wwPDB X-ray Structure Validation Summary Report (i)

Aug 8, 2023 – 12:42 PM EDT

PDB ID	:	1Q9Y
Title	:	CRYSTAL STRUCTURE OF ENTEROBACTERIA PHAGE RB69 GP43 DNA POLYMERASE COMPLEXED WITH 8-OXOGUANOSINE CONTAINING DNA
Authors	:	Freisinger, E.; Grollman, A.P.; Miller, H.; Kisker, C.
Deposited on	:	2003-08-26
Resolution	:	2.80 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

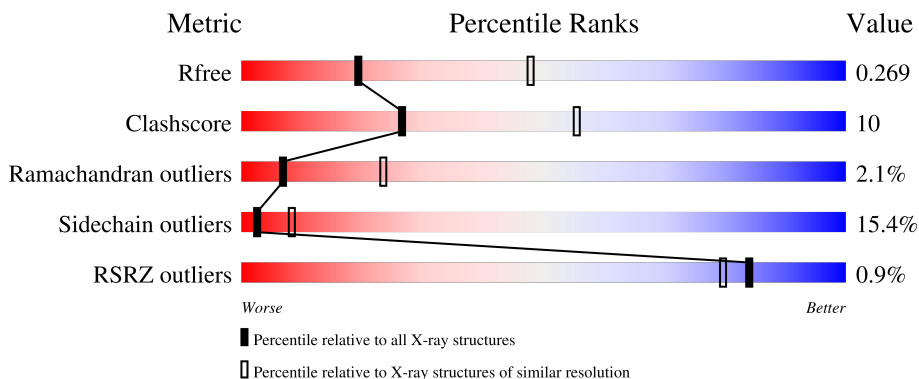|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & 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 2.80 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	Percentile Ranks	Value
Rfree		0.269
Clashscore		10
Ramachandran outliers		2.1%
Sidechain outliers		15.4%
RSRZ outliers		0.9%

Worse   *Better*

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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% 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	T	18	6% 67% 28%
2	P	14	57% 43%
3	A	906	% 68% 24% 6% ·

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 8123 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'-AC(8-OXOGUANOSINE)GGTAAGCAGTCCGCG-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	T	18	Total	C	N	O	P	0	0	0
			371	175	74	105	17			

- Molecule 2 is a DNA chain called 5'-GCGGACTGCTTAC(DIDEOXYCYTIDINE)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	14	Total	C	N	O	P	0	0	0
			281	135	51	82	13			

- Molecule 3 is a protein called DNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	902	Total	C	N	O	S	0	0	0
			7356	4721	1227	1376	32			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	HIS	-	cloning artifact	UNP Q38087
A	2	GLY	-	cloning artifact	UNP Q38087
A	3	SER	-	cloning artifact	UNP Q38087
A	225	ALA	ASP	engineered mutation	UNP Q38087
A	330	ALA	ASP	engineered mutation	UNP Q38087

- Molecule 4 is 2'-DEOXYCYTIDINE-5'-TRIPHOSPHATE (three-letter code: DCP) (formula: $C_9H_{16}N_3O_{13}P_3$).

DCP
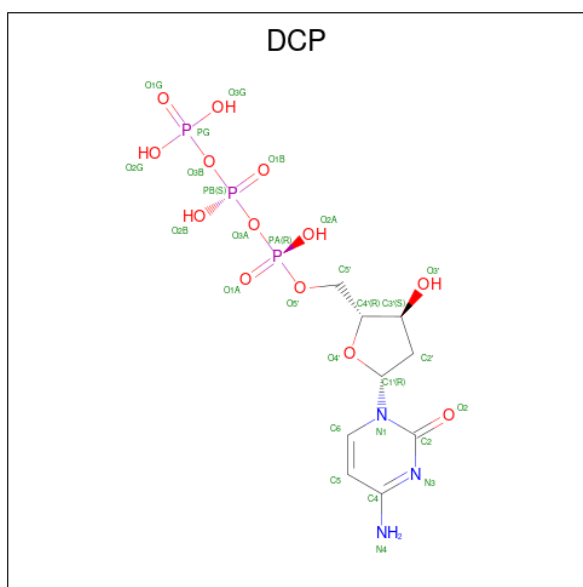
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
4	P	1	28	9	3	13	3	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
5	A	3	3	3	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	T	5	Total	O	0	0
			5	5		
6	P	2	Total	O	0	0
			2	2		
6	A	77	Total	O	0	0
			77	77		

3 Residue-property plots ⓘ

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 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.
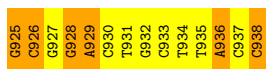
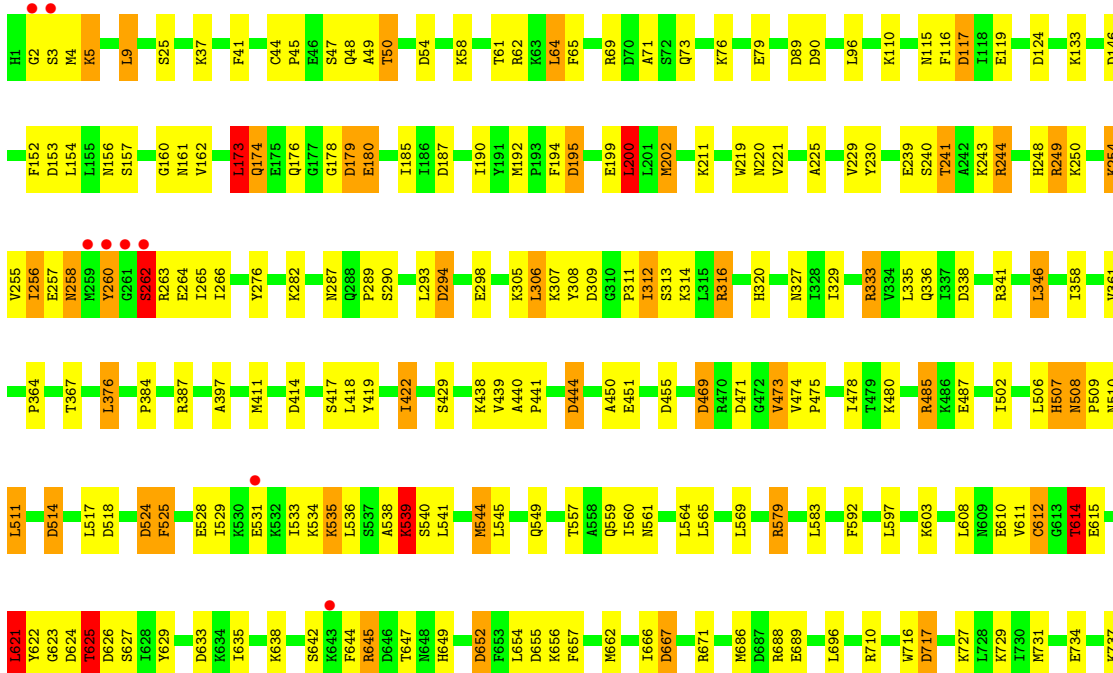• Molecule 1: 5'-AC(8-OXOGUANOSINE)GGTAAGCAGTCCGCG-3'

Chain T:
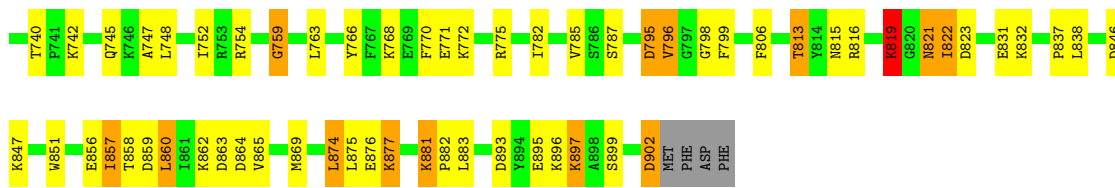

• Molecule 2: 5'-GCGGACTGCTTAC(DIDEOXYCYTIDINE)-3'

Chain P:

• Molecule 3: DNA POLYMERASE

Chain A:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	80.82Å 118.56Å 127.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.06 – 2.80 46.06 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.5 (46.06-2.80) 91.0 (46.06-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	2.51 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.204 , 0.271 0.203 , 0.269	Depositor DCC
R_{free} test set	1579 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	47.4	Xtriage
Anisotropy	0.079	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 35.9	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	8123	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 4.38% 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: 8OG, DOC, DCP, CA

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	T	1.40	0/390	2.49	35/598 (5.9%)
2	P	1.38	1/294 (0.3%)	2.61	27/452 (6.0%)
3	A	0.75	0/7536	0.97	31/10184 (0.3%)
All	All	0.82	1/8220 (0.0%)	1.21	93/11234 (0.8%)

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
3	A	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	933	DC	C3'-O3'	-5.57	1.36	1.44

The worst 5 of 93 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	P	931	DT	O4'-C1'-N1	11.90	116.33	108.00
2	P	928	DG	O4'-C1'-N9	11.50	116.05	108.00
3	A	469	ASP	CB-CG-OD2	10.44	127.69	118.30
2	P	936	DA	O5'-P-OP2	-10.27	96.46	105.70
1	T	921	DC	O4'-C4'-C3'	-9.00	100.60	106.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	3	SER	Peptide

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	T	371	0	202	5	0
2	P	281	0	159	7	0
3	A	7356	0	7251	153	0
4	P	28	0	12	1	0
5	A	3	0	0	0	0
6	A	77	0	0	10	0
6	P	2	0	0	0	0
6	T	5	0	0	0	0
All	All	8123	0	7624	161	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 161 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:202:MET:CE	3:A:202:MET:SD	2.01	1.48
3:A:174:GLN:HE21	3:A:174:GLN:HA	1.30	0.97
3:A:897:LYS:HA	3:A:897:LYS:NZ	1.91	0.86
3:A:508:ASN:O	3:A:508:ASN:CG	2.19	0.80
3:A:173:LEU:HD22	3:A:173:LEU:H	1.47	0.79

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	900/906 (99%)	813 (90%)	68 (8%)	19 (2%)	7	23

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	256	ILE
3	A	258	ASN
3	A	306	LEU
3	A	625	THR
3	A	759	GLY

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	798/802 (100%)	675 (85%)	123 (15%)	2	8

5 of 123 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	451	GLU
3	A	846	ASP
3	A	535	LYS
3	A	838	LEU
3	A	881	LYS

Sometimes 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	567	ASN
3	A	649	HIS
3	A	789	ASN
3	A	764	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	A	336	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	$\#|Z| > 2$
2	DOC	P	938	2,1	16,19,20	1.05	1 (6%)	20,26,29	0.92	0
1	8OG	T	909	1	22,25,26	1.61	2 (9%)	30,37,40	3.12	6 (20%)

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	DOC	P	938	2,1	-	2/7/18/19	0/2/2/2
1	8OG	T	909	1	-	2/7/21/22	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	T	909	8OG	C8-N7	-5.84	1.27	1.38
1	T	909	8OG	C8-N9	-3.87	1.33	1.40
2	P	938	DOC	C4-N3	-2.58	1.29	1.34

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	T	909	8OG	O4'-C1'-N9	9.40	117.75	108.29
1	T	909	8OG	N7-C8-N9	8.35	116.49	106.58
1	T	909	8OG	C2'-C1'-N9	-7.72	106.88	116.01
1	T	909	8OG	O8-C8-N9	-6.31	117.18	125.99
1	T	909	8OG	C5-N7-C8	-3.23	104.82	109.47

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	T	909	8OG	O4'-C1'-N9-C8
1	T	909	8OG	C2'-C1'-N9-C4
2	P	938	DOC	C3'-C4'-C5'-O5'
2	P	938	DOC	O4'-C4'-C5'-O5'

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	P	938	DOC	1	0
1	T	909	8OG	1	0

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

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	#$|Z| > 2$
4	DCP	P	939	5	25,29,29	1.04	1 (4%)	37,45,45	1.64	7 (18%)

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
4	DCP	P	939	5	-	4/22/34/34	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	P	939	DCP	C6-C5	2.92	1.41	1.35

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	P	939	DCP	O4'-C1'-N1	3.70	114.48	107.86
4	P	939	DCP	O2-C2-N3	-3.38	116.83	122.33
4	P	939	DCP	C5-C4-N4	2.97	125.25	120.57
4	P	939	DCP	C2'-C1'-N1	-2.84	107.22	113.77
4	P	939	DCP	N1-C2-N3	2.81	123.92	118.81

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	P	939	DCP	PB-O3B-PG-O2G
4	P	939	DCP	PA-O3A-PB-O2B
4	P	939	DCP	O4'-C4'-C5'-O5'
4	P	939	DCP	PB-O3B-PG-O1G

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	P	939	DCP	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be

highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand DCP P 939

Bond lengths   Bond angles

Torsions   Rings
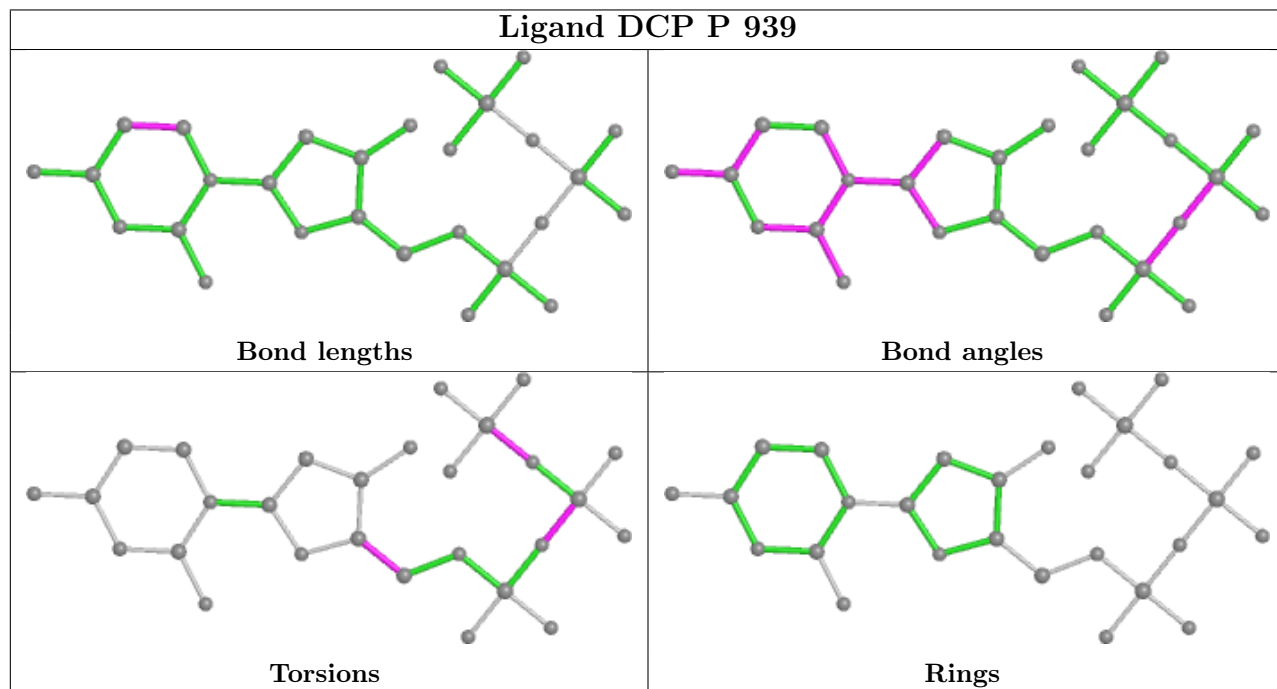
5.7 Other polymers ⓘ

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, 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	T	17/18 (94%)	-0.49	0	100	100	17, 32, 51, 60	0
2	P	13/14 (92%)	-0.35	0	100	100	20, 33, 51, 56	0
3	A	902/906 (99%)	-0.36	8 (0%)	84	80	14, 35, 64, 90	0
All	All	932/938 (99%)	-0.36	8 (0%)	84	80	14, 35, 64, 90	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	260	TYR	4.3
3	A	3	SER	3.0
3	A	262	SER	2.6
3	A	261	GLY	2.5
3	A	2	GLY	2.3

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

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.

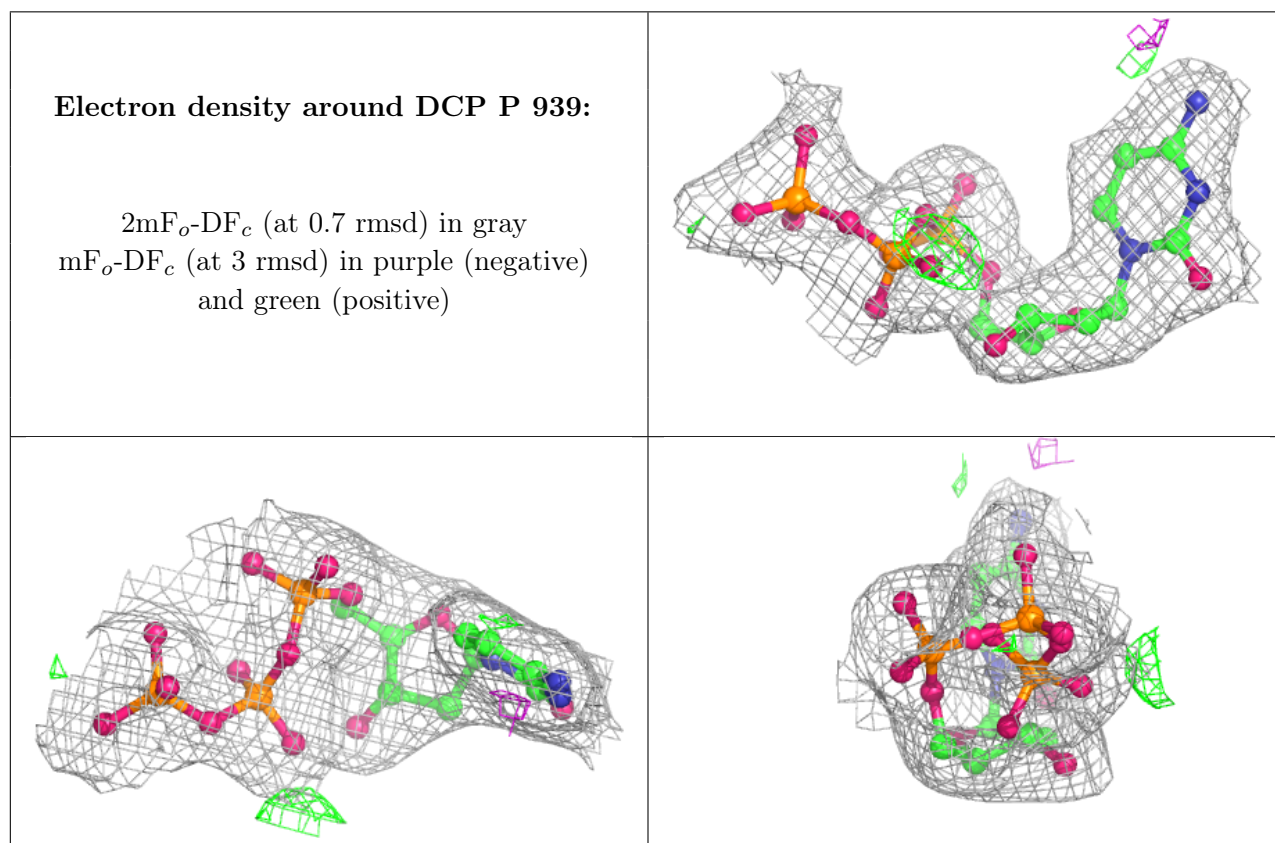| Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	DOC	P	938	18/19	0.95	0.19	12,21,33,36	0
1	8OG	T	909	23/24	0.99	0.15	9,14,18,22	0

6.3   Carbohydrates ⓘ

There are no monosaccharides 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, 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
5	CA	A	940	1/1	0.70	0.29	55,55,55,55	0
5	CA	A	941	1/1	0.88	0.29	51,51,51,51	0
4	DCP	P	939	28/28	0.98	0.17	1,19,25,31	0
5	CA	A	939	1/1	0.98	0.11	30,30,30,30	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around DCP P 939:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.