



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 07:03 pm GMT

PDB ID : 2W2H  
Title : STRUCTURAL BASIS OF TRANSCRIPTION ACTIVATION BY THE CYCLIN T1-TAT-TAR RNA COMPLEX FROM EIAV  
Authors : Anand, K.; Geyer, M.  
Deposited on : 2008-10-30  
Resolution : 3.25 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

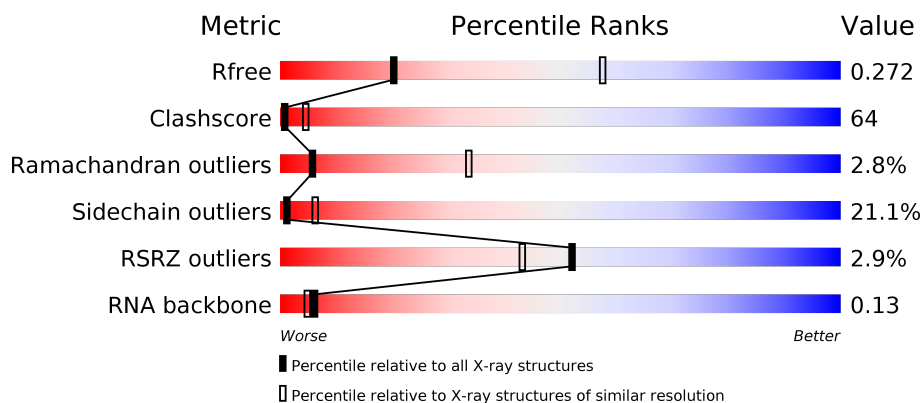
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.25 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))
$R_{free}$	100719	1852 (3.32-3.20)
Clashscore	112137	2036 (3.32-3.20)
Ramachandran outliers	110173	2000 (3.32-3.20)
Sidechain outliers	110143	1998 (3.32-3.20)
RSRZ outliers	101464	1861 (3.32-3.20)
RNA backbone	2435	1085 (3.72-2.80)

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	264	<div> <div>3%</div> <div>35% 52% 13%</div> </div>
1	B	264	<div> <div>3%</div> <div>38% 43% 14%</div> </div>
2	C	29	<div> <div>3%</div> <div>17% 62% 21%</div> </div>
2	D	29	<div> <div></div> <div>24% 52% 24%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	R	22	<div><div></div><div>23%55%23%</div></div>
3	S	22	<div><div></div><div>23%55%23%</div></div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 5593 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-T1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	263	Total	C	N	O	S	7	0	0
			2102	1338	373	381	10			
1	B	260	Total	C	N	O	S	11	0	0
			2084	1329	367	378	10			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	77	ARG	GLN	CONFLICT	UNP Q9XT26
A	110	THR	ALA	CONFLICT	UNP Q9XT26
A	256	TRP	ARG	CONFLICT	UNP Q9XT26
B	77	ARG	GLN	CONFLICT	UNP Q9XT26
B	110	THR	ALA	CONFLICT	UNP Q9XT26
B	256	TRP	ARG	CONFLICT	UNP Q9XT26

- Molecule 2 is a protein called PROTEIN TAT.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	29	Total	C	N	O	0	0	0
			233	150	41	42			
2	D	29	Total	C	N	O	0	0	0
			230	148	40	42			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	R	22	Total	C	N	O	P	0	0	0
			468	208	80	158	22			
3	S	22	Total	C	N	O	P	0	0	0
			468	208	80	158	22			

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	R	1	Total	Mn	0	0
			1	1		
4	A	1	Total	Mn	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	3	Total	O	0	0
			3	3		
5	B	3	Total	O	0	0
			3	3		

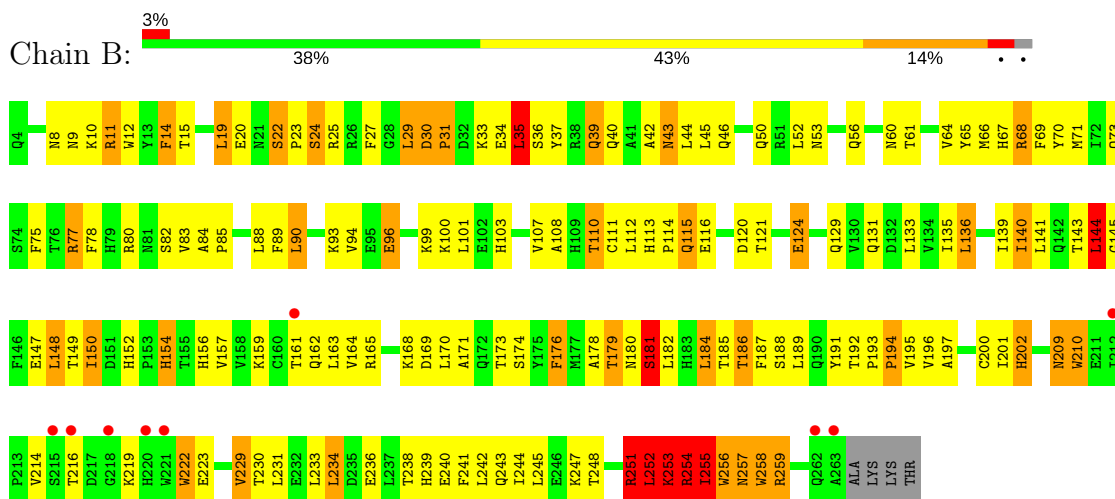
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-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. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

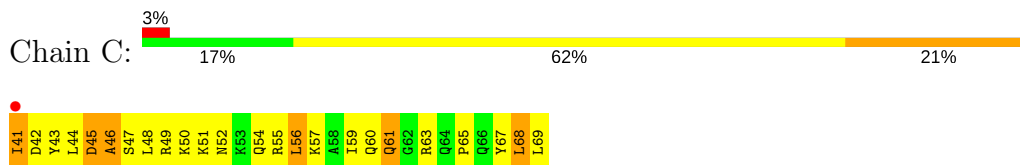
#### • Molecule 1: CYCLIN-T1



#### • Molecule 1: CYCLIN-T1



#### • Molecule 2: PROTEIN TAT



- Molecule 2: PROTEIN TAT

Chain D:  24% 52% 24%



- Molecule 3: 5'-R(\*GP\*CP\*UP\*CP\*AP\*GP\*AP\*UP\*CP\*UP \*GP\*CP\*GP\*GP\*UP\*CP\*UP\*GP\*AP\*GP\*C)-3'

Chain R:  23% 55% 23%



- Molecule 3: 5'-R(\*GP\*CP\*UP\*CP\*AP\*GP\*AP\*UP\*CP\*UP \*GP\*CP\*GP\*GP\*UP\*CP\*UP\*GP\*AP\*GP\*C)-3'

Chain S:  23% 55% 23%



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	149.46Å 149.46Å 129.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.93 – 3.25 29.93 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.9 (29.93-3.25) 99.9 (29.93-3.25)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.52 (at 3.24Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.243 , 0.278 0.231 , 0.272	Depositor DCC
$R_{free}$ test set	1189 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	103.9	Xtriage
Anisotropy	0.240	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 99.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	0.429 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5593	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	137.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.52% 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

### 5.1 Standard geometry

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

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	0.68	2/2153 (0.1%)	0.84	3/2934 (0.1%)
1	B	0.65	1/2135 (0.0%)	0.82	4/2910 (0.1%)
2	C	0.72	0/235	1.22	2/312 (0.6%)
2	D	0.77	0/232	1.30	2/309 (0.6%)
3	R	1.35	5/521 (1.0%)	2.54	46/808 (5.7%)
3	S	1.30	4/521 (0.8%)	2.45	44/808 (5.4%)
All	All	0.83	12/5797 (0.2%)	1.37	101/8081 (1.2%)

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 outliers	#Planarity outliers
1	A	0	3
1	B	0	3
2	C	0	2
2	D	0	2
All	All	0	10

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	R	3	G	OP3-P	-11.21	1.47	1.61
3	S	3	G	OP3-P	-11.15	1.47	1.61
1	A	5	ARG	CZ-NH1	7.42	1.42	1.33
3	S	13	U	N1-C2	-6.74	1.32	1.38
3	R	16	G	C5-C6	-6.62	1.35	1.42

The worst 5 of 101 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	R	16	G	C1'-O4'-C4'	-15.57	97.45	109.90
3	R	16	G	N1-C6-O6	15.47	129.18	119.90
3	S	16	G	C1'-O4'-C4'	-15.46	97.53	109.90
3	R	16	G	C4-C5-N7	13.58	116.23	110.80
1	A	5	ARG	NE-CZ-NH2	-13.51	113.55	120.30

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	252	LEU	Peptide
1	A	263	ALA	Peptide
1	A	264	ALA	Peptide
1	B	252	LEU	Peptide
1	B	253	LYS	Peptide

## 5.2 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 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	2102	0	2044	219	0
1	B	2084	0	2036	228	0
2	C	233	0	239	84	0
2	D	230	0	230	80	0
3	R	468	0	237	73	0
3	S	468	0	237	79	0
4	A	1	0	0	0	0
4	R	1	0	0	0	0
5	A	3	0	0	0	0
5	B	3	0	0	0	0
All	All	5593	0	5023	680	0

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 64.

The worst 5 of 680 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)
1:B:252:LEU:HD13	1:B:254:ARG:HG2	1.24	1.16
1:B:254:ARG:HD2	1:B:255:ILE:HG12	1.31	1.12
1:B:180:ASN:HB3	1:B:184:LEU:HD11	1.29	1.12
1:A:180:ASN:HB3	1:A:184:LEU:HD11	1.30	1.09
2:C:55:ARG:NH2	3:R:12:C:H42	1.49	1.09

There are no symmetry-related clashes.

## 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	261/264 (99%)	200 (77%)	55 (21%)	6 (2%)	7	38
1	B	258/264 (98%)	196 (76%)	52 (20%)	10 (4%)	3	25
2	C	27/29 (93%)	21 (78%)	6 (22%)	0	100	100
2	D	27/29 (93%)	21 (78%)	6 (22%)	0	100	100
All	All	573/586 (98%)	438 (76%)	119 (21%)	16 (3%)	6	34

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	181	SER
1	B	255	ILE
1	B	258	TRP
1	A	181	SER
1	A	26	ARG

### 5.3.2 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 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	225/241 (93%)	178 (79%)	47 (21%)	1	6
1	B	226/241 (94%)	176 (78%)	50 (22%)	1	5
2	C	22/26 (85%)	19 (86%)	3 (14%)	4	20
2	D	21/26 (81%)	17 (81%)	4 (19%)	2	9
All	All	494/534 (92%)	390 (79%)	104 (21%)	1	6

5 of 104 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	11	ARG
1	B	61	THR
1	B	259	ARG
1	B	14	PHE
1	B	24	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	152	HIS
1	A	156	HIS
1	B	152	HIS
1	A	109	HIS
1	A	115	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	R	21/22 (95%)	10 (47%)	0
3	S	21/22 (95%)	12 (57%)	0
All	All	42/44 (95%)	22 (52%)	0

5 of 22 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	R	4	C
3	R	5	U
3	R	11	U

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	R	12	C
3	R	14	G

There are no RNA pucker outliers to report.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

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 [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

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	263/264 (99%)	-0.22	8 (3%)	51	41	65, 121, 242, 282	10 (3%)
1	B	260/264 (98%)	-0.17	9 (3%)	44	34	64, 120, 239, 269	11 (4%)
2	C	29/29 (100%)	-0.17	1 (3%)	46	36	100, 125, 165, 186	0
2	D	29/29 (100%)	-0.27	0	100	100	94, 118, 151, 181	0
3	R	22/22 (100%)	-0.40	0	100	100	92, 149, 202, 208	0
3	S	22/22 (100%)	-0.46	0	100	100	105, 148, 196, 215	0
All	All	625/630 (99%)	-0.21	18 (2%)	52	43	64, 124, 238, 282	21 (3%)

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	262	GLN	5.6
1	A	218	GLY	5.0
1	B	220	HIS	4.2
1	A	217	ASP	4.1
1	A	220	HIS	4.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 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(Å <sup>2</sup> )	Q<0.9
4	MN	R	1025	1/1	0.39	0.41	-	351,351,351,351	0
4	MN	A	1268	1/1	0.45	0.34	-	344,344,344,344	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.