



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 09:13 am BST

PDB ID : 1PHJ  
Title : CRYSTAL STRUCTURE OF THE OXYTRICHA NOVA TELOMERE  
END-BINDING PROTEIN COMPLEXED WITH NONCOGNATE SSDNA  
GG(3DR)GTTTTGGGG  
Authors : Theobald, D.L.; Schultz, S.C.  
Deposited on : 2003-05-29  
Resolution : 2.50 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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
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

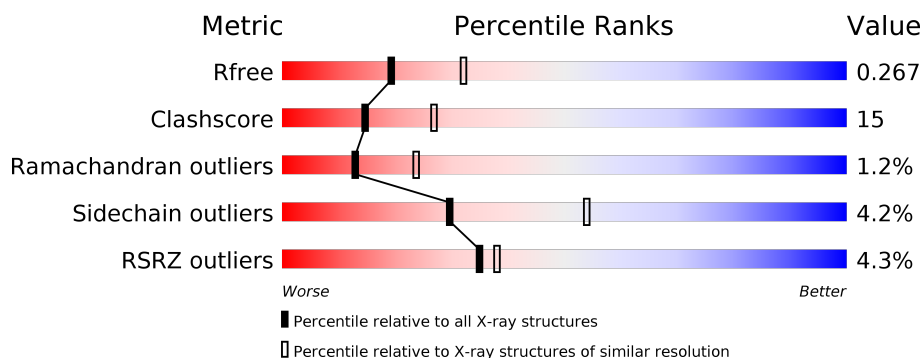
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	G	13	<div> <div>23%</div> <div>54%</div> <div>15%</div> <div>8%</div> </div>
1	H	13	<div> <div>85%</div> <div>8%</div> <div>8%</div> </div>
2	D	12	<div> <div>33%</div> <div>67%</div> </div>
3	A	461	<div> <div>2%</div> <div>73%</div> <div>20%</div> <div>• •</div> </div>
4	B	216	<div> <div>9%</div> <div>73%</div> <div>23%</div> <div>•</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6410 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 DNA chain called 5'-D(\*GP\*GP\*GP\*GP\*TP\*TP\*TP\*TP\*GP\*GP\*GP\*GP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	G	12	Total	C	N	O	P	0	0	0
			253	120	48	74	11			
1	H	12	Total	C	N	O	P	0	0	0
			253	120	48	74	11			

- Molecule 2 is a DNA chain called 5'-D(\*GP\*GP\*(3DR)P\*GP\*TP\*TP\*TP\*TP\*GP\*GP\*GP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	12	Total	C	N	O	P	0	0	0
			242	115	43	73	11			

- Molecule 3 is a protein called Telomere-binding protein alpha subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	449	Total	C	N	O	S	0	0	0
			3636	2312	625	697	2			

- Molecule 4 is a protein called Telomere-binding protein beta subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	215	Total	C	N	O	S	0	0	0
			1723	1107	292	323	1			

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	G	5	Total	O	0	0
			5	5		
5	D	18	Total	O	0	0
			18	18		

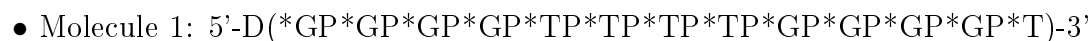
*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	5	Total 5	O 5	0	0
5	A	232	Total 232	O 232	0	0
5	B	43	Total 43	O 43	0	0

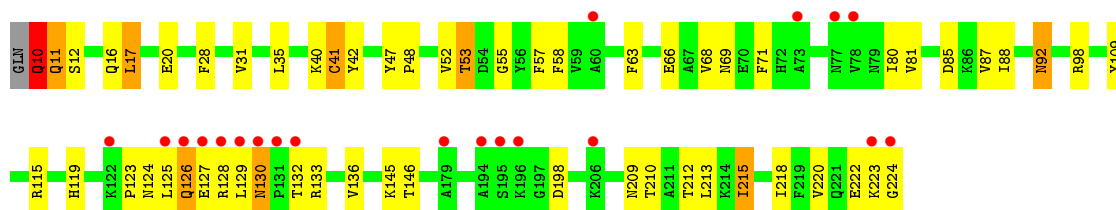


- Molecule 1: 5'-D(\*GP\*GP\*GP\*GP\*TP\*TP\*TP\*TP\*GP\*GP\*GP\*GP\*T)-3'



## ● Molecule 4: Telomere-binding protein beta subunit

Chain B:  9% 73% 23%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.48 Å 93.48 Å 423.49 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	27.31 – 2.50 27.30 – 2.25	Depositor EDS
% Data completeness (in resolution range)	92.9 (27.31-2.50) 90.0 (27.30-2.25)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.96 (at 2.26 Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.246 , 0.273 0.238 , 0.267	Depositor DCC
$R_{free}$ test set	4755 reflections (9.92%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	31.3	Xtriage
Anisotropy	0.373	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 34.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	6410	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.88% 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: 3DR

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	G	0.57	0/284	1.00	3/440 (0.7%)
1	H	0.32	0/284	0.74	0/440
2	D	0.41	1/258 (0.4%)	0.73	0/397
3	A	0.58	14/3706 (0.4%)	0.79	14/5010 (0.3%)
4	B	0.88	4/1761 (0.2%)	0.72	4/2379 (0.2%)
All	All	0.66	19/6293 (0.3%)	0.78	21/8666 (0.2%)

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
4	B	0	1

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	10	GLN	N-CA	27.08	2.00	1.46
4	B	10	GLN	CA-C	17.64	1.98	1.52
3	A	492	LYS	CA-C	10.19	1.79	1.52
3	A	401	LYS	CB-CG	9.64	1.78	1.52
4	B	11	GLN	N-CA	8.31	1.62	1.46

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	10	GLN	N-CA-C	16.73	156.18	111.00
4	B	10	GLN	CB-CA-C	-13.14	84.13	110.40
3	A	400	LEU	CA-CB-CG	-9.81	92.74	115.30

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	406	LYS	N-CA-C	9.73	137.28	111.00
3	A	401	LYS	N-CA-C	-9.00	86.70	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	B	10	GLN	Mainchain

## 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	G	253	0	138	12	0
1	H	253	0	138	1	0
2	D	242	0	135	7	0
3	A	3636	0	3633	91	0
4	B	1723	0	1710	74	0
5	A	232	0	0	4	0
5	B	43	0	0	0	0
5	D	18	0	0	1	0
5	G	5	0	0	0	0
5	H	5	0	0	0	0
All	All	6410	0	5754	182	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 15.

The worst 5 of 182 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)
3:A:401:LYS:CB	3:A:401:LYS:CG	1.78	1.56
3:A:492:LYS:C	3:A:492:LYS:CA	1.79	1.49
4:B:10:GLN:CA	4:B:10:GLN:C	1.98	1.31
4:B:10:GLN:CA	4:B:10:GLN:N	2.00	1.24
1:G:5:DT:H2"	1:G:6:DT:H5"	1.20	1.10

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	
3	A	447/461 (97%)	421 (94%)	21 (5%)	5 (1%)	14	26
4	B	213/216 (99%)	189 (89%)	21 (10%)	3 (1%)	11	20
All	All	660/677 (98%)	610 (92%)	42 (6%)	8 (1%)	13	24

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	93	ASP
3	A	321	VAL
3	A	406	LYS
4	B	126	GLN
4	B	130	ASN

### 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	
3	A	403/409 (98%)	385 (96%)	18 (4%)	27	51
4	B	188/189 (100%)	181 (96%)	7 (4%)	34	60
All	All	591/598 (99%)	566 (96%)	25 (4%)	30	54

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	324	LEU
3	A	353	LEU
4	B	92	ASN
3	A	337	GLU
3	A	398	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	372	GLN
4	B	149	GLN
4	B	92	ASN
3	A	331	ASN
4	B	10	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 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	$\# Z  > 2$
2	3DR	D	3	2	8,11,12	0.34	0	9,14,17	0.74	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
2	3DR	D	3	2	-	1/3/15/16	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	3	3DR	O4'-C4'-C5'-O5'

There are no ring outliers.

No monomer is involved in short contacts.

## 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	G	12/13 (92%)	0.04	0 100 100	34, 43, 49, 50	0
1	H	12/13 (92%)	0.18	0 100 100	40, 46, 50, 50	0
2	D	11/12 (91%)	-0.47	0 100 100	25, 34, 52, 52	0
3	A	449/461 (97%)	-0.09	10 (2%) 62 65	17, 29, 49, 60	0
4	B	215/216 (99%)	0.49	20 (9%) 8 8	23, 46, 61, 70	0
All	All	699/715 (97%)	0.09	30 (4%) 35 38	17, 33, 57, 70	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	407	GLY	9.8
4	B	127	GLU	7.6
3	A	406	LYS	6.6
4	B	130	ASN	6.3
4	B	126	GLN	5.9

### 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. 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(Å <sup>2</sup> )	Q<0.9
2	3DR	D	3	11/12	0.79	0.29	50,53,63,68	0

### 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.