



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 05:21 am GMT

PDB ID : 1SL1
Title : Binary 5' complex of T7 DNA polymerase with a DNA primer/template containing a cis-syn thymine dimer on the template
Authors : Li, Y.; Dutta, S.; Doublié, S.; Bdour, H.M.; Taylor, J.S.; Ellenberger, T.
Deposited on : 2004-03-05
Resolution : 2.20 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

<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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : trunk28620
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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) : recalc28949

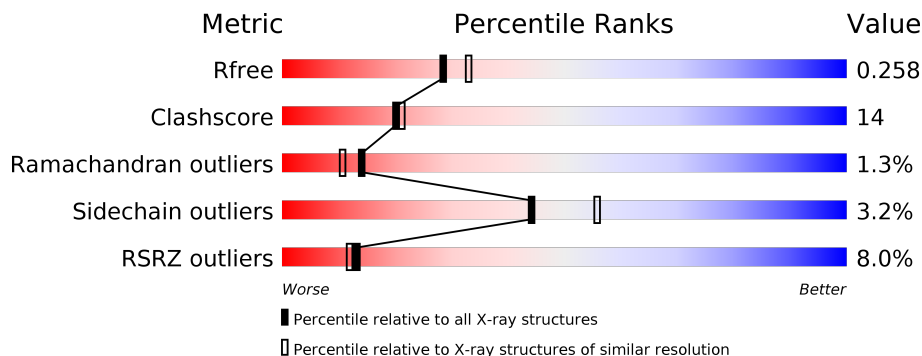
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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}	100719	4002 (2.20-2.20)
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)
RSRZ outliers	101464	4033 (2.20-2.20)

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. 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	P	22	
2	T	25	
3	A	698	
4	B	108	

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	MG	A	4003	-	-	-	X

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 6654 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(*CP*GP*AP*AP*AP*AP*CP*GP*AP*CP*GP*GP*CP*CP*AP*GP*TP*GP*CP*CP*TP*(2DA))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	P	9	Total	C	N	O	P	0	0	0
			172	82	28	53	9			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	10	Total	C	N	O	P	0	0	0
			223	107	42	64	10			

- Molecule 3 is a protein called DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	678	Total	C	N	O	S	0	0	0
			5243	3340	907	972	24			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	LYS	DELETION	UNP P00581
A	?	-	ARG	DELETION	UNP P00581
A	?	-	PHE	DELETION	UNP P00581
A	?	-	GLY	DELETION	UNP P00581
A	?	-	SER	DELETION	UNP P00581
A	?	-	HIS	DELETION	UNP P00581

- Molecule 4 is a protein called Thioredoxin 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	105	Total	C	N	O	S	0	0	0
			776	503	122	148	3			

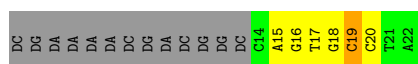
- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total 1	Mg 1	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	213	Total 213	O 213	0	0
6	B	20	Total 20	O 20	0	0
6	T	6	Total 6	O 6	0	0

● Molecule 1: 5'-D(*CP*GP*AP*AP*AP*AP*CP*GP*AP*CP*GP*GP*CP*CP*AP*GP*TP*GP*CP*CP*TP*(2DA))-3'

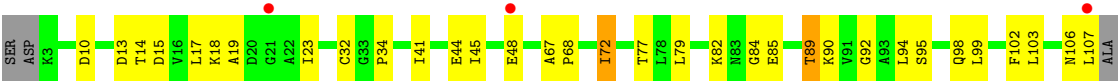


DC	DC	DC	T5	A6	G7	G8	C9	A10	G14	DC	DC	DG	DT	DC	DG	DT	DT	DT	DT	DC	DG
----	----	----	----	----	----	----	----	-----	-----	----	----	----	----	----	----	----	----	----	----	----	----

Glu	A513	Y420	P301	M167	W1
	E514	N421	LYS		
	L515	P422	ASN	Y170	F18
	P516		LYS	M171	
	T517	A425	ALA	V172	T28
	B518		GLN	Q173	A29
		R429	ARG		E50
	A521	A430	GLU	V176	
	I525	T431	GLY	V177	P35
	G527	H432	ARG		R36
D597	F528	A433	E311	L181	E49
	L529	M436	P312		
	Y530	Q439	C313	L185	
	G531	D439	D316		G53
	A532	L440		P193	G54
	G533	P441	V321	P194	
	D534	G442	A322	T204	V57
	E535	P446	G323	T205	H62
	I537	Y447	A324	E209	Q76
	G538	G448	P325	S210	R79
E621	G539		P326	L211	E80
	I540	R452			
		G456	R339	A220	R85
	A543	A457		A221	
	G544	A458	K344	L223	I89
	K545	H459		L224	D90
	E546		Q347		T91
	E547	K467	E346	F232	L92
	K549	Q471	P353	P233	V93
	E550		T354	F234	L94
M649	L551	L474	K355	D235	
	K552		V356		R111
		E480	T357	T239	
	F555	L481	D358	G270	Y130
	L556		K359	R131	
	E557	M488	G360	M273	E134
	D654	A489	A361	F274	
	H558	R490	P362	C275	E138
	T559	F491	V363	C276	
	P560	D492	M276	P277	D341
G659	A561	L496	D365	R278	
	I562				R145
		R566	V368	P284	
		E499	L369		E149
	I569	F500	E370	R288	Q150
		L501	G371	L289	G151
	L573	D504	V372	K290	
	E574	I505	X389	T291	D156
		R596			G157
	SER	T507	K394	V294	M158
L689	SER	K508	G296		
		GLN		C296	
	TRP	N509	K404	L297	W161
		VAL			
	M696	D510	F298		N164
					E165
	G687	T511	V410	K299	F655
		I512		V200	



● Molecule 4: Thioredoxin 1



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	104.38Å 214.78Å 52.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20 36.90 – 2.20	Depositor EDS
% Data completeness (in resolution range)	98.4 (50.00-2.20) 98.6 (36.90-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.03 (at 2.20Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.242 , 0.262 0.240 , 0.258	Depositor DCC
R_{free} test set	2814 reflections (4.72%)	DCC
Wilson B-factor (Å ²)	36.1	Xtriage
Anisotropy	0.256	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 45.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	6654	wwPDB-VP
Average B, all atoms (Å ²)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.41% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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, 2DA, TTD

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	P	0.50	0/180	0.79	0/275
2	T	0.38	0/211	0.76	0/324
3	A	0.34	0/5374	0.57	0/7301
4	B	0.30	0/791	0.55	0/1079
All	All	0.34	0/6556	0.58	0/8979

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	P	0	1
2	T	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	P	19	DC	Sidechain
2	T	8	DG	Sidechain

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	P	172	0	97	11	0
2	T	223	0	122	10	0
3	A	5243	0	4982	141	0
4	B	776	0	761	22	0
5	A	1	0	0	0	0
6	A	213	0	0	4	0
6	B	20	0	0	0	0
6	T	6	0	0	0	0
All	All	6654	0	5962	174	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 14.

The worst 5 of 174 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:131:ARG:HH11	3:A:131:ARG:HB3	1.20	1.07
4:B:19:ALA:HB3	4:B:23:ILE:HD11	1.43	0.99
3:A:642:TRP:HH2	3:A:649:MET:HE2	1.29	0.95
3:A:18:PHE:H	3:A:76:GLN:HE22	1.23	0.86
3:A:501:LEU:HD21	3:A:689:LEU:HB3	1.67	0.77

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	
3	A	672/698 (96%)	641 (95%)	22 (3%)	9 (1%)	14	11
4	B	103/108 (95%)	99 (96%)	3 (3%)	1 (1%)	18	16
All	All	775/806 (96%)	740 (96%)	25 (3%)	10 (1%)	14	11

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	156	ASP
3	A	531	GLY
3	A	360	GLY
3	A	514	GLU
3	A	544	GLY

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	
3	A	520/579 (90%)	504 (97%)	16 (3%)	45	57
4	B	78/87 (90%)	75 (96%)	3 (4%)	38	47
All	All	598/666 (90%)	579 (97%)	19 (3%)	44	56

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	481	LEU
3	A	518	ARG
3	A	686	PHE
3	A	429	ARG
4	B	10	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	343	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	B	98	GLN
3	A	347	GLN
3	A	164	ASN
3	A	510	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues 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	$\# Z > 2$
1	2DA	P	22	1	7,10,23	0.64	0	6,12,34	0.43	0
2	TTD	T	5	2	39,40,46	3.96	9 (23%)	57,67,77	3.88	23 (40%)

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
1	2DA	P	22	1	-	0/3/12/19	0/1/1/3
2	TTD	T	5	2	-	0/19/101/110	0/3/6/6

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	5	TTD	C5T-C6T	-19.52	1.32	1.55
2	T	5	TTD	C5-C6	-10.25	1.43	1.55
2	T	5	TTD	C2-N3	-3.76	1.31	1.38

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	5	TTD	O4R-C1R	-2.11	1.37	1.42
2	T	5	TTD	C5M-C5T	2.86	1.59	1.53

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	5	TTD	O4'-C1'-N1	-9.65	97.22	108.65
2	T	5	TTD	O5R-PB-O5P	-7.16	80.36	109.25
2	T	5	TTD	O4'-C1'-C2'	-6.77	97.12	107.11
2	T	5	TTD	C5A-C5-C4	-6.27	97.41	108.22
2	T	5	TTD	O4P-PB-O5R	-5.22	83.51	108.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	T	5	TTD	5	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is 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, 95th 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	P	8/22 (36%)	-0.16	0 100 100	42, 49, 60, 68	0
2	T	9/25 (36%)	0.36	0 100 100	35, 49, 64, 68	0
3	A	678/698 (97%)	0.33	61 (8%) 10 9	19, 35, 64, 71	0
4	B	105/108 (97%)	0.13	3 (2%) 52 50	31, 45, 58, 60	0
All	All	800/853 (93%)	0.30	64 (8%) 13 12	19, 37, 63, 71	0

The worst 5 of 64 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	544	GLY	7.5
3	A	548	GLY	6.2
3	A	513	ALA	6.0
3	A	298	PHE	5.0
3	A	313	CYS	4.9

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. 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, 95th 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(Å ²)	Q<0.9
1	2DA	P	22	10/21	0.88	0.14	-	59,62,64,64	0
2	TTD	T	5	35/41	0.67	0.31	-	71,73,79,80	0

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. 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, 95th 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(\AA^2)	Q<0.9
5	MG	A	4003	1/1	0.55	0.34	8.97	47,47,47,47	0

6.5 Other polymers [i](#)

There are no such residues in this entry.