



# Full wwPDB X-ray Structure Validation Report ⓘ

May 18, 2020 – 11:44 pm BST

PDB ID : 1HU0  
Title : CRYSTAL STRUCTURE OF AN HOGG1-DNA BOROHYDRIDE  
TRAPPED INTERMEDIATE COMPLEX  
Authors : Fromme, J.C.; Bruner, S.D.; Yang, W.; Karplus, M.; Verdine, G.L.  
Deposited on : 2001-01-03  
Resolution : 2.35 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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) : 1.13  
EDS : 2.11  
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) : 2.11

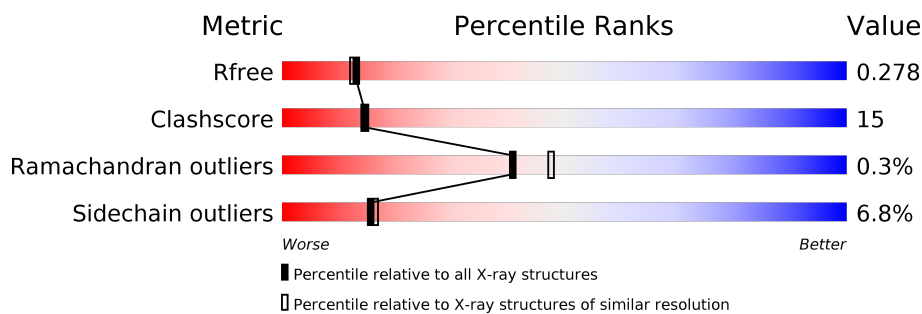
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	D	15	<div> <div style="width: 33%; background-color: green;"></div> <div style="width: 67%; background-color: yellow;"></div> </div> <div>33% 67%</div>
2	E	15	<div> <div style="width: 60%; background-color: green;"></div> <div style="width: 40%; background-color: yellow;"></div> </div> <div>60% 40%</div>
3	A	324	<div> <div style="width: 71%; background-color: green;"></div> <div style="width: 22%; background-color: yellow;"></div> <div style="width: 7%; background-color: orange;"></div> <div style="width: 2%; background-color: grey;"></div> </div> <div>71% 22% . .</div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3200 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(\*GP\*GP\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	15	Total	C	N	O	P	0	0	0
			308	146	61	87	14			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	15	Total	C	N	O	P	0	0	0
			290	139	49	88	14			

- Molecule 3 is a protein called 8-OXOGUANINE DNA GLYCOSYLASE 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	314	Total	C	N	O	S	0	0	0
			2491	1579	448	453	11			

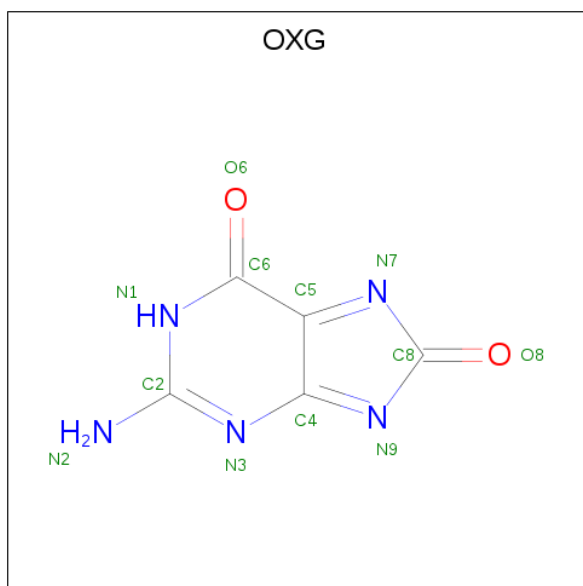
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	4	ALA	-	CLONING ARTIFACT	UNP O15527
A	5	MET	-	CLONING ARTIFACT	UNP O15527
A	6	ALA	-	CLONING ARTIFACT	UNP O15527
A	7	ASP	-	CLONING ARTIFACT	UNP O15527
A	8	ILE	-	CLONING ARTIFACT	UNP O15527
A	9	GLY	-	CLONING ARTIFACT	UNP O15527
A	10	SER	-	CLONING ARTIFACT	UNP O15527
A	11	GLU	-	CLONING ARTIFACT	UNP O15527

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total	Ca	0	0
			1	1		

- Molecule 5 is 8-OXOGUANINE (three-letter code: OXG) (formula:  $C_5H_3N_5O_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			12	5	5	2		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	2	Total	O	0	0
			2	2		
6	E	5	Total	O	0	0
			5	5		
6	A	91	Total	O	0	0
			91	91		

### 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.

- Molecule 1: 5'-D(\*GP\*GP\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*C)-3'

Chain D: 



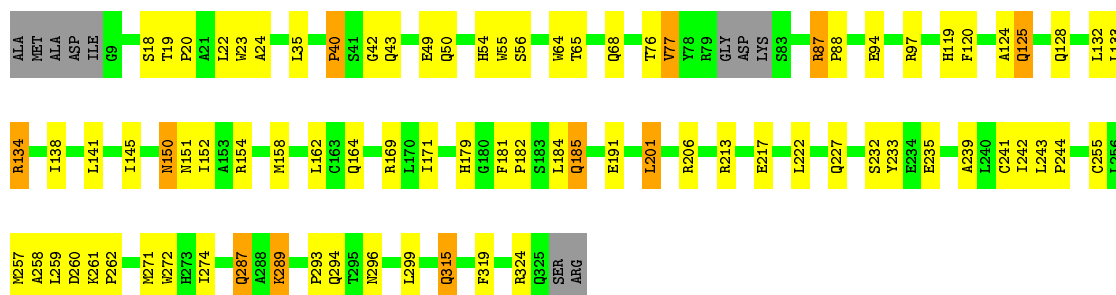
- Molecule 2: 5'-D(\*GP\*CP\*GP\*TP\*CP\*CP\*AP\*(PED)P\*GP\*TP\*CP\*TP\*AP\*CP\*C)-3',

Chain E: 



- Molecule 3: 8-OXOGUANINE DNA GLYCOSYLASE 1

Chain A: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	91.99 Å 91.99 Å 211.62 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.14 – 2.35 42.18 – 2.33	Depositor EDS
% Data completeness (in resolution range)	81.5 (42.14-2.35) 90.0 (42.18-2.33)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.75 (at 2.34 Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.210 , 0.245 0.253 , 0.278	Depositor DCC
$R_{free}$ test set	1860 reflections (8.13%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	37.8	Xtriage
Anisotropy	0.057	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 38.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	3200	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.95% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PED, CA, OXG

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	D	0.29	0/346	0.65	0/533
2	E	0.28	0/310	0.64	0/473
3	A	0.37	0/2559	0.62	0/3482
All	All	0.36	0/3215	0.63	0/4488

There are no bond length outliers.

There are no bond angle outliers.

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	D	308	0	169	5	1
2	E	290	0	167	4	0
3	A	2491	0	2423	82	1
4	E	1	0	0	0	0
5	A	12	0	3	1	0
6	A	91	0	0	2	0
6	D	2	0	0	0	0
6	E	5	0	0	0	0
All	All	3200	0	2762	90	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 15.

All (90) 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:87:ARG:NE	3:A:87:ARG:HA	1.65	1.09
3:A:87:ARG:NH2	3:A:88:PRO:HD3	1.80	0.96
3:A:87:ARG:CZ	3:A:88:PRO:HD3	2.07	0.84
3:A:185:GLN:H	3:A:185:GLN:NE2	1.75	0.82
3:A:76:THR:HA	3:A:87:ARG:HH22	1.46	0.78
3:A:289:LYS:NZ	3:A:289:LYS:HB3	2.04	0.72
1:D:3:DT:H2''	1:D:4:DA:C8	2.25	0.71
3:A:23:TRP:N	3:A:87:ARG:HD2	2.05	0.70
3:A:35:LEU:H	3:A:68:GLN:HE22	1.40	0.70
3:A:233:TYR:HA	3:A:261:LYS:HD2	1.73	0.69
3:A:87:ARG:NE	3:A:87:ARG:CA	2.52	0.69
3:A:272:TRP:HE1	3:A:296:ASN:ND2	1.92	0.68
3:A:35:LEU:H	3:A:68:GLN:NE2	1.94	0.65
3:A:289:LYS:HD2	3:A:289:LYS:H	1.62	0.65
3:A:87:ARG:HA	3:A:87:ARG:CZ	2.28	0.63
3:A:119:HIS:HE1	6:A:341:HOH:O	1.82	0.62
3:A:134:ARG:NH2	3:A:260:ASP:H	1.98	0.61
3:A:138:ILE:HD12	3:A:184:LEU:HD13	1.82	0.61
3:A:43:GLN:HE22	3:A:133:LEU:H	1.49	0.61
3:A:141:LEU:O	3:A:145:ILE:HG13	2.01	0.60
3:A:233:TYR:CA	3:A:261:LYS:HD2	2.33	0.59
2:E:28:DA:H2''	2:E:29:DC:O5'	2.04	0.58
3:A:64:TRP:CZ3	3:A:77:VAL:HG22	2.38	0.58
2:E:23:PED:O1P	3:A:152:ILE:HG13	2.03	0.57
3:A:22:LEU:C	3:A:87:ARG:CZ	2.74	0.56
3:A:94:GLU:OE2	3:A:97:ARG:HD3	2.05	0.56
3:A:87:ARG:HA	3:A:87:ARG:HE	1.62	0.56
3:A:185:GLN:H	3:A:185:GLN:HE21	1.52	0.55
3:A:43:GLN:NE2	3:A:133:LEU:H	2.03	0.55
3:A:289:LYS:HZ3	3:A:289:LYS:HB3	1.71	0.55
1:D:2:DG:H2''	1:D:3:DT:OP2	2.07	0.55
3:A:22:LEU:HA	3:A:87:ARG:NE	2.22	0.55
3:A:22:LEU:O	3:A:87:ARG:NH2	2.39	0.55
3:A:50:GLN:NE2	3:A:50:GLN:HA	2.22	0.54
3:A:271:MET:HE3	3:A:319:PHE:HD1	1.72	0.54
3:A:134:ARG:HH22	3:A:258:ALA:C	2.11	0.53
2:E:18:DG:H2''	2:E:19:DT:O5'	2.09	0.53
3:A:87:ARG:NH2	3:A:88:PRO:CD	2.65	0.52

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:150:ASN:HD22	3:A:151:ASN:H	1.57	0.52
3:A:206:ARG:HG3	3:A:206:ARG:HH21	1.75	0.52
3:A:134:ARG:NE	3:A:260:ASP:OD1	2.42	0.52
2:E:27:DT:H2"	2:E:28:DA:C8	2.45	0.52
3:A:287:GLN:NE2	3:A:287:GLN:H	2.08	0.52
3:A:77:VAL:HG23	3:A:88:PRO:HG3	1.92	0.51
3:A:289:LYS:HD2	3:A:289:LYS:N	2.23	0.51
3:A:255:CYS:HB3	3:A:261:LYS:HG3	1.93	0.51
3:A:22:LEU:O	3:A:87:ARG:CZ	2.59	0.51
3:A:287:GLN:H	3:A:287:GLN:CD	2.14	0.51
3:A:23:TRP:HA	3:A:87:ARG:NH2	2.27	0.50
3:A:87:ARG:HH21	3:A:88:PRO:HD3	1.74	0.49
3:A:42:GLY:O	5:A:328:OXG:N7	2.46	0.48
3:A:255:CYS:HA	3:A:259:LEU:HB2	1.94	0.48
1:D:12:DA:H2"	1:D:13:DC:O5'	2.13	0.48
3:A:239:ALA:O	3:A:242:ILE:HB	2.13	0.48
3:A:23:TRP:HA	3:A:87:ARG:HH21	1.77	0.48
3:A:150:ASN:ND2	3:A:151:ASN:H	2.11	0.48
3:A:232:SER:OG	3:A:235:GLU:HB2	2.13	0.48
3:A:293:PRO:HG2	3:A:294:GLN:NE2	2.29	0.48
3:A:162:LEU:CD2	3:A:182:PRO:HG2	2.46	0.46
3:A:18:SER:C	3:A:20:PRO:HD3	2.35	0.46
3:A:191:GLU:HA	6:A:393:HOH:O	2.15	0.45
3:A:24:ALA:HB2	3:A:88:PRO:HD2	1.98	0.45
1:D:8:DC:H2"	1:D:9:DT:O5'	2.17	0.45
3:A:56:SER:HA	3:A:64:TRP:O	2.15	0.45
3:A:55:TRP:O	3:A:65:THR:HA	2.16	0.45
3:A:19:THR:N	3:A:20:PRO:HD3	2.30	0.45
3:A:120:PHE:CE2	3:A:124:ALA:HB2	2.53	0.44
3:A:181:PHE:CD2	3:A:182:PRO:HD2	2.53	0.44
3:A:23:TRP:N	3:A:87:ARG:CD	2.80	0.44
3:A:132:LEU:HD12	3:A:257:MET:HG2	2.00	0.43
3:A:324:ARG:N	3:A:324:ARG:HD2	2.33	0.43
3:A:206:ARG:HG3	3:A:206:ARG:NH2	2.34	0.43
3:A:164:GLN:HA	3:A:179:HIS:CD2	2.53	0.43
3:A:289:LYS:HZ2	3:A:289:LYS:HB3	1.83	0.43
3:A:185:GLN:N	3:A:185:GLN:NE2	2.57	0.43
3:A:64:TRP:CE2	3:A:77:VAL:HG13	2.54	0.43
3:A:87:ARG:HD3	3:A:88:PRO:HD3	2.01	0.43
3:A:293:PRO:HG2	3:A:294:GLN:HE22	1.82	0.43
3:A:271:MET:CE	3:A:274:ILE:HD12	2.49	0.42

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:50:GLN:HE21	3:A:50:GLN:HA	1.84	0.42
3:A:260:ASP:O	3:A:262:PRO:HD3	2.20	0.42
1:D:5:DG:H2''	1:D:6:DA:C8	2.55	0.42
3:A:125:GLN:HB3	3:A:125:GLN:HE21	1.58	0.42
3:A:158:MET:HG2	3:A:201:LEU:HD22	2.02	0.42
3:A:171:ILE:C	3:A:171:ILE:HD12	2.40	0.41
3:A:243:LEU:HA	3:A:244:PRO:HD3	1.81	0.41
3:A:49:GLU:HA	3:A:54:HIS:O	2.20	0.41
3:A:35:LEU:N	3:A:68:GLN:HE22	2.15	0.41
3:A:271:MET:HE3	3:A:319:PHE:CD1	2.54	0.41
3:A:315:GLN:HE21	3:A:315:GLN:HB3	1.69	0.40

All (2) 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)
1:D:15:DC:O3'	1:D:15:DC:O3'[11_655]	0.81	1.39
3:A:213:ARG:NH1	3:A:217:GLU:OE2[11_555]	1.88	0.32

## 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	
3	A	310/324 (96%)	297 (96%)	12 (4%)	1 (0%)	41	47

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	40	PRO

### 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	264/271 (97%)	246 (93%)	18 (7%)	16	16

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	40	PRO
3	A	77	VAL
3	A	87	ARG
3	A	125	GLN
3	A	128	GLN
3	A	134	ARG
3	A	150	ASN
3	A	154	ARG
3	A	169	ARG
3	A	185	GLN
3	A	201	LEU
3	A	222	LEU
3	A	227	GLN
3	A	241	CYS
3	A	287	GLN
3	A	289	LYS
3	A	299	LEU
3	A	315	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (16) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	43	GLN
3	A	50	GLN
3	A	68	GLN
3	A	101	GLN
3	A	108	GLN
3	A	119	HIS
3	A	125	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	A	150	ASN
3	A	172	GLN
3	A	185	GLN
3	A	219	GLN
3	A	226	GLN
3	A	227	GLN
3	A	273	HIS
3	A	276	GLN
3	A	296	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 1 is 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	$\# Z  > 2$
5	OXG	A	328	-	10,13,13	1.68	2 (20%)	9,19,19	3.88	4 (44%)

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
5	OXG	A	328	-	-	-	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	328	OXG	C6-N1	3.74	1.39	1.33
5	A	328	OXG	C8-N7	-2.82	1.30	1.39

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	328	OXG	C5-C6-N1	-8.84	111.34	123.43
5	A	328	OXG	C6-N1-C2	5.94	125.36	115.93
5	A	328	OXG	C4-N3-C2	-3.06	111.86	115.36
5	A	328	OXG	C5-C4-N3	2.36	128.64	124.81

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
5	A	328	OXG	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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.