



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 03:36 pm BST

PDB ID : 2PIS  
Title : Efforts toward Expansion of the Genetic Alphabet: Structure and Replication of Unnatural Base Pairs  
Authors : Matsuda, S.; Fillo, J.D.; Henry, A.A.; Wilkins, S.J.; Rai, P.; Dwyer, T.J.; Geierstanger, B.H.; Wemmer, D.E.; Schultz, P.G.; Spraggon, G.; Romesberg, F.E.  
Deposited on : 2007-04-13  
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](mailto: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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

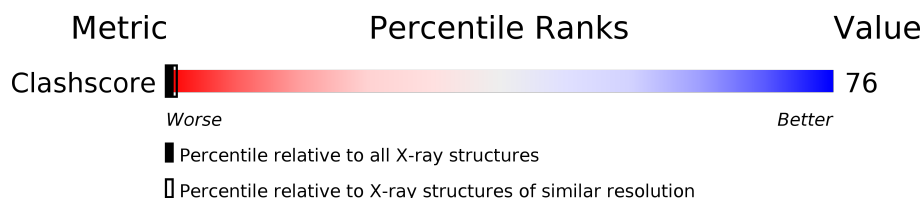
# 1 Overall quality at a glance

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	3569 (2.80-2.80)

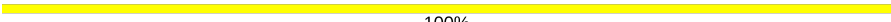

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria 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  $\leq 5\%$

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	13	8% 85% 8%
1	B	13	85% 15%
1	C	13	8% 69% 23%
1	D	13	8% 62% 31%
1	E	13	15% 69% 15%
1	F	13	8% 62% 31%
1	G	13	31% 38% 31%
1	H	13	23% 69% 8%
1	I	13	85% 15%
1	J	13	100%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	K	13	 100%
1	L	13	 85% 15%

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
1	FFD	H	20	-	-	X	-
1	CBR	J	16	-	-	X	-
1	FFD	J	20	-	-	X	-
1	CBR	K	3	-	-	X	-
1	CBR	L	16	-	-	X	-

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 3150 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 DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3').

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	A	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	B	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	C	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	D	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	E	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	F	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	G	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	H	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	I	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	J	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	K	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			
1	L	13	Total	Br	C	F	N	O	P	0	0	0
			262	1	127	1	46	75	12			

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	3	Total	Mg	0	0
			3	3		
2	D	1	Total	Mg	0	0
			1	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	F	1	Total 1	Mg 1	0	0
2	E	1	Total 1	Mg 1	0	0

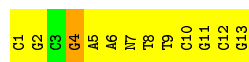
### 3 Residue-property plots [i](#)

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. 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.


Note EDS was not executed.

- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain A: 

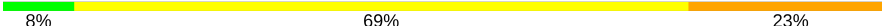


- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain B: 




- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain C: 



- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain D: 

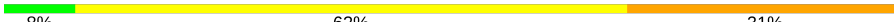


- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain E: 



- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain F:  8% 62% 31%



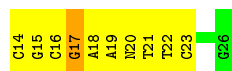
- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain G:  31% 38% 31%




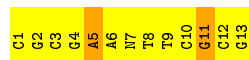
- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain H:  23% 69% 8%



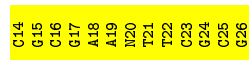
- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain I:  85% 15%

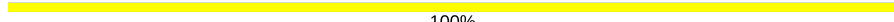


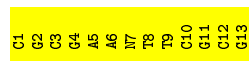
- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain J:  100%

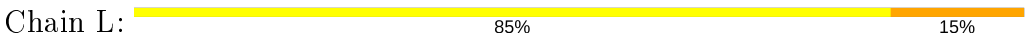


- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')

Chain K:  100%



- Molecule 1: DNA (5'-D(\*CP\*GP\*(CBR)P\*GP\*AP\*AP\*(FFD)P\*TP\*TP\*CP\*GP\*CP\*G)-3')



C14	G15	C16	G17	A18	A19	N20	T21	T22	C23	G24	C25	G26
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----



## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	I 41 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	146.03Å 146.03Å 93.14Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	100.00 – 2.80	Depositor
% Data completeness (in resolution range)	91.8 (100.00-2.80)	Depositor
$R_{merge}$	0.05	Depositor
$R_{sym}$	0.05	Depositor
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.224 , 0.303	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3150	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	wwPDB-VP

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CBR, MG, FFD

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	1.54	2/250 (0.8%)	1.31	2/382 (0.5%)
1	B	1.66	2/250 (0.8%)	1.26	0/382
1	C	1.60	4/250 (1.6%)	1.33	1/382 (0.3%)
1	D	1.42	1/250 (0.4%)	1.46	4/382 (1.0%)
1	E	1.29	0/250	1.14	2/382 (0.5%)
1	F	1.61	3/250 (1.2%)	1.35	4/382 (1.0%)
1	G	1.58	2/250 (0.8%)	1.37	5/382 (1.3%)
1	H	1.28	0/250	1.29	2/382 (0.5%)
1	I	0.97	0/250	1.27	2/382 (0.5%)
1	J	0.99	0/250	0.99	0/382
1	K	0.91	0/250	1.05	1/382 (0.3%)
1	L	1.03	0/250	1.17	2/382 (0.5%)
All	All	1.35	14/3000 (0.5%)	1.26	25/4584 (0.5%)

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	15	DG	C3'-O3'	-8.58	1.32	1.44
1	C	11	DG	C3'-O3'	-7.33	1.34	1.44
1	A	4	DG	C3'-O3'	-6.74	1.35	1.44
1	C	4	DG	C2-N3	6.48	1.38	1.32
1	C	8	DT	C3'-O3'	-6.32	1.35	1.44

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	19	DA	C1'-O4'-C4'	-9.82	100.28	110.10
1	H	17	DG	O4'-C1'-N9	9.43	114.60	108.00
1	D	19	DA	O4'-C1'-N9	8.26	113.78	108.00
1	G	4	DG	O4'-C1'-N9	8.07	113.65	108.00
1	A	2	DG	O4'-C1'-N9	-7.71	102.60	108.00

There are no chirality outliers.

There are no planarity outliers.

## 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	262	0	146	10	1
1	B	262	0	146	37	0
1	C	262	0	146	30	0
1	D	262	0	146	28	0
1	E	262	0	146	32	0
1	F	262	0	146	38	1
1	G	262	0	146	18	1
1	H	262	0	146	30	0
1	I	262	0	146	45	0
1	J	262	0	146	46	0
1	K	262	0	146	46	0
1	L	262	0	146	45	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
2	G	3	0	0	0	0
All	All	3150	0	1752	351	2

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 76.

The worst 5 of 351 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:K:2:DG:C2'	1:K:3:CBR:BR	2.28	1.36
1:K:2:DG:H2'	1:K:3:CBR:BR	1.79	1.35
1:F:22:DT:H2''	1:F:23:DC:C5'	1.63	1.26
1:K:4:DG:H2'	1:K:5:DA:C8	1.74	1.22
1:K:11:DG:C8	1:K:12:DC:C5	2.33	1.16

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:4:DG:O6	1:G:1:DC:C5[8_657]	2.09	0.11
1:F:14:DC:C5	1:F:17:DG:O6[16_556]	2.14	0.06

### 5.3 Torsion angles [i](#)

#### 5.3.1 Protein backbone [i](#)

There are no protein molecules in this entry.

#### 5.3.2 Protein sidechains [i](#)

There are no protein molecules in this entry.

#### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

24 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	CBR	L	16	1	15,21,22	1.31	2 (13%)	18,30,33	1.34	3 (16%)
1	FFD	L	20	1	16,19,20	0.99	1 (6%)	22,26,29	1.08	2 (9%)
1	FFD	I	7	1	16,19,20	1.14	1 (6%)	22,26,29	0.88	1 (4%)
1	CBR	A	3	1	15,21,22	1.67	3 (20%)	18,30,33	1.09	2 (11%)
1	CBR	E	3	1	15,21,22	1.70	4 (26%)	18,30,33	1.15	2 (11%)
1	FFD	E	7	1	16,19,20	1.10	1 (6%)	22,26,29	1.32	2 (9%)
1	CBR	C	3	1	15,21,22	1.42	2 (13%)	18,30,33	1.49	4 (22%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
1	FFD	D	20	1	16,19,20	1.28	1 (6%)	22,26,29	0.88	2 (9%)
1	FFD	K	7	1	16,19,20	0.96	1 (6%)	22,26,29	0.97	2 (9%)
1	FFD	A	7	1	16,19,20	1.62	1 (6%)	22,26,29	0.98	1 (4%)
1	CBR	G	3	1	15,21,22	1.70	4 (26%)	18,30,33	1.31	2 (11%)
1	CBR	K	3	1	15,21,22	1.20	2 (13%)	18,30,33	3.28	3 (16%)
1	FFD	F	20	1	16,19,20	1.09	1 (6%)	22,26,29	1.01	3 (13%)
1	FFD	G	7	1	16,19,20	1.26	1 (6%)	22,26,29	1.08	3 (13%)
1	CBR	B	16	1	15,21,22	1.49	2 (13%)	18,30,33	1.46	4 (22%)
1	CBR	D	16	1	15,21,22	1.42	2 (13%)	18,30,33	1.11	2 (11%)
1	CBR	F	16	1	15,21,22	1.59	3 (20%)	18,30,33	1.31	3 (16%)
1	CBR	H	16	1	15,21,22	1.36	2 (13%)	18,30,33	1.34	3 (16%)
1	CBR	J	16	1	15,21,22	1.25	2 (13%)	18,30,33	4.23	3 (16%)
1	FFD	C	7	1	16,19,20	1.46	1 (6%)	22,26,29	1.14	2 (9%)
1	FFD	H	20	1	16,19,20	1.11	2 (12%)	22,26,29	1.07	2 (9%)
1	FFD	J	20	1	16,19,20	1.00	1 (6%)	22,26,29	0.95	1 (4%)
1	CBR	I	3	1	15,21,22	1.49	2 (13%)	18,30,33	1.17	2 (11%)
1	FFD	B	20	1	16,19,20	1.20	1 (6%)	22,26,29	0.95	1 (4%)

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	CBR	L	16	1	-	2/4/21/22	0/2/2/2
1	FFD	L	20	1	-	2/7/21/22	0/2/2/2
1	FFD	I	7	1	-	2/7/21/22	0/2/2/2
1	CBR	A	3	1	-	1/4/21/22	0/2/2/2
1	CBR	E	3	1	-	3/4/21/22	0/2/2/2
1	FFD	E	7	1	-	4/7/21/22	0/2/2/2
1	CBR	C	3	1	-	2/4/21/22	0/2/2/2
1	FFD	D	20	1	-	2/7/21/22	0/2/2/2
1	FFD	K	7	1	-	1/7/21/22	0/2/2/2
1	FFD	A	7	1	-	0/7/21/22	0/2/2/2
1	CBR	G	3	1	-	1/4/21/22	0/2/2/2
1	CBR	K	3	1	-	3/4/21/22	0/2/2/2
1	FFD	F	20	1	-	0/7/21/22	0/2/2/2

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	FFD	G	7	1	-	2/7/21/22	0/2/2/2
1	CBR	B	16	1	-	1/4/21/22	0/2/2/2
1	CBR	D	16	1	-	3/4/21/22	0/2/2/2
1	CBR	F	16	1	-	0/4/21/22	0/2/2/2
1	CBR	H	16	1	-	2/4/21/22	0/2/2/2
1	CBR	J	16	1	-	0/4/21/22	0/2/2/2
1	FFD	C	7	1	-	3/7/21/22	0/2/2/2
1	FFD	H	20	1	-	6/7/21/22	0/2/2/2
1	FFD	J	20	1	-	2/7/21/22	0/2/2/2
1	CBR	I	3	1	-	1/4/21/22	0/2/2/2
1	FFD	B	20	1	-	2/7/21/22	0/2/2/2

The worst 5 of 43 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	7	FFD	O4'-C1'	-5.97	1.37	1.44
1	C	7	FFD	O4'-C1'	-4.89	1.39	1.44
1	G	7	FFD	O4'-C1'	-3.98	1.40	1.44
1	E	3	CBR	BR-C5	-3.93	1.80	1.89
1	I	7	FFD	O4'-C1'	-3.83	1.40	1.44

The worst 5 of 55 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	J	16	CBR	BR-C5-C6	-17.21	77.77	117.34
1	K	3	CBR	BR-C5-C6	-13.01	87.44	117.34
1	E	7	FFD	C3'-C2'-C1'	-4.04	98.91	102.74
1	L	16	CBR	C2-N3-C4	3.58	120.34	116.02
1	J	16	CBR	C5-C4-N4	-3.51	120.04	122.94

There are no chirality outliers.

5 of 45 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	L	20	FFD	O4'-C4'-C5'-O5'
1	I	7	FFD	O4'-C4'-C5'-O5'
1	I	7	FFD	C3'-C4'-C5'-O5'
1	A	3	CBR	O4'-C1'-N1-C6
1	E	3	CBR	O4'-C1'-N1-C6

There are no ring outliers.

22 monomers are involved in 102 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	L	16	CBR	8	0
1	L	20	FFD	6	0
1	I	7	FFD	6	0
1	E	3	CBR	6	0
1	E	7	FFD	6	0
1	C	3	CBR	4	0
1	D	20	FFD	2	0
1	K	7	FFD	5	0
1	A	7	FFD	2	0
1	K	3	CBR	7	0
1	F	20	FFD	2	0
1	G	7	FFD	4	0
1	B	16	CBR	5	0
1	D	16	CBR	1	0
1	F	16	CBR	1	0
1	H	16	CBR	6	0
1	J	16	CBR	7	0
1	C	7	FFD	1	0
1	H	20	FFD	12	0
1	J	20	FFD	9	0
1	I	3	CBR	1	0
1	B	20	FFD	5	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 6 are 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section is therefore empty.