone acetyltransferase PCAF

Chain A: 14% 47% 18% 20%

G715 S716 H717 M718 S719 K720 E721 P722 R723 D724 Q727 L728 Y729 S730 T731 L732 K733 S734 I735 L736 Q737 Q738 V739 K740 S741 H742 Q743 S744 A745 W746 P747 F748 M749 E750 P751 V752 K753 R754 T755 E756 A757 P758 G759 Y760 Y761 E762 V763 I764 R765 F766 P767 M768 D769 L770 K771 T772 M773 S774

L777 K778 N779 R780 Y781 Y782 V783 S784 K785 K786 L787 F788 M789 A790 D791 L792 Q793 R794 V795 F796 T797 N798 C799 K800 E801 Y802 N803 P804 P805 E806 Y809 Y810 K811 C812 A813 L816 E817 F820 F821 S822 K823 I824 K825 E826 A827 G828 L829 I830 D831 K832

- Molecule 2: Histone H3

Chain B: 100%

A1
R2
T3
K4
Q5
T6
A7
R8
K9
S10
T11
G12
G13
K14
A15

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, 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
CNS	structure solution	2.1
CNS	refinement	2.1

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

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

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	0.49±0.01	0±0/813 (0.0±0.0%)	0.65±0.02	0±1/1093 (0.0±0.0%)
All	All	0.49	0/16260 (0.0%)	0.65	4/21860 (0.0%)

There are no bond-length outliers.

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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	802	TYR	CB-CG-CD2	-7.39	116.57	121.00	14	3
1	A	802	TYR	CB-CG-CD1	6.17	124.70	121.00	14	1

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	792	796	795	129±5
2	B	0	0	0	0±0
All	All	15840	15920	15900	2572

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 81.

5 of 405 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:824:ILE:HD12	1:A:830:ILE:HD13	1.15	1.16	17	20
1:A:824:ILE:HG22	1:A:829:LEU:HD12	0.90	1.42	15	19
1:A:745:ALA:HB1	1:A:770:LEU:HD11	0.88	1.46	11	8
1:A:733:LYS:HA	1:A:777:LEU:HD21	0.88	1.45	18	20
1:A:821:PHE:HA	1:A:824:ILE:HD11	0.85	1.48	15	20

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	94/118 (80%)	71±2 (76±2%)	18±3 (19±3%)	5±1 (5±1%)	5 26
2	B	0	-	-	-	-
All	All	1880/2660 (71%)	1424 (76%)	360 (19%)	96 (5%)	5 26

5 of 11 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	830	ILE	20
1	A	829	LEU	19
1	A	747	PRO	18
1	A	790	ALA	16
1	A	724	ASP	15

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	89/110 (81%)	65±3 (73±3%)	24±3 (27±3%)	2 22
2	B	0	-	-	-
All	All	1780/2380 (75%)	1295 (73%)	485 (27%)	2 22

5 of 52 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	771	LYS	20
1	A	816	LEU	20
1	A	824	ILE	20
1	A	810	TYR	20
1	A	792	LEU	20

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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	ALY	B	9	2	9,11,12	0.62±0.05	0±0 (0±0%)

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	ALY	B	9	2	10,12,14	1.07±0.06	0±0 (0±0%)

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	ALY	B	9	2	-	2±0 ,8,10,12	0±0,0,0,0

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique torsion outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	B	9	ALY	OH-CH-NZ-CE	9
2	B	9	ALY	CH3-CH-NZ-CE	8

There are no ring outliers.

6.5 Carbohydrates (i)

There are no carbohydrates 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