wwPDB NMR Structure Validation Summary Report (i)

Nov 5, 2023 – 03:39 PM EST

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

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 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	Percentile Ranks	Value
Clashscore		42
Ramachandran outliers		5.0%
Sidechain outliers		27.9%

Worse   *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all NMR structures
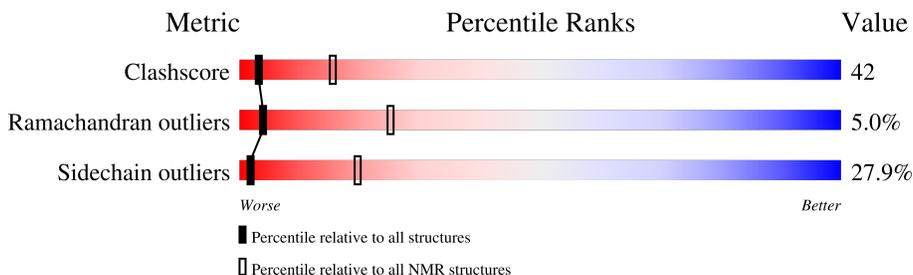
Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	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)	0.73	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 (i)

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
			Total	C	H	N	O	S	
1	A	11	214	60	111	26	16	1	0

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

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

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, 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: HIV-1 Tat Peptide

Chain A:

• Molecule 2: P300/CBP-associated Factor

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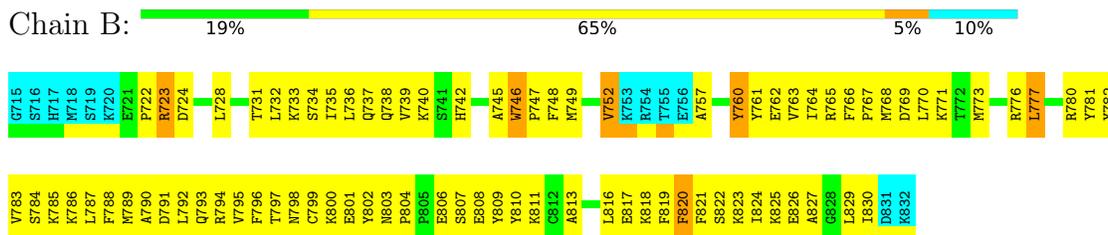
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

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

• Molecule 1: HIV-1 Tat Peptide

Chain A:

• Molecule 2: P300/CBP-associated Factor

Chain B:

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 L829 I830 D831 K832

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	

No chemical shift data was provided.

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(^{o})	Ideal(^{o})	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 592 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models 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 ⓘ

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	0	-	-	-	-	
2	B	106/118 (90%)	84±3 (79±2%)	17±3 (16±3%)	5±1 (5±1%)	4	25
All	All	2650/3225 (82%)	2105 (79%)	413 (16%)	132 (5%)	4	25

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 ⓘ

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	19
All	All	2475/2975 (83%)	1784 (72%)	691 (28%)	2	19

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	733	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	10,11,12	0.80±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	7,12,14	1.34±0.03	<mark>1±0 (14±2%)</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
1	ALY	A	50	1	-	0±0,9,10,12	-

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	50	ALY	CD-CE-NZ	2.78	104.27	112.21	14	25
1	A	50	ALY	CE-NZ-CH	2.20	119.18	122.56	12	1

There are no chirality outliers.

All unique torsion outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	A	50	ALY	CH3-CH-NZ-CE	2

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 ⓘ

No chemical shift data were provided