



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 22, 2020 – 03:28 PM BST

PDB ID : 7BU2  
Title : Structure of alcohol dehydrogenase YjgB from Escherichia coli  
Authors : Nguyen, G.T.; Kim, Y.-G.; Ahn, J.-W.; Chang, J.H.  
Deposited on : 2020-04-03  
Resolution : 1.55 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) : 1.13  
EDS : 2.13.1  
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) : 2.13.1

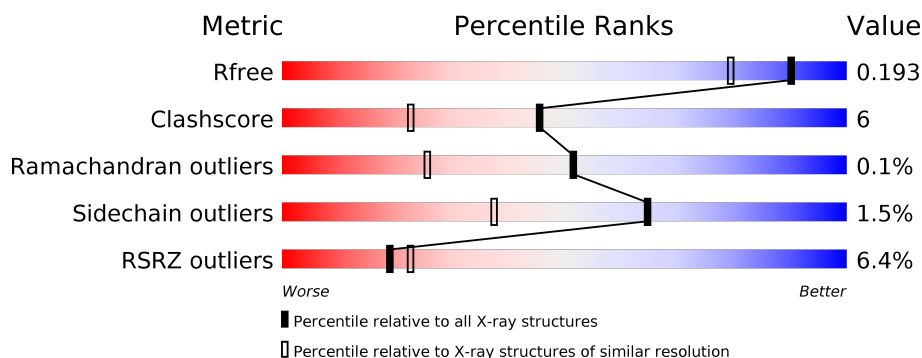
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.55 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	1483 (1.56-1.56)
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)
RSRZ outliers	127900	1465 (1.56-1.56)

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  $\leq 5\%$ . 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	347	<div> <div>7%</div> <div> <div></div> <div>86%</div> <div>10%</div> <div>••</div> </div> </div>
1	B	347	<div> <div>5%</div> <div> <div></div> <div>89%</div> <div>8%</div> <div>••</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6093 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 Alcohol dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	337	Total	C	N	O	S	0	9	0
			2613	1645	458	493	17			
1	B	339	Total	C	N	O	S	0	8	0
			2636	1664	462	493	17			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	340	LEU	-	expression tag	UNP A0A024L8S1
A	341	ALA	-	expression tag	UNP A0A024L8S1
A	342	HIS	-	expression tag	UNP A0A024L8S1
A	343	HIS	-	expression tag	UNP A0A024L8S1
A	344	HIS	-	expression tag	UNP A0A024L8S1
A	345	HIS	-	expression tag	UNP A0A024L8S1
A	346	HIS	-	expression tag	UNP A0A024L8S1
A	347	HIS	-	expression tag	UNP A0A024L8S1
B	340	LEU	-	expression tag	UNP A0A024L8S1
B	341	ALA	-	expression tag	UNP A0A024L8S1
B	342	HIS	-	expression tag	UNP A0A024L8S1
B	343	HIS	-	expression tag	UNP A0A024L8S1
B	344	HIS	-	expression tag	UNP A0A024L8S1
B	345	HIS	-	expression tag	UNP A0A024L8S1
B	346	HIS	-	expression tag	UNP A0A024L8S1
B	347	HIS	-	expression tag	UNP A0A024L8S1

- Molecule 2 is NITRATE ION (three-letter code: NO3) (formula: NO<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	N	O	0	0
			4	1	3		
2	B	1	Total	N	O	0	0
			4	1	3		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	B	1	Total	C	O	0	1
			12	6	6		
3	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by author).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	Zn	0	0
			2	2		
4	A	2	Total	Zn	0	0
			2	2		

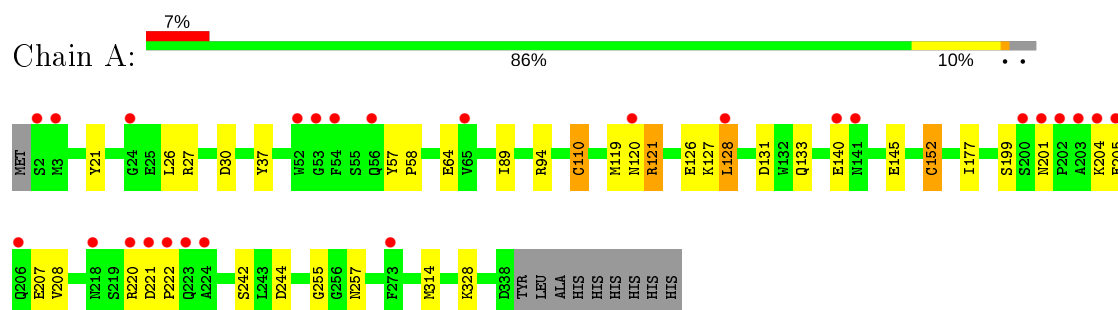
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	370	Total	O	0	0
			370	370		
5	B	420	Total	O	0	0
			420	420		

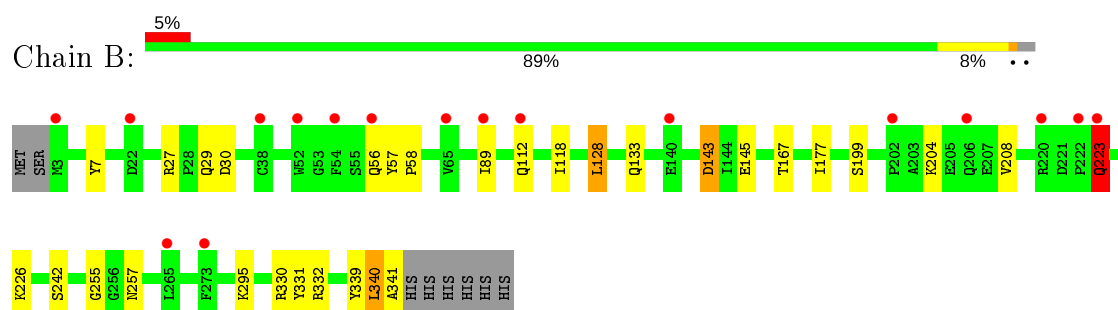
### 3 Residue-property plots

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: Alcohol dehydrogenase



#### • Molecule 1: Alcohol dehydrogenase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	133.14Å 64.47Å 81.66Å 90.00° 106.14° 90.00°	Depositor
Resolution (Å)	25.02 – 1.55 25.02 – 1.55	Depositor EDS
% Data completeness (in resolution range)	99.7 (25.02-1.55) 99.7 (25.02-1.55)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.41 (at 1.55Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.161 , 0.194 0.162 , 0.193	Depositor DCC
$R_{free}$ test set	4781 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.6	Xtriage
Anisotropy	0.349	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 48.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	6093	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.01% 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: GOL, ZN, NO3

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	0.59	3/2666 (0.1%)	0.67	2/3616 (0.1%)
1	B	0.51	0/2691	0.69	2/3650 (0.1%)
All	All	0.55	3/5357 (0.1%)	0.68	4/7266 (0.1%)

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
1	B	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	152	CYS	CB-SG	-13.28	1.59	1.82
1	A	152	CYS	CA-CB	7.05	1.69	1.53
1	A	110	CYS	CB-SG	-6.65	1.71	1.82

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	152	CYS	CA-CB-SG	-7.15	101.14	114.00
1	B	143	ASP	CB-CG-OD1	6.80	124.42	118.30
1	B	223	GLN	CA-CB-CG	6.78	128.31	113.40
1	A	119	MET	CG-SD-CE	-5.52	91.36	100.20

There are no chirality outliers.

All (2) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	B	118	ILE	Mainchain

## 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	2613	0	2575	35	1
1	B	2636	0	2606	33	1
2	A	4	0	0	0	0
2	B	4	0	0	0	0
3	A	24	0	32	0	0
3	B	18	0	24	1	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
5	A	370	0	0	11	4
5	B	420	0	0	14	5
All	All	6093	0	5237	68	8

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 6.

The worst 5 of 68 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:145:GLU:OE2	5:A:501:HOH:O	1.66	1.11
1:B:226:LYS:NZ	5:B:501:HOH:O	1.98	0.96
1:A:133:GLN:NE2	5:A:503:HOH:O	2.05	0.89
1:A:133:GLN:OE1	5:A:502:HOH:O	1.96	0.84
1:B:223:GLN:HE21	1:B:223:GLN:CA	1.92	0.83

The worst 5 of 8 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)
5:B:802:HOH:O	5:B:802:HOH:O[2_456]	1.80	0.40
5:A:842:HOH:O	5:B:834:HOH:O[4_456]	1.93	0.27

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:B:625:HOH:O	5:B:825:HOH:O[2_456]	1.99	0.21
5:A:505:HOH:O	5:B:512:HOH:O[4_456]	2.07	0.13
5:B:602:HOH:O	5:B:802:HOH:O[2_456]	2.09	0.11

## 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	344/347 (99%)	331 (96%)	12 (4%)	1 (0%)	41	19
1	B	346/347 (100%)	335 (97%)	11 (3%)	0	100	100
All	All	690/694 (99%)	666 (96%)	23 (3%)	1 (0%)	51	26

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	121	ARG

### 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	277/277 (100%)	270 (98%)	7 (2%)	47	18
1	B	278/277 (100%)	275 (99%)	3 (1%)	73	53
All	All	555/554 (100%)	545 (98%)	10 (2%)	65	31

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	242[B]	SER
1	A	314[A]	MET
1	B	128	LEU
1	A	242[A]	SER
1	A	314[B]	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	56	GLN
1	B	223	GLN

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 13 ligands modelled in this entry, 4 are monoatomic - leaving 9 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	$\# Z  > 2$
3	GOL	A	404	-	5,5,5	0.28	0	5,5,5	0.49	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	GOL	B	402[B]	-	5,5,5	0.27	0	5,5,5	0.31	0
3	GOL	B	402[A]	-	5,5,5	0.34	0	5,5,5	0.26	0
3	GOL	A	405	-	5,5,5	0.25	0	5,5,5	0.40	0
3	GOL	A	402	-	5,5,5	0.44	0	5,5,5	0.35	0
2	NO3	A	401	-	1,3,3	0.94	0	0,3,3	0.00	-
3	GOL	B	403	-	5,5,5	0.41	0	5,5,5	0.50	0
2	NO3	B	401	-	1,3,3	0.69	0	0,3,3	0.00	-
3	GOL	A	403	-	5,5,5	0.67	0	5,5,5	0.66	0

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
3	GOL	A	404	-	-	2/4/4/4	-
3	GOL	B	402[B]	-	-	0/4/4/4	-
3	GOL	B	402[A]	-	-	0/4/4/4	-
3	GOL	A	405	-	-	0/4/4/4	-
3	GOL	A	402	-	-	1/4/4/4	-
3	GOL	B	403	-	-	0/4/4/4	-
3	GOL	A	403	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	404	GOL	C1-C2-C3-O3
3	A	402	GOL	O1-C1-C2-C3
3	A	403	GOL	O1-C1-C2-C3
3	A	404	GOL	O2-C2-C3-O3
3	A	403	GOL	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	403	GOL	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, 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	337/347 (97%)	0.51	26 (7%) 13 16	12, 20, 39, 60	0
1	B	339/347 (97%)	0.33	17 (5%) 28 33	12, 18, 35, 47	0
All	All	676/694 (97%)	0.42	43 (6%) 19 23	12, 19, 37, 60	0

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	204	LYS	8.3
1	A	203	ALA	8.0
1	A	201	ASN	6.0
1	A	202	PRO	5.7
1	A	53	GLY	5.1

### 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 monosaccharides in this entry.

### 6.4 Ligands [i](#)

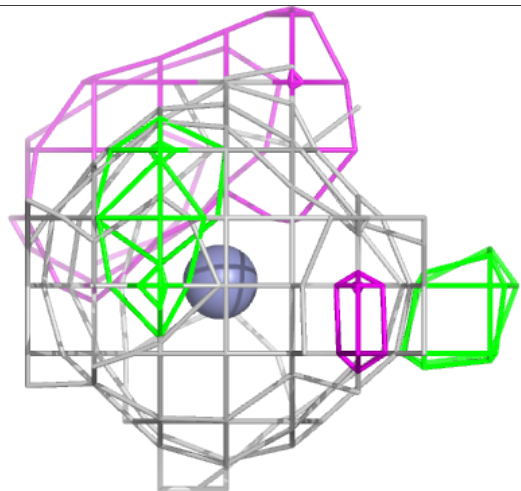
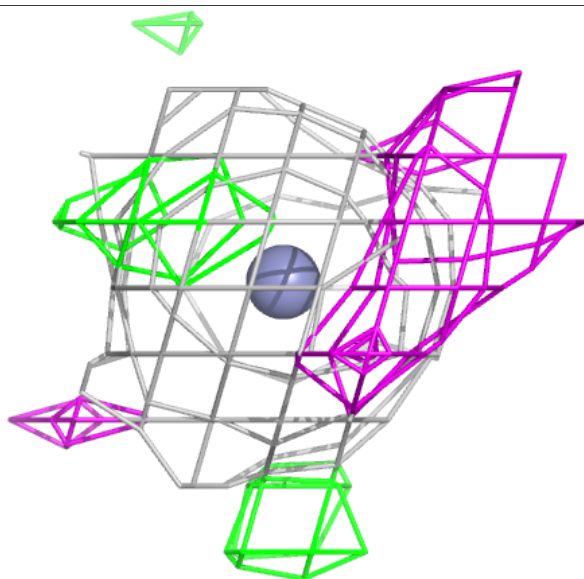
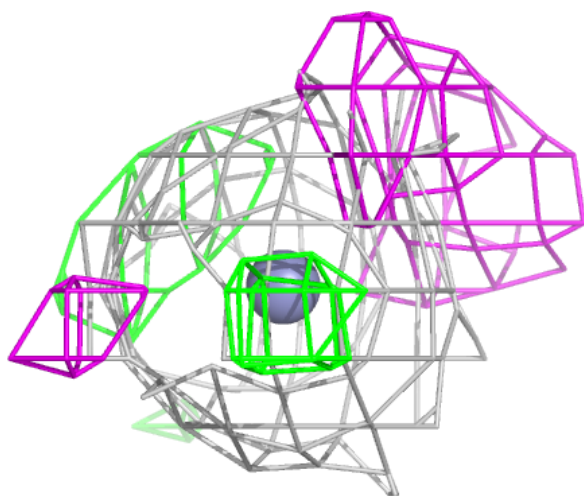
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NO3	A	401	4/4	0.77	0.24	39,44,46,47	0
3	GOL	A	403	6/6	0.85	0.21	26,34,38,39	0
2	NO3	B	401	4/4	0.86	0.17	23,25,32,34	0
3	GOL	A	402	6/6	0.87	0.14	28,32,35,40	0
3	GOL	B	403	6/6	0.90	0.12	29,32,35,37	0
3	GOL	A	405	6/6	0.93	0.10	17,18,22,24	0
3	GOL	A	404	6/6	0.95	0.12	15,38,45,47	0
3	GOL	B	402[B]	6/6	0.97	0.09	14,16,17,19	6
3	GOL	B	402[A]	6/6	0.97	0.09	14,15,17,19	6
4	ZN	A	407	1/1	1.00	0.06	13,13,13,13	0
4	ZN	B	404	1/1	1.00	0.05	14,14,14,14	0
4	ZN	B	405	1/1	1.00	0.05	17,17,17,17	0
4	ZN	A	406	1/1	1.00	0.04	18,18,18,18	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 ZN A 407:**

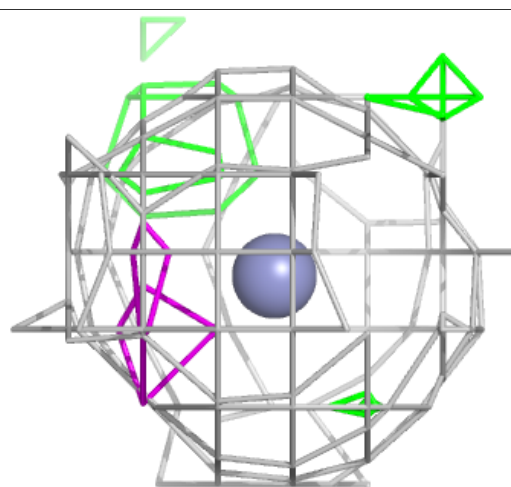
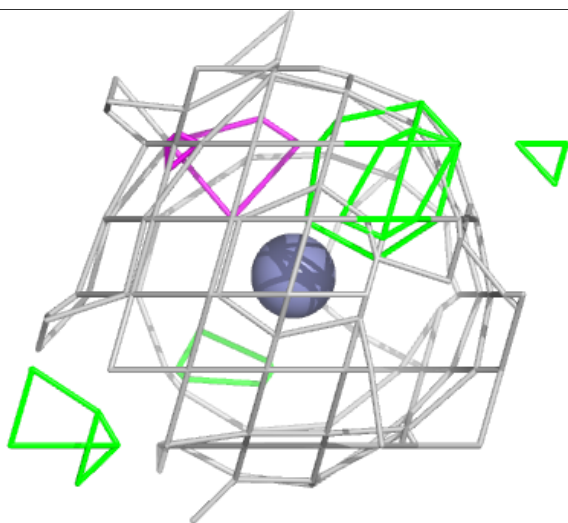
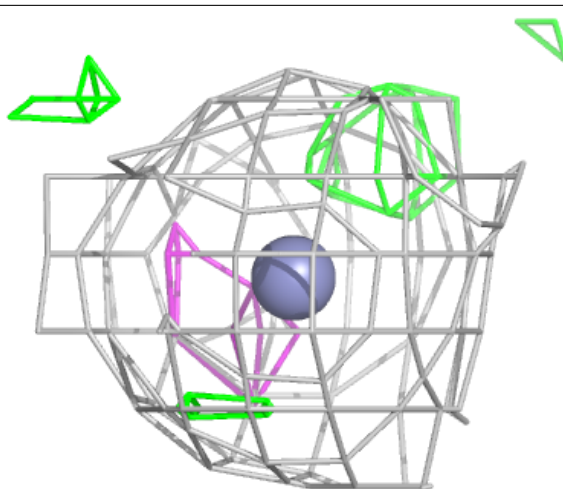
$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)





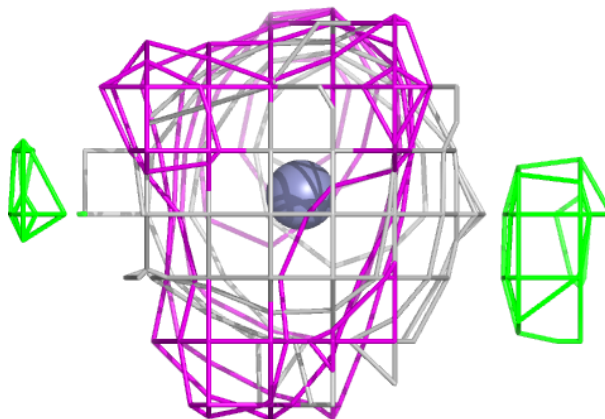
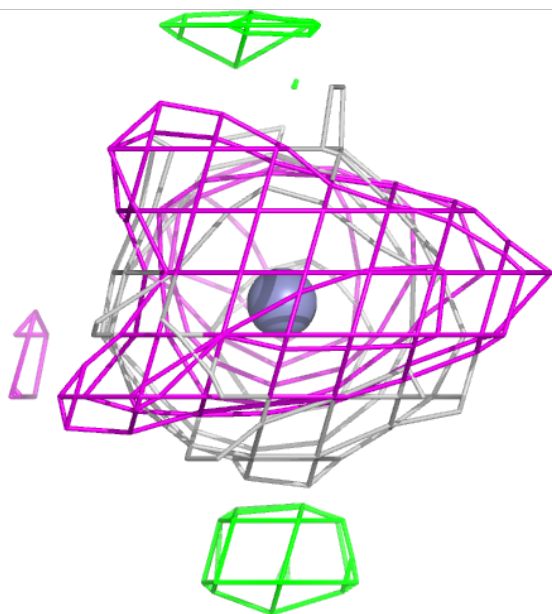
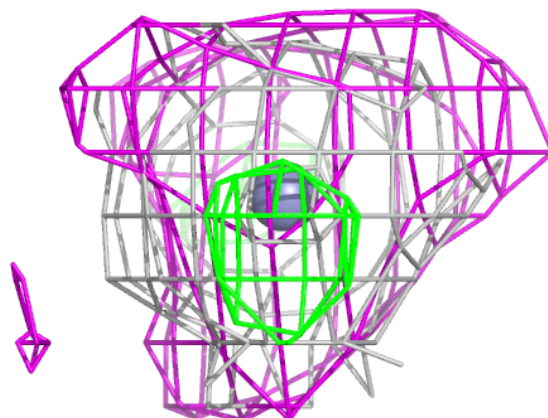
**Electron density around ZN B 404:**

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



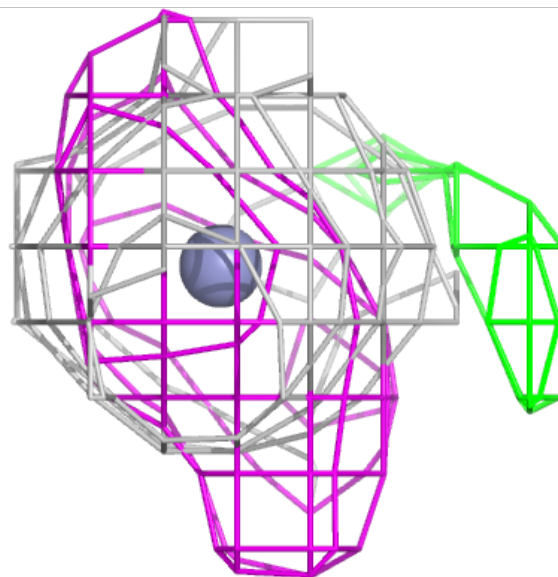
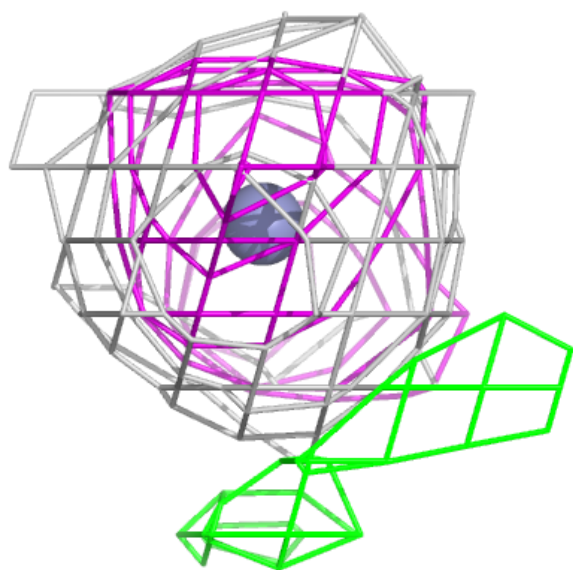
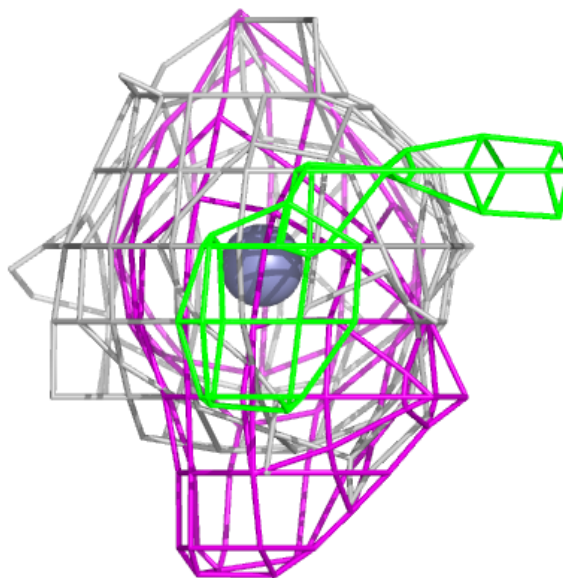
**Electron density around ZN B 405:**

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



**Electron density around ZN A 406:**

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



## 6.5 Other polymers ⓘ

There are no such residues in this entry.