



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 02:40 pm GMT

PDB ID : 4B3P  
Title : Structures of HIV-1 RT and RNA-DNA Complex Reveal a Unique RT Con-  
formation and Substrate Interface  
Authors : Lapkouski, M.; Tian, L.; Miller, J.T.; Le Grice, S.F.J.; Yang, W.  
Deposited on : 2012-07-25  
Resolution : 4.84 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

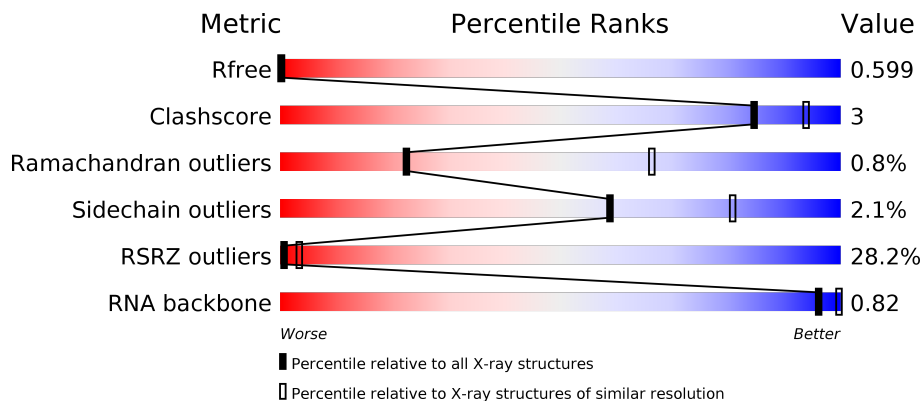
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.84 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	1038 (6.02-3.66)
Clashscore	112137	1071 (5.98-3.70)
Ramachandran outliers	110173	1007 (5.98-3.70)
Sidechain outliers	110143	1008 (6.00-3.68)
RSRZ outliers	101464	1009 (6.00-3.68)
RNA backbone	2435	1042 (6.76-2.90)

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	560	
2	B	454	
3	D	29	
4	R	34	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8162 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 REVERSE TRANSCRIPTASE/RIBONUCLEASE H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	519	Total	C	N	O	S	0	0	0
			4054	2612	669	766	7			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	68	GLY	SER	ENGINEERED MUTATION	UNP P04585
A	83	LYS	ARG	ENGINEERED MUTATION	UNP P04585
A	411	VAL	ILE	ENGINEERED MUTATION	UNP P04585
A	447	SER	ASN	ENGINEERED MUTATION	UNP P04585
A	461	LYS	ARG	ENGINEERED MUTATION	UNP P04585
A	483	HIS	TYR	ENGINEERED MUTATION	UNP P04585
A	498	ALA	ASP	ENGINEERED MUTATION	UNP P04585
A	559	ILE	VAL	ENGINEERED MUTATION	UNP P04585

- Molecule 2 is a protein called P51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	404	Total	C	N	O	S	0	0	0
			3230	2102	524	597	7			

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-13	MET	-	EXPRESSION TAG	UNP P04585
B	-12	ARG	-	EXPRESSION TAG	UNP P04585
B	-11	GLY	-	EXPRESSION TAG	UNP P04585
B	-10	SER	-	EXPRESSION TAG	UNP P04585
B	-9	HIS	-	EXPRESSION TAG	UNP P04585
B	-8	HIS	-	EXPRESSION TAG	UNP P04585
B	-7	HIS	-	EXPRESSION TAG	UNP P04585
B	-6	HIS	-	EXPRESSION TAG	UNP P04585

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-5	HIS	-	EXPRESSION TAG	UNP P04585
B	-4	HIS	-	EXPRESSION TAG	UNP P04585
B	-3	GLY	-	EXPRESSION TAG	UNP P04585
B	-2	SER	-	EXPRESSION TAG	UNP P04585
B	-1	GLN	-	EXPRESSION TAG	UNP P04585
B	0	LEU	-	EXPRESSION TAG	UNP P04585
B	68	GLY	SER	ENGINEERED MUTATION	UNP P04585
B	83	LYS	ARG	ENGINEERED MUTATION	UNP P04585
B	411	VAL	ILE	ENGINEERED MUTATION	UNP P04585

- Molecule 3 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	21	Total 433	C 207	N 75	O 130	P 21	0	0	0

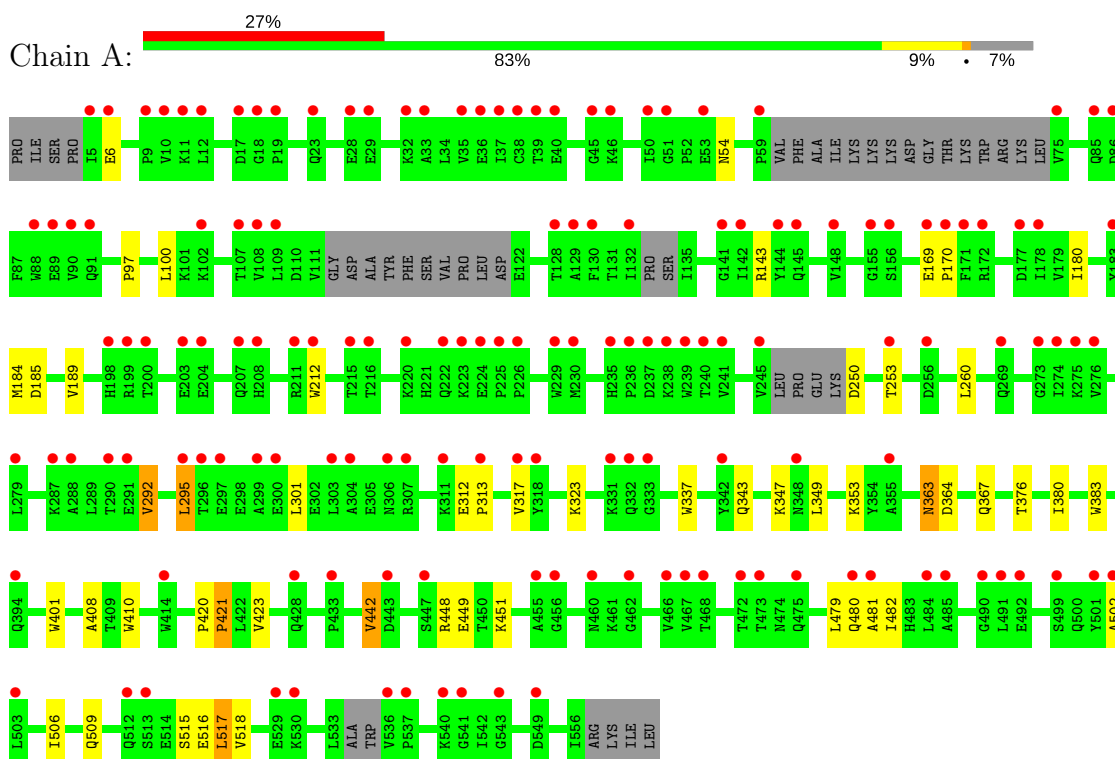
- Molecule 4 is a RNA chain called RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	R	21	Total 445	C 200	N 81	O 143	P 21	0	0	0

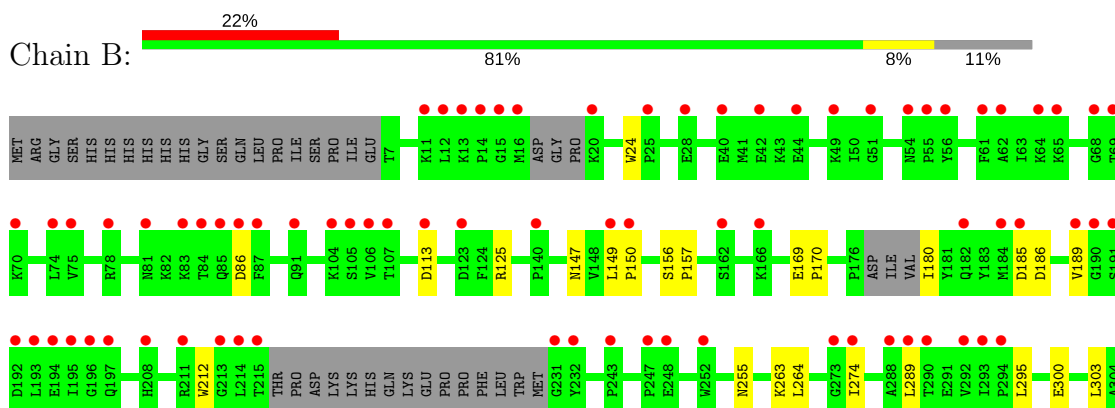
### 3 Residue-property plots [i](#)

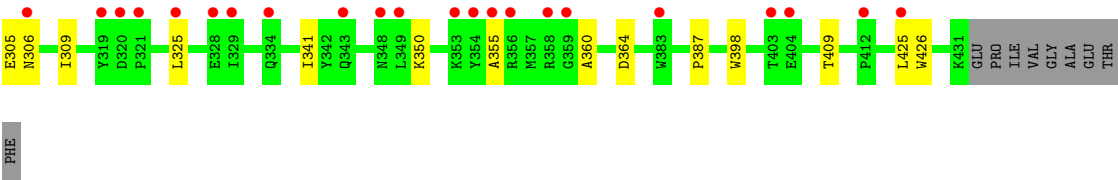
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: REVERSE TRANSCRIPTASE/RIBONUCLEASE H

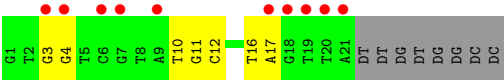


#### • Molecule 2: P51 RT

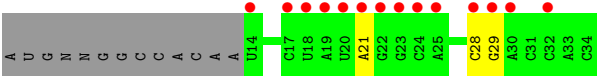
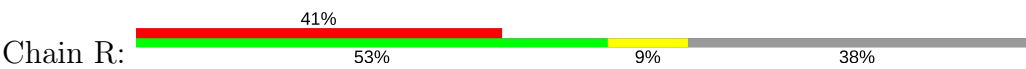




● Molecule 3: DNA



● Molecule 4: RNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	163.00Å 163.00Å 229.60Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.10 – 4.84 48.38 – 4.84	Depositor EDS
% Data completeness (in resolution range)	99.2 (46.10-4.84) 99.4 (48.38-4.84)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.16	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.25 (at 4.85Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.364 , 0.404 0.607 , 0.599	Depositor DCC
$R_{free}$ test set	231 reflections (2.85%)	DCC
Wilson B-factor (Å <sup>2</sup> )	112.9	Xtriage
Anisotropy	0.683	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 139.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.26$ , $\langle L^2 \rangle = 0.10$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.60	EDS
Total number of atoms	8162	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	189.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.55% 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](#)

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	0.49	4/4146 (0.1%)	0.47	0/5648
2	B	0.53	4/3315 (0.1%)	0.47	0/4513
3	D	0.30	0/484	0.70	0/746
4	R	0.29	1/497 (0.2%)	0.65	0/771
All	All	0.49	9/8442 (0.1%)	0.50	0/11678

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	410	TRP	CD2-CE2	5.06	1.47	1.41
1	A	383	TRP	CD2-CE2	5.05	1.47	1.41
2	B	426	TRP	CD2-CE2	5.03	1.47	1.41
2	B	24	TRP	CD2-CE2	5.03	1.47	1.41
2	B	212	TRP	CD2-CE2	5.03	1.47	1.41

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	4054	0	3965	23	0
2	B	3230	0	3186	14	0
3	D	433	0	240	5	0
4	R	445	0	228	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	8162	0	7619	43	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 3.

The worst 5 of 43 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:A:480:GLN:HG2	1:A:517:LEU:HD11	1.69	0.74
1:A:337:TRP:HE1	1:A:367:GLN:HE21	1.45	0.63
3:D:3:DG:H2'	3:D:4:DG:H8	1.69	0.58
1:A:482:ILE:HD13	1:A:506:ILE:HD11	1.85	0.57
3:D:16:DT:H2'	3:D:17:DA:C8	2.40	0.56

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	507/560 (90%)	473 (93%)	29 (6%)	5 (1%)	18	61
2	B	396/454 (87%)	383 (97%)	11 (3%)	2 (0%)	32	74
All	All	903/1014 (89%)	856 (95%)	40 (4%)	7 (1%)	22	66

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	360	ALA
1	A	6	GLU
1	A	363	ASN
2	B	355	ALA

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	184	MET

### 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	426/498 (86%)	415 (97%)	11 (3%)	51	75
2	B	344/411 (84%)	339 (98%)	5 (2%)	70	86
All	All	770/909 (85%)	754 (98%)	16 (2%)	59	81

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	442	VAL
1	A	448	ARG
2	B	113	ASP
1	A	353	LYS
2	B	186	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	222	GLN
1	A	367	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	R	20/34 (58%)	0	0

There are no RNA backbone outliers to report.

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](#)

There are no ligands in this entry.

## 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	519/560 (92%)	1.49	150 (28%)	1 3	167, 189, 198, 207	0
2	B	404/454 (88%)	1.21	98 (24%)	1 4	172, 189, 196, 201	0
3	D	21/29 (72%)	2.66	10 (47%)	0 2	187, 193, 197, 201	0
4	R	21/34 (61%)	2.37	14 (66%)	0 2	187, 192, 195, 198	0
All	All	965/1077 (89%)	1.42	272 (28%)	1 3	167, 189, 197, 207	0

The worst 5 of 272 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	20	DT	10.8
1	A	307	ARG	8.4
1	A	276	VAL	7.9
2	B	358	ARG	7.9
3	D	19	DT	7.9

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.