



# wwPDB NMR Structure Validation Summary Report ⓘ

Apr 26, 2016 – 03:50 PM BST

PDB ID : 1JM4  
Title : NMR Structure of P/CAF Bromodomain in Complex with HIV-1 Tat Peptide  
Authors : Mujtaba, S.; He, Y.; Zeng, L.; Farooq, A.; Carlson, J.E.; Ott, M.; Verdin, E.; Zhou, M.-M.  
Deposited on : 2001-07-17

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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:

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

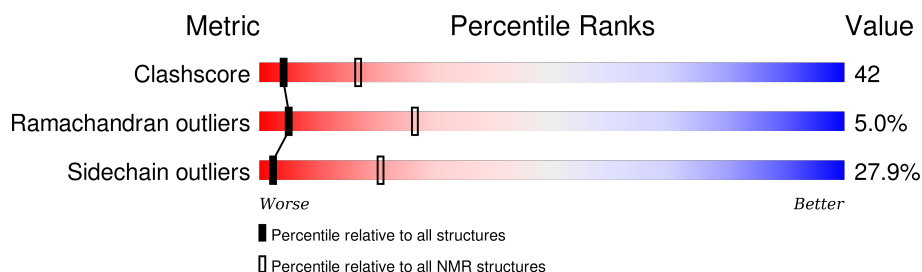
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 77%.

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	11	100%
2	B	118	19% 65% 5% 10%

## 2 Ensemble composition and analysis

This entry contains 25 models. Model 13 is the overall representative, medoid model (most similar to other models).

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	B:721-B:752, B:757-B:830 (106)	0.42	13

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 7 single-model clusters were found.

Cluster number	Models
1	2, 10, 11, 16, 19
2	3, 9, 12, 13, 18
3	5, 7
4	8, 14
5	4, 6
6	20, 23
Single-model clusters	1; 15; 17; 21; 22; 24; 25

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2188 atoms, of which 1098 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called HIV-1 Tat Peptide.

Mol	Chain	Residues	Atoms						Trace
1	A	11	Total	C	H	N	O	S	0
			214	60	111	26	16	1	

- Molecule 2 is a protein called P300/CBP-associated Factor.

Mol	Chain	Residues	Atoms						Trace
2	B	118	Total	C	H	N	O	S	0
			1974	636	987	164	180	7	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	715	GLY	-	CLONING ARTIFACT	UNP Q92831
B	716	SER	-	CLONING ARTIFACT	UNP Q92831
B	717	HIS	-	CLONING ARTIFACT	UNP Q92831
B	718	MET	-	CLONING ARTIFACT	UNP Q92831

## 4 Residue-property plots

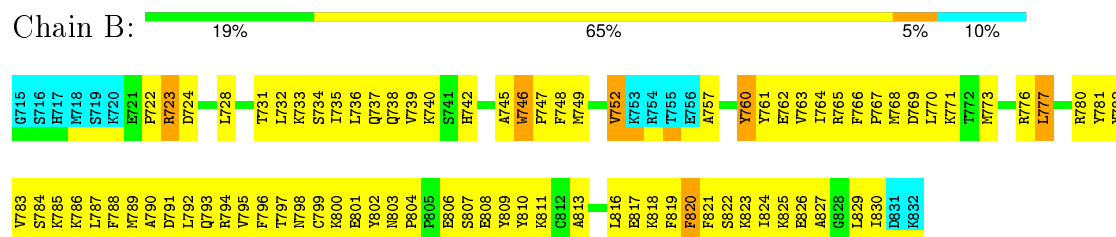
### 4.1 Average score per residue in the NMR ensemble

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: HIV-1 Tat Peptide



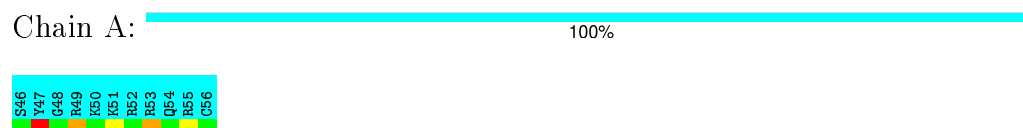
- Molecule 2: P300/CBP-associated Factor



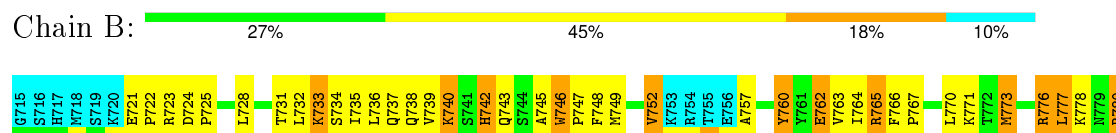
### 4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 13. Colouring as in section 4.1 above.

- Molecule 1: HIV-1 Tat Peptide



- Molecule 2: P300/CBP-associated Factor



Y781	Y782	Y783	S784	K785	L786	L787	F788	M789	A790	D791	L792	Q793	K794	W795	F796	L797	N798	C799	K800	E801	Y802	M803	P804	P805	E806	S807	E808	Y809	Y810	K811	G812	A813	L816	E817	K818	F819	F820	F821	S822	K823	L824	K825	E826	L829	I830	D831	F832
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *Distance Geometry/Simulated Annealing*.

Of the 100 calculated structures, 25 were deposited, based on the following criterion: *Back Calculated Data Agree with Experimental NOESY Spectrum, Structures with Acceptable Covalent Geometry, Structures with Favorable Non-bond Energy, Structures with the Least Restraint Violations, Structures with the Lowest Energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR/ARIA	structure solution	2
X-PLOR 3.851	refinement	

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)	BMRB entry 4312
Number of chemical shift lists	1
Total number of shifts	1197
Number of shifts mapped to atoms	1190
Number of unparsed shifts	0
Number of shifts with mapping errors	7
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	77%

No validations of the models with respect to experimental NMR restraints is performed at this time.

## 6 Model quality

### 6.1 Standard geometry

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
2	B	1.01±0.00	0±0/916 (0.0±0.0%)	1.27±0.01	0±0/1236 (0.0±0.0%)
All	All	1.01	0/22900 (0.0%)	1.27	3/30900 (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(°)	Ideal(°)	Models	
								Worst	Total
2	B	749	MET	N-CA-CB	-5.74	100.26	110.60	23	2
2	B	757	ALA	N-CA-CB	-5.12	102.94	110.10	20	1

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	0	0	0	0±0
2	B	890	887	887	75±10
All	All	22250	22175	22175	1863

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 42.

5 of 607 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:824:ILE:HD12	2:B:830:ILE:HD13	1.06	1.24	10	2
2:B:796:PHE:CE1	2:B:816:LEU:HD13	0.94	1.97	17	5
2:B:760:TYR:CE2	2:B:764:ILE:HG21	0.92	1.98	7	11
2:B:732:LEU:HD22	2:B:788:PHE:CG	0.91	2.00	13	14
2:B:824:ILE:HG22	2:B:829:LEU:HD12	0.90	1.41	23	10

## 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	0	-	-	-	-	
2	B	106/118 (90%)	84±3 (79±2%)	17±3 (16±3%)	5±1 (5±1%)	5	26
All	All	2650/3225 (82%)	2105 (79%)	413 (16%)	132 (5%)	5	26

5 of 28 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	723	ARG	22
2	B	767	PRO	16
2	B	752	VAL	13
2	B	760	TYR	13
2	B	761	TYR	10

### 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	0	-	-	-

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	99/110 (90%)	71±3 (72±3%)	28±3 (28±3%)	2	21
All	All	2475/2975 (83%)	1784 (72%)	691 (28%)	2	21

5 of 75 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)
2	B	746	TRP	25
2	B	771	LYS	25
2	B	820	PHE	24
2	B	777	LEU	23
2	B	740	LYS	21

### 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
1	ALY	A	50	1	9,11,12	0.83±0.02	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
1	ALY	A	50	1	10,12,14	1.42±0.03	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
1	ALY	A	50	1	-	0±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)
1	A	50	ALY	OH-CH-NZ-CE	10
1	A	50	ALY	CH3-CH-NZ-CE	10

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

The completeness of assignment taking into account all chemical shift lists is 77% for the well-defined parts and 65% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 4312

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	1197
Number of shifts mapped to atoms	1190
Number of unparsed shifts	0
Number of shifts with mapping errors	7
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

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

- Residue not found in structure. First 5 (of 7) occurrences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
B	52	SER	HB2	2.91	-1.0	1
B	52	SER	HA	4.98	-1.0	1
B	52	SER	HB3	3.03	-1.0	1
B	52	SER	CA	54.62	-1.0	1
B	52	SER	H	8.39	-1.0	1

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	109	$-0.63 \pm 0.21$	Should be applied
$^{13}\text{C}_\beta$	107	$0.10 \pm 0.15$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	0	—	—

*Continued on next page...*

Continued from previous page...

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{15}\text{N}$	101	$0.92 \pm 0.44$	Should be applied

### 7.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 77%, i.e. 1084 atoms were assigned a chemical shift out of a possible 1413. 13 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	395/514 (77%)	198/204 (97%)	102/212 (48%)	95/98 (97%)
Sidechain	564/753 (75%)	336/451 (75%)	223/267 (84%)	5/35 (14%)
Aromatic	125/146 (86%)	65/77 (84%)	59/67 (88%)	1/2 (50%)
Overall	1084/1413 (77%)	599/732 (82%)	384/546 (70%)	101/135 (75%)

### 7.1.4 Statistically unusual chemical shifts [i](#)

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.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
2	B	747	PRO	HG2	-0.93	3.48 – 0.38	-9.2
2	B	747	PRO	HB2	-0.78	3.82 – 0.32	-8.1
2	B	760	TYR	HD1	5.12	8.44 – 5.44	-6.1
2	B	805	PRO	HA	2.42	6.05 – 2.75	-6.0
2	B	747	PRO	HD2	1.57	5.45 – 1.85	-5.8
2	B	744	SER	H	11.67	11.23 – 5.33	5.7
2	B	747	PRO	HG3	0.23	3.56 – 0.26	-5.1

### 7.1.5 Random Coil Index (RCI) plots [i](#)

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.

Random coil index (RCI) for chain B:

