



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 2, 2023 – 08:13 PM EDT

PDB ID : 3RMC
Title : Crystal Structure of a replicative DNA polymerase bound to DNA containing Thymine Glycol
Authors : Aller, P.; Duclos, S.; Wallace, S.S.; Doublet, S.
Deposited on : 2011-04-20
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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.35
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) : 2.35

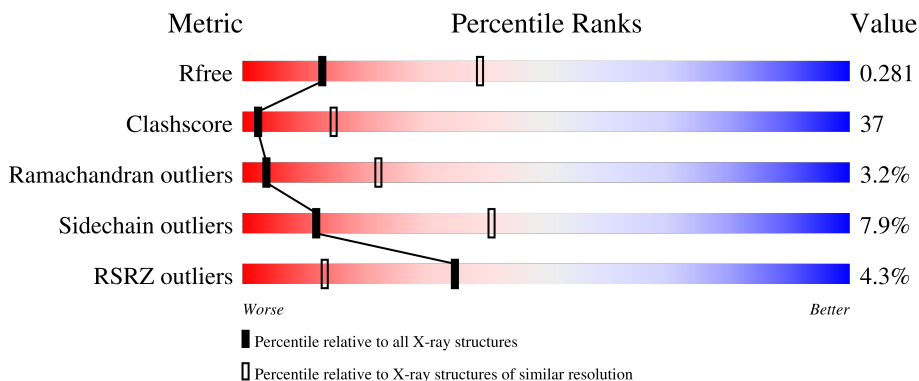
1 Overall quality at a glance

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.



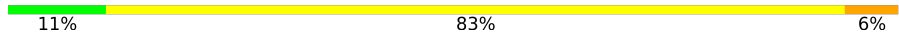
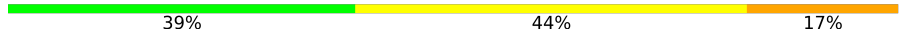

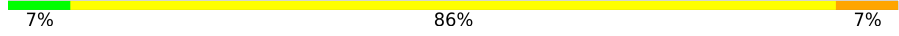


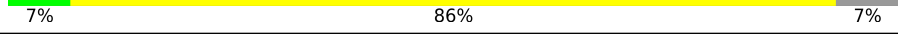
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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 $\leq 5\%$. 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	906	
1	B	906	
1	C	906	
1	D	906	
2	E	18	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	18	 11% 83% 6%
2	I	18	 39% 44% 17%
2	K	18	 6% 44% 33% 22%
3	F	14	 7% 86% 7%
3	H	14	 21% 79%
3	J	14	 50% 50%
3	L	14	 7% 7% 86% 7%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 31561 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 DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	902	Total 7323	C 4704	N 1220	O 1367	S 32	0	0	0
1	B	902	Total 7246	C 4643	N 1202	O 1368	S 33	13	0	0
1	C	901	Total 7339	C 4712	N 1220	O 1374	S 33	0	0	0
1	D	890	Total 7027	C 4507	N 1155	O 1334	S 31	15	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	222	ALA	ASP	engineered mutation	UNP Q38087
A	327	ALA	ASP	engineered mutation	UNP Q38087
A	904	HIS	-	expression tag	UNP Q38087
A	905	HIS	-	expression tag	UNP Q38087
A	906	HIS	-	expression tag	UNP Q38087
B	222	ALA	ASP	engineered mutation	UNP Q38087
B	327	ALA	ASP	engineered mutation	UNP Q38087
B	904	HIS	-	expression tag	UNP Q38087
B	905	HIS	-	expression tag	UNP Q38087
B	906	HIS	-	expression tag	UNP Q38087
C	222	ALA	ASP	engineered mutation	UNP Q38087
C	327	ALA	ASP	engineered mutation	UNP Q38087
C	904	HIS	-	expression tag	UNP Q38087
C	905	HIS	-	expression tag	UNP Q38087
C	906	HIS	-	expression tag	UNP Q38087
D	222	ALA	ASP	engineered mutation	UNP Q38087
D	327	ALA	ASP	engineered mutation	UNP Q38087
D	904	HIS	-	expression tag	UNP Q38087
D	905	HIS	-	expression tag	UNP Q38087
D	906	HIS	-	expression tag	UNP Q38087

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	18	Total	C	N	O	P	0	0	0
			370	175	71	107	17			
2	G	18	Total	C	N	O	P	0	0	0
			370	175	71	107	17			
2	I	18	Total	C	N	O	P	0	0	0
			370	175	71	107	17			
2	K	14	Total	C	N	O	P	0	0	0
			287	136	59	79	13			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	14	Total	C	N	O	P	0	0	0
			282	136	50	83	13			
3	H	14	Total	C	N	O	P	0	0	0
			282	136	50	83	13			
3	J	14	Total	C	N	O	P	0	0	0
			282	136	50	83	13			
3	L	13	Total	C	N	O	P	0	0	0
			262	126	45	79	12			

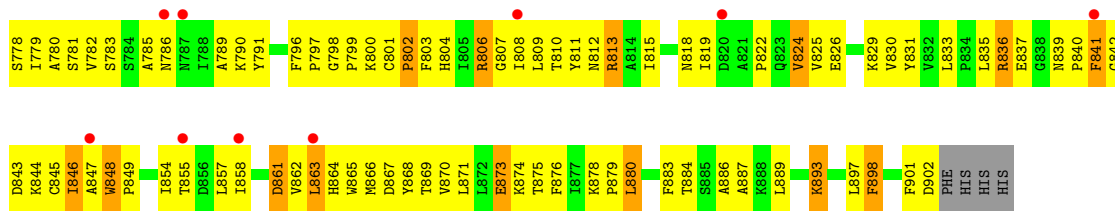
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	43	Total	O	0	0
			43	43		
4	B	33	Total	O	0	0
			33	33		
4	C	32	Total	O	0	0
			32	32		
4	D	2	Total	O	0	0
			2	2		
4	E	1	Total	O	0	0
			1	1		
4	G	3	Total	O	0	0
			3	3		
4	H	3	Total	O	0	0
			3	3		
4	I	3	Total	O	0	0
			3	3		

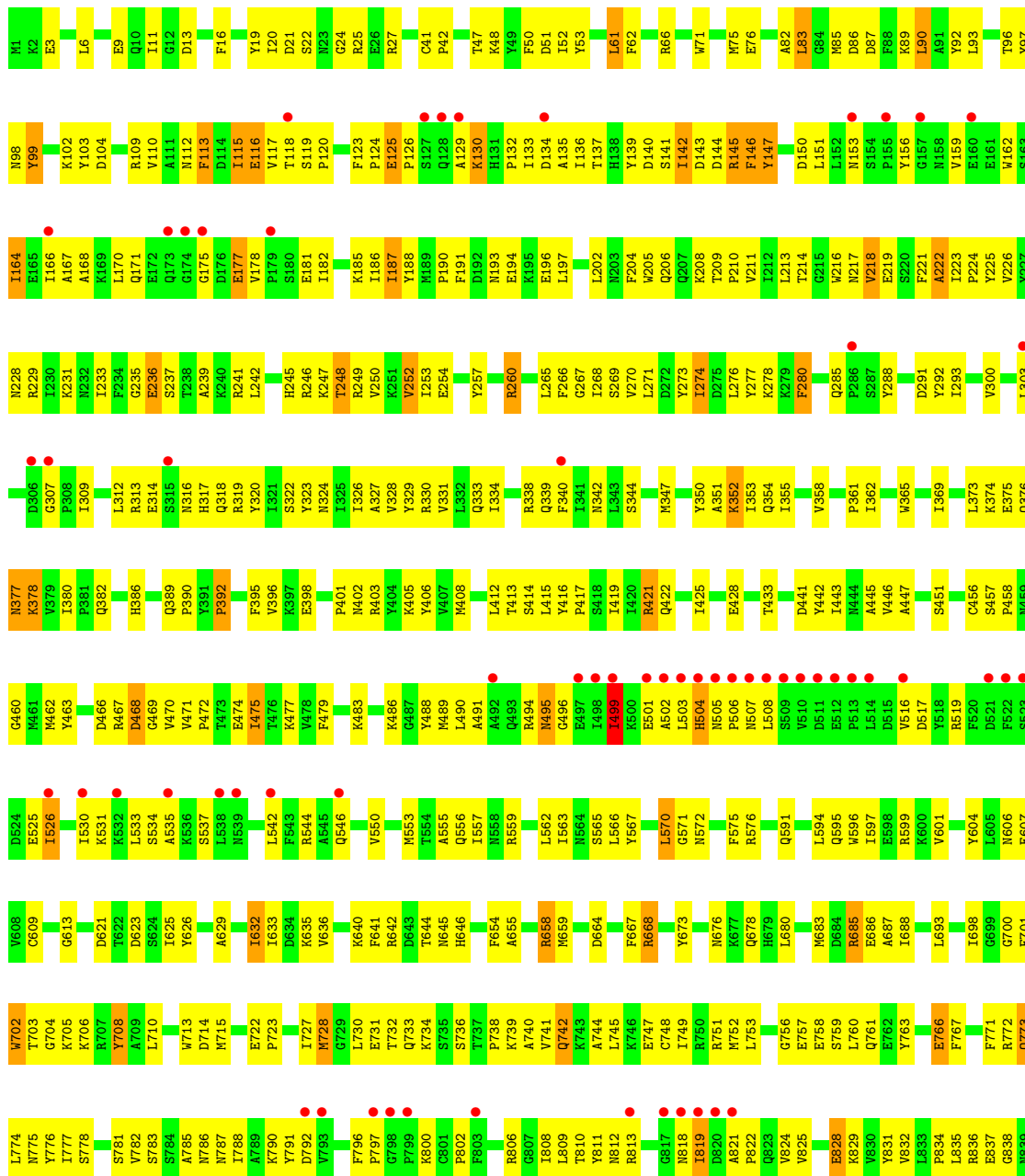
Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	J	1	Total	O	0	0
			1	1		

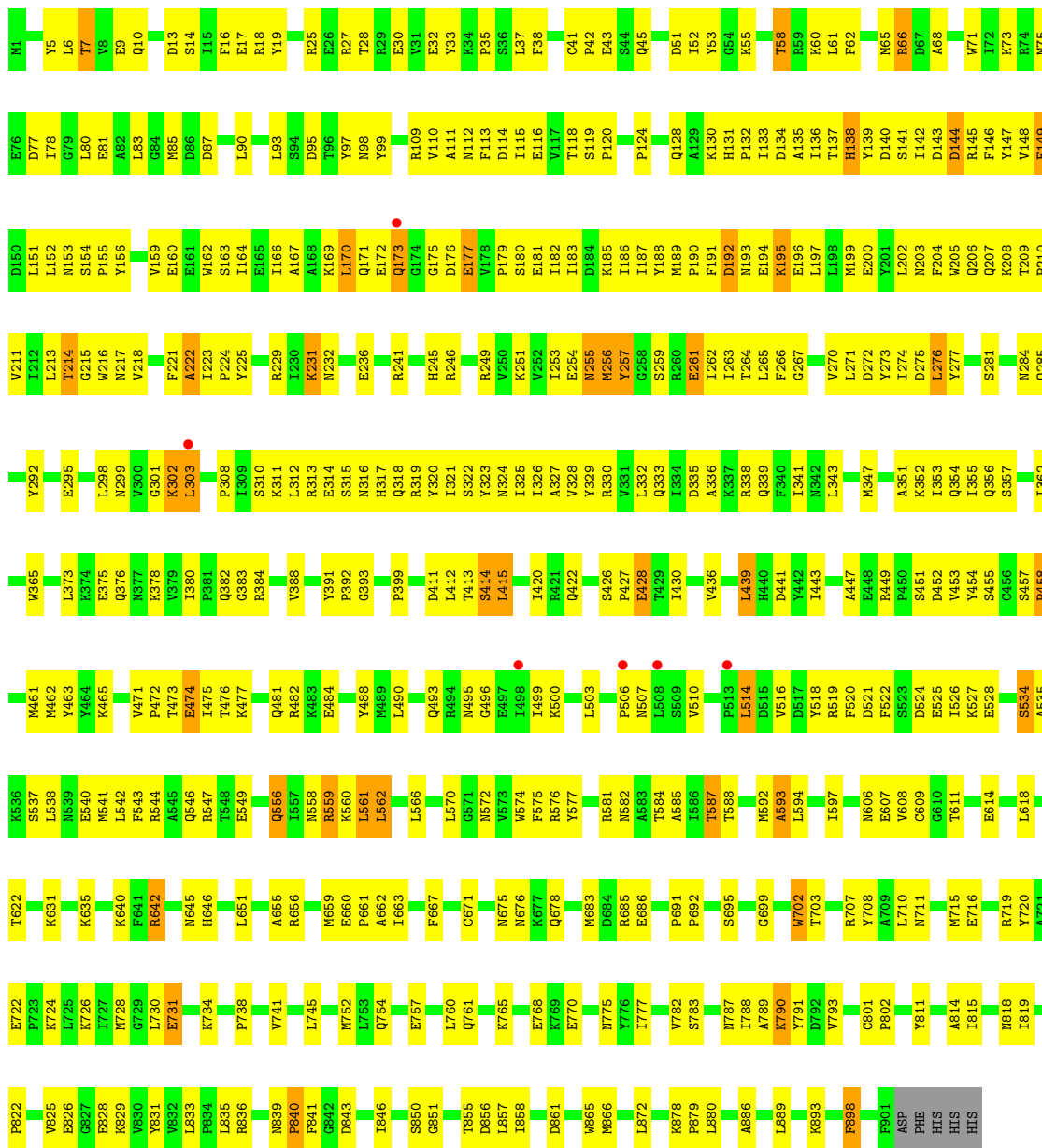


● Molecule 1: DNA polymerase

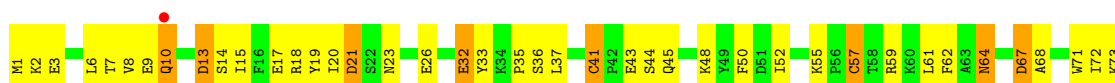


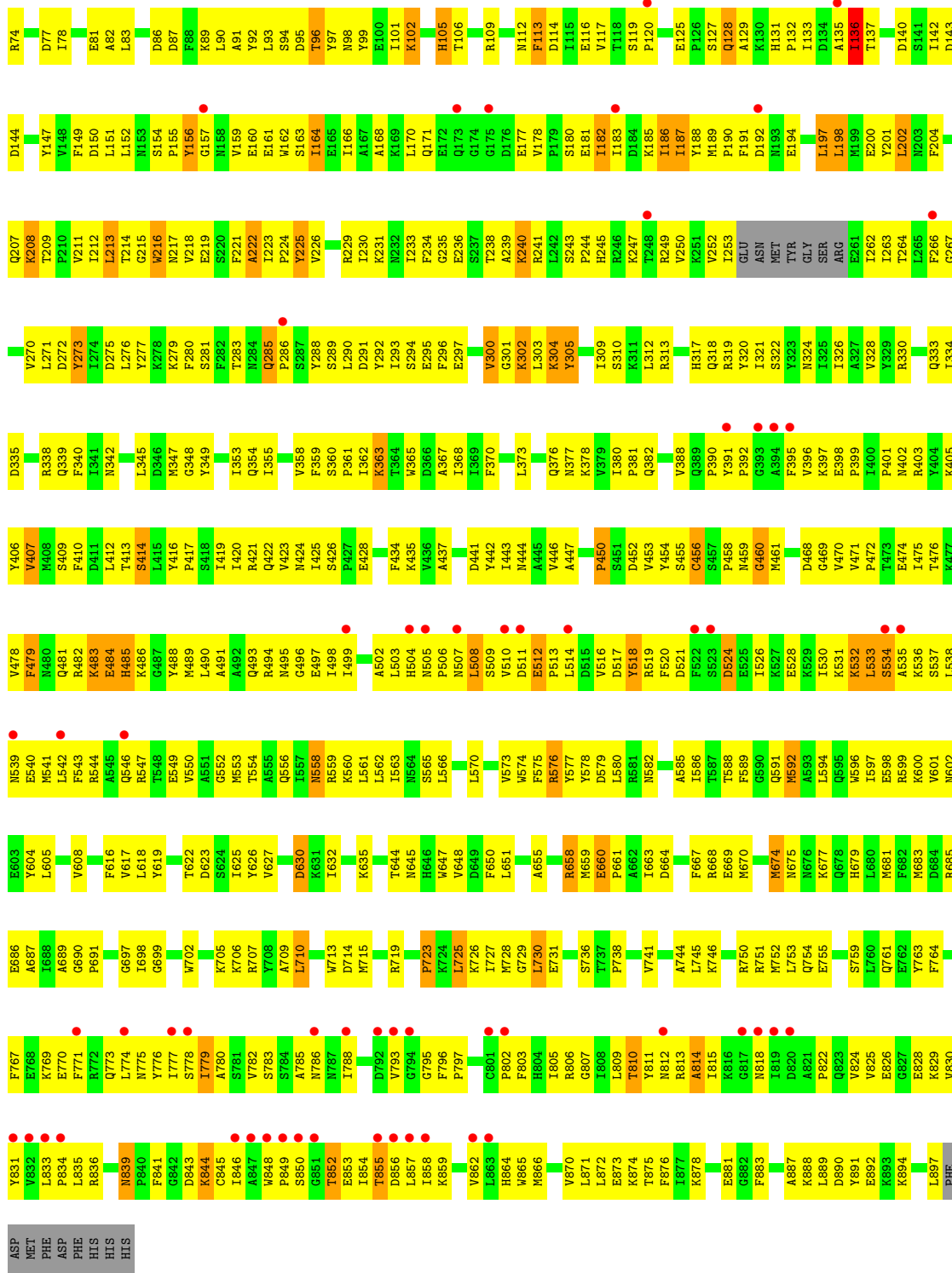


• Molecule 1: DNA polymerase

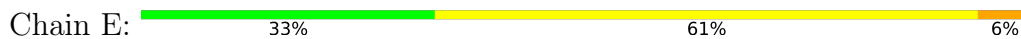


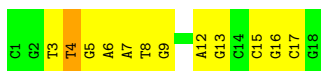
• Molecule 1: DNA polymerase





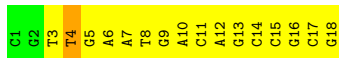
• Molecule 2: DNA (5'-D(*CP*GP*TP*(CTG)P*GP*AP*AP*TP*GP*AP*CP*AP*GP*CP*C
P*GP*CP*G)-3')





- Molecule 2: DNA (5'-D(*CP*GP*TP*(CTG)P*GP*AP*AP*TP*GP*AP*CP*AP*GP*CP*C P*GP*CP*G)-3')

Chain G: 11% 83% 6%



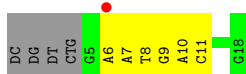
- Molecule 2: DNA (5'-D(*CP*GP*TP*(CTG)P*GP*AP*AP*TP*GP*AP*CP*AP*GP*CP*C P*GP*CP*G)-3')

Chain I: 39% 44% 17%



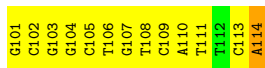
- Molecule 2: DNA (5'-D(*CP*GP*TP*(CTG)P*GP*AP*AP*TP*GP*AP*CP*AP*GP*CP*C P*GP*CP*G)-3')

Chain K: 6% 44% 33% 22%



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

Chain F: 7% 86% 7%



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

Chain H: 21% 79%

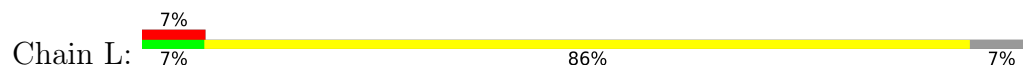


- Molecule 3: DNA (5'-D(*GP*CP*GP*GP*CP*TP*GP*TP*CP*AP*TP*TP*CP*A)-3')

Chain J: 50% 50%



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	133.99Å 123.77Å 163.64Å 90.00° 95.57° 90.00°	Depositor
Resolution (Å)	50.00 – 3.00 49.30 – 3.00	Depositor EDS
% Data completeness (in resolution range)	90.5 (50.00-3.00) 96.5 (49.30-3.00)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.01 (at 3.01Å)	Xtrriage
Refinement program	CNS	Depositor
R, R_{free}	0.222 , 0.282 0.219 , 0.281	Depositor DCC
R_{free} test set	19675 reflections (9.40%)	wwPDB-VP
Wilson B-factor (Å ²)	68.4	Xtrriage
Anisotropy	0.120	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 69.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	31561	wwPDB-VP
Average B, all atoms (Å ²)	92.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.59% 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: CTG

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	0.46	0/7503	0.68	0/10148
1	B	0.42	0/7421	0.62	0/10054
1	C	0.46	0/7519	0.66	0/10166
1	D	0.34	0/7198	0.59	1/9779 (0.0%)
2	E	0.40	0/390	0.71	0/598
2	G	0.39	0/390	0.69	0/598
2	I	0.60	0/390	0.77	0/598
2	K	0.31	0/323	0.67	0/497
3	F	0.33	0/315	0.77	0/484
3	H	0.30	0/315	0.70	0/484
3	J	0.62	0/315	0.89	0/484
3	L	0.28	0/292	0.64	0/449
All	All	0.42	0/32371	0.65	1/44339 (0.0%)

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	A	0	1
2	I	0	2
3	F	0	1
All	All	0	4

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	D	533	LEU	CA-CB-CG	5.08	126.97	115.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	92	TYR	Sidechain
3	F	114	DA	Sidechain
2	I	11	DC	Sidechain
2	I	7	DA	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	A	7323	0	7185	536	0
1	B	7246	0	7027	512	0
1	C	7339	0	7210	456	0
1	D	7027	0	6703	622	0
2	E	370	0	205	19	0
2	G	370	0	205	34	0
2	I	370	0	205	14	0
2	K	287	0	157	11	0
3	F	282	0	158	27	0
3	H	282	0	158	19	0
3	J	282	0	158	21	0
3	L	262	0	149	18	0
4	A	43	0	0	3	0
4	B	33	0	0	2	0
4	C	32	0	0	3	0
4	D	2	0	0	0	0
4	E	1	0	0	0	0
4	G	3	0	0	0	0
4	H	3	0	0	1	0
4	I	3	0	0	0	0
4	J	1	0	0	0	0
All	All	31561	0	29520	2248	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 37.

The worst 5 of 2248 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:D:202:LEU:HD21	1:D:241:ARG:HD3	1.26	1.14
1:D:533:LEU:HD13	1:D:534:SER:H	1.05	1.12
1:D:116:GLU:HB2	1:D:135:ALA:HB3	1.32	1.12
1:C:642:ARG:HE	1:C:646:HIS:CD2	1.67	1.12
3:J:104:DG:H2''	3:J:105:DC:H5''	1.20	1.12

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	900/906 (99%)	745 (83%)	129 (14%)	26 (3%)	4	24
1	B	900/906 (99%)	761 (85%)	119 (13%)	20 (2%)	6	31
1	C	899/906 (99%)	748 (83%)	125 (14%)	26 (3%)	4	24
1	D	886/906 (98%)	710 (80%)	133 (15%)	43 (5%)	2	13
All	All	3585/3624 (99%)	2964 (83%)	506 (14%)	115 (3%)	4	22

5 of 115 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	300	VAL
1	A	790	LYS
1	A	863	LEU
1	B	819	ILE
1	C	177	GLU

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
1	A	789/803 (98%)	715 (91%)	74 (9%)	8 32
1	B	773/803 (96%)	717 (93%)	56 (7%)	14 45
1	C	794/803 (99%)	745 (94%)	49 (6%)	18 52
1	D	737/803 (92%)	672 (91%)	65 (9%)	10 36
All	All	3093/3212 (96%)	2849 (92%)	244 (8%)	12 41

5 of 244 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	668	ARG
1	D	520	PHE
1	C	200	GLU
1	D	511	ASP
1	D	725	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 98 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	255	ASN
1	C	761	GLN
1	C	285	GLN
1	C	558	ASN
1	D	128	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

3 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
2	CTG	E	4	2,3	19,23,24	0.81	1 (5%)	21,35,38	1.10	1 (4%)
2	CTG	G	4	2,3	19,23,24	0.91	2 (10%)	21,35,38	1.13	2 (9%)
2	CTG	I	4	2,3	19,23,24	1.07	2 (10%)	21,35,38	1.24	3 (14%)

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	CTG	E	4	2,3	-	1/7/45/46	0/2/2/2
2	CTG	G	4	2,3	-	0/7/45/46	0/2/2/2
2	CTG	I	4	2,3	-	0/7/45/46	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	4	CTG	C1'-N1	2.76	1.49	1.45
2	E	4	CTG	C1'-N1	2.42	1.48	1.45
2	I	4	CTG	O5'-C5'	-2.28	1.39	1.44
2	G	4	CTG	C1'-N1	2.27	1.48	1.45
2	G	4	CTG	C5-C4	2.08	1.54	1.52

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	I	4	CTG	C2'-C1'-N1	-3.59	110.74	115.59
2	E	4	CTG	C2'-C1'-N1	-3.30	111.13	115.59
2	G	4	CTG	C2'-C1'-N1	-3.22	111.23	115.59
2	G	4	CTG	N3-C2-N1	-2.49	114.11	116.69
2	I	4	CTG	N3-C2-N1	-2.15	114.46	116.69

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	E	4	CTG	O4'-C4'-C5'-O5'

There are no ring outliers.

3 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	4	CTG	3	0
2	G	4	CTG	4	0
2	I	4	CTG	4	0

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 [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	A	902/906 (99%)	-0.13	23 (2%) 57 29	29, 63, 155, 176	0
1	B	902/906 (99%)	0.18	67 (7%) 14 4	30, 84, 172, 189	2 (0%)
1	C	901/906 (99%)	-0.23	6 (0%) 87 69	23, 64, 127, 148	0
1	D	890/906 (98%)	0.29	61 (6%) 16 5	72, 132, 174, 188	3 (0%)
2	E	17/18 (94%)	0.13	0 100 100	75, 106, 144, 148	0
2	G	17/18 (94%)	0.16	0 100 100	76, 113, 137, 147	0
2	I	17/18 (94%)	-0.26	0 100 100	39, 53, 124, 140	0
2	K	14/18 (77%)	0.52	1 (7%) 16 5	69, 156, 169, 170	0
3	F	14/14 (100%)	-0.11	0 100 100	85, 125, 157, 161	0
3	H	14/14 (100%)	-0.10	0 100 100	89, 132, 155, 156	0
3	J	14/14 (100%)	-0.55	0 100 100	35, 55, 123, 127	0
3	L	13/14 (92%)	0.48	1 (7%) 13 4	146, 165, 169, 175	0
All	All	3715/3752 (99%)	0.02	159 (4%) 35 13	23, 83, 165, 189	5 (0%)

The worst 5 of 159 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	504	HIS	8.2
1	D	847	ALA	7.7
1	D	862	VAL	7.4
1	D	535	ALA	6.5
1	B	503	LEU	6.0

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, 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	B-factors(Å ²)	Q<0.9
2	CTG	G	4	22/23	0.74	0.27	122,128,129,129	0
2	CTG	E	4	22/23	0.86	0.22	109,113,124,125	0
2	CTG	I	4	22/23	0.93	0.21	72,78,81,81	0

6.3 Carbohydrates [i](#)

There are no monosaccharides 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.