Full wwPDB NMR Structure Validation Report (i)

Feb 14, 2022 – 08:16 AM EST

PDB ID	:	1JFW
Title	:	HOMONUCLEAR AND HETERONUCLEAR 1H-13C NUCLEAR MAGNETIC RESONANCE ASSIGNMENT AND STRUCTURAL CHARACTERIZATION OF A HIV-1 TAT PROTEIN
Authors	:	Peloponese, J.M.; Gregoire, C.; Opi, S.; Esquieu, D.
Deposited on	:	2001-06-22

This is a Full wwPDB NMR Structure Validation 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 following versions of software and data (see references (i)) were used in the production of this report:

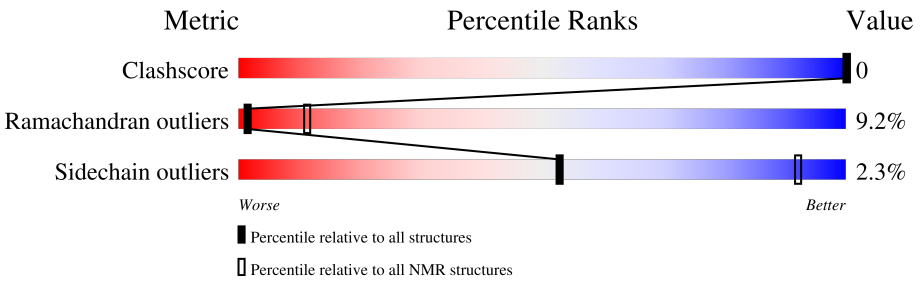
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	86	72% 26% •

2 Ensemble composition and analysis (i)

This entry contains 11 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 8 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	A:2-A:85 (84)	0.47	8

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 3 clusters and 2 single-model clusters were found.

Cluster number	Models
1	4, 5, 8, 9, 10
2	6, 7
3	1, 3
Single-model clusters	2; 11

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1354 atoms, of which 673 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TAT PROTEIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	86	1354	418	673	134	121	8	0

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: TAT PROTEIN

Chain A: 72% 26%

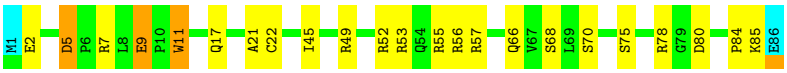
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: TAT PROTEIN

Chain A: 71% 23%



4.2.2 Score per residue for model 2

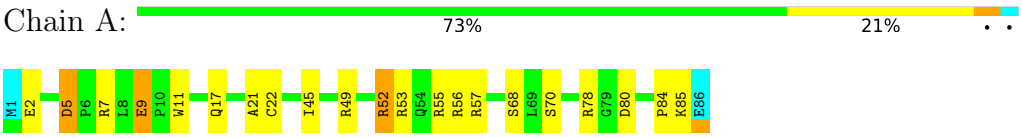
• Molecule 1: TAT PROTEIN

Chain A: 74% 22%

4.2.3 Score per residue for model 3

- Molecule 1: TAT PROTEIN

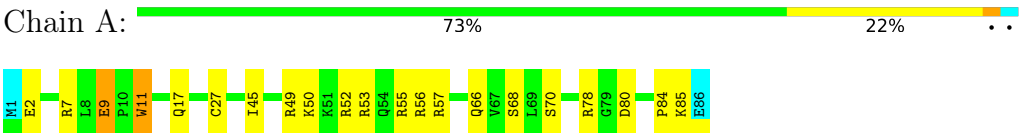
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: TAT PROTEIN

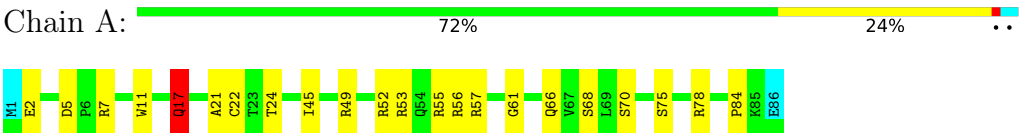
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: TAT PROTEIN

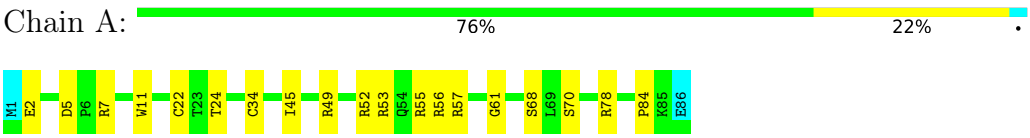
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: TAT PROTEIN

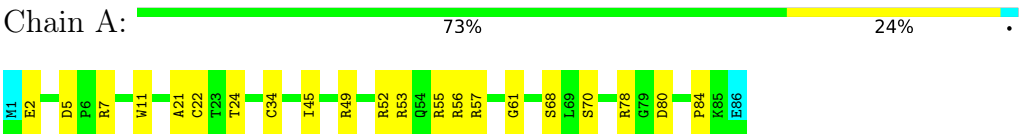
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: TAT PROTEIN

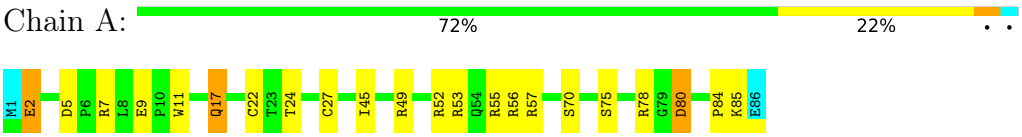
Chain A:



4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: TAT PROTEIN

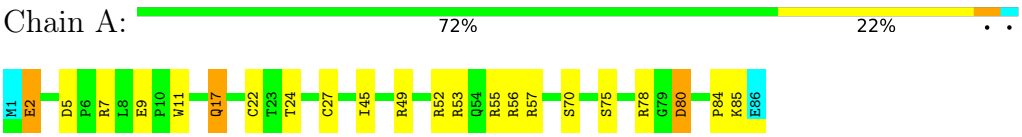
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: TAT PROTEIN

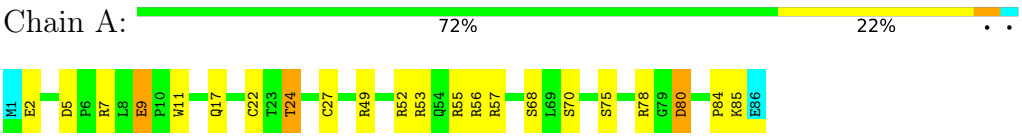
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: TAT PROTEIN

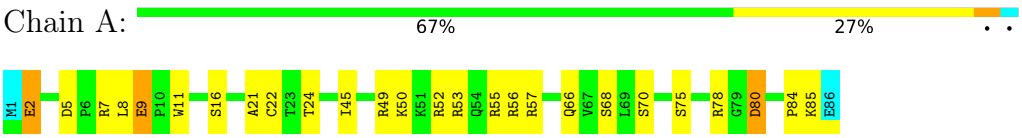
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: TAT PROTEIN

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *SIMULATED ANNEALING, RANDOM LOOP RESEARCH VALIDATED WITH NOE BACK-CALCULATION ENERGY MINIMIZATION AND MOLECULAR DYNAMIC.*

Of the 100 calculated structures, 11 were deposited, based on the following criterion: *BACK CALCULATED DATA AGREE WITH EXPERIMEN TAL NOESY SPECTRUM, STRUCTURES WITH ACCEPTABLE COVALENT GEOMETRY,STRUCTURES WITH FAVORABLE NON-BOND ENERGY,STRUCTURES WITH THE LEAST RESTRAINT VIOLATIONS,STRUCTURES WITH THE LOWEST ENERGY, TARGET FUNCTION.*

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

Software name	Classification	Version
DISCOVER 95,	refinement	97
Felix	structure solution	95
97 AND	structure solution	98
NMR REFINE	structure solution	97
HOMOLOGY	structure solution	95
97	structure solution	
98	structure solution	

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	1.37±0.06	1±1/683 (0.2± 0.1%)	1.73±0.04	20±4/923 (2.2± 0.4%)
All	All	1.37	14/7513 (0.2%)	1.73	223/10153 (2.2%)

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	1.2±1.1
All	All	0	13

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	9	GLU	CD-OE2	9.93	1.36	1.25	9	7
1	A	2	GLU	CD-OE2	9.92	1.36	1.25	9	7

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	57	ARG	NE-CZ-NH1	9.81	125.20	120.30	7	11
1	A	7	ARG	NE-CZ-NH1	9.68	125.14	120.30	7	11
1	A	56	ARG	NE-CZ-NH1	9.53	125.06	120.30	6	11
1	A	52	ARG	NE-CZ-NH1	9.29	124.94	120.30	5	11
1	A	53	ARG	NE-CZ-NH1	9.15	124.88	120.30	6	11
1	A	78	ARG	NE-CZ-NH1	8.69	124.65	120.30	3	11
1	A	55	ARG	NE-CZ-NH1	8.62	124.61	120.30	9	11
1	A	49	ARG	NE-CZ-NH1	8.57	124.59	120.30	5	11

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	7	ARG	NE-CZ-NH2	-7.80	116.40	120.30	10	9
1	A	5	ASP	CB-CG-OD2	-7.40	111.64	118.30	9	6
1	A	52	ARG	NE-CZ-NH2	-7.11	116.75	120.30	11	9
1	A	5	ASP	CB-CG-OD1	6.67	124.30	118.30	10	6
1	A	57	ARG	NE-CZ-NH2	-6.64	116.98	120.30	11	10
1	A	56	ARG	NE-CZ-NH2	-6.59	117.00	120.30	6	10
1	A	47	TYR	CB-CG-CD2	-6.53	117.08	121.00	2	1
1	A	80	ASP	CB-CG-OD2	-6.44	112.50	118.30	11	6
1	A	53	ARG	NE-CZ-NH2	-6.34	117.13	120.30	5	10
1	A	80	ASP	CB-CG-OD1	6.28	123.95	118.30	11	7
1	A	49	ARG	NE-CZ-NH2	-6.14	117.23	120.30	5	10
1	A	11	TRP	CD1-NE1-CE2	-5.93	103.67	109.00	7	11
1	A	55	ARG	NE-CZ-NH2	-5.92	117.34	120.30	9	10
1	A	84	PRO	O-C-N	-5.75	113.49	122.70	3	9
1	A	85	LYS	O-C-N	-5.66	113.64	122.70	10	5
1	A	78	ARG	NE-CZ-NH2	-5.59	117.50	120.30	3	9
1	A	17	GLN	CB-CA-C	5.54	121.47	110.40	5	3
1	A	24	THR	N-CA-CB	5.50	120.75	110.30	10	1
1	A	2	GLU	CB-CA-C	5.37	121.14	110.40	11	3

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	21	ALA	Peptide	5
1	A	11	TRP	Peptide	3
1	A	84	PRO	Peptide	1
1	A	79	GLY	Peptide	1
1	A	85	LYS	Peptide	1
1	A	8	LEU	Peptide	1
1	A	16	SER	Peptide	1

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
All	All	7293	7217	7359	-

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 -.

There are no clashes.

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	84/86 (98%)	45±3 (54±3%)	31±3 (37±3%)	8±1 (9±2%)	1 11
All	All	924/946 (98%)	495 (54%)	344 (37%)	85 (9%)	1 11

All 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	22	CYS	10
1	A	45	ILE	10
1	A	70	SER	10
1	A	68	SER	9
1	A	17	GLN	8
1	A	75	SER	7
1	A	24	THR	7
1	A	2	GLU	6
1	A	27	CYS	5
1	A	80	ASP	5
1	A	61	GLY	3
1	A	50	LYS	2
1	A	34	CYS	2
1	A	65	HIS	1

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	76/78 (97%)	74±1 (98±2%)	2±1 (2±2%)	53	92
All	All	836/858 (97%)	817 (98%)	19 (2%)	53	92

All 7 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	5	ASP	5
1	A	9	GLU	5
1	A	66	GLN	4
1	A	85	LYS	2
1	A	69	LEU	1
1	A	52	ARG	1
1	A	17	GLN	1

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