



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 7, 2017 – 08:59 PM EDT

PDB ID : 4F4Y
Title : Y-family DNA polymerase chimera Dbh-Dpo4-Dbh
Authors : Pata, J.D.; Wilson, R.C.
Deposited on : unknown
Resolution : 2.34 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	:	rb-20029824
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)	:	rb-20029824

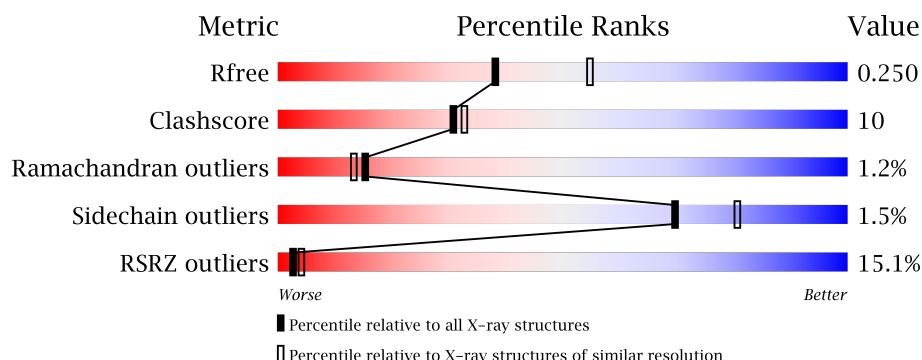
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.34 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	1570 (2.36-2.32)
Clashscore	112137	1673 (2.36-2.32)
Ramachandran outliers	110173	1654 (2.36-2.32)
Sidechain outliers	110143	1655 (2.36-2.32)
RSRZ outliers	101464	1576 (2.36-2.32)

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	A	362	<div> <div>13%</div> <div> <div></div> <div>74%</div> <div>20%</div> <div>• 5%</div> </div> </div>
1	B	362	<div> <div>18%</div> <div> <div></div> <div>68%</div> <div>25%</div> <div>• 5%</div> </div> </div>
2	C	16	<div> <div>6%</div> <div> <div></div> <div>63%</div> <div>25%</div> <div>13%</div> </div> </div>
2	P	16	<div> <div></div> <div> <div></div> <div>56%</div> <div>25%</div> <div>19%</div> </div> </div>
3	D	19	<div> <div>5%</div> <div> <div></div> <div>63%</div> <div>21%</div> <div>11%</div> <div>5%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	T	19	 <div>63% 26% 5% 5%</div>

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	CA	B	401	-	-	-	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7148 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 IV.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	343	Total	C	N	O	S	0	1	0
			2736	1758	476	494	8			
1	B	343	Total	C	N	O	S	0	0	0
			2730	1753	475	494	8			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	355	GLY	-	EXPRESSION TAG	UNP Q4JB80
A	356	GLY	-	EXPRESSION TAG	UNP Q4JB80
A	357	HIS	-	EXPRESSION TAG	UNP Q4JB80
A	358	HIS	-	EXPRESSION TAG	UNP Q4JB80
A	359	HIS	-	EXPRESSION TAG	UNP Q4JB80
A	360	HIS	-	EXPRESSION TAG	UNP Q4JB80
A	361	HIS	-	EXPRESSION TAG	UNP Q4JB80
A	362	HIS	-	EXPRESSION TAG	UNP Q4JB80
B	355	GLY	-	EXPRESSION TAG	UNP Q4JB80
B	356	GLY	-	EXPRESSION TAG	UNP Q4JB80
B	357	HIS	-	EXPRESSION TAG	UNP Q4JB80
B	358	HIS	-	EXPRESSION TAG	UNP Q4JB80
B	359	HIS	-	EXPRESSION TAG	UNP Q4JB80
B	360	HIS	-	EXPRESSION TAG	UNP Q4JB80
B	361	HIS	-	EXPRESSION TAG	UNP Q4JB80
B	362	HIS	-	EXPRESSION TAG	UNP Q4JB80

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	16	Total	C	N	O	P	0	0	0
			325	154	62	94	15			

Continued on next page...

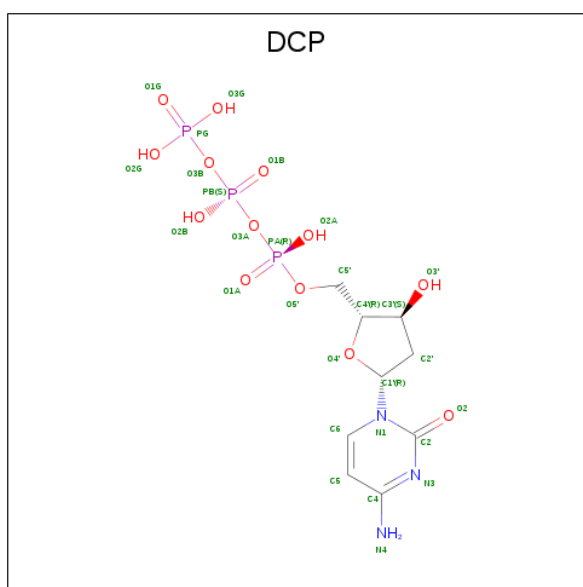
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	16	Total	C	N	O	P	0	0	0
			325	154	62	94	15			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	18	Total	C	N	O	P	0	0	0
			364	173	64	109	18			
3	T	18	Total	C	N	O	P	0	0	0
			364	173	64	109	18			

- Molecule 4 is 2'-DEOXYCYTIDINE-5'-TRIPHOSPHATE (three-letter code: DCP) (formula: C₉H₁₆N₃O₁₃P₃).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			24	9	3	10	2		
4	C	1	Total	C	N	O	P	0	0
			24	9	3	10	2		

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	2	Total	Ca	0	0
			2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Ca	0	0
			2	2		

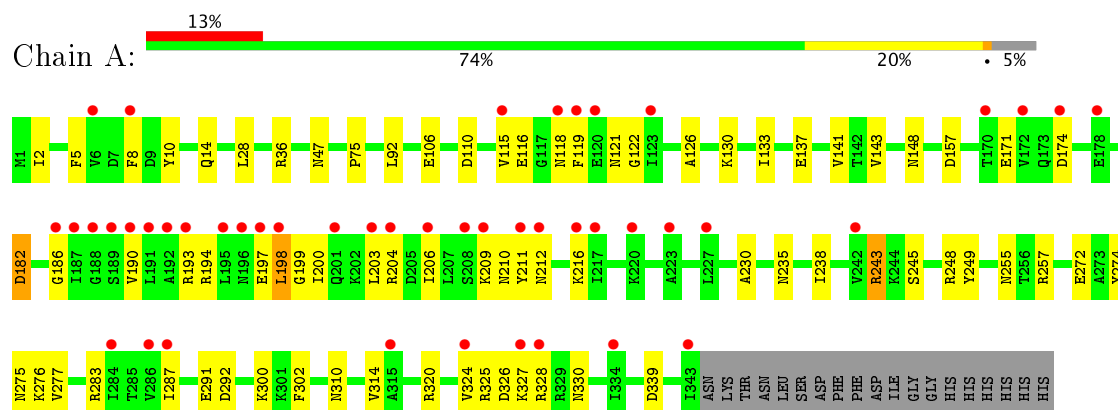
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	114	Total	O	0	0
			114	114		
6	B	77	Total	O	0	0
			77	77		
6	C	18	Total	O	0	0
			18	18		
6	D	13	Total	O	0	0
			13	13		
6	P	17	Total	O	0	0
			17	17		
6	T	13	Total	O	0	0
			13	13		

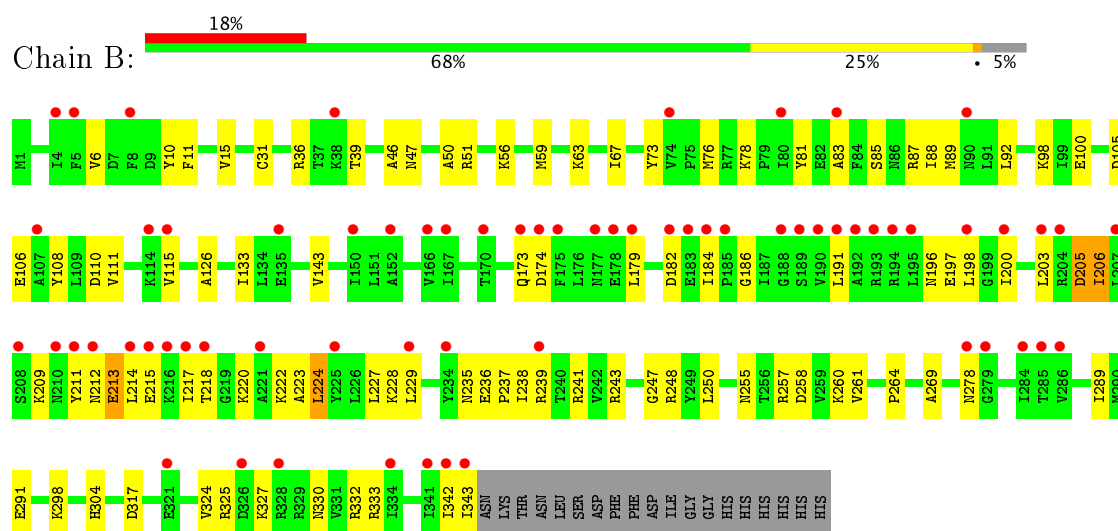
3 Residue-property plots

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.

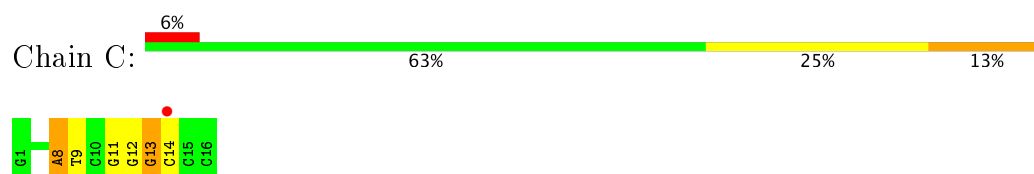
• Molecule 1: DNA polymerase IV



• Molecule 1: DNA polymerase IV

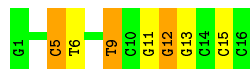


• Molecule 2: DNA (5'-D(*GP*GP*CP*AP*CP*TP*GP*AP*TP*CP*GP*GP*GP*CP*CP*C)-3')



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

Chain P: 



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

Chain D: 



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

Chain T: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	53.08 Å 103.39 Å 112.99 Å 90.00° 101.76° 90.00°	Depositor
Resolution (Å)	29.25 – 2.34 29.25 – 2.34	Depositor EDS
% Data completeness (in resolution range)	94.1 (29.25-2.34) 94.2 (29.25-2.34)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.14 (at 2.34 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.222 , 0.259 0.212 , 0.250	Depositor DCC
R_{free} test set	1888 reflections (3.97%)	DCC
Wilson B-factor (Å ²)	31.3	Xtriage
Anisotropy	0.598	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 56.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.027 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	7148	wwPDB-VP
Average B, all atoms (Å ²)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 25.57 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.0937e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical 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: CA, DCP

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	0.32	0/2777	0.51	0/3741
1	B	0.31	0/2768	0.52	1/3730 (0.0%)
2	C	0.55	0/364	1.29	4/560 (0.7%)
2	P	0.63	0/364	1.50	6/560 (1.1%)
3	D	0.55	0/406	1.37	2/623 (0.3%)
3	T	0.60	0/406	1.38	4/623 (0.6%)
All	All	0.39	0/7085	0.81	17/9837 (0.2%)

There are no bond length outliers.

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	T	5	DG	O4'-C1'-N9	-13.59	98.49	108.00
3	D	5	DG	O4'-C1'-N9	-11.58	99.89	108.00
3	D	6	DC	O4'-C1'-N1	-8.68	101.92	108.00
2	P	12	DG	O4'-C1'-N9	8.13	113.69	108.00
2	P	13	DG	O4'-C1'-N9	-7.04	103.07	108.00

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	A	2736	0	2917	54	0
1	B	2730	0	2904	66	0
2	C	325	0	180	7	0
2	P	325	0	180	6	0
3	D	364	0	203	9	0
3	T	364	0	203	3	1
4	A	24	0	12	0	0
4	C	24	0	12	1	0
5	A	2	0	0	0	0
5	B	2	0	0	0	0
6	A	114	0	0	13	0
6	B	77	0	0	8	1
6	C	18	0	0	3	0
6	D	13	0	0	6	0
6	P	17	0	0	2	0
6	T	13	0	0	0	0
All	All	7148	0	6611	137	1

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 137 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:A:198:LEU:O	6:A:588:HOH:O	1.96	0.82
1:B:325:ARG:NH2	2:C:9:DT:OP1	2.13	0.81
2:C:9:DT:O4	6:C:216:HOH:O	1.99	0.80
1:B:182:ASP:O	6:B:560:HOH:O	2.00	0.79
1:A:36:ARG:NH2	1:A:255:ASN:OD1	2.15	0.75

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:T:2:DT:O4	6:B:515:HOH:O[1_554]	2.12	0.08

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	342/362 (94%)	315 (92%)	22 (6%)	5 (2%)	12	9
1	B	341/362 (94%)	318 (93%)	20 (6%)	3 (1%)	20	20
All	All	683/724 (94%)	633 (93%)	42 (6%)	8 (1%)	15	13

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	10	TYR
1	A	211	TYR
1	A	210	ASN
1	A	198	LEU
1	A	327	LYS

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	299/315 (95%)	297 (99%)	2 (1%)	87	93
1	B	298/315 (95%)	291 (98%)	7 (2%)	56	67
All	All	597/630 (95%)	588 (98%)	9 (2%)	70	80

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	205	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	324	VAL
1	B	213	GLU
1	B	39	THR
1	B	206	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 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	A	401	5	21,25,29	1.63	5 (23%)	24,38,45	1.42	4 (16%)
4	DCP	C	101	5	21,25,29	1.53	4 (19%)	24,38,45	1.41	3 (12%)

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	A	401	5	-	0/12/28/34	0/2/2/2
4	DCP	C	101	5	-	0/12/28/34	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	401	DCP	C3'-C4'	-3.76	1.42	1.53
4	C	101	DCP	C3'-C4'	-3.34	1.43	1.53
4	A	401	DCP	O3'-C3'	-2.32	1.38	1.43
4	C	101	DCP	C2-N3	-2.28	1.33	1.38
4	A	401	DCP	C2'-C1'	-2.28	1.46	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	C	101	DCP	O2B-PB-O1B	-3.47	96.92	110.50
4	A	401	DCP	O2B-PB-O1B	-3.32	97.51	110.50
4	A	401	DCP	O2A-PA-O1A	-2.61	98.75	112.28
4	A	401	DCP	C2'-C1'-N1	-2.57	108.17	114.23
4	C	101	DCP	O2A-PA-O1A	-2.49	99.42	112.28

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	101	DCP	1	0

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	343/362 (94%)	0.93	46 (13%) 4 6	19, 38, 89, 112	0
1	B	343/362 (94%)	1.24	66 (19%) 1 2	25, 43, 91, 118	0
2	C	16/16 (100%)	0.69	1 (6%) 21 29	46, 59, 70, 71	0
2	P	16/16 (100%)	0.43	0 100 100	37, 48, 57, 72	0
3	D	18/19 (94%)	0.81	1 (5%) 25 35	27, 59, 72, 75	0
3	T	18/19 (94%)	0.48	0 100 100	32, 50, 64, 79	0
All	All	754/794 (94%)	1.04	114 (15%) 3 4	19, 42, 90, 118	0

The worst 5 of 114 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	211	TYR	9.6
1	B	217	ILE	8.2
1	B	343	ILE	7.3
1	B	215	GLU	6.6
1	B	191	LEU	6.6

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

There are no non-standard protein/DNA/RNA residues in this entry.

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(Å ²)	Q<0.9
5	CA	B	401	1/1	0.96	0.29	3.01	47,47,47,47	0
5	CA	A	403	1/1	0.89	0.29	1.98	36,36,36,36	0
5	CA	A	402	1/1	0.97	0.30	1.47	30,30,30,30	0
5	CA	B	402	1/1	0.92	0.29	1.32	35,35,35,35	0
4	DCP	A	401	24/28	0.95	0.18	-0.20	7,21,34,143	0
4	DCP	C	101	24/28	0.95	0.16	-0.62	22,32,45,53	0

6.5 Other polymers [i](#)

There are no such residues in this entry.