



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 06:36 am BST

PDB ID : 5C9H  
Title : Structural Basis of Template Boundary Definition in Tetrahymena Telomerase  
Authors : Jansson, L.I.; Akiyama, B.M.; Ooms, A.; Lu, C.; Rubin, S.M.; Stone, M.D.  
Deposited on : 2015-06-26  
Resolution : 3.00 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

<https://www.wwpdb.org/validation/2017/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)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

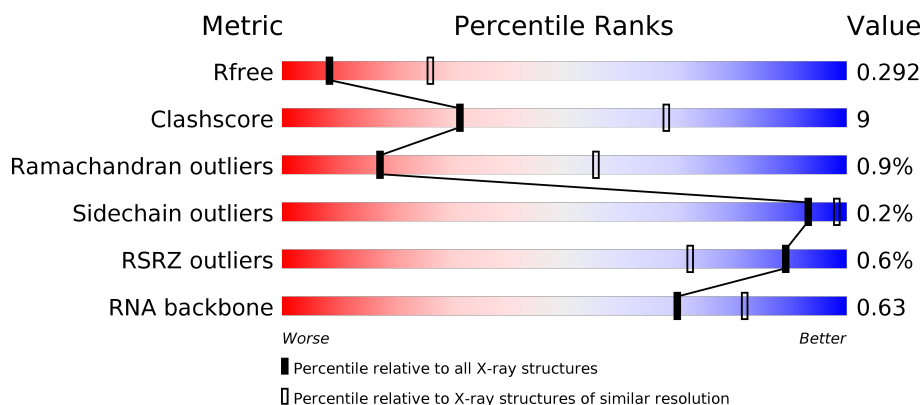
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*





The reported resolution of this entry is 3.00 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))
$R_{free}$	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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 respectively. 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	303	
1	B	303	
2	D	13	
2	E	13	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	C	13	<div><div></div><div>54%</div><div>31%</div><div>8%</div><div>8%</div></div>
3	G	13	<div><div>8%</div><div>38%</div><div>38%</div><div>8%</div><div>15%</div></div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 4793 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 Telomerase reverse transcriptase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	245	Total	C	N	O	S	0	0	0
			2021	1330	330	352	9			
1	B	225	Total	C	N	O	S	0	0	0
			1864	1224	304	328	8			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	214	GLY	-	expression tag	UNP O77448
A	215	HIS	-	expression tag	UNP O77448
A	216	MET	-	expression tag	UNP O77448
B	214	GLY	-	expression tag	UNP O77448
B	215	HIS	-	expression tag	UNP O77448
B	216	MET	-	expression tag	UNP O77448

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	10	Total	C	N	O	P	0	0	0
			196	86	35	65	10			
2	E	10	Total	C	N	O	P	0	0	0
			212	95	37	70	10			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	12	Total	C	N	O	P	0	0	0
			247	111	36	88	12			
3	G	11	Total	C	N	O	P	0	0	0
			227	102	34	80	11			


- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	G	1	Total 1	Mg 1	0	0
4	C	1	Total 1	Mg 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	13	Total 13	O 13	0	0
5	B	6	Total 6	O 6	0	0
5	D	2	Total 2	O 2	0	0
5	C	2	Total 2	O 2	0	0
5	G	1	Total 1	O 1	0	0

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 ( $\text{RSRZ} > 2$ ). 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.

- Chain A: 
- GLY HIS MET ASN Q218 P222 R226 S227 Q228 C232 M233 E234 M235 Q236 R237 E238 P239 Q240 V241 F242 K243 M248 Y249 SER GLU ILE LYS LYS GLY PHE PHE GLN VAL ILE ILE GLN GLU LYS LEU GLN GLY ARG ARG PHE ILE ASN SER ASP LYS ILE LYS PRAQ HIS P80 GLN THR ILE LYS LYS THR LEU LEU LYS GLU Y293 E303 E312 K313 L314 V315 F318 L327 K328 K332 L333 P334 K335 K336 Y344 K346 K346 T347 V348 GLN SER SER GLU ASN K353 K354 N355 K356 Q357 E360 L380 C387 F408 M411 Q417 Y422 Y426 D436 N439 Q440 K441 F442 T443 Q444 K445 K452 I460 V461 M465 K466 I467 V468 I469 P470 V471 L472 K473 Y474 R492 K493 P494 K497 L498 E509 GLU ASN LEU GLU GLU VAL GLU

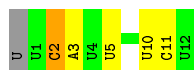
- Chain B:
- 
- 57% 17% 26%
- GLY HIS MET ASN Q218 Y219 I220 Y221 P222 R226 S227 Q228 G236 R237 Y241 F242 F247 N248 Y249 SER GLU ILE LYS LYS GLY PHE GLN PHE LYS VAL ILE GLN GLU LYS LEU GLN GLY ARG GLN LYS ILE LYS PRO ASP HIS PRO THR ILE LYS
- LYS THR LEU L290 Q294 S295 K296 C300 R301 E302 E303 R304 L308 T311 E312 R313 L314 V315 L327 K328 C331 P334 E335 R336 K345 GLN ILE VAL GLN SER SER GLU ASN LYS ALA ASN GLN GLN SER CYS SER GLU R361 L366 Y367 D368 T369 E370 I371 Q375 L376
- C387 M411 Y422 V435 D436 L437 K438 H439 Q440 LYS PHE THR GLN K445 Y448 I460 V461 M465 I469 L472 K473 Y474 K481 HIS LYS E484 R492 K493 P494 S500 T503 ILE VAL LYS LYS LEU GLU GLU LYS ASN LEU GLU VAL GLU

- Chain D: 
- 

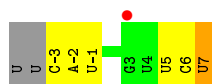
- Chain E:  46% 15% 15% 23%



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



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	58.53Å 117.77Å 131.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.51 – 3.00 41.51 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.1 (41.51-3.00) 90.0 (41.51-3.00)	Depositor EDS
$R_{merge}$	0.23	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.54 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.239 , 0.290 0.240 , 0.292	Depositor DCC
$R_{free}$ test set	1865 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	42.5	Xtriage
Anisotropy	0.602	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 43.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	4793	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.85% 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: MG

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	0.27	0/2067	0.47	0/2787
1	B	0.25	0/1906	0.46	0/2568
2	D	0.17	0/218	0.73	0/338
2	E	0.19	0/236	0.77	0/365
3	C	0.19	0/273	0.94	0/421
3	G	0.20	0/251	0.79	0/387
All	All	0.25	0/4951	0.56	0/6866

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	2021	0	1976	44	0
1	B	1864	0	1794	31	0
2	D	196	0	97	2	0
2	E	212	0	108	2	0
3	C	247	0	127	4	0
3	G	227	0	117	4	0
4	C	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	G	1	0	0	0	0
5	A	13	0	0	0	0
5	B	6	0	0	0	0
5	C	2	0	0	0	0
5	D	2	0	0	0	0
5	G	1	0	0	0	0
All	All	4793	0	4219	81	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 9.

The worst 5 of 81 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:357:GLN:NE2	1:A:360:GLU:OE1	1.94	1.01
1:A:357:GLN:HE21	1:A:360:GLU:HB3	1.47	0.78
1:A:222:PRO:HB2	1:A:461:VAL:HG21	1.71	0.72
1:B:371:ILE:HD11	1:B:492:ARG:HG2	1.73	0.69
1:B:228:GLN:O	1:B:473:ARG:NH1	2.27	0.68

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	239/303 (79%)	222 (93%)	15 (6%)	2 (1%)	19	57
1	B	215/303 (71%)	201 (94%)	12 (6%)	2 (1%)	17	55
All	All	454/606 (75%)	423 (93%)	27 (6%)	4 (1%)	17	55

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	347	ILE
1	A	241	VAL
1	B	295	SER
1	B	220	ILE

### 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	218/294 (74%)	218 (100%)	0	100	100
1	B	199/294 (68%)	198 (100%)	1 (0%)	88	96
All	All	417/588 (71%)	416 (100%)	1 (0%)	93	98

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	336	ASN

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	248	ASN
1	A	357	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	D	8/13 (61%)	2 (25%)	0
2	E	9/13 (69%)	2 (22%)	0
3	C	11/13 (84%)	2 (18%)	0
3	G	10/13 (76%)	2 (20%)	0
All	All	38/52 (73%)	8 (21%)	0

5 of 8 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	D	2	G
2	D	9	C
3	C	2	C
3	C	3	A
2	E	6	C

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	245/303 (80%)	-0.40	1 (0%) 92 79	20, 48, 96, 163	0
1	B	225/303 (74%)	-0.29	1 (0%) 92 79	26, 63, 113, 163	0
2	D	10/13 (76%)	0.03	0 100 100	39, 70, 96, 105	0
2	E	10/13 (76%)	0.29	0 100 100	66, 78, 122, 177	0
3	C	12/13 (92%)	-0.04	0 100 100	38, 54, 102, 120	0
3	G	11/13 (84%)	0.48	1 (9%) 9 3	77, 88, 118, 141	0
All	All	513/658 (77%)	-0.30	3 (0%) 89 72	20, 55, 109, 177	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	500	SER	2.9
3	G	3	G	2.1
1	A	355	ASN	2.1

### 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
4	MG	C	101	1/1	0.87	0.07	46,46,46,46	0
4	MG	G	101	1/1	0.93	0.27	84,84,84,84	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.