



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:21 PM GMT

PDB ID : 4OL8  
Title : Ty3 reverse transcriptase bound to DNA/RNA  
Authors : Nowak, E.; Miller, J.T.; Bona, M.K.; Studnicka, J.; Szczepanowski, R.H.;  
Jurkowski, J.; Le Grice, S.F.J.; Nowotny, M.  
Deposited on : 2014-01-23  
Resolution : 3.10 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

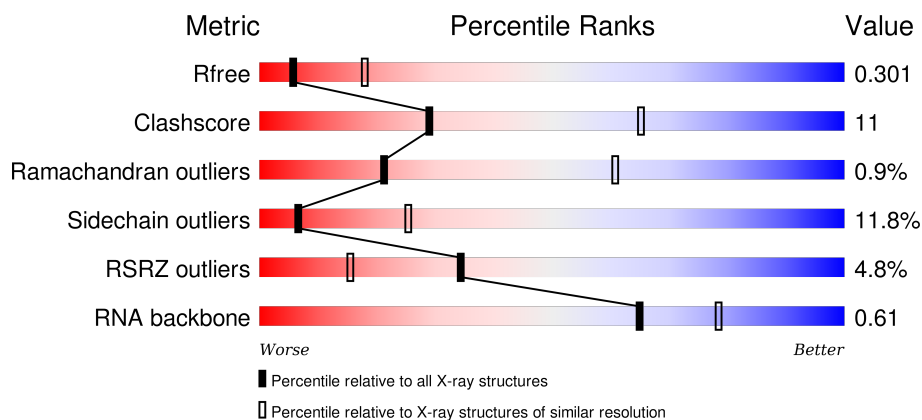
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)
RNA backbone	2183	1010 (3.52-2.68)

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	478	
1	B	478	
1	E	478	
1	F	478	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	C	18	 72% 22% 6%
2	G	18	 61% 28% 6% 6%
3	D	16	 44% 38% 13% 6%
3	H	16	 50% 44% 6%

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 14105 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	390	Total	C	N	O	S	Se	0	0	0
			3004	1938	504	552	5	5			
1	B	452	Total	C	N	O	S	Se	0	0	0
			3547	2291	604	639	8	5			
1	F	427	Total	C	N	O	S	Se	0	0	0
			3310	2129	561	608	7	5			
1	E	381	Total	C	N	O	S	Se	0	0	0
			2873	1839	484	540	5	5			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	EXPRESSION TAG	UNP Q99315
A	0	PRO	-	EXPRESSION TAG	UNP Q99315
A	426	ASN	ASP	ENGINEERED MUTATION	UNP Q99315
B	-1	GLY	-	EXPRESSION TAG	UNP Q99315
B	0	PRO	-	EXPRESSION TAG	UNP Q99315
B	426	ASN	ASP	ENGINEERED MUTATION	UNP Q99315
F	-1	GLY	-	EXPRESSION TAG	UNP Q99315
F	0	PRO	-	EXPRESSION TAG	UNP Q99315
F	426	ASN	ASP	ENGINEERED MUTATION	UNP Q99315
E	-1	GLY	-	EXPRESSION TAG	UNP Q99315
E	0	PRO	-	EXPRESSION TAG	UNP Q99315
E	426	ASN	ASP	ENGINEERED MUTATION	UNP Q99315

- Molecule 2 is a RNA chain called 5'-R(\*CP\*UP\*GP\*AP\*GP\*AP\*GP\*AP\*GP\*AP\*GP\*GP\*AP\*AP\*GP\*AP\*UP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	17	Total	C	N	O	P	0	0	0
			378	168	79	114	17			

*Continued on next page...*

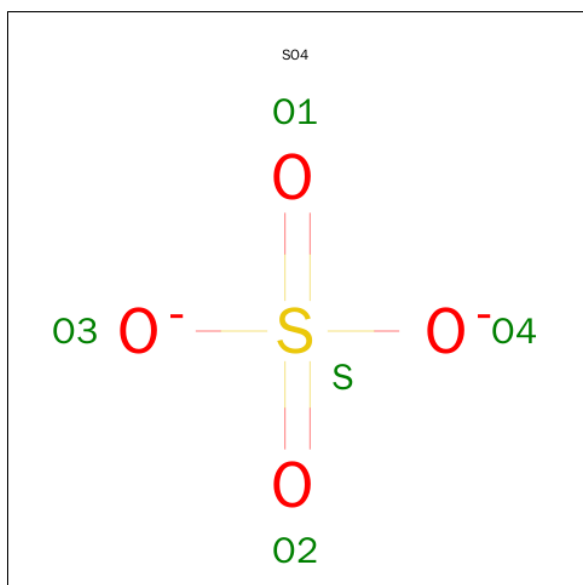
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	17	Total	C	N	O	P	0	0	0
			378	168	79	114	17			

- Molecule 3 is a DNA chain called 5'-D(\*CP\*AP\*TP\*CP\*TP\*TP\*CP\*CP\*TP\*CP\*TP\*CP\*TP\*CP\*TP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	15	Total	C	N	O	P	0	0	0
			294	143	40	96	15			
3	H	16	Total	C	N	O	P	0	0	0
			310	152	43	100	15			

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	O	S	0	0
			5	4	1		
4	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	O	0	0
			1	1		





VAL  
VAL  
ALA  
ASP  
ALA  
ILE  
SER  
ARG  
ALA  
VAL  
TYR

- Molecule 2: 5'-R(\*CP\*UP\*GP\*AP\*GP\*AP\*GP\*AP\*GP\*AP\*GP\*GP\*AP\*AP\*GP\*AP\*UP\*G)-3'

Chain C:  72% 22% 6%

C  
U2  
G3  
A4  
G5  
G11  
G18

- Molecule 2: 5'-R(\*CP\*UP\*GP\*AP\*GP\*AP\*GP\*AP\*GP\*AP\*GP\*GP\*AP\*AP\*GP\*AP\*UP\*G)-3'

Chain G:  61% 28% 6% 6%

C  
U2  
G3  
A4  
G5  
A6  
G7  
A8  
A16  
U17  
G18

- Molecule 3: 5'-D(\*CP\*AP\*TP\*CP\*TP\*TP\*CP\*CP\*TP\*CP\*TP\*CP\*TP\*CP\*TP\*C)-3'

Chain D:  44% 38% 13% 6%

DC  
A33  
T34  
C35  
T36  
T37  
C38  
C39  
T44  
C45  
T46  
C47

- Molecule 3: 5'-D(\*CP\*AP\*TP\*CP\*TP\*TP\*CP\*CP\*TP\*CP\*TP\*CP\*TP\*CP\*TP\*C)-3'

Chain H:  50% 44% 6%

C32  
A33  
T34  
T37  
C38  
C41  
T44  
C45  
T46  
C47



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	320.70Å 75.06Å 108.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.90 – 3.10 48.90 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.1 (48.90-3.10) 99.6 (48.90-3.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.10 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.227 , 0.296 0.230 , 0.301	Depositor DCC
$R_{free}$ test set	2420 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	78.5	Xtriage
Anisotropy	0.243	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 98.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 53285 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	14105	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	131.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.89% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: SO4

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	0.32	0/3069	0.54	0/4169
1	B	0.42	0/3629	0.60	0/4928
1	E	0.29	0/2933	0.51	0/3990
1	F	0.30	0/3382	0.49	0/4597
2	C	0.37	0/426	0.75	0/665
2	G	0.35	0/426	0.80	0/665
3	D	0.62	0/324	1.47	4/495 (0.8%)
3	H	0.60	0/342	1.35	3/523 (0.6%)
All	All	0.36	0/14531	0.63	7/20032 (0.0%)

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	B	0	1

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	35	DC	O4'-C4'-C3'	-8.14	101.11	106.00
3	H	41	DC	O4'-C4'-C3'	-6.58	101.87	104.50
3	D	36	DT	O4'-C4'-C3'	-5.88	102.15	104.50
3	H	44	DT	N3-C4-O4	5.66	123.30	119.90
3	H	44	DT	C5-C4-O4	-5.57	121.00	124.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	475	VAL	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	3004	0	2838	85	0
1	B	3547	0	3500	84	0
1	E	2873	0	2637	88	0
1	F	3310	0	3193	50	0
2	C	378	0	186	2	0
2	G	378	0	186	6	0
3	D	294	0	173	6	0
3	H	310	0	185	8	0
4	B	10	0	0	1	0
5	B	1	0	0	0	0
All	All	14105	0	12898	307	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 11.

The worst 5 of 307 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:E:154:SER:HB3	1:E:157:HIS:HB2	1.57	0.86
1:F:50:LYS:HG2	1:F:160:PRO:HG2	1.67	0.76
1:B:205:LEU:HD21	1:B:235:ARG:HH12	1.51	0.75
1:A:349:ASN:HD21	1:B:72:LYS:HE3	1.50	0.75
1:E:153:HIS:HB3	1:E:244:LYS:HB2	1.68	0.73

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 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 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	374/478 (78%)	338 (90%)	36 (10%)	0	100	100
1	B	446/478 (93%)	398 (89%)	45 (10%)	3 (1%)	26	65
1	E	361/478 (76%)	307 (85%)	45 (12%)	9 (2%)	7	32
1	F	415/478 (87%)	376 (91%)	37 (9%)	2 (0%)	34	72
All	All	1596/1912 (84%)	1419 (89%)	163 (10%)	14 (1%)	21	61

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	205	LEU
1	E	264	ILE
1	E	251	ALA
1	B	318	CYS
1	F	372	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	304/426 (71%)	272 (90%)	32 (10%)	8	31
1	B	377/426 (88%)	325 (86%)	52 (14%)	4	19
1	E	286/426 (67%)	247 (86%)	39 (14%)	5	19
1	F	347/426 (82%)	315 (91%)	32 (9%)	11	40
All	All	1314/1704 (77%)	1159 (88%)	155 (12%)	6	25

5 of 155 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	378	VAL
1	F	69	VAL
1	E	326	LYS
1	B	381	VAL
1	B	428	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	35	ASN
1	A	349	ASN
1	E	288	GLN
1	E	327	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	16/18 (88%)	2 (12%)	0
2	G	16/18 (88%)	1 (6%)	0
All	All	32/36 (88%)	3 (9%)	0

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	C	5	G
2	C	11	G
2	G	5	G

There are no RNA pucker outliers to report.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

2 ligands are 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	$\# Z  > 2$
4	SO4	B	501	-	4,4,4	0.22	0	6,6,6	0.20	0
4	SO4	B	502	-	4,4,4	0.21	0	6,6,6	0.09	0

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
4	SO4	B	501	-	-	0/0/0/0	0/0/0/0
4	SO4	B	502	-	-	0/0/0/0	0/0/0/0

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	501	SO4	1	0

## 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	385/478 (80%)	0.13	10 (2%) 59 35	81, 132, 173, 195	1 (0%)
1	B	447/478 (93%)	-0.20	5 (1%) 82 66	55, 87, 137, 174	2 (0%)
1	E	376/478 (78%)	0.53	39 (10%) 8 3	94, 158, 204, 236	0
1	F	422/478 (88%)	0.25	28 (6%) 22 8	75, 142, 197, 239	0
2	C	17/18 (94%)	-0.25	0 100 100	101, 147, 167, 189	0
2	G	17/18 (94%)	-0.15	0 100 100	120, 146, 202, 216	0
3	D	15/16 (93%)	-0.09	0 100 100	125, 147, 196, 224	0
3	H	16/16 (100%)	-0.33	0 100 100	126, 144, 216, 234	0
All	All	1695/1980 (85%)	0.15	82 (4%) 34 15	55, 132, 193, 239	3 (0%)

The worst 5 of 82 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	438	GLU	6.9
1	E	242	ILE	5.9
1	F	299	TYR	5.7
1	E	29	TYR	4.9
1	F	274	CYS	4.5

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

### 6.3 Carbohydrates ⓘ

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
4	SO4	B	501	5/5	0.92	0.25	-	154,156,157,158	0
4	SO4	B	502	5/5	0.86	0.24	-	174,175,175,175	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.