Full wwPDB X-ray Structure Validation Report ⓘ

Oct 11, 2021 – 03:44 PM EDT

PDB ID	:	2H7G
Title	:	Structure of variola topoisomerase non-covalently bound to DNA
Authors	:	Perry, K.; Hwang, Y.; Bushman, F.D.; Van Duyne, G.D.
Deposited on	:	2006-06-02
Resolution	:	1.90 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ symbol.

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

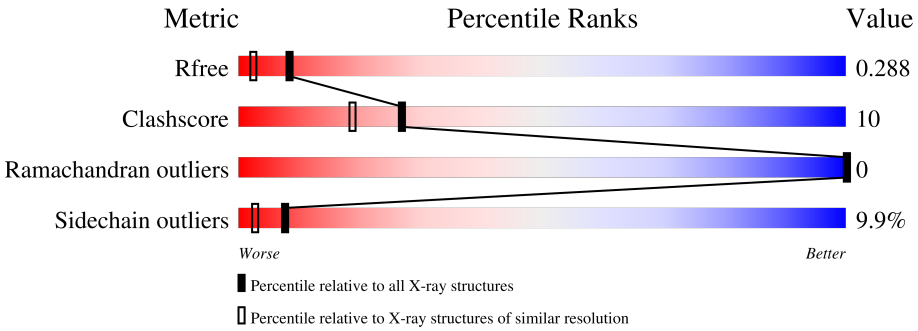
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)

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 of 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%

Mol	Chain	Length	Quality of chain
1	Y	12	33% 58% 8%
2	Z	14	21% 57% 21%
3	X	314	73% 23%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3547 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(*TP*TP*GP*TP*CP*GP*CP*CP*CP*TP*TP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	Y	12	Total	C	N	O	P	0	0	0
			221	106	32	72	11			

- Molecule 2 is a DNA chain called 5'-D(*TP*AP*AP*TP*AP*AP*GP*GP*GP*CP*GP*AP*CP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	Z	14	Total	C	N	O	P	0	1	0
			310	148	65	83	14			

- Molecule 3 is a protein called DNA topoisomerase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	X	312	Total	C	N	O	S	0	7	0
			2621	1694	446	472	9			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
X	100	SER	CYS	engineered mutation	UNP P32989
X	211	SER	CYS	engineered mutation	UNP P32989

- Molecule 4 is water.

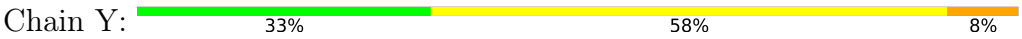
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	Y	30	Total	O	0	0
			30	30		
4	Z	51	Total	O	0	0
			51	51		
4	X	314	Total	O	0	0
			314	314		

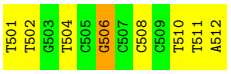
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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(*TP*TP*GP*TP*CP*GP*CP*CP*CP*TP*TP*A)-3'

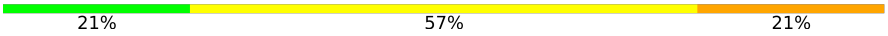
Chain Y:





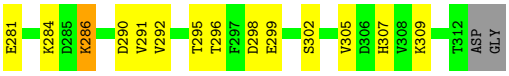
• Molecule 2: 5'-D(*TP*AP*AP*TP*AP*AP*GP*GP*GP*CP*GP*AP*CP*A)-3'

Chain Z:



• Molecule 3: DNA topoisomerase 1

Chain X:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	66.18Å 133.65Å 112.95Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 1.90 40.90 – 1.90	Depositor EDS
% Data completeness (in resolution range)	96.2 (50.00-1.90) 96.1 (40.90-1.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.97 (at 1.89Å)	Xtriage
Refinement program	REFMAC refmac_5.2.0005	Depositor
R, R_{free}	0.197 , 0.243 0.245 , 0.288	Depositor DCC
R_{free} test set	1915 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	35.5	Xtriage
Anisotropy	0.138	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 39.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3547	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.82% 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 (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	Y	1.44	1/244 (0.4%)	2.34	11/375 (2.9%)
2	Z	1.30	1/350 (0.3%)	2.17	17/539 (3.2%)
3	X	0.65	0/2677	0.74	4/3606 (0.1%)
All	All	0.83	2/3271 (0.1%)	1.21	32/4520 (0.7%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	Z	527	DC	O3'-P	-7.03	1.52	1.61
1	Y	510	DT	C5-C7	5.68	1.53	1.50

All (32) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	Y	501	DT	O4'-C1'-N1	-14.13	98.11	108.00
2	Z	520	DA	O4'-C1'-N9	9.57	114.70	108.00
2	Z	517[A]	DA	O4'-C1'-N9	9.29	114.50	108.00
2	Z	517[B]	DA	O4'-C1'-N9	9.29	114.50	108.00
2	Z	524	DC	O4'-C1'-N1	8.97	114.28	108.00
2	Z	523	DG	O4'-C1'-N9	-8.40	102.12	108.00
1	Y	506	DG	C5-C6-O6	-7.75	123.95	128.60
2	Z	518	DT	O4'-C1'-N1	7.54	113.28	108.00
2	Z	527	DC	O4'-C4'-C3'	-7.54	101.48	106.00
1	Y	511	DT	P-O3'-C3'	7.36	128.53	119.70
1	Y	511	DT	C6-C5-C7	-6.88	118.78	122.90
2	Z	526	DA	O4'-C1'-N9	-6.84	103.21	108.00
1	Y	511	DT	C4-C5-C7	6.67	123.00	119.00
2	Z	523	DG	P-O3'-C3'	6.65	127.68	119.70
1	Y	508	DC	P-O3'-C3'	6.55	127.56	119.70
3	X	67	ARG	NE-CZ-NH2	-6.50	117.05	120.30
2	Z	516	DA	O4'-C1'-C2'	6.29	110.93	105.90
1	Y	504	DT	O4'-C1'-N1	-6.28	103.61	108.00
2	Z	526	DA	OP2-P-O3'	5.92	118.22	105.20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	Y	512	DA	O5'-P-OP2	-5.75	100.53	105.70
2	Z	522	DG	C2-N3-C4	5.74	114.77	111.90
1	Y	502	DT	N3-C2-O2	-5.71	118.88	122.30
1	Y	510	DT	C4'-C3'-C2'	5.70	108.23	103.10
2	Z	525	DG	O4'-C1'-N9	-5.70	104.01	108.00
2	Z	527	DC	P-O3'-C3'	5.61	126.44	119.70
2	Z	517[A]	DA	C1'-O4'-C4'	-5.54	104.56	110.10
2	Z	517[B]	DA	C1'-O4'-C4'	-5.54	104.56	110.10
3	X	121	MET	CG-SD-CE	-5.54	91.34	100.20
3	X	80	ARG	NE-CZ-NH2	-5.44	117.58	120.30
2	Z	515	DT	C1'-O4'-C4'	-5.39	104.71	110.10
3	X	245	LEU	CA-CB-CG	5.31	127.51	115.30
1	Y	501	DT	O4'-C1'-C2'	5.09	109.97	105.90

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	Y	221	0	127	2	0
2	Z	310	0	169	3	0
3	X	2621	0	2688	58	0
4	X	314	0	0	10	0
4	Y	30	0	0	0	0
4	Z	51	0	0	1	0
All	All	3547	0	2984	63	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.

All (63) 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:X:149[A]:LYS:H	3:X:152:HIS:HD2	1.06	0.95

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:X:11[B]:LEU:HD21	3:X:31:LEU:HD11	1.54	0.90
3:X:149[B]:LYS:H	3:X:152:HIS:HD2	1.07	0.89
3:X:155:ILE:HD11	3:X:190:THR:HG21	1.65	0.77
3:X:149[A]:LYS:H	3:X:152:HIS:CD2	1.98	0.77
3:X:149[B]:LYS:H	3:X:152:HIS:CD2	1.98	0.76
3:X:116:GLN:HE22	3:X:202:LYS:H	1.38	0.70
3:X:298:ASP:HB3	4:X:422:HOH:O	1.92	0.69
3:X:239:VAL:HG21	3:X:292:VAL:CG1	2.25	0.67
3:X:295:THR:HB	3:X:299:GLU:HG3	1.78	0.64
3:X:239:VAL:HG21	3:X:292:VAL:HG12	1.81	0.62
3:X:291:VAL:O	3:X:295:THR:HG23	1.99	0.61
3:X:11[B]:LEU:HD21	3:X:31:LEU:CD1	2.29	0.60
3:X:260:ALA:HB2	3:X:270:SER:HB2	1.83	0.60
3:X:126:MET:HE3	4:X:397:HOH:O	2.03	0.57
3:X:296:THR:HG22	3:X:299:GLU:HG2	1.86	0.57
3:X:206:ARG:O	3:X:210:GLU:HG2	2.05	0.56
3:X:260:ALA:HB2	3:X:270:SER:CB	2.35	0.56
3:X:296:THR:HG23	3:X:299:GLU:H	1.70	0.56
3:X:128:PHE:CD2	3:X:226:GLY:HA3	2.42	0.54
3:X:252:ILE:HD13	3:X:280:LEU:HD21	1.88	0.54
3:X:286:LYS:O	3:X:286:LYS:HE3	2.07	0.54
2:Z:516:DA:H4'	2:Z:517[B]:DA:OP1	2.07	0.54
3:X:60:VAL:HG22	3:X:68:ARG:NH2	2.23	0.54
1:Y:506:DG:O5'	1:Y:506:DG:H2'	2.08	0.53
3:X:252:ILE:HD13	3:X:280:LEU:CD2	2.38	0.53
2:Z:516:DA:H4'	2:Z:517[A]:DA:OP1	2.07	0.53
3:X:117:LEU:O	3:X:121:MET:HG2	2.09	0.53
3:X:210:GLU:HG3	4:X:359:HOH:O	2.09	0.53
3:X:235:PHE:HD2	4:X:558:HOH:O	1.92	0.52
3:X:269:ILE:HG21	4:X:575:HOH:O	2.10	0.52
3:X:266:THR:OG1	3:X:269:ILE:HG22	2.10	0.52
3:X:97:ARG:HG2	3:X:97:ARG:HH11	1.76	0.50
3:X:239:VAL:HG21	3:X:292:VAL:HG11	1.92	0.50
3:X:247:SER:HB2	3:X:248:PRO:HD2	1.93	0.50
3:X:122:LEU:HG	3:X:182:LEU:HG	1.93	0.49
3:X:163:LYS:HD2	4:X:471:HOH:O	2.13	0.48
3:X:236:TRP:HA	4:X:558:HOH:O	2.13	0.48
3:X:243:SER:HB3	3:X:244:PRO:HD3	1.95	0.47
3:X:100:SER:O	3:X:104:LYS:HD3	2.14	0.47
3:X:75:MET:HG2	4:X:443:HOH:O	2.13	0.47
3:X:150:ASN:HA	3:X:153:ILE:HD12	1.96	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:X:281:GLU:O	3:X:284:LYS:HB2	2.15	0.47
3:X:50:TRP:CZ2	3:X:54:LEU:HD13	2.51	0.46
4:Z:118:HOH:O	3:X:135:LYS:HE3	2.15	0.46
3:X:156:SER:O	3:X:157:PRO:C	2.54	0.45
3:X:172:HIS:HE1	4:X:336:HOH:O	1.98	0.45
3:X:201:ASN:N	3:X:201:ASN:HD22	2.14	0.45
3:X:243:SER:CB	3:X:244:PRO:HD3	2.48	0.44
3:X:276:ALA:O	3:X:280:LEU:HD23	2.19	0.43
3:X:156:SER:HB2	3:X:157:PRO:CD	2.49	0.43
3:X:201:ASN:HD22	3:X:201:ASN:H	1.67	0.43
3:X:281:GLU:OE2	3:X:307:HIS:HE1	2.03	0.42
2:Z:515:DT:H2'	2:Z:516:DA:C5	2.54	0.42
3:X:219:ILE:O	3:X:222:LEU:HB2	2.19	0.42
3:X:305:VAL:O	3:X:309:LYS:HG3	2.20	0.42
3:X:235:PHE:CD2	4:X:558:HOH:O	2.57	0.41
1:Y:506:DG:O5'	1:Y:506:DG:C2'	2.68	0.41
3:X:63:ASP:OD2	3:X:67:ARG:CD	2.69	0.41
3:X:63:ASP:OD2	3:X:67:ARG:HD3	2.21	0.41
3:X:156:SER:HB2	3:X:157:PRO:HD2	2.03	0.40
3:X:194:SER:HB3	3:X:197:GLU:HG3	2.03	0.40
3:X:296:THR:HG22	3:X:299:GLU:CG	2.52	0.40

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	X	317/314 (101%)	309 (98%)	8 (2%)	0	100 100

There are no Ramachandran outliers to report.

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	X	299/293 (102%)	267 (89%)	32 (11%)	6 2

All (32) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	X	2	ARG
3	X	7	LYS
3	X	11[A]	LEU
3	X	11[B]	LEU
3	X	15	ASN
3	X	32	GLN
3	X	60	VAL
3	X	67	ARG
3	X	96	LYS
3	X	104	LYS
3	X	110	SER
3	X	111	THR
3	X	117	LEU
3	X	121	MET
3	X	128	PHE
3	X	137	LEU
3	X	145	LEU
3	X	146	LEU
3	X	148	LEU
3	X	154[A]	GLU
3	X	154[B]	GLU
3	X	182	LEU
3	X	188	LYS
3	X	192	ASP
3	X	201	ASN
3	X	206	ARG
3	X	222	LEU
3	X	257	LYS
3	X	286	LYS
3	X	290	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	X	302[A]	SER
3	X	302[B]	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
3	X	15	ASN
3	X	78	GLN
3	X	93	ASN
3	X	116	GLN
3	X	152	HIS
3	X	172	HIS
3	X	201	ASN
3	X	234	ASN
3	X	307	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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 monosaccharides 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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.