



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 15, 2023 – 03:43 AM EDT

PDB ID : 7UXD  
Title : Crystal structure of APOBEC3G Catalytic domain complex with ssDNA containing 2'-deoxy Zebularine.  
Authors : Maiti, A.; Matsuo, H.  
Deposited on : 2022-05-05  
Resolution : 1.50 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 : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtrriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

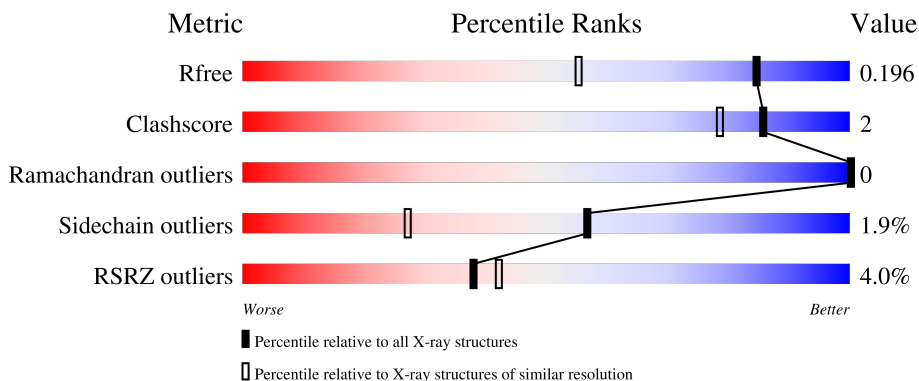
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.50 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	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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	199	 4% (poor fit), 89% (0-1 outliers), 6% (2 outliers), 5% (3+ outliers), 0% (not modelled)
2	B	9	 78% (0-1 outliers), 22% (2+ outliers)

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 1931 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 dC->dU-editing enzyme APOBEC-3G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	190	1539	976	268	283	12	0	6	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	186	GLY	-	expression tag	UNP Q9HC16
A	187	PRO	-	expression tag	UNP Q9HC16
A	188	LEU	-	expression tag	UNP Q9HC16
A	189	GLY	-	expression tag	UNP Q9HC16
A	190	SER	-	expression tag	UNP Q9HC16
A	200	ALA	PRO	engineered mutation	UNP Q9HC16
A	234	LYS	LEU	engineered mutation	UNP Q9HC16
A	236	ALA	ASN	engineered mutation	UNP Q9HC16
A	243	ALA	CYS	engineered mutation	UNP Q9HC16
A	247	LYS	PRO	engineered mutation	UNP Q9HC16
A	310	LYS	PHE	engineered mutation	UNP Q9HC16
A	318	LYS	GLN	engineered mutation	UNP Q9HC16
A	321	ALA	CYS	engineered mutation	UNP Q9HC16
A	322	ALA	GLN	engineered mutation	UNP Q9HC16
A	356	ALA	CYS	engineered mutation	UNP Q9HC16

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	9	179	87	35	49	8	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Zn 1	0	0


- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	193	Total 193	O 193	0	0
4	B	19	Total 19	O 19	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 dC->dU-editing enzyme APOBEC-3G

Chain A: 



- Molecule 2: DNA (5'-D(\*AP\*AP\*TP\*CP\*CP\*(DDN)P\*AP\*AP\*A)-3')

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	47.49Å 47.12Å 52.04Å 90.00° 103.18° 90.00°	Depositor
Resolution (Å)	34.51 – 1.50 34.51 – 1.50	Depositor EDS
% Data completeness (in resolution range)	93.3 (34.51-1.50) 93.3 (34.51-1.50)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.63 (at 1.50Å)	Xtrriage
Refinement program	PHENIX 1.16_3549	Depositor
R, $R_{free}$	0.172 , 0.194 0.172 , 0.196	Depositor DCC
$R_{free}$ test set	1655 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	23.0	Xtrriage
Anisotropy	0.092	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 43.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	1931	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.09% 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: ZN, DDN

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	0.31	0/1601	0.51	0/2170
2	B	0.59	0/179	0.83	0/271
All	All	0.35	0/1780	0.55	0/2441

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 [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	1539	0	1421	7	0
2	B	179	0	102	1	0
3	A	1	0	0	0	0
4	A	193	0	0	1	0
4	B	19	0	0	0	0
All	All	1931	0	1523	7	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 2.

All (7) 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:227:MET:SD	1:A:230:ASP:HA	2.37	0.65
1:A:213:ARG:HG2	2:B:-2:DC:H5''	1.89	0.55
1:A:238:ARG:NH1	1:A:269:TRP:HA	2.31	0.45
1:A:209:GLU:HG2	4:A:545:HOH:O	2.17	0.45
1:A:303:LYS:HB2	1:A:303:LYS:HE3	1.84	0.45
1:A:338:MET:HA	1:A:342:GLU:OE1	2.18	0.42
1:A:259:GLU:HG2	1:A:295[B]:MET:HE1	2.02	0.41

There are no symmetry-related clashes.

## 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	194/199 (98%)	191 (98%)	3 (2%)	0	<a href="#">100</a> <a href="#">100</a>

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	160/172 (93%)	157 (98%)	3 (2%)	<a href="#">57</a> <a href="#">27</a>

All (3) residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	A	230	ASP
1	A	252	PHE
1	A	304	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	228	HIS

### 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	DDN	B	0	2,3	18,20,21	4.34	10 (55%)	18,28,31	1.25	2 (11%)

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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	DDN	B	0	2,3	-	5/7/34/35	0/2/2/2

All (10) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	0	DDN	C2-N3	10.04	1.48	1.34

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	0	DDN	C3'-C4'	-7.60	1.32	1.53
2	B	0	DDN	O4'-C4'	7.53	1.61	1.45
2	B	0	DDN	C2-N1	5.78	1.47	1.38
2	B	0	DDN	C6-C5	5.69	1.46	1.33
2	B	0	DDN	O4'-C1'	-4.17	1.33	1.42
2	B	0	DDN	C6-N1	3.15	1.45	1.38
2	B	0	DDN	O3'-C3'	2.98	1.49	1.43
2	B	0	DDN	C4-N3	2.94	1.53	1.45
2	B	0	DDN	O5'-C5'	-2.40	1.38	1.44

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	0	DDN	N3-C2-N1	3.38	120.91	116.84
2	B	0	DDN	O2-C2-N1	-2.48	119.49	122.79

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	0	DDN	C2'-C1'-N1-C6
2	B	0	DDN	O4'-C1'-N1-C2
2	B	0	DDN	C2'-C1'-N1-C2
2	B	0	DDN	O4'-C1'-N1-C6
2	B	0	DDN	O4'-C4'-C5'-O5'

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is 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

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	190/199 (95%)	0.34	8 (4%) 36 40	17, 31, 62, 87	0
2	B	8/9 (88%)	-0.02	0 100 100	29, 42, 49, 56	0
All	All	198/208 (95%)	0.33	8 (4%) 38 42	17, 31, 62, 87	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	231	THR	3.6
1	A	268	PHE	3.6
1	A	229	ASN	3.5
1	A	380	GLN	2.9
1	A	251	GLY	2.7
1	A	228	HIS	2.3
1	A	381	ASN	2.3
1	A	383	GLU	2.3

### 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	DDN	B	0	19/20	0.97	0.08	22,24,26,27	0

### 6.3 Carbohydrates [i](#)

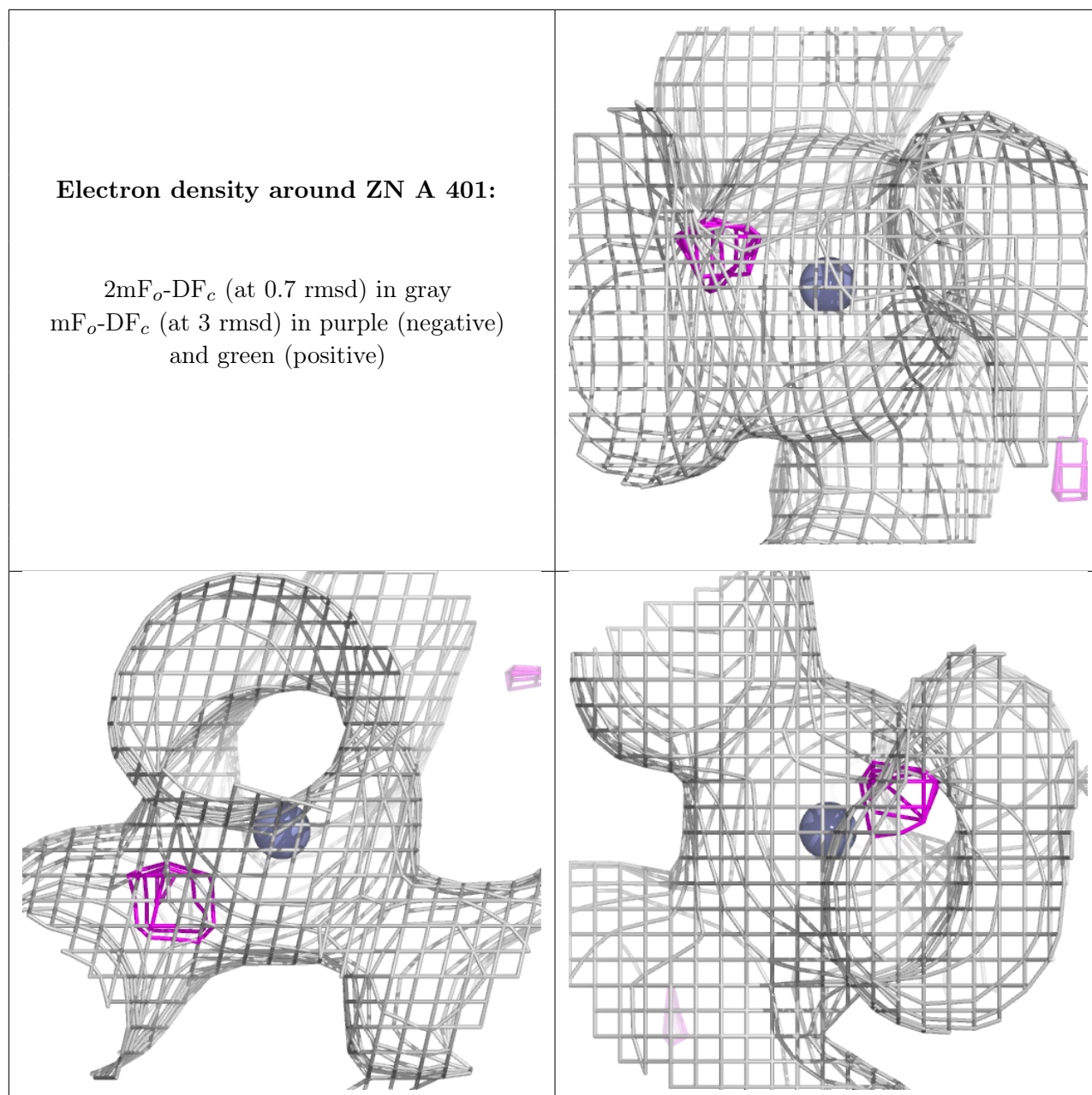
There are no monosaccharides in this entry.

## 6.4 Ligands

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
3	ZN	A	401	1/1	0.99	0.06	27,27,27,27	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.



## 6.5 Other polymers ⓘ

There are no such residues in this entry.