

wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 05:50 pm BST

PDB ID : 4WXG

Title : Crystal structure of L-Serine Hydroxymethyltransferase in complex with a

mixture of L-Threonine and Glycine

Authors: Hernandez, K.; Zelen, I.; Petrillo, G.; Uson, I.; Wandtke, C.; Bujons, J.; Joglar,

J.; Parella, T.; Clapes, P.

Deposited on : 2014-11-13

Resolution : 2.00 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
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) 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

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)

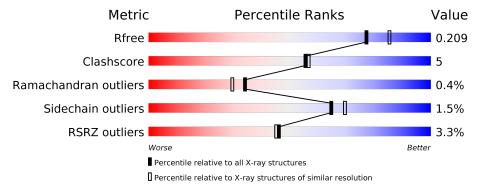
Validation Pipeline (wwPDB-VP) : 2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: X- $RAY\ DIFFRACTION$

The reported resolution of this entry is 2.00 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	$\begin{array}{c} \text{Whole archive} \\ (\#\text{Entries}) \end{array}$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries,\ resolution\ range(\AA)}) \end{array}$
R_{free}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 >=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 <=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	428	87%	8%			
1	С	428	6% 85%	11%			

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
4	GOL	A	503	_	X	_	-



2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6742 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 Serine hydroxymethyltransferase.

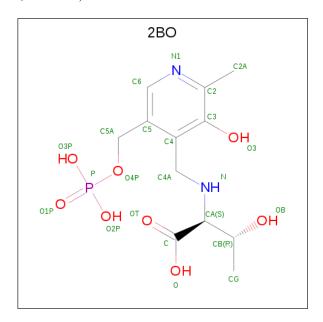
Mo	Chain	Residues		Atoms				ZeroOcc	AltConf	Trace
1	Α	410	Total	С	N	О	S	0	4	0
1	1 A	410	3143	1991	538	605	9	0		
1	C	410	Total	С	N	О	S	0	4	0
1		C 410	3143	1991	538	605	9		'	U

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MET	-	initiating methionine	UNP Q5M0B4
A	-10	ARG	-	expression tag	UNP Q5M0B4
A	-9	GLY	1	expression tag	UNP Q5M0B4
A	-8	SER	-	expression tag	UNP Q5M0B4
A	-7	HIS	_	expression tag	UNP Q5M0B4
A	-6	HIS	=	expression tag	UNP Q5M0B4
A	-5	HIS	=	expression tag	UNP Q5M0B4
A	-4	HIS	_	expression tag	UNP Q5M0B4
A	-3	HIS	=	expression tag	UNP Q5M0B4
A	-2	HIS	_	expression tag	UNP Q5M0B4
A	-1	GLY	_	expression tag	UNP Q5M0B4
A	0	SER	-	expression tag	UNP Q5M0B4
С	-11	MET	_	initiating methionine	UNP Q5M0B4
С	-10	ARG	-	expression tag	UNP Q5M0B4
С	-9	GLY	_	expression tag	UNP Q5M0B4
С	-8	SER	_	expression tag	UNP Q5M0B4
С	-7	HIS	_	expression tag	UNP Q5M0B4
С	-6	HIS	_	expression tag	UNP Q5M0B4
С	-5	HIS	=	expression tag	UNP Q5M0B4
С	-4	HIS		expression tag	UNP Q5M0B4
С	-3	HIS	=	expression tag	UNP Q5M0B4
С	-2	HIS	-	expression tag	UNP Q5M0B4
С	-1	GLY	-	expression tag	UNP Q5M0B4
С	0	SER	-	expression tag	UNP Q5M0B4



• Molecule 2 is N-($\{3-\text{hydroxy-}2-\text{methyl-}5-[(\text{phosphonooxy})\text{methyl}]$ pyridin-4-yl $\}$ methyl)-L-t hreonine (three-letter code: 2BO) (formula: $C_{12}H_{19}N_2O_8P$) (labeled as "Ligand of Interest" by author).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf		
9	Λ	1	Total	С	N	О	Р	0	1
	2 A	1	27	14	2	10	1	U	1
9	2 C	1	Total	С	N	О	Р	0	1
		1	27	14	2	10	1	0	

• Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na) (labeled as "Ligand of Interest" by author).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Na 1 1	0	0
3	С	1	Total Na 1 1	0	0

• Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$) (labeled as "Ligand of Interest" by author).





Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
1	Λ	1	Total	С	О	0	0
4	Λ	1	6	3	3	0	0

• Molecule 5 is water.

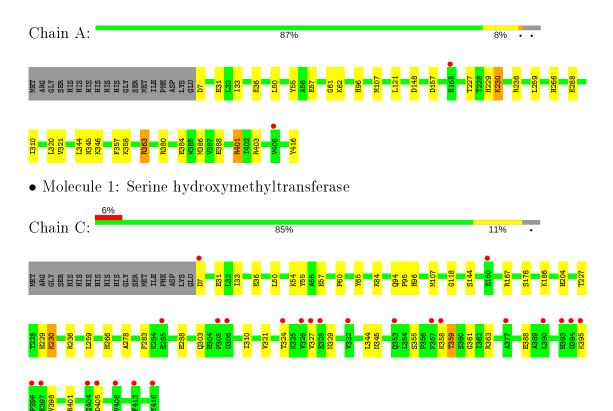
Mol	Chain	Residues	${f Atoms}$	ZeroOcc	\mid AltConf \mid
5	A	229	Total O 230 230	0	1
5	C	163	Total O 164 164	0	1



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 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: Serine hydroxymethyltransferase





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants	115.43Å 115.43Å 191.48Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.43 - 2.00	Depositor
Resolution (A)	45.44 - 2.00	EDS
% Data completeness	98.6 (49.43-2.00)	Depositor
(in resolution range)	98.6 (45.44-2.00)	EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	1.94 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
P. P.	0.171 , 0.203	Depositor
R, R_{free}	0.181 , 0.209	DCC
R_{free} test set	4344 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	32.5	Xtriage
Anisotropy	0.010	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.37, 42.6	EDS
L-test for twinning ²	$ < L >=0.49, < L^2>=0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	6742	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.25% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical values of <|L|>, $< L^2>$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



¹Intensities estimated from amplitudes.

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: 2BO, GOL, NA

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	Boı	nd lengths	Bond angles		
Wioi Chain		RMSZ	# Z > 5	RMSZ	# Z >5	
1	A	0.96	$2/3225 \ (0.1\%)$	1.02	$14/4377 \ (0.3\%)$	
1	С	0.92	$2/3225 \ (0.1\%)$	0.98	$7/4377 \ (0.2\%)$	
All	All	0.94	4/6450 (0.1%)	1.00	$21/8754 \ (0.2\%)$	

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\operatorname{\AA})$	Ideal(A)
1	A	288	GLU	CD-OE1	-6.42	1.18	1.25
1	A	416	TYR	CE1-CZ	6.04	1.46	1.38
1	С	65	TYR	CE1-CZ	5.30	1.45	1.38
1	С	65	TYR	CG-CD2	5.06	1.45	1.39

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^o)$
1	С	236	ARG	NE-CZ-NH2	-10.78	114.91	120.30
1	A	401	ARG	NE-CZ-NH1	-9.57	115.51	120.30
1	A	236	ARG	NE-CZ-NH2	-9.22	115.69	120.30
1	A	229[A]	HIS	N-CA-C	8.18	133.09	111.00
1	A	229[B]	HIS	N-CA-C	8.18	133.09	111.00

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



Mol	Chain	Non-H	H(model)	$\mathbf{H}(\mathbf{added})$	Clashes	Symm-Clashes
1	A	3143	0	3114	31	0
1	С	3143	0	3114	32	0
2	A	27	0	8	0	0
2	С	27	0	8	0	0
3	A	1	0	0	0	0
3	С	1	0	0	0	0
4	A	6	0	8	0	0
5	A	230	0	0	8	1
5	С	164	0	0	7	0
All	All	6742	0	6252	64	1

the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

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.

The worst 5 of 64 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{array}{l} ext{Interatomic} \ ext{distance } (ext{Å}) \end{array}$	$egin{array}{c} { m Clash} \\ { m overlap} \ ({ m \AA}) \end{array}$
1:C:355:SER:HB3	1:C:358:LYS:HD3	1.35	1.07
1:C:388:GLU:OE2	1:C:401:ARG:NH1	2.02	0.93
1:C:310:ILE:HG12	5:C:651:HOH:O	1.79	0.82
1:A:31[A]:GLU:HG3	5:A:604:HOH:O	1.79	0.80
1:C:7:ASP:OD1	5:C:601:HOH:O	2.05	0.74

All (1) 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-1 Atom-2		$egin{array}{c} { m Clash} \\ { m overlap} \ ({ m \AA}) \end{array}$
5:A:802:HOH:O	5:A:802:HOH:O[7_555]	2.05	0.15

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 Allower		Allowed	Outliers	Perce	ntiles
1	A	412/428 (96%)	400 (97%)	10 (2%)	2 (0%)	29	23
1	С	412/428 (96%)	398 (97%)	11 (3%)	3 (1%)	22	16
All	All	824/856 (96%)	798 (97%)	21 (2%)	5 (1%)	34	19

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	230[A]	LYS
1	A	230[B]	LYS
1	С	230[A]	LYS
1	С	230[B]	LYS
1	С	359	THR

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	331/343 (96%)	328 (99%)	3 (1%)	78 83		
1	С	331/343 (96%)	324 (98%)	7 (2%)	53 57		
All	All	662/686 (96%)	652 (98%)	10 (2%)	65 69		

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

Mol	Chain	Res	Type
1	С	55	TYR
1	С	60	PRO
1	С	288	GLU
1	С	50	LEU
1	С	84	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	347	ASN

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type
1	A	353	GLN
1	С	335	ASN
1	A	345	ASN
1	С	266	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)

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 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	Mol Type Chain Res		Link	Во	ond leng	ths	Е	ond ang	gles	
WIOI	Mol Type Chain	Chain	rtes	Lilik	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	2BO	A	501[B]	-	20,20,23	5.19	4 (20%)	26,28,33	3.76	10 (38%)
4	GOL	A	503	-	5,5,5	1.79	2 (40%)	5, 5, 5	1.68	1 (20%)
2	2BO	A	501[A]	-	20,23,23	1.83	4 (20%)	26,33,33	4.18	11 (42%)
2	2BO	С	501[A]	-	20,23,23	1.68	4 (20%)	26,33,33	3.56	11 (42%)
2	2BO	С	501[B]	-	20,20,23	4.18	4 (20%)	26,28,33	2.93	10 (38%)

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.



, ,	moone	no	outlions	$\alpha \mathbf{f}$	that	kind	moro	identified.
_	means	110	oumers	OI	unau	MIIIU	were	identined.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	2BO	A	501[B]	-	-	10/15/12/19	0/1/1/1
4	GOL	A	503	-	1	4/4/4/4	_
2	2BO	A	501[A]	-	-	10/15/19/19	0/1/1/1
2	2BO	С	501[A]	-	-	6/15/19/19	0/1/1/1
2	2BO	С	501[B]	-	-	6/15/12/19	0/1/1/1

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(\operatorname{\AA})$	$\operatorname{Ideal}(ext{\AA})$
2	A	501[B]	2BO	CB-CA	21.96	2.18	1.53
2	С	501[B]	2BO	CB-CA	17.29	2.05	1.53
2	A	501[B]	2BO	C4A-N	-5.28	1.31	1.46
2	A	501[A]	2BO	C4A-N	-5.28	1.31	1.46
2	С	501[A]	2BO	C4A-N	-5.20	1.31	1.46

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

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^o)$
2	A	501[A]	2BO	C4A-N-CA	16.86	137.13	113.78
2	С	501[A]	2BO	C4A-N-CA	14.99	134.54	113.78
2	A	501[B]	2BO	CB-CA-N	-13.96	76.48	111.73
2	С	501[B]	2BO	CB-CA-N	-10.41	85.43	111.73
2	A	501[B]	2BO	CG-CB-CA	7.04	126.55	112.30

There are no chirality outliers.

5 of 36 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	503	GOL	C1-C2-C3-O3
2	A	501[B]	2BO	C5-C4-C4A-N
2	A	501[B]	2BO	C5A-O4P-P-O2P
2	A	501[B]	2BO	C5A-O4P-P-O3P
2	A	501[B]	2BO	N-CA-CB-OB

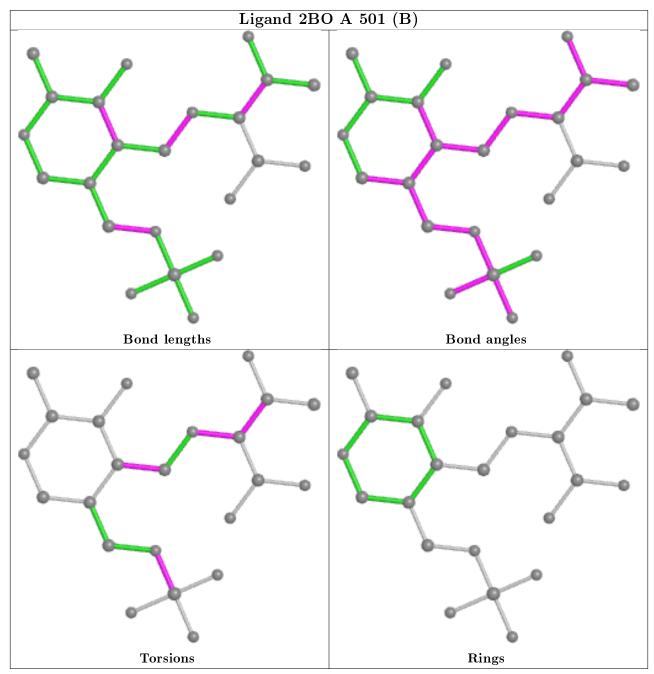
There are no ring outliers.

No monomer is involved in short contacts.

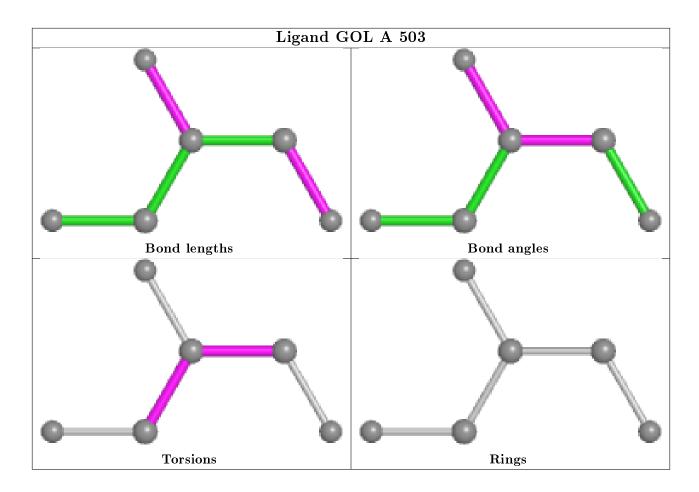
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will



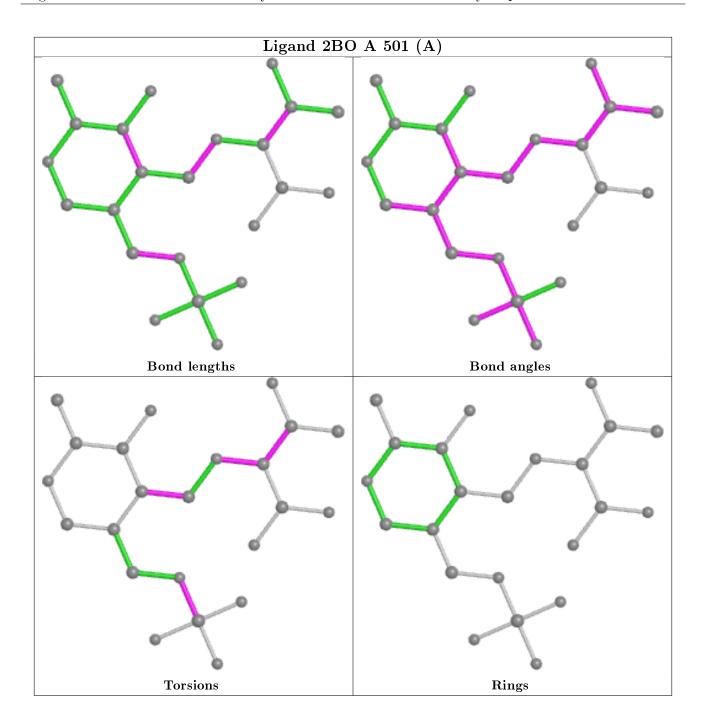
also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



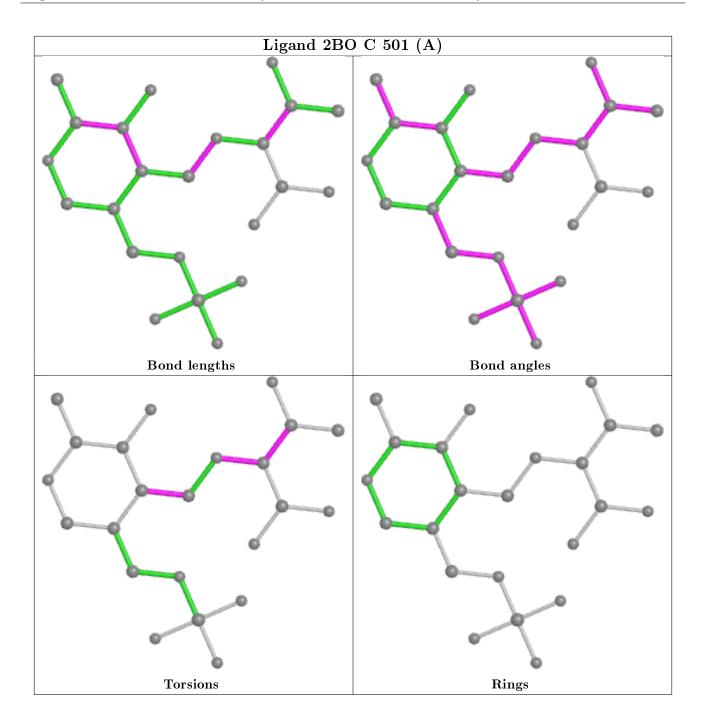




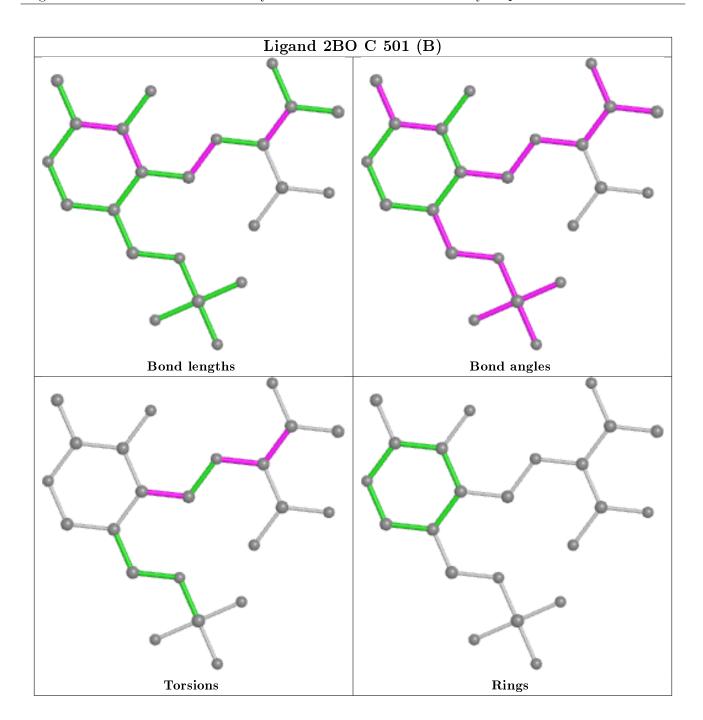












5.7 Other polymers (i)

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^{th} 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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(m \AA^2)$	Q < 0.9
1	A	410/428 (95%)	-0.19	2 (0%) 91 90	22, 30, 50, 74	0
1	С	410/428 (95%)	0.07	25 (6%) 21 20	22, 36, 65, 86	0
All	All	820/856 (95%)	-0.06	27 (3%) 46 45	22, 33, 59, 86	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	С	324	THR	3.7
1	С	394	ASP	3.7
1	С	357	PHE	3.2
1	С	328	GLU	3.2
1	С	396	PRO	3.1

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 carbohydrates in this entry.

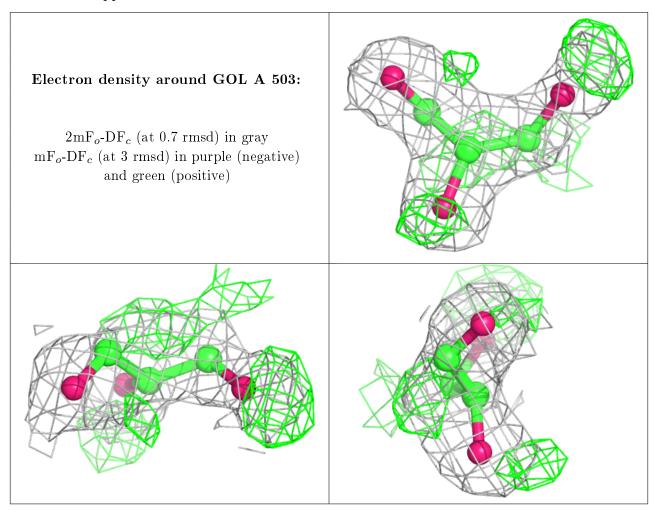
6.4 Ligands (i)

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^{th} 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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
4	GOL	A	503	6/6	0.83	0.24	36,42,49,54	0
3	NA	С	502	1/1	0.91	0.23	60,60,60,60	0
2	2BO	A	501[B]	20/23	0.97	0.17	25,32,41,45	4
2	2BO	A	501[A]	23/23	0.97	0.17	25,32,38,45	7
2	2BO	С	501[A]	23/23	0.97	0.14	26,36,43,49	7
2	2BO	С	501[B]	20/23	0.97	0.14	26,35,40,43	4
3	NA	A	502	1/1	0.97	0.11	40,40,40,40	0

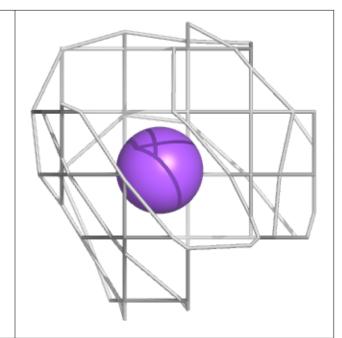
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

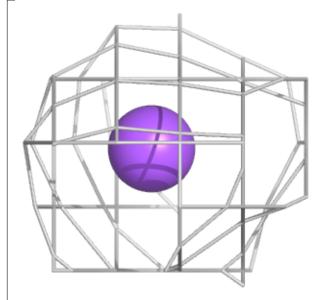


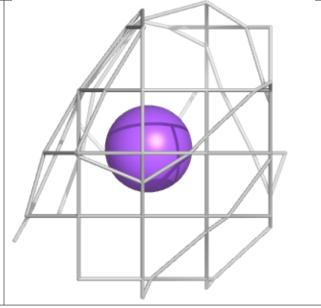


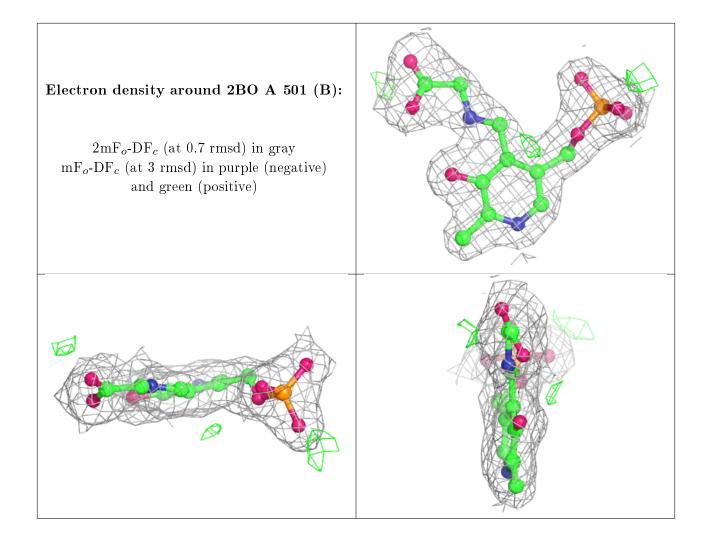
Electron density around NA C 502:

 $2 \mathrm{mF}_o\text{-}\mathrm{DF}_c$ (at 0.7 rmsd) in gray $\mathrm{mF}_o\text{-}\mathrm{DF}_c$ (at 3 rmsd) in purple (negative) and green (positive)

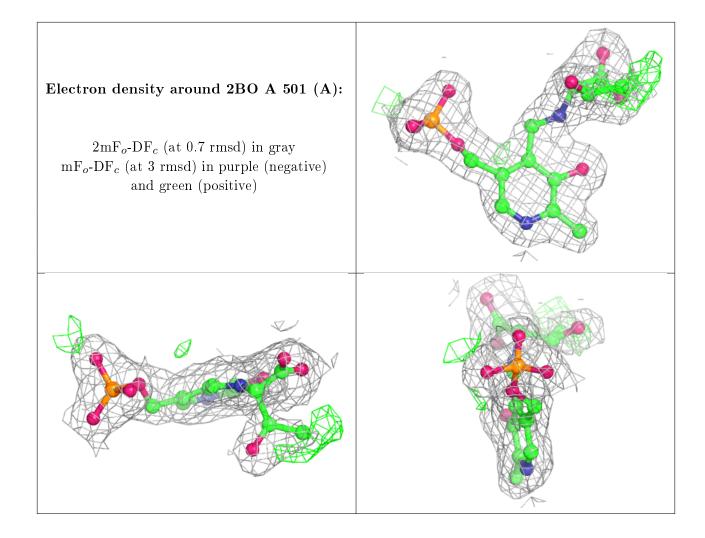




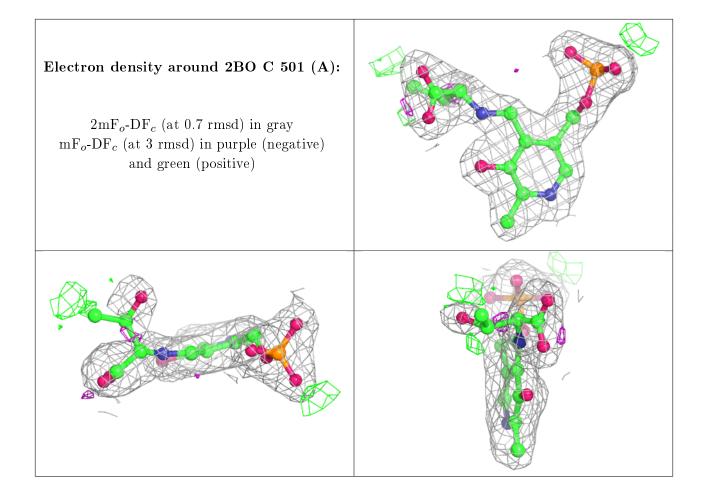




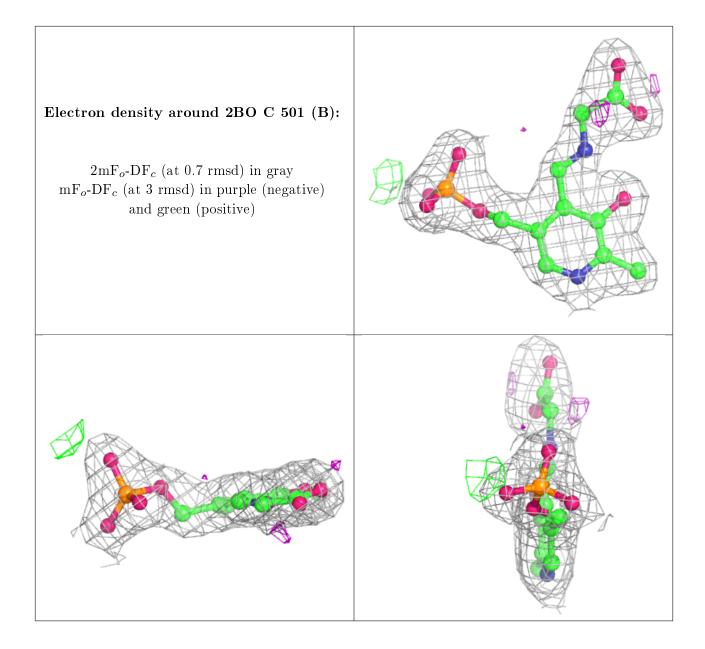




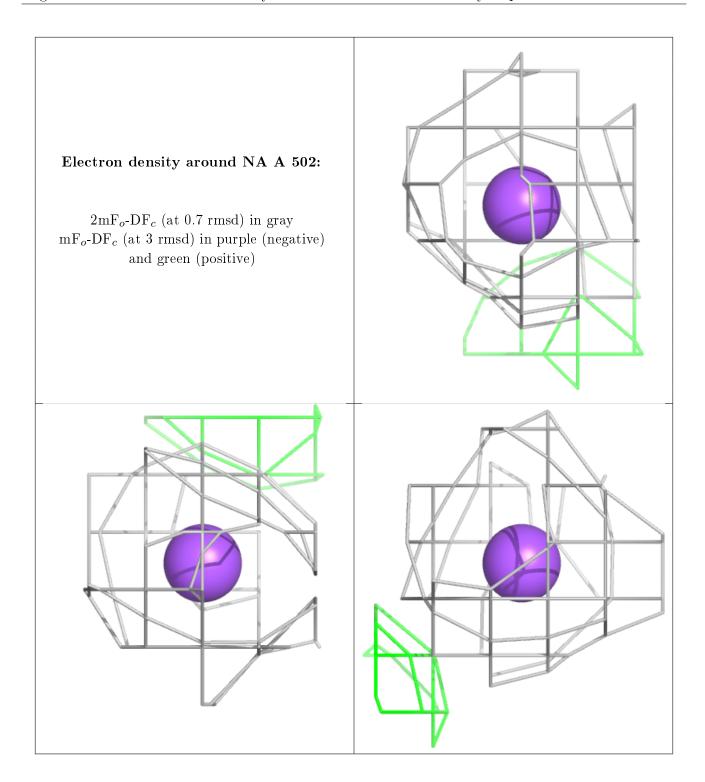












6.5 Other polymers (i)

There are no such residues in this entry.

