wwPDB X-ray Structure Validation Summary Report (i)

May 28, 2020 – 08:19 pm BST

PDB ID	:	1PH6
Title	:	Crystal Structure of THE OXYTRICHA NOVA TELOMERE END-BINDING PROTEIN COMPLEXED WITH NONCOGNATE SSDNA GGGGTTTTGTGG
Authors	:	Theobald, D.L.; Schultz, S.C.
Deposited on	:	2003-05-29
Resolution	:	2.10 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

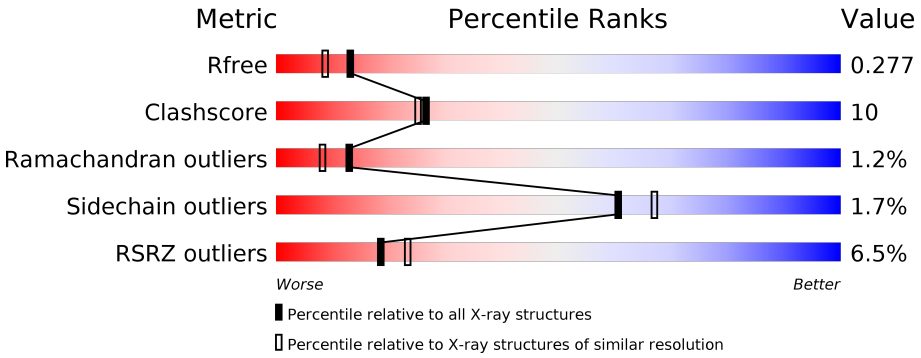
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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}	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 >=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 <=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	38% 54% 8%
1	H	13	77% 15% 8%
2	D	11	9% 55% 45%
3	A	461	4% 79% 17% ••
4	B	217	12% 76% 22% •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	NA	A	496	-	-	-	X
5	NA	A	497	-	-	-	X
5	NA	A	498	-	-	-	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6448 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*GP*GP*TP*TP*TP*TP*GP*TP*GP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	11	Total	C	N	O	P	0	0	0
			229	110	40	69	10			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	452	Total	C	N	O	S	0	0	0
			3665	2333	628	702	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	217	Total	C	N	O	S	0	0	0
			1740	1117	295	327	1			

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	5	Total	Na	0	0
			5	5		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	G	5	Total O 5 5	0	0
6	D	18	Total O 18 18	0	0
6	H	5	Total O 5 5	0	0
6	A	230	Total O 230 230	0	0
6	B	45	Total O 45 45	0	0

3 Residue-property plots (i)

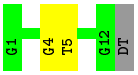
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: 5'-D(*GP*GP*GP*GP*TP*TP*TP*TP*GP*GP*GP*GP*T)-3'

Chain G:

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

Chain H:

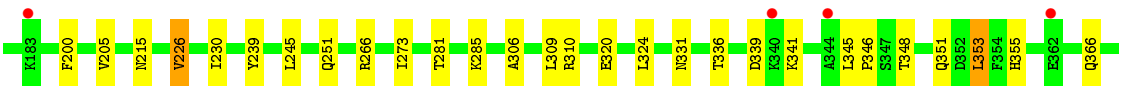


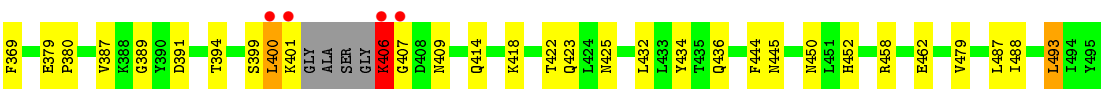
- Molecule 2: 5'-D(*GP*GP*GP*GP*TP*TP*TP*TP*GP*TP*GP*G)-3'

Chain D:

- Molecule 3: Telomere-binding protein alpha subunit

Chain A:

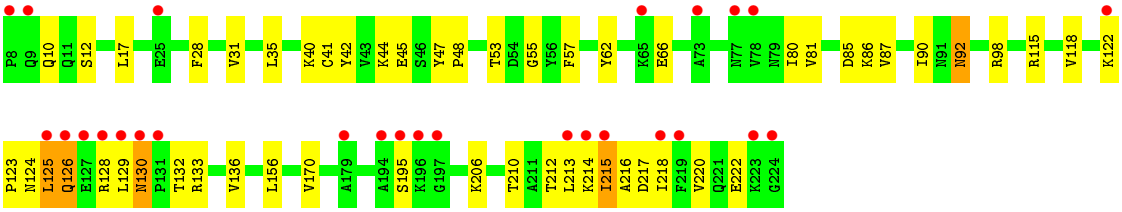




- Molecule 4: Telomere-binding protein beta subunit

Chain B:


12%
76%
22%


P8
Q9
Q10
Q11
S12
L17
E25
F28
V31
L35
K40
C41
Y42
V43
K44
E45
S46
Y47
P48
T53
D54
G55
Y56
F57
Y62
K65
E66
A73
N77
V78
N79
I80
V81
D85
K86
V87
I90
N91
N92
R98
R115
V118
K122
P123
N124
L125
Q126
E127
R128
L129
N130
P131
T132
R133
V136
L156
V170
A179
A194
S195
K196
G197
K206
T210
A211
T212
L213
K214
I215
A216
D217
I218
F219
V220
Q221
E222
K223
G224

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	93.47Å 93.47Å 423.94Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.90 – 2.10 29.90 – 2.00	Depositor EDS
% Data completeness (in resolution range)	92.5 (29.90-2.10) 90.8 (29.90-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 2.00Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.242 , 0.276 0.242 , 0.277	Depositor DCC
R_{free} test set	6779 reflections (9.88%)	wwPDB-VP
Wilson B-factor (Å^2)	27.6	Xtriage
Anisotropy	0.302	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 40.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6448	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.71% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: NA

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	0.54	0/284	0.73	0/440
1	H	0.45	0/284	0.65	0/440
2	D	0.73	0/256	0.83	0/396
3	A	0.55	0/3736	0.73	3/5050 (0.1%)
4	B	0.40	0/1779	0.54	0/2402
All	All	0.52	0/6339	0.68	3/8728 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	86	LYS	N-CA-C	-6.82	92.58	111.00
3	A	406	LYS	N-CA-C	5.60	126.11	111.00
3	A	266	ARG	NE-CZ-NH2	-5.23	117.69	120.30

There are no chirality outliers.

There are no planarity outliers.

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	7	0
1	H	253	0	138	1	0
2	D	229	0	128	8	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	3665	0	3664	64	0
4	B	1740	0	1726	47	0
5	A	5	0	0	0	0
6	A	230	0	0	1	0
6	B	45	0	0	1	0
6	D	18	0	0	0	0
6	G	5	0	0	0	0
6	H	5	0	0	0	0
All	All	6448	0	5794	123	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 10.

The worst 5 of 123 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)
4:B:80:ILE:HD12	4:B:218:ILE:HD11	1.52	0.91
4:B:10:GLN:HE21	4:B:12:SER:H	1.19	0.91
1:G:11:DG:H2"	1:G:12:DG:H5'	1.53	0.89
2:D:6:DT:H2"	2:D:7:DT:H5"	1.57	0.85
2:D:6:DT:H2"	2:D:7:DT:C5'	2.17	0.75

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	450/461 (98%)	429 (95%)	16 (4%)	5 (1%)	14 9
4	B	215/217 (99%)	194 (90%)	18 (8%)	3 (1%)	11 6
All	All	665/678 (98%)	623 (94%)	34 (5%)	8 (1%)	13 8

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	93	ASP
3	A	406	LYS
4	B	130	ASN
4	B	126	GLN
4	B	210	THR

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	406/409 (99%)	400 (98%)	6 (2%)	65 71
4	B	190/190 (100%)	186 (98%)	4 (2%)	53 59
All	All	596/599 (100%)	586 (98%)	10 (2%)	60 67

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	353	LEU
3	A	493	LEU
4	B	92	ASN
3	A	309	LEU
4	B	66	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	331	ASN
3	A	372	GLN
4	B	126	GLN
3	A	327	ASN
4	B	92	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

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

Of 5 ligands modelled in this entry, 5 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

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^{th} 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^2)	Q<0.9
1	G	12/13 (92%)	0.24	0 100 100	32, 42, 50, 53	0
1	H	12/13 (92%)	0.09	0 100 100	40, 45, 52, 56	0
2	D	11/11 (100%)	-0.32	1 (9%) 9 12	23, 31, 47, 51	0
3	A	452/461 (98%)	0.08	18 (3%) 38 44	16, 27, 54, 69	0
4	B	217/217 (100%)	0.84	27 (12%) 4 5	20, 46, 74, 89	0
All	All	704/715 (98%)	0.31	46 (6%) 18 23	16, 32, 67, 89	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	127	GLU	10.2
3	A	406	LYS	9.7
3	A	407	GLY	8.6
4	B	8	PRO	7.6
4	B	126	GLN	6.7

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

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^{th} 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^2)	Q<0.9
5	NA	A	498	1/1	-0.06	1.92	94,94,94,94	0
5	NA	A	499	1/1	0.23	0.36	85,85,85,85	0
5	NA	A	496	1/1	0.38	0.99	99,99,99,99	0
5	NA	A	497	1/1	0.57	0.69	101,101,101,101	0
5	NA	A	500	1/1	0.80	0.66	85,85,85,85	0

6.5 Other polymers (i)

There are no such residues in this entry.