



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 24, 2016 – 07:48 PM EDT

PDB ID : 5L1Z  
Title : TAR complex with HIV-1 Tat-AFF4-P-TEFb  
Authors : Schulze-Gahmen, U.; Hurley, J.  
Deposited on : 2016-07-29  
Resolution : 5.90 Å(reported)

This is a wwPDB X-ray 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/XrayValidationReportHelp>

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:

MolProbity : 4.02b-467  
Mogul : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027939  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027939

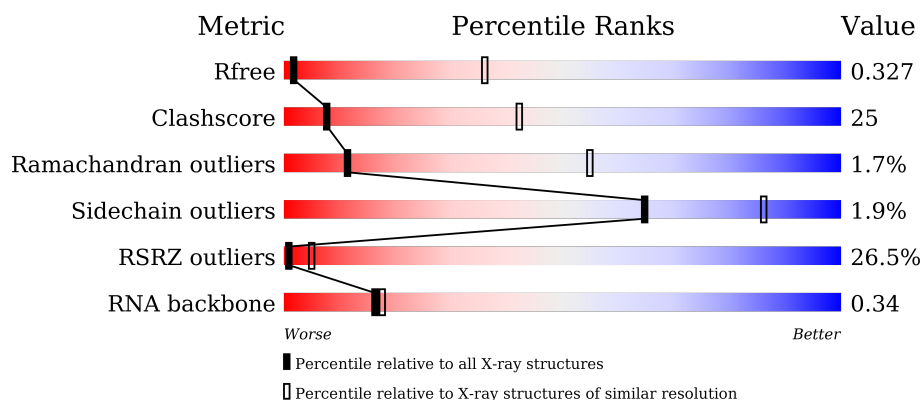
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 5.90 Å.

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)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1000 (8.00-3.66)
Clashscore	102246	1048 (8.00-3.70)
Ramachandran outliers	100387	1021 (8.00-3.66)
Sidechain outliers	100360	1010 (8.00-3.64)
RSRZ outliers	91569	1015 (8.00-3.64)
RNA backbone	2183	1103 (8.70-2.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	330	<div> <div>21%</div> <div> <div>53%</div> <div>40%</div> <div>6%</div> </div> </div>
2	B	264	<div> <div>33%</div> <div> <div>51%</div> <div>45%</div> <div>6%</div> </div> </div>
3	C	36	<div> <div>25%</div> <div> <div>39%</div> <div>44%</div> <div>11%</div> <div>6%</div> </div> </div>
4	D	58	<div> <div>14%</div> <div> <div>34%</div> <div>48%</div> <div>14%</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	N	23	<div><div></div><div></div><div></div><div></div><div></div></div>

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 5713 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Cyclin-dependent kinase 9.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	311	Total	C	N	O	P	S	0	0	0
			2515	1611	434	454	1	15			

- Molecule 2 is a protein called Cyclin-T1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	255	Total	C	N	O	S	0	0	0
			2085	1333	362	380	10			

- Molecule 3 is a protein called AF4/FMR2 family member 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	34	Total	C	N	O	S	0	0	0
			276	175	43	56	2			

- Molecule 4 is a protein called Protein Tat.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	50	Total	C	N	O	S	0	0	0
			388	246	68	66	8			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	ACE	-	acetylation	UNP P69697

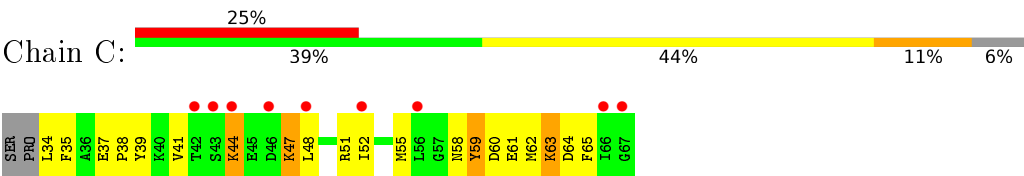
- Molecule 5 is a RNA chain called RNA (5'-R(P\*AP\*GP\*AP\*UP\*CP\*UP\*GP\*AP\*GP\*CP\*CP\*UP\*GP\*GP\*GP\*AP\*GP\*CP\*UP\*CP\*UP\*CP\*U)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	N	21	Total	C	N	O	P	0	0	0
			447	199	78	149	21			

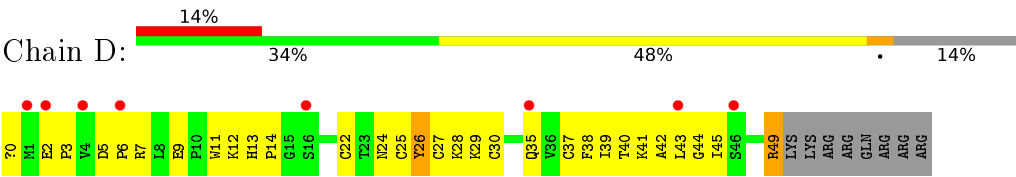
- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	2	Total 2	Zn 2	0	0

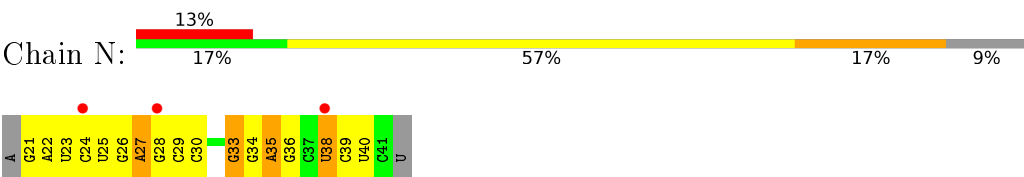




● Molecule 4: Protein Tat



● Molecule 5: RNA (5'-R(P\*AP\*GP\*AP\*UP\*CP\*UP\*GP\*AP\*GP\*CP\*CP\*UP\*GP\*GP\*GP\*A  
P\*GP\*CP\*UP\*CP\*UP\*CP\*U)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	146.87Å 146.87Å 103.75Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	500.00 – 5.90 48.07 – 5.90	Depositor EDS
% Data completeness (in resolution range)	100.0 (500.00–5.90) 100.0 (48.07–5.90)	Depositor EDS
$R_{merge}$	0.26	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.14 (at 5.73Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.217 , 0.313 0.237 , 0.327	Depositor DCC
$R_{free}$ test set	345 reflections (9.67%)	DCC
Wilson B-factor (Å <sup>2</sup> )	360.9	Xtriage
Anisotropy	0.463	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 380.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	0.105 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.84	EDS
Total number of atoms	5713	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	424.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.93% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: TPO, ZN, ACE

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 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	0.23	0/2552	0.44	0/3440
2	B	0.24	0/2139	0.45	0/2915
3	C	0.29	0/280	0.47	0/371
4	D	0.25	0/398	0.43	0/538
5	N	0.27	0/498	0.66	0/774
All	All	0.24	0/5867	0.47	0/8038

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2515	0	2572	106	1
2	B	2085	0	2060	118	0
3	C	276	0	271	35	0
4	D	388	0	377	47	0
5	N	447	0	227	23	1
6	D	2	0	0	0	0
All	All	5713	0	5507	284	2

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 25.

The worst 5 of 284 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:D:49:ARG:H	4:D:49:ARG:NE	1.69	0.90
2:B:248:THR:HG21	4:D:43:LEU:HD22	1.53	0.89
4:D:25:CYS:C	5:N:33:G:H22	1.78	0.87
1:A:268:VAL:HG21	1:A:271:GLN:HG2	1.58	0.84
2:B:251:ARG:H	2:B:251:ARG:HD3	1.43	0.83

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:N:21:G:N2	5:N:21:G:N2[4_558]	2.02	0.18
1:A:261:LEU:CD2	1:A:261:LEU:CD2[4_557]	2.16	0.04

## 5.3 Torsion angles [i](#)

### 5.3.1 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 X-ray entries followed by that with respect to entries of similar resolution.

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	304/330 (92%)	268 (88%)	30 (10%)	6 (2%)	9	51
2	B	253/264 (96%)	237 (94%)	15 (6%)	1 (0%)	39	80
3	C	32/36 (89%)	23 (72%)	6 (19%)	3 (9%)	1	16
4	D	48/58 (83%)	43 (90%)	4 (8%)	1 (2%)	9	50
All	All	637/688 (93%)	571 (90%)	55 (9%)	11 (2%)	11	55

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	149	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	177	ALA
1	A	190	VAL
3	C	59	TYR
3	C	60	ASP

### 5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution.

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	275/291 (94%)	273 (99%)	2 (1%)	88	94
2	B	233/239 (98%)	230 (99%)	3 (1%)	76	89
3	C	31/33 (94%)	28 (90%)	3 (10%)	10	40
4	D	44/52 (85%)	41 (93%)	3 (7%)	20	57
All	All	583/615 (95%)	572 (98%)	11 (2%)	65	86

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	251	ARG
3	C	44	LYS
4	D	24	ASN
2	B	43	ASN
3	C	63	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	39	GLN
2	B	73	GLN
3	C	58	ASN
2	B	21	ASN
4	D	17	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
5	N	20/23 (86%)	6 (30%)	0

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
5	N	24	C
5	N	25	U
5	N	27	A
5	N	33	G
5	N	35	A

There are no RNA pucker outliers to report.

## 5.4 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 (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or 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 length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
1	TPO	A	186	1	7,10,11	1.42	1 (14%)	10,14,16	4.88	6 (60%)

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	TPO	A	186	1	-	0/8/11/13	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	186	TPO	P-O1P	2.18	1.57	1.50

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	186	TPO	O3P-P-OG1	-9.31	78.78	106.62
1	A	186	TPO	O3P-P-O1P	-8.15	84.04	110.63
1	A	186	TPO	O3P-P-O2P	-5.06	88.86	107.44
1	A	186	TPO	O2P-P-O1P	2.31	118.16	110.63
1	A	186	TPO	O2P-P-OG1	3.10	115.89	106.62

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	186	TPO	2	0

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	310/330 (93%)	1.09	69 (22%) <b>1</b> <b>5</b>	285, 383, 455, 513	0
2	B	255/264 (96%)	1.59	88 (34%) <b>0</b> <b>4</b>	315, 427, 534, 641	0
3	C	34/36 (94%)	1.13	9 (26%) <b>1</b> <b>5</b>	372, 453, 531, 562	0
4	D	49/58 (84%)	0.78	8 (16%) <b>2</b> <b>8</b>	363, 448, 532, 564	0
5	N	21/23 (91%)	0.96	3 (14%) <b>4</b> <b>9</b>	468, 535, 605, 664	0
All	All	669/711 (94%)	1.25	177 (26%) <b>1</b> <b>5</b>	285, 410, 531, 664	0

The worst 5 of 177 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	85	PRO	8.1
2	B	84	ALA	7.7
2	B	86	ALA	7.4
1	A	222	MET	6.8
2	B	225	VAL	6.7

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
1	TPO	A	186	11/12	0.57	0.46	-	318,388,417,431	0

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
6	ZN	D	101	1/1	0.96	0.07	-1.39	637,637,637,637	0
6	ZN	D	102	1/1	0.91	0.07	-2.19	533,533,533,533	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.