



# Full wwPDB X-ray Structure Validation Report ⓘ

Dec 22, 2025 – 12:09 PM EST

PDB ID : 9YHY / pdb\_00009yhy  
Title : DNA ligase 1 wild-type in complex with nick containing 3'-8oxodG:C captured at pre-catalytic stage  
Authors : KanalElamparithi, B.; Caglayan, M.  
Deposited on : 2025-10-01  
Resolution : 2.76 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 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.47

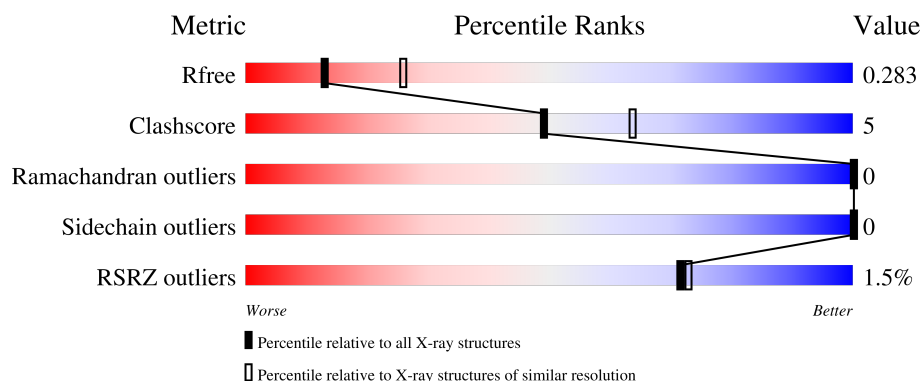
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.76 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))
$R_{free}$	164625	1606 (2.78-2.74)
Clashscore	180529	1689 (2.78-2.74)
Ramachandran outliers	177936	1665 (2.78-2.74)
Sidechain outliers	177891	1665 (2.78-2.74)
RSRZ outliers	164620	1606 (2.78-2.74)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	668	<div> <div></div> <div>84%</div> <div>9%</div> <div>7%</div> </div>
2	B	11	<div> <div>55%</div> <div>36%</div> <div>9%</div> </div>
3	C	8	<div> <div>62%</div> <div>38%</div> </div>
4	D	18	<div> <div>72%</div> <div>28%</div> </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
2	8OG	B	11	X	-	-	-

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5015 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 ligase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	621	Total	C	N	O	S	0	0	0
			4230	2674	737	805	14			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	919	ALA	-	expression tag	UNP P18858
A	920	ALA	-	expression tag	UNP P18858
A	921	ALA	-	expression tag	UNP P18858
A	922	LEU	-	expression tag	UNP P18858
A	923	GLU	-	expression tag	UNP P18858
A	924	HIS	-	expression tag	UNP P18858
A	925	HIS	-	expression tag	UNP P18858
A	926	HIS	-	expression tag	UNP P18858
A	927	HIS	-	expression tag	UNP P18858
A	928	HIS	-	expression tag	UNP P18858
A	929	HIS	-	expression tag	UNP P18858

- Molecule 2 is a DNA chain called DNA (5'-D(P\*GP\*CP\*TP\*GP\*AP\*TP\*GP\*CP\*GP\*TP\*(8OG))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	P	0	0	0
			227	108	42	67	10			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	8	Total	C	N	O	P	0	0	0
			167	78	33	48	8			

- Molecule 4 is a DNA chain called DNA (5'-D(P\*GP\*TP\*CP\*CP\*GP\*AP\*CP\*CP\*AP\*CP

\*GP\*CP\*AP\*TP\*CP\*AP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	18	Total 361	C 172	N 68	O 104	P 17	0	0	0

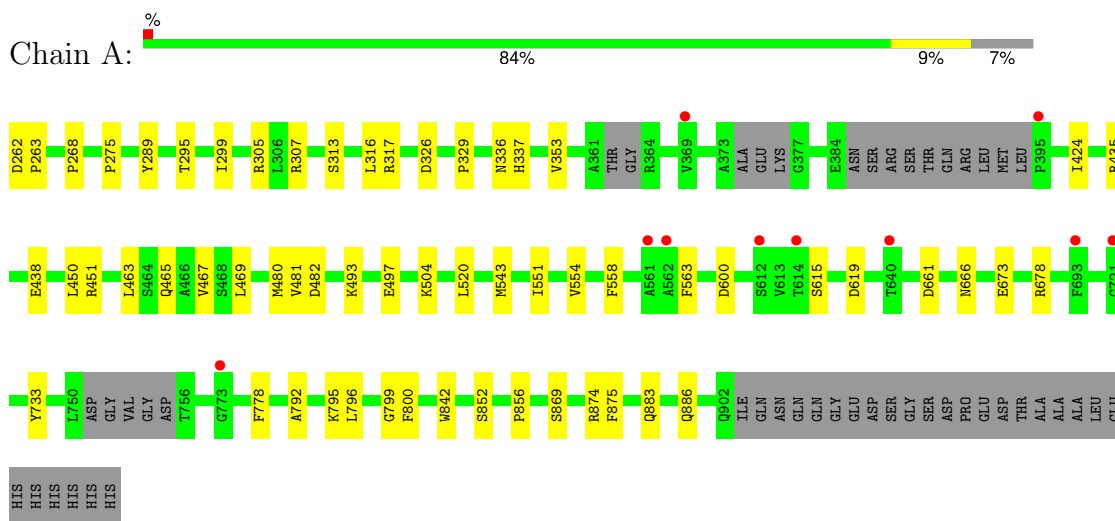
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	16	Total 16	O 16	0	0
5	B	4	Total 4	O 4	0	0
5	C	5	Total 5	O 5	0	0
5	D	5	Total 5	O 5	0	0

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

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.

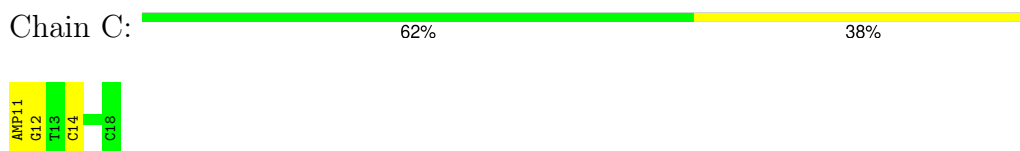
- Molecule 1: DNA ligase 1



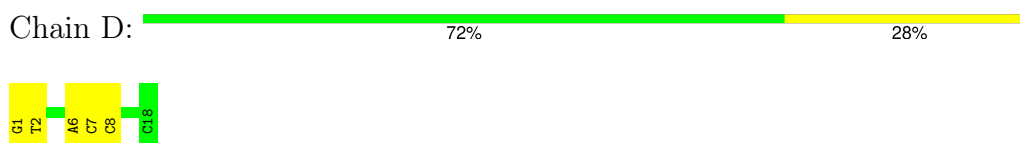
- Molecule 2: DNA (5'-D(P\*GP\*CP\*TP\*GP\*AP\*TP\*GP\*CP\*GP\*TP\*(8OG))-3')



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



- Molecule 4: DNA (5'-D(P\*GP\*TP\*CP\*CP\*GP\*AP\*CP\*CP\*AP\*CP\*GP\*CP\*AP\*TP\*CP\*AP\*GP\*C)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	63.51Å 115.59Å 126.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.76 – 2.76 24.76 – 2.76	Depositor EDS
% Data completeness (in resolution range)	99.2 (24.76-2.76) 93.3 (24.76-2.76)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.78 (at 2.76Å)	Xtriage
Refinement program	PHENIX (1.21.1_5286: ???)	Depositor
R, $R_{free}$	0.228 , 0.283 0.230 , 0.283	Depositor DCC
$R_{free}$ test set	1991 reflections (8.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	70.3	Xtriage
Anisotropy	0.894	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 88.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5015	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	99.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.69% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 8OG, AMP

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	0.14	0/4316	0.32	0/5897
2	B	0.24	0/228	0.39	0/351
3	C	0.26	0/162	0.40	0/248
4	D	0.24	0/404	0.41	0/620
All	All	0.16	0/5110	0.33	0/7116

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
2	B	2	0

There are no bond length outliers.

There are no bond angle outliers.

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	11	8OG	C1',C3'

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	4230	0	3601	40	0
2	B	227	0	126	3	0
3	C	167	0	91	3	0
4	D	361	0	202	5	0
5	A	16	0	0	3	0
5	B	4	0	0	1	0
5	C	5	0	0	0	0
5	D	5	0	0	0	0
All	All	5015	0	4020	47	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 5.

All (47) 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:317:ARG:NH2	1:A:481:VAL:O	2.14	0.80
1:A:313:SER:O	1:A:317:ARG:HG2	1.93	0.69
1:A:883:GLN:H	1:A:886:GLN:HG3	1.58	0.69
2:B:7:DG:OP1	5:B:101:HOH:O	2.13	0.67
1:A:480:MET:HE2	1:A:482:ASP:H	1.61	0.65
1:A:275:PRO:HD2	1:A:481:VAL:HG13	1.83	0.61
1:A:796:LEU:HD21	1:A:875:PHE:HB2	1.82	0.60
3:C:11:AMP:P	3:C:12:DG:P	2.99	0.59
1:A:262:ASP:HB3	1:A:263:PRO:HD3	1.86	0.58
1:A:316:LEU:HD23	1:A:463:LEU:HA	1.86	0.56
1:A:558:PHE:HD1	1:A:563:PHE:HB3	1.74	0.53
4:D:1:DG:H1'	4:D:2:DT:H5'	1.89	0.53
1:A:504:LYS:NZ	5:A:1002:HOH:O	2.34	0.53
2:B:10:DT:H2'	2:B:11:8OG:O4'	2.09	0.52
1:A:543:MET:HE1	1:A:733:TYR:HA	1.92	0.51
1:A:852:SER:HB2	1:A:869:SER:HB2	1.94	0.50
1:A:795:LYS:HB3	4:D:8:DC:H5''	1.93	0.49
1:A:305:ARG:NH2	5:A:1003:HOH:O	2.45	0.49
2:B:3:DT:H2''	2:B:4:DG:C8	2.46	0.49
1:A:435:ARG:H	1:A:438:GLU:HB2	1.78	0.48
1:A:673:GLU:O	1:A:678:ARG:NH1	2.45	0.48
1:A:600:ASP:OD1	1:A:600:ASP:N	2.46	0.48
1:A:336:ASN:HB2	1:A:450:LEU:HD12	1.96	0.47
1:A:493:LYS:NZ	1:A:497:GLU:OE2	2.47	0.47
1:A:289:TYR:HD2	1:A:438:GLU:HG2	1.80	0.47
1:A:326:ASP:C	1:A:329:PRO:HD2	2.40	0.46

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:353:VAL:HG13	1:A:424:ILE:HG12	1.97	0.46
1:A:451:ARG:HD2	4:D:6:DA:OP1	2.17	0.45
1:A:465:GLN:O	1:A:469:LEU:HG	2.17	0.44
1:A:778:PHE:HD2	1:A:842:TRP:HZ3	1.66	0.44
1:A:295:THR:O	1:A:299:ILE:HG13	2.18	0.44
1:A:435:ARG:N	1:A:438:GLU:OE1	2.50	0.44
4:D:7:DC:H2'	4:D:8:DC:H6	1.83	0.44
1:A:543:MET:HB2	1:A:543:MET:HE2	1.68	0.43
1:A:874:ARG:HD3	3:C:14:DC:OP1	2.19	0.43
1:A:792:ALA:O	1:A:856:PRO:HD2	2.20	0.42
1:A:551:ILE:HA	1:A:554:VAL:HG12	2.01	0.42
1:A:799:GLY:HA2	3:C:14:DC:H4'	2.01	0.42
1:A:615:SER:O	1:A:666:ASN:N	2.39	0.42
4:D:7:DC:H2'	4:D:8:DC:C6	2.55	0.41
1:A:467:VAL:HG11	1:A:520:LEU:HD21	2.03	0.41
1:A:619:ASP:HB3	1:A:661:ASP:HB3	2.02	0.41
1:A:289:TYR:CD2	1:A:438:GLU:HG2	2.56	0.41
1:A:307:ARG:HA	1:A:307:ARG:HD2	1.85	0.41
1:A:337:HIS:HD2	5:A:1015:HOH:O	2.03	0.41
1:A:800:PHE:HD2	1:A:875:PHE:HB3	1.87	0.40
1:A:268:PRO:HG3	1:A:295:THR:OG1	2.21	0.40

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	
1	A	611/668 (92%)	583 (95%)	28 (5%)	0	100	100

There are no Ramachandran outliers to report.

### 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	331/567 (58%)	331 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	274	HIS
1	A	336	ASN
1	A	594	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](#)

1 non-standard protein/DNA/RNA residue is 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	8OG	B	11	4,2	22,25,26	0.77	1 (4%)	26,37,40	1.04	1 (3%)

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	8OG	B	11	4,2	2/2/4/4	6/7/21/22	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	11	8OG	C8-N9	-2.82	1.35	1.40

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	11	8OG	C3'-C2'-C1'	-2.47	96.56	102.60

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	11	8OG	C1'
2	B	11	8OG	C3'

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	11	8OG	C2'-C1'-N9-C8
2	B	11	8OG	C3'-C4'-C5'-O5'
2	B	11	8OG	O4'-C4'-C5'-O5'
2	B	11	8OG	O4'-C1'-N9-C8
2	B	11	8OG	C4'-C5'-O5'-P
2	B	11	8OG	O4'-C1'-N9-C4

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	11	8OG	1	0

## 5.5 Carbohydrates

There are no oligosaccharides 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	11:AMP	O3'	12:DG	P	4.23

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	621/668 (92%)	0.38	10 (1%) 70 71	57, 102, 140, 160	25 (4%)
2	B	10/11 (90%)	-0.39	0 100 100	67, 76, 92, 93	1 (10%)
3	C	7/8 (87%)	-0.39	0 100 100	60, 70, 76, 80	0
4	D	18/18 (100%)	-0.60	0 100 100	61, 69, 88, 89	3 (16%)
All	All	656/705 (93%)	0.33	10 (1%) 71 73	57, 100, 139, 160	29 (4%)

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	773	GLY	4.7
1	A	562	ALA	3.1
1	A	612	SER	2.6
1	A	614	THR	2.4
1	A	561	ALA	2.4
1	A	693	PHE	2.4
1	A	721	GLY	2.3
1	A	395	PRO	2.1
1	A	369	VAL	2.1
1	A	640	THR	2.0

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	8OG	B	11	23/24	0.94	0.09	55,76,93,124	0

### 6.3 Carbohydrates [i](#)

There are no oligosaccharides 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.