

# wwPDB X-ray Structure Validation Summary Report (i)

#### Nov 21, 2023 – 03:14 PM JST

:	7VFL
:	Crystal structure of SdgB (UDP, NAG, and O-glycosylated SD peptide-binding
	form)
:	Kim, DG.; Baek, I.; Lee, Y.; Kim, H.S.
	2021-09-13
:	2.45  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 types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

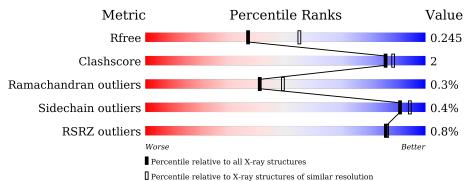
MolProbity Mogul Xtriage (Phenix) EDS	:	4.02b-467 1.8.5 (274361), CSD as541be (2020) 1.13 2.36
buster-report Percentile statistics Refmac	: : :	1.1.7 (2018) 20191225.v01 (using entries in the PDB archive December 25th 2019) 5.8.0158 7.0.044 (Gargrove)
Ideal geometry (DNA, RNA) Validation Pipeline (wwPDB-VP)		Parkinson et al. (1996) 2.36

# 1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.45 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R <sub>free</sub>	130704	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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 >=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							
1	А	505	% • 93%		6% •				
1	С	505	% 91%		8% •				
2	D	9	89%		11%				
2	G	9	11%	11%	33%				

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 crite-



ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	NAG	D	100	X	-	-	-
4	NAG	D	101	Х	-	-	-
4	NAG	G	101	Х	-	-	-



# 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8579 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	Δ	500	Total	С	Ν	0	$\mathbf{S}$	0	0	0
	A	500	4167	2689	694	773	11	0		
1	C	500	Total	С	Ν	0	S	0	0	0
		500	4167	2689	694	773	11	0	0	0

• Molecule 1 is a protein called Glycosyl transferase, group 1 family protein.

Chain	Residue	Modelled	Actual	Comment	Reference
А	0	ALA	-	expression tag	UNP A0A0H2XGN0
А	497	LEU	-	expression tag	UNP A0A0H2XGN0
А	498	GLU	-	expression tag	UNP A0A0H2XGN0
А	499	HIS	-	expression tag	UNP A0A0H2XGN0
А	500	HIS	-	expression tag	UNP A0A0H2XGN0
А	501	HIS	-	expression tag	UNP A0A0H2XGN0
А	502	HIS	-	expression tag	UNP A0A0H2XGN0
А	503	HIS	-	expression tag	UNP A0A0H2XGN0
А	504	HIS	-	expression tag	UNP A0A0H2XGN0
С	0	ALA	-	expression tag	UNP A0A0H2XGN0
С	497	LEU	-	expression tag	UNP A0A0H2XGN0
С	498	GLU	-	expression tag	UNP A0A0H2XGN0
С	499	HIS	-	expression tag	UNP A0A0H2XGN0
С	500	HIS	-	expression tag	UNP A0A0H2XGN0
С	501	HIS	-	expression tag	UNP A0A0H2XGN0
С	502	HIS	-	expression tag	UNP A0A0H2XGN0
С	503	HIS	-	expression tag	UNP A0A0H2XGN0
С	504	HIS	-	expression tag	UNP A0A0H2XGN0

There are 18 discrepancies between the modelled and reference sequences:

• Molecule 2 is a protein called SER-ASP-SER-ASP-SER-ASP.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace		
2	D	8	Total 57	C 28	N 8	O 21	0	0	0

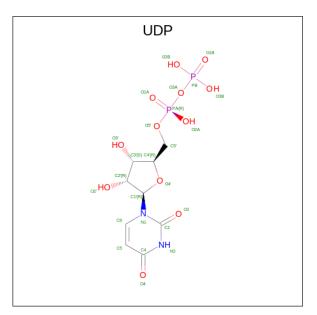
Continued on next page...



Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace		
2	G	6	Total 42	C 21	N 6	0 15	0	0	0

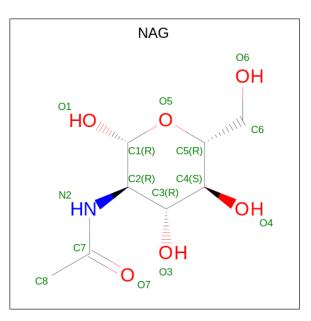
• Molecule 3 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula:  $C_9H_{14}N_2O_{12}P_2$ ) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
3	Δ	1	Total	С	Ν	Ο	Р	0	Ο	
5	Л	1	25	9	2	12	2	0	0	
2	С	1	Total	С	Ν	Ο	Р	0	0	
5	U		25	9	2	12	2		U	

• Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ) (labeled as "Ligand of Interest" by depositor).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	1	Total C N O 15 8 1 6	0	0
4	С	1	Total         C         N         O           15         8         1         6	0	0
4	D	1	Total         C         N         O           14         8         1         5	0	0
4	D	1	Total         C         N         O           14         8         1         5	0	0
4	G	1	Total         C         N         O           14         8         1         5	0	0

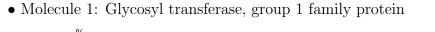
• Molecule 5 is water.

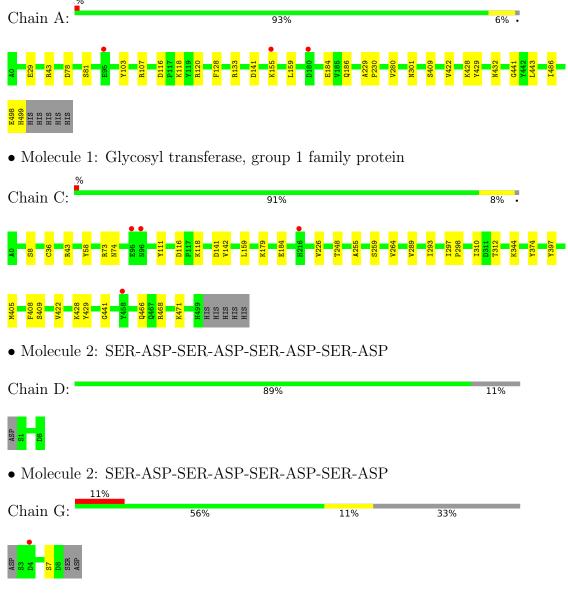
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	13	Total         O           13         13	0	0
5	С	11	Total         O           11         11	0	0



## 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.







## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	70.20Å 130.57Å 189.85Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	47.00 - 2.45	Depositor
Resolution (A)	47.00 - 2.45	EDS
% Data completeness	93.1 (47.00-2.45)	Depositor
(in resolution range)	93.2 (47.00-2.45)	EDS
R <sub>merge</sub>	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.24 (at 2.45 \text{\AA})$	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
D D.	0.182 , $0.240$	Depositor
$R, R_{free}$	0.193 , $0.245$	DCC
$R_{free}$ test set	3041 reflections $(5.01%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	39.1	Xtriage
Anisotropy	0.147	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.38 , $38.5$	EDS
L-test for twinning <sup>2</sup>	$ \langle L  \rangle = 0.48, \langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	8579	wwPDB-VP
Average B, all atoms $(Å^2)$	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.88% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities 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: UDP, NAG

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	А	0.79	0/4273	0.76	0/5783	
1	С	0.79	0/4273	0.77	0/5783	
2	D	0.67	0/56	0.75	0/73	
2	G	0.71	0/41	0.72	0/54	
All	All	0.79	0/8643	0.76	0/11693	

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	А	4167	0	4077	18	0
1	С	4167	0	4077	20	0
2	D	57	0	36	0	0
2	G	42	0	25	1	0
3	А	25	0	11	0	0
3	С	25	0	11	0	0
4	А	15	0	15	0	0
4	С	15	0	15	0	0
4	D	28	0	26	0	0

Continued on next page...



0011111	Continueu front prettous page										
Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes					
4	G	14	0	13	0	0					
5	А	13	0	0	0	0					
5	С	11	0	0	0	0					
All	All	8579	0	8306	37	0					

Continued from previous page...

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.

The worst 5 of 37 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:43:ARG:HB2	1:A:141:ASP:HA	1.85	0.58
1:C:468:ARG:HA	1:C:471:LYS:CE	2.35	0.56
1:C:310:ILE:HG22	1:C:312:THR:H	1.75	0.50
1:C:248:THR:OG1	1:C:405:MET:HE2	2.11	0.49
1:A:428:LYS:HD3	1:A:429:TYR:CE1	2.47	0.49

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	Percer	ntiles
1	А	498/505~(99%)	480 (96%)	17 (3%)	1 (0%)	47	57
1	$\mathbf{C}$	498/505~(99%)	484 (97%)	12 (2%)	2~(0%)	34	41
2	D	6/9~(67%)	6 (100%)	0	0	100	100
2	G	4/9~(44%)	4 (100%)	0	0	100	100
All	All	1006/1028~(98%)	974 (97%)	29~(3%)	3(0%)	41	49

All (3) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	А	409	SER
1	С	408	PHE
1	С	409	SER

#### 5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent side chain 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	ain Analysed Rotameric Outliers		Perce	ntiles	
1	А	460/465~(99%)	459 (100%)	1 (0%)	93	96
1	С	460/465~(99%)	457 (99%)	3 (1%)	84	90
2	D	8/9~(89%)	8 (100%)	0	100	100
2	G	6/9~(67%)	6 (100%)	0	100	100
All	All	934/948~(98%)	930 (100%)	4 (0%)	91	94

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	155	LYS
1	С	8	SER
1	С	74	ASN
1	С	118	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	А	372	GLN
1	А	467	GLN
1	С	115	GLN
1	С	51	ASN
1	А	156	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)

7 ligands are 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	Turne	oe Chain Res Link		Bo	Bond lengths			Bond angles		
	Type	Chain	nes		Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	NAG	А	602	-	$15,\!15,\!15$	0.47	0	21,21,21	0.68	0
4	NAG	D	100	2	14,14,15	0.90	1 (7%)	17,19,21	0.76	0
4	NAG	D	101	2	14,14,15	0.77	0	17,19,21	1.18	2 (11%)
4	NAG	G	101	2	14,14,15	0.77	0	17,19,21	1.38	3 (17%)
4	NAG	С	602	-	$15,\!15,\!15$	0.48	0	21,21,21	0.72	0
3	UDP	А	601	-	24,26,26	1.27	3 (12%)	37,40,40	1.56	5 (13%)
3	UDP	С	601	-	24,26,26	1.29	3 (12%)	37,40,40	1.57	5 (13%)

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
4	NAG	С	602	-	-	2/6/26/26	0/1/1/1
4	NAG	D	100	2	1/1/5/7	0/6/23/26	0/1/1/1
4	NAG	D	101	2	1/1/5/7	2/6/23/26	0/1/1/1
4	NAG	G	101	2	1/1/5/7	2/6/23/26	0/1/1/1
4	NAG	А	602	-	-	0/6/26/26	0/1/1/1
3	UDP	С	601	-	-	3/16/32/32	0/2/2/2

Continued on next page...



Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	UDP	А	601	-	-	4/16/32/32	0/2/2/2

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\operatorname{\AA})$	$\mathrm{Ideal}(\mathrm{\AA})$
3	С	601	UDP	C4-N3	-3.17	1.32	1.38
3	А	601	UDP	C4-N3	-3.00	1.33	1.38
3	С	601	UDP	C2-N3	-2.72	1.33	1.38
4	D	100	NAG	O5-C1	-2.60	1.39	1.43
3	А	601	UDP	C2-N3	-2.53	1.33	1.38

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
3	С	601	UDP	C4-N3-C2	-4.44	120.72	126.58
3	А	601	UDP	C4-N3-C2	-4.35	120.84	126.58
3	А	601	UDP	N3-C2-N1	4.10	120.33	114.89
4	G	101	NAG	C1-O5-C5	3.95	117.54	112.19
3	С	601	UDP	N3-C2-N1	3.88	120.04	114.89

All (3) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	D	100	NAG	C1
4	D	101	NAG	C1
4	G	101	NAG	C1

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	А	601	UDP	PA-O3A-PB-O2B
3	С	601	UDP	C5'-O5'-PA-O1A
4	D	101	NAG	C8-C7-N2-C2
4	D	101	NAG	O7-C7-N2-C2
4	G	101	NAG	C8-C7-N2-C2

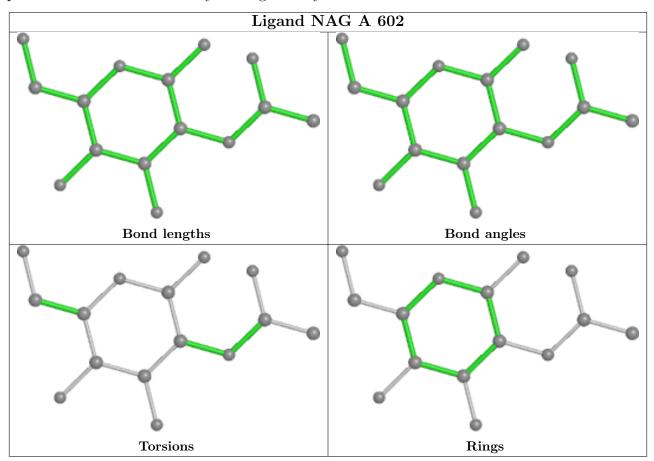
There are no ring outliers.

No monomer is involved in short contacts.

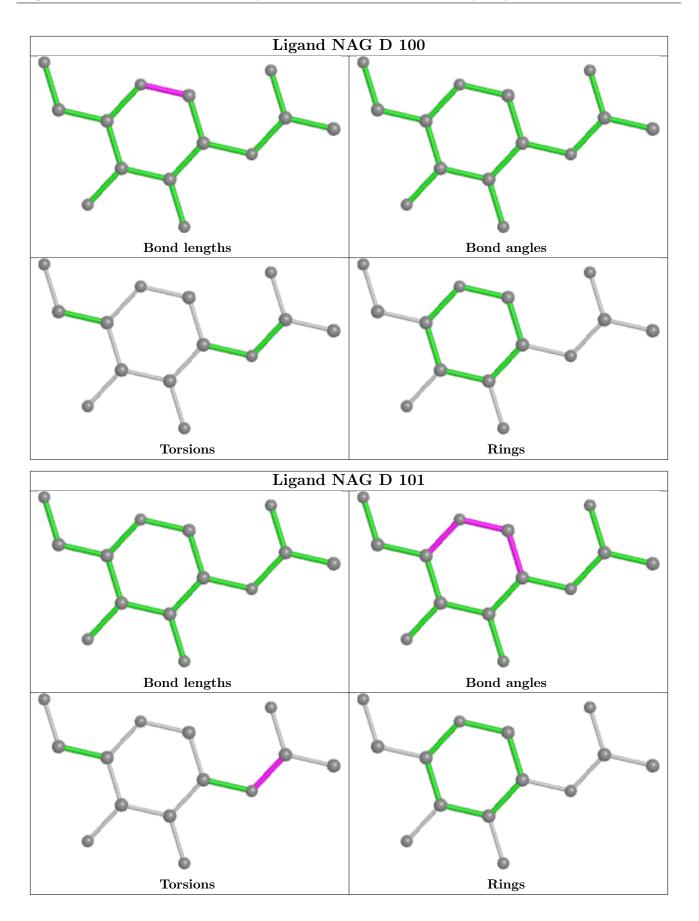
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will



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.

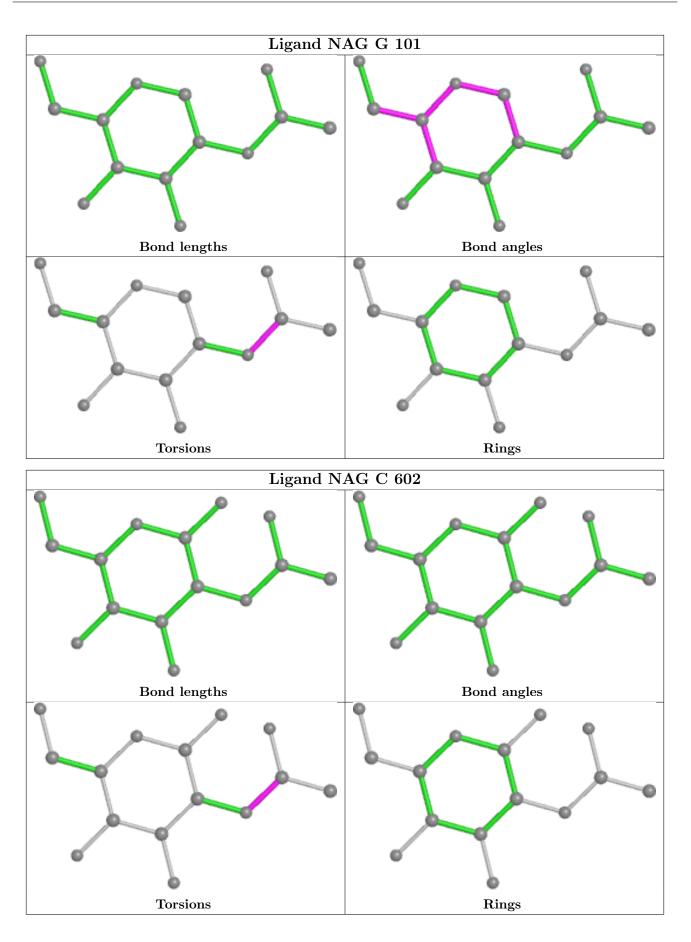






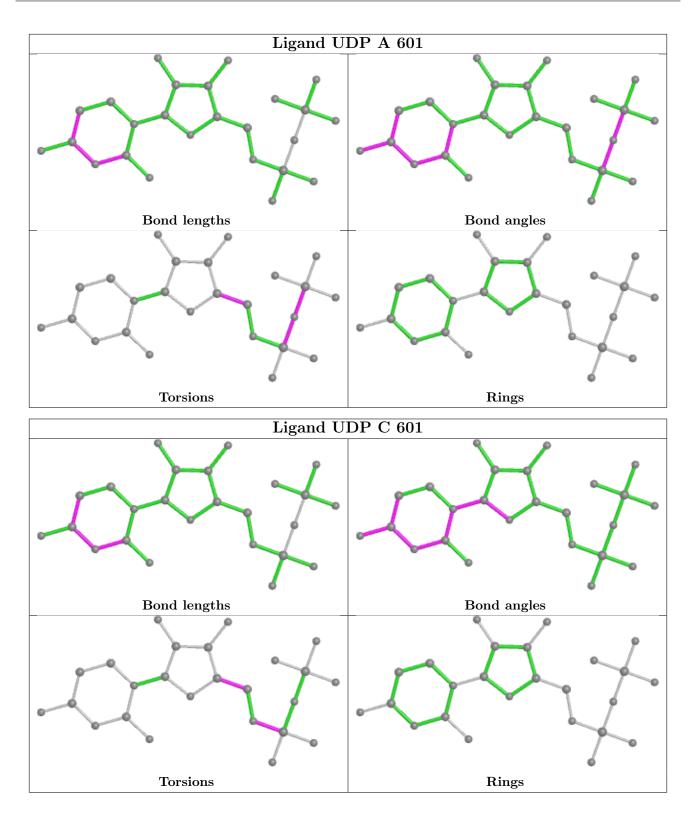












### 5.7 Other polymers (i)

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	$\langle RSRZ \rangle$	$\mathbf{RSRZ} > $ # $\mathbf{RSRZ} > 2$		Q < 0.9
1	А	500/505~(99%)	-0.40	3 (0%) 89 89	23, 37, 63, 87	0
1	С	500/505~(99%)	-0.33	4 (0%) 86 86	25, 38, 61, 86	0
2	D	8/9~(88%)	0.19	0 100 100	39,51,54,57	0
2	G	6/9~(66%)	0.50	1 (16%) 1 1	41, 49, 65, 82	0
All	All	1014/1028~(98%)	-0.35	8 (0%) 86 86	23, 37, 62, 87	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	С	96	ASN	3.0
2	G	4	ASP	2.9
1	А	95	GLU	2.7
1	С	95	GLU	2.6
1	А	155	LYS	2.5

## 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.

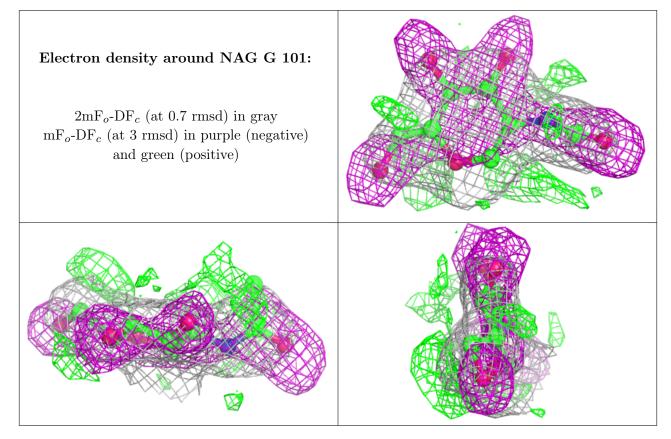
## 6.4 Ligands (i)

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,  $95^{th}$  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.

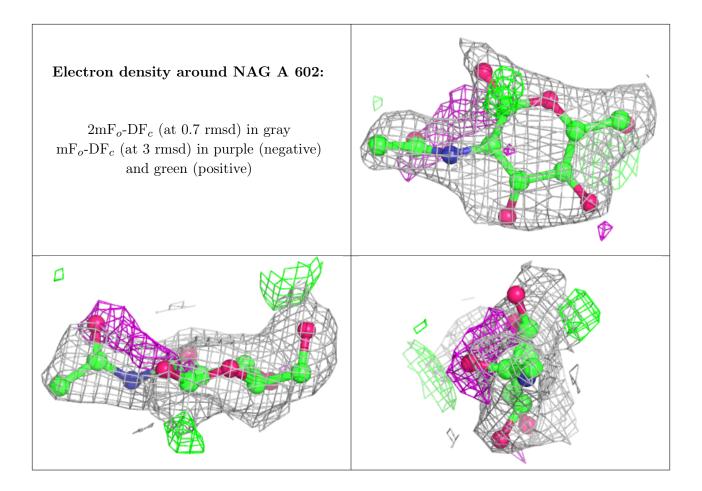


Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-factors}(\mathrm{\AA}^2)$	Q<0.9
4	NAG	G	101	14/15	0.82	0.36	20,20,20,20	0
4	NAG	А	602	15/15	0.86	0.23	37,49,61,61	0
4	NAG	D	101	14/15	0.88	0.28	20,20,20,20	0
4	NAG	С	602	15/15	0.93	0.16	40,55,63,64	0
4	NAG	D	100	14/15	0.94	0.17	20,20,20,20	0
3	UDP	С	601	25/25	0.97	0.13	25,34,42,56	0
3	UDP	А	601	25/25	0.98	0.11	27,35,47,54	0

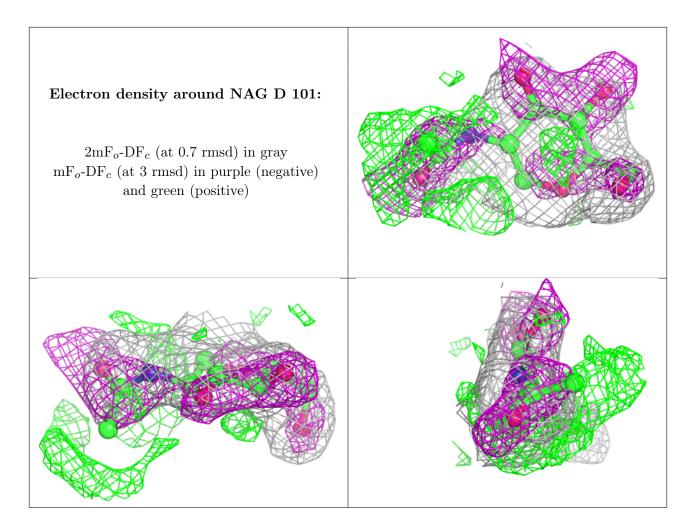
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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



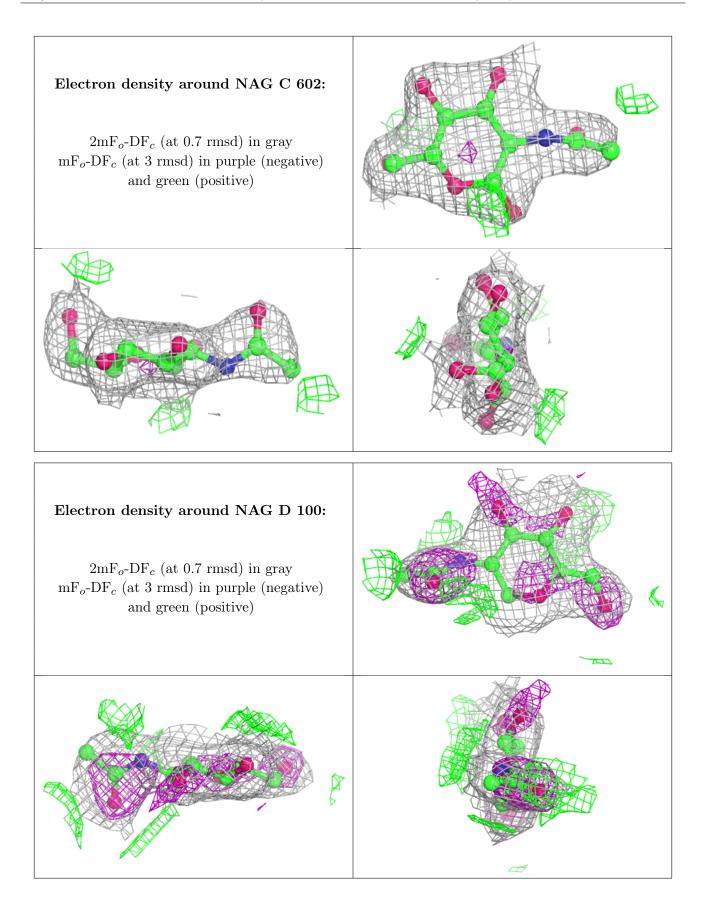




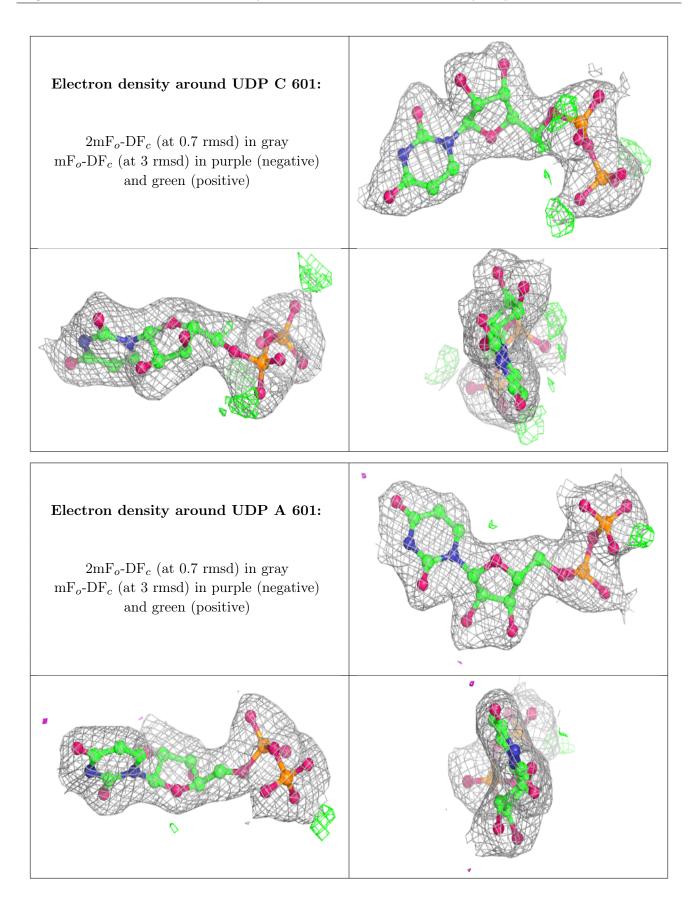














## 6.5 Other polymers (i)

There are no such residues in this entry.

