



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 30, 2025 – 04:15 am BST

PDB ID : 9EVJ / pdb\_00009evj  
Title : Crystal Structure of human Collagen Hydroxylysine Galactosyltransferase  
GLT25D1/COLGALT1: complex with Mn<sup>2+</sup> and UDP-Gal  
Authors : De Marco, M.; Scietti, L.; Rai, S.R.; Mattoteia, D.; Pinnola, A.; Forneris, F.  
Deposited on : 2024-03-30  
Resolution : 2.70 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 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 : ?? (???), CSD ??CSD?? (????)  
Xtriage (Phenix) : 1.13  
EDS : 3.0  
buster-report : 1.1.7 (2018)  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.003 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.42

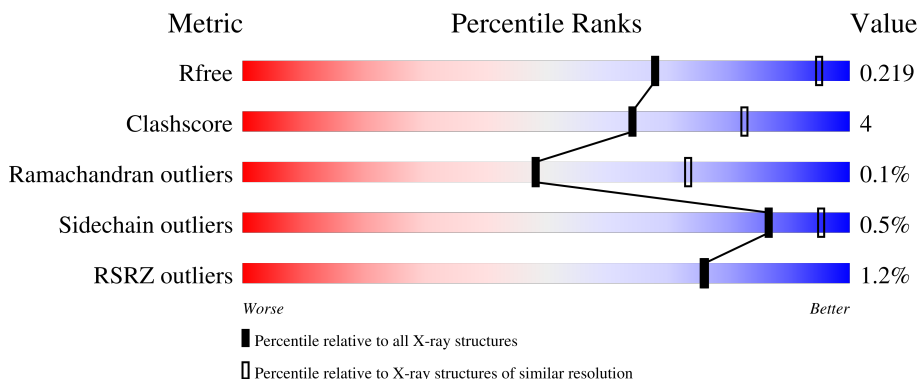
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	3333 (2.70-2.70)
Clashscore	180529	3684 (2.70-2.70)
Ramachandran outliers	177936	3633 (2.70-2.70)
Sidechain outliers	177891	3633 (2.70-2.70)
RSRZ outliers	164620	3333 (2.70-2.70)

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	594	<div> <div>2%</div> <div>81%</div> <div>9%</div> <div>10%</div> </div>
1	B	594	<div> <div>%</div> <div>78%</div> <div>10%</div> <div>12%</div> </div>

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 8923 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.

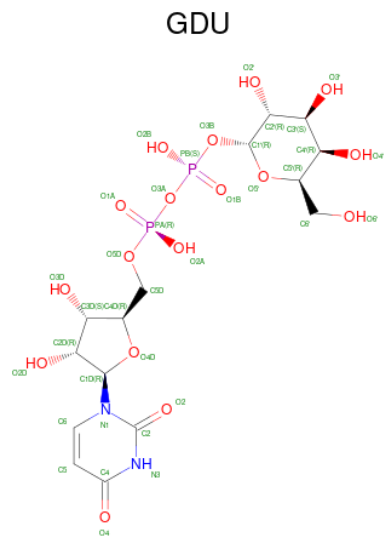
- Molecule 1 is a protein called Procollagen galactosyltransferase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	537	Total	C	N	O	S	0	0	0
			4385	2798	772	790	25			
1	B	521	Total	C	N	O	S	0	0	0
			4271	2732	757	757	25			

There are 10 discrepancies between the modelled and reference sequences:

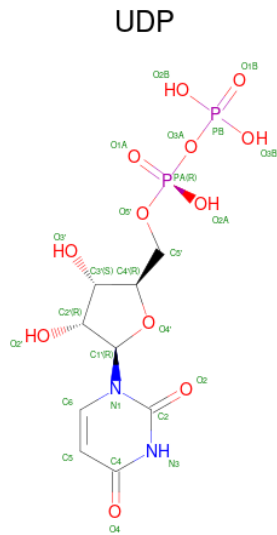
Chain	Residue	Modelled	Actual	Comment	Reference
A	28	GLY	-	expression tag	UNP Q8NBJ5
A	29	SER	-	expression tag	UNP Q8NBJ5
A	619	ALA	-	expression tag	UNP Q8NBJ5
A	620	ALA	-	expression tag	UNP Q8NBJ5
A	621	ALA	-	expression tag	UNP Q8NBJ5
B	28	GLY	-	expression tag	UNP Q8NBJ5
B	29	SER	-	expression tag	UNP Q8NBJ5
B	619	ALA	-	expression tag	UNP Q8NBJ5
B	620	ALA	-	expression tag	UNP Q8NBJ5
B	621	ALA	-	expression tag	UNP Q8NBJ5

- Molecule 2 is GALACTOSE-URIDINE-5'-DIPHOSPHATE (CCD ID: GDU) (formula: C<sub>15</sub>H<sub>24</sub>N<sub>2</sub>O<sub>17</sub>P<sub>2</sub>) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 36	C 15	N 2	O 17	P 2	0	0
2	B	1	Total 36	C 15	N 2	O 17	P 2	0	0

- Molecule 3 is URIDINE-5'-DIPHOSPHATE (CCD ID: UDP) (formula:  $\text{C}_9\text{H}_{14}\text{N}_2\text{O}_{12}\text{P}_2$ ) (labeled as "Ligand of Interest" by depositor).

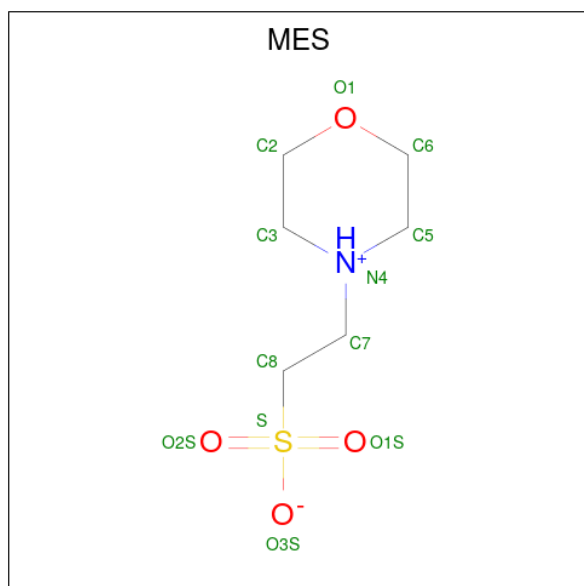


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total 25	C 9	N 2	O 12	P 2	0	0

- Molecule 4 is MANGANESE (II) ION (CCD ID: MN) (formula: Mn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Mn	0	0
			1	1		

- Molecule 5 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (CCD ID: MES) (formula: C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
5	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 6 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	3	Total	Cl	0	0
			3	3		
6	B	3	Total	Cl	0	0
			3	3		

- Molecule 7 is CALCIUM ION (CCD ID: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	1	Total	Ca	0	0
			1	1		
7	B	1	Total	Ca	0	0
			1	1		

- Molecule 8 is SODIUM ION (CCD ID: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	B	1	Total	Na	0	0
			1	1		

- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	66	Total	O	0	0
			66	66		
9	B	70	Total	O	0	0
			70	70		



- Molecule 1: Procollagen galactosyltransferase 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	220.85Å 220.85Å 220.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.38 – 2.70 49.38 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.8 (49.38-2.70) 99.9 (49.38-2.70)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.41 (at 2.69Å)	Xtriage
Refinement program	PHENIX (1.20_4459: ???)	Depositor
R, $R_{free}$	0.185 , 0.223 0.182 , 0.219	Depositor DCC
$R_{free}$ test set	2449 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	60.1	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 47.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.023 for -l,-k,-h	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	8923	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.84% 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: CL, MES, UDP, MN, CA, NA, GDU

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	0.26	0/4504	0.50	0/6115
1	B	0.25	0/4389	0.50	0/5956
All	All	0.25	0/8893	0.50	0/12071

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	4385	0	4321	32	0
1	B	4271	0	4232	33	0
2	A	36	0	21	3	0
2	B	36	0	21	1	0
3	A	25	0	11	0	0
4	A	1	0	0	0	0
5	A	24	0	24	1	0
6	A	3	0	0	0	0
6	B	3	0	0	0	0
7	A	1	0	0	0	0
7	B	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	B	1	0	0	0	0
9	A	66	0	0	3	0
9	B	70	0	0	2	0
All	All	8923	0	8630	66	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 4.

The worst 5 of 66 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:B:139:ARG:HE	2:B:701:GDU:H5	1.32	0.92
1:A:139:ARG:HE	2:A:701:GDU:H5	1.36	0.91
1:B:104:TRP:HB2	1:B:328:ILE:HD11	1.64	0.80
1:A:390:ILE:HD13	1:A:410:LEU:HD11	1.79	0.65
1:B:304:GLU:OE2	9:B:801:HOH:O	2.16	0.61

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	535/594 (90%)	518 (97%)	17 (3%)	0	100	100
1	B	519/594 (87%)	502 (97%)	16 (3%)	1 (0%)	44	68
All	All	1054/1188 (89%)	1020 (97%)	33 (3%)	1 (0%)	48	73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	236	SER

### 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	474/518 (92%)	474 (100%)	0	100	100
1	B	460/518 (89%)	455 (99%)	5 (1%)	70	87
All	All	934/1036 (90%)	929 (100%)	5 (0%)	86	95

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	102	ARG
1	B	203	CYS
1	B	348	LEU
1	B	405	LEU
1	B	438	LEU

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	578	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](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

## 5.6 Ligand geometry

Of 15 ligands modelled in this entry, 10 are monoatomic - leaving 5 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

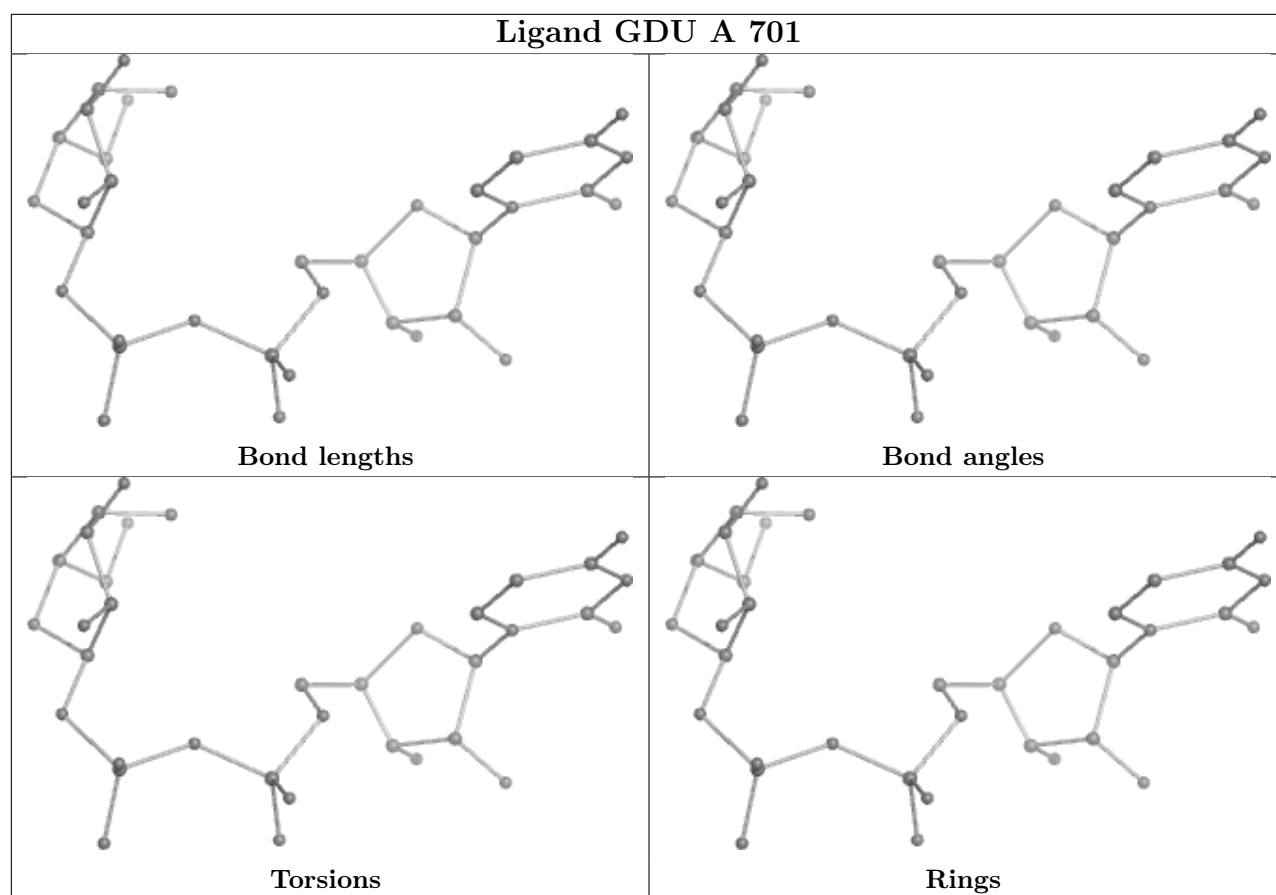
There are no chirality outliers.

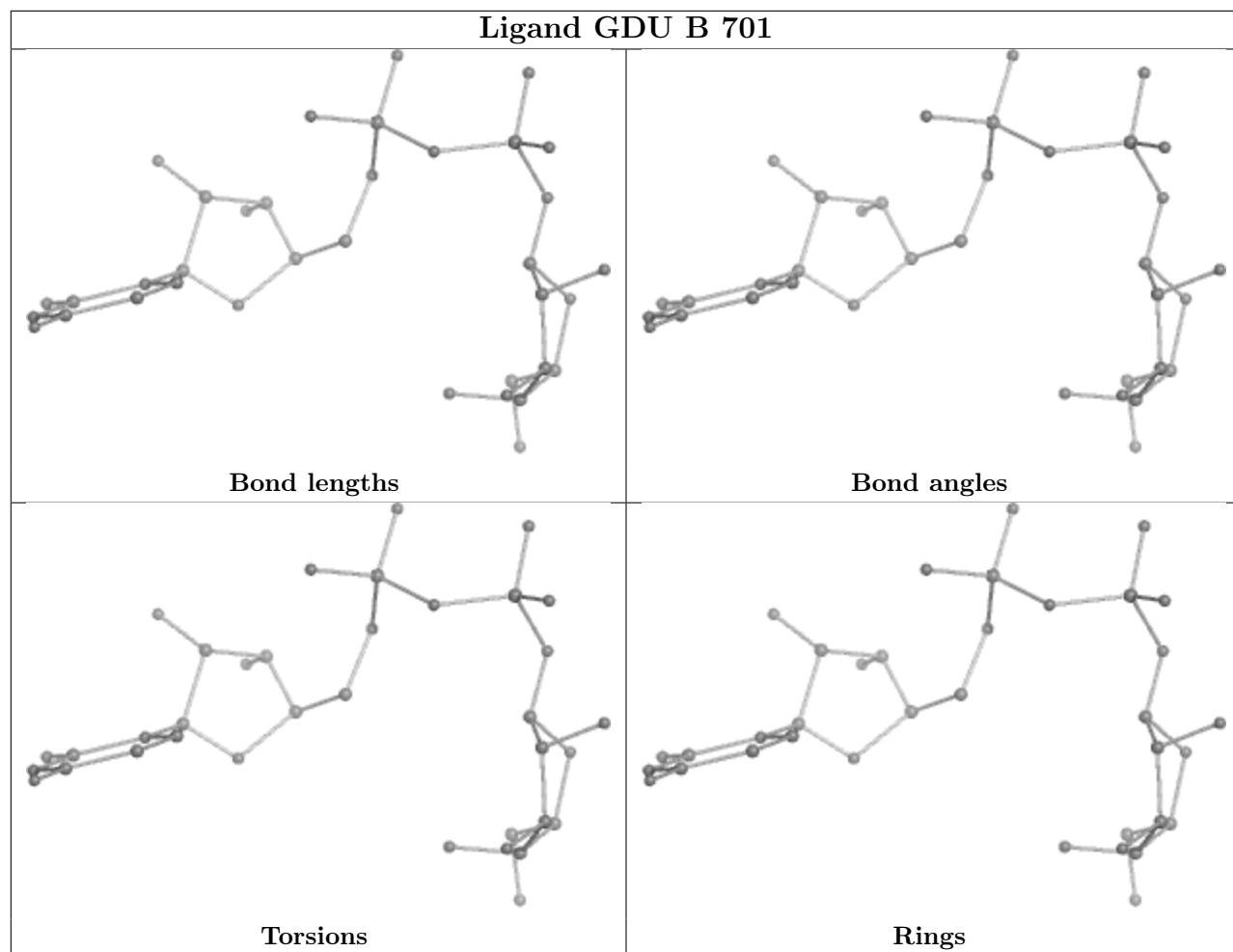
There are no torsion outliers.

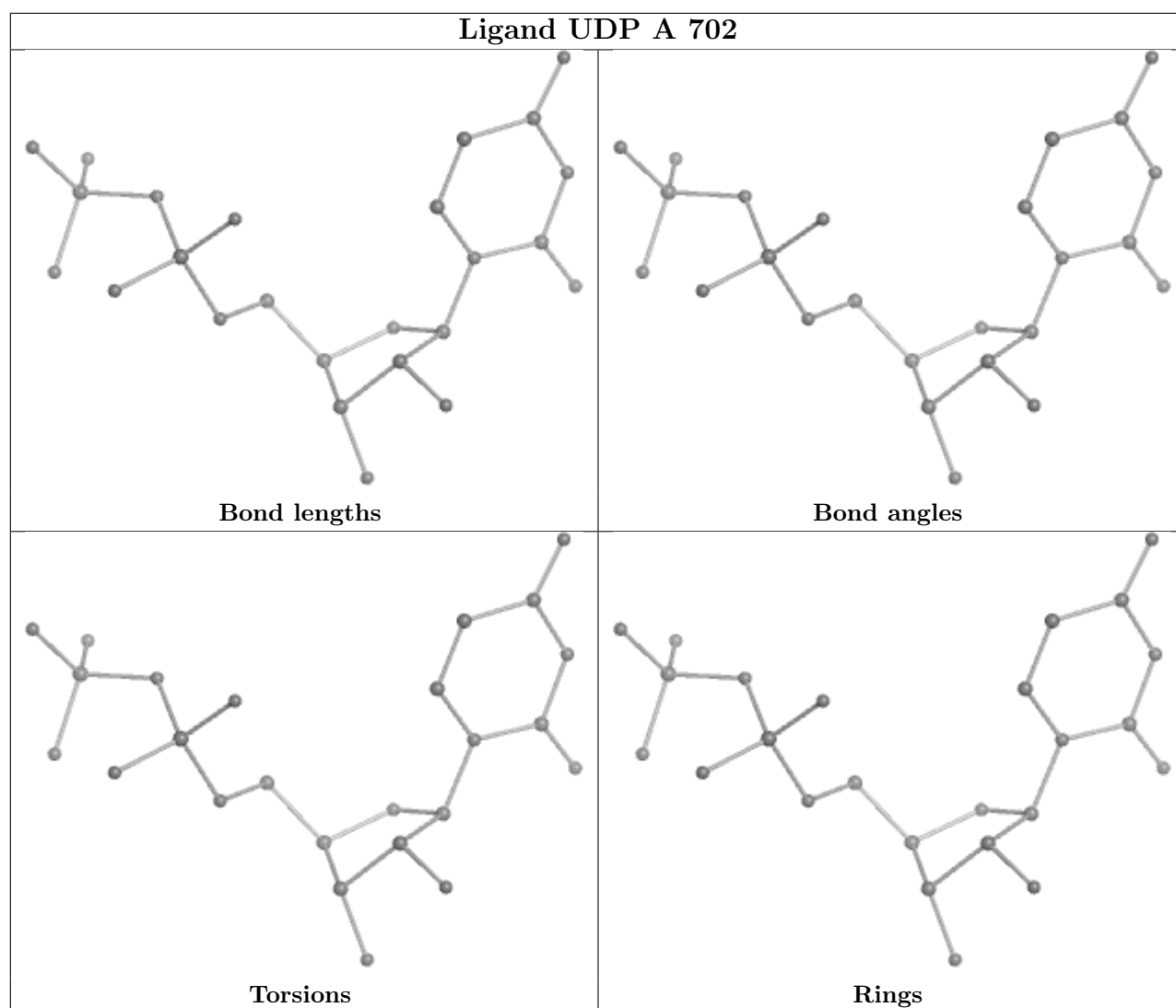
There are no ring outliers.

No monomer is involved in short contacts.

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 than 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<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	537/594 (90%)	-0.29	10 (1%) 66 65	41, 70, 131, 173	0
1	B	521/594 (87%)	-0.38	3 (0%) 85 85	38, 65, 118, 153	0
All	All	1058/1188 (89%)	-0.33	13 (1%) 76 76	38, 68, 128, 173	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	561	TYR	3.7
1	A	44	SER	3.4
1	B	562	THR	3.3
1	A	494	TYR	3.2
1	A	576	VAL	2.8

### 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.

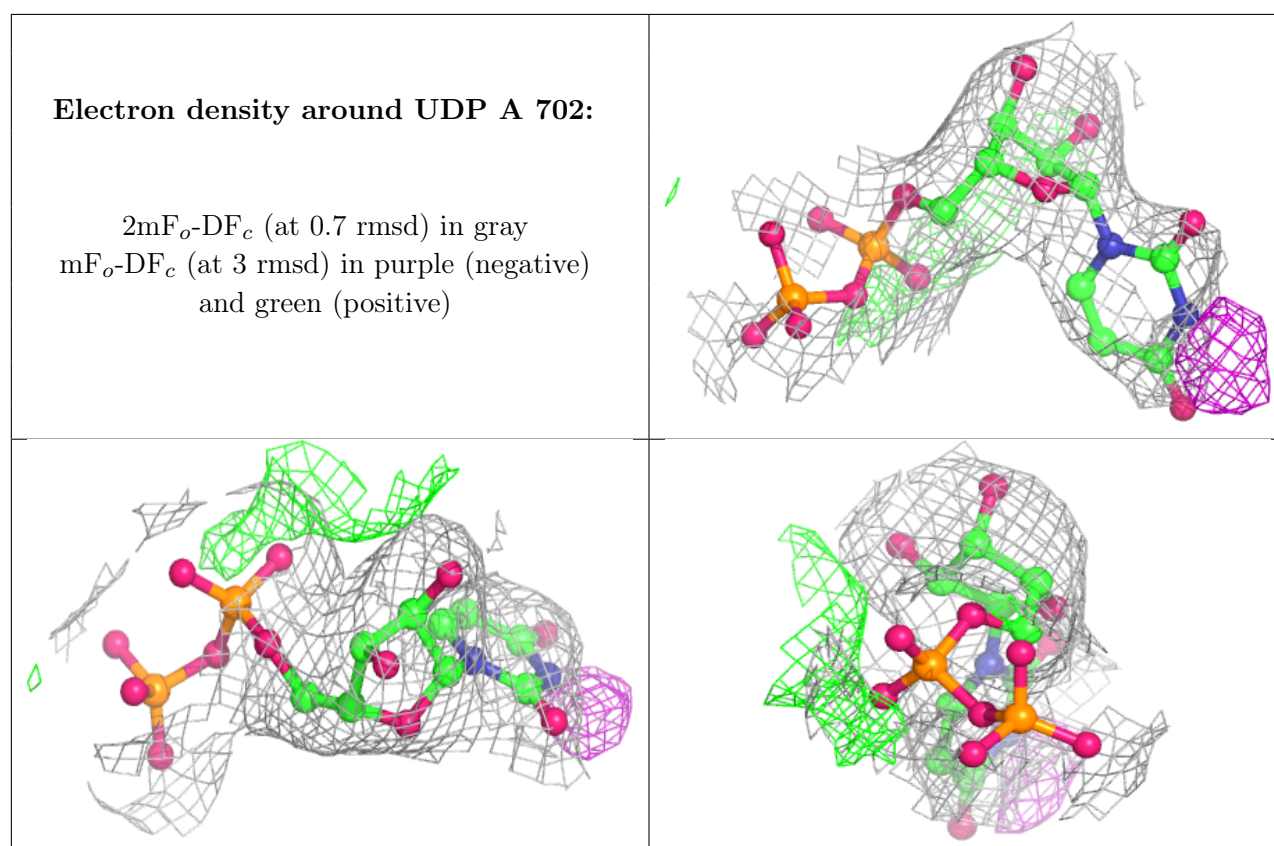
### 6.4 Ligands [i](#)

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.



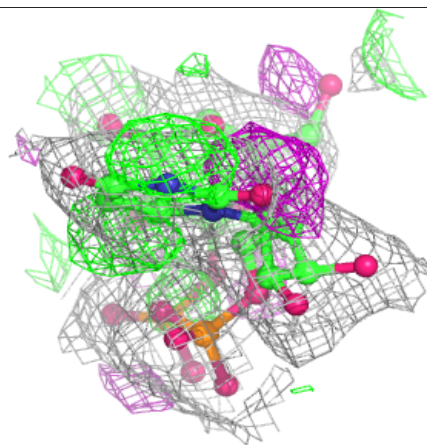
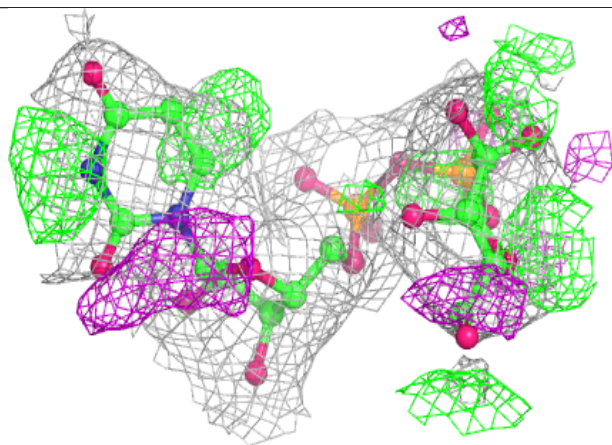
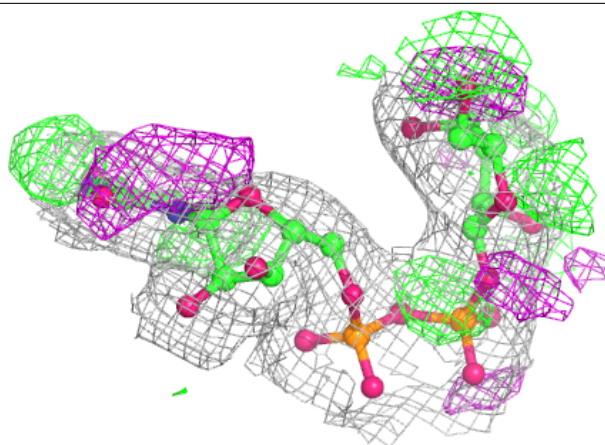
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
6	CL	B	704	1/1	0.73	0.15	118,118,118,118	0
5	MES	A	705	12/12	0.83	0.18	85,106,129,229	0
6	CL	B	702	1/1	0.88	0.20	74,74,74,74	0
8	NA	B	705	1/1	0.88	0.20	46,46,46,46	0
3	UDP	A	702	25/25	0.89	0.08	89,114,122,129	0
2	GDU	B	701	36/36	0.92	0.11	33,50,71,89	0
5	MES	A	704	12/12	0.93	0.11	55,63,83,85	0
2	GDU	A	701	36/36	0.93	0.11	28,52,78,99	0
6	CL	A	706	1/1	0.93	0.18	89,89,89,89	0
6	CL	A	707	1/1	0.95	0.09	93,93,93,93	0
7	CA	B	706	1/1	0.96	0.04	62,62,62,62	0
4	MN	A	703	1/1	0.97	0.05	92,92,92,92	0
6	CL	B	703	1/1	0.98	0.14	72,72,72,72	0
6	CL	A	708	1/1	0.98	0.09	74,74,74,74	0
7	CA	A	709	1/1	0.99	0.04	49,49,49,49	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.



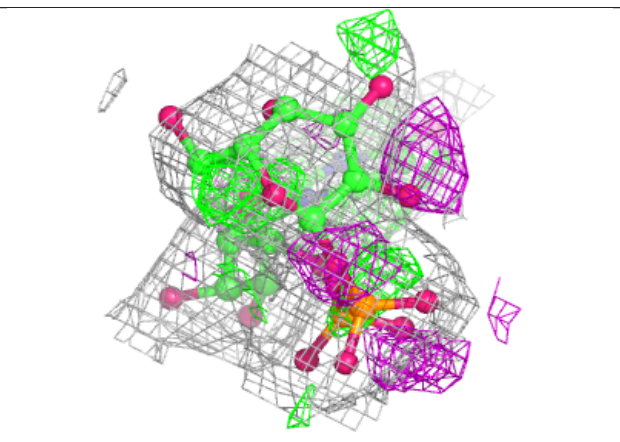
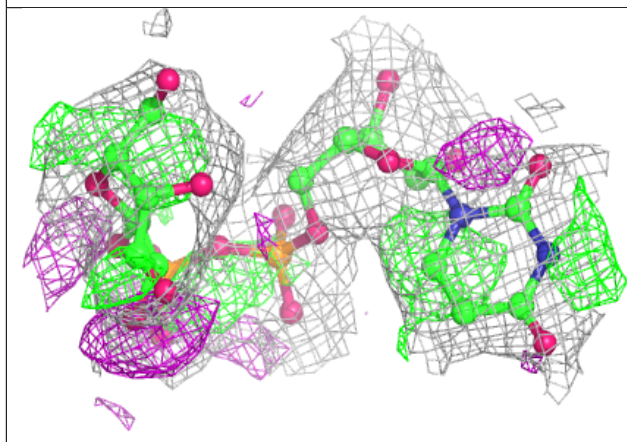
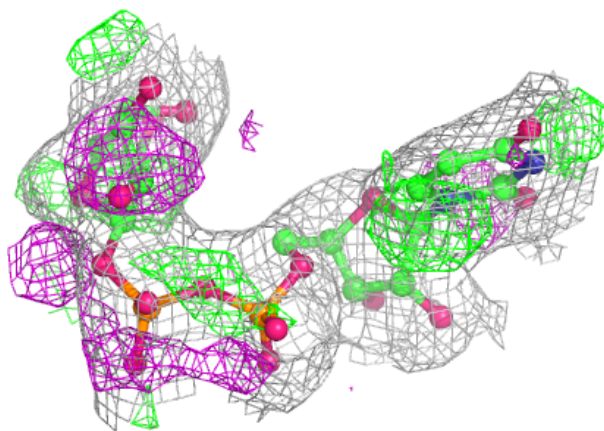
**Electron density around GDU B 701:**

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



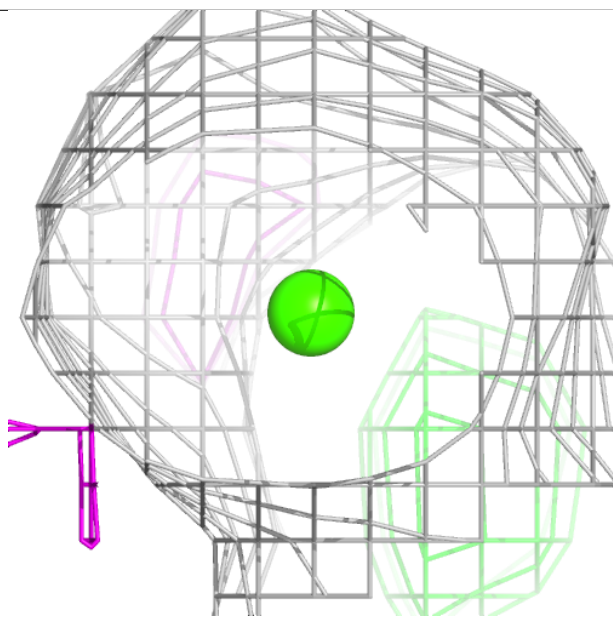
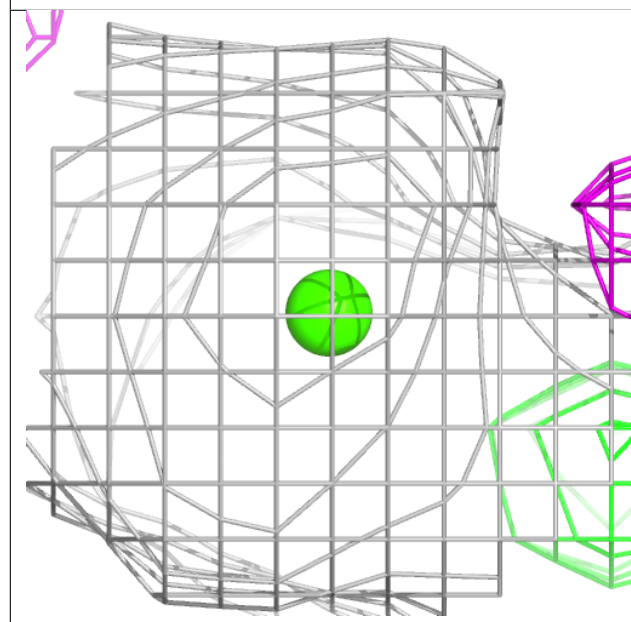
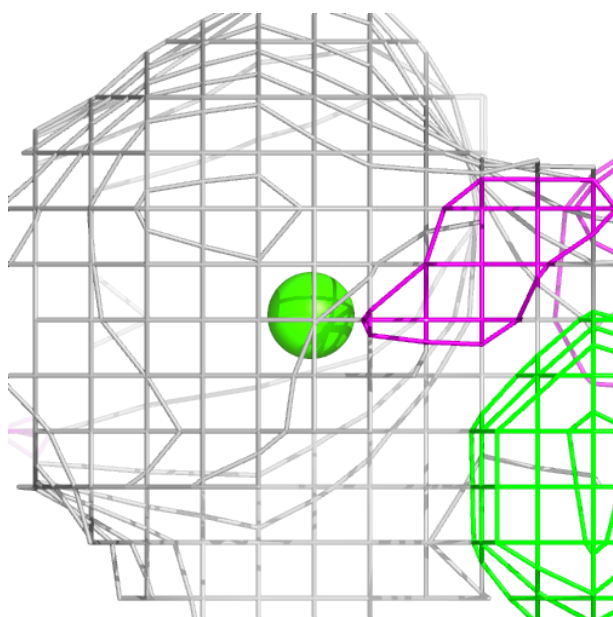
**Electron density around GDU A 701:**

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



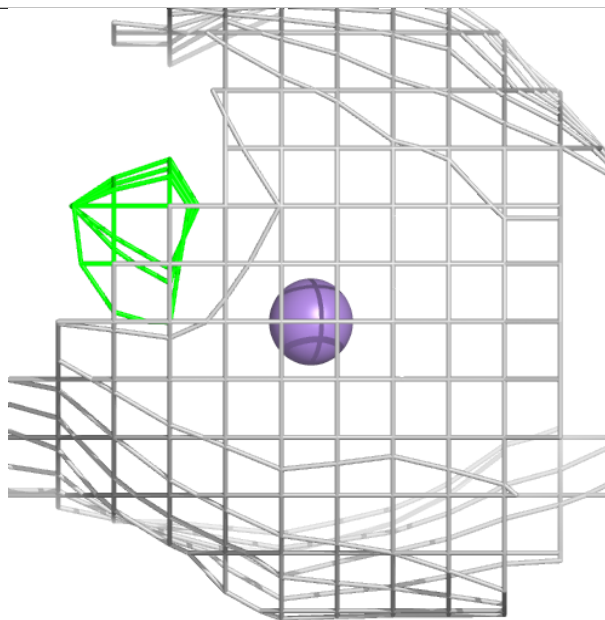
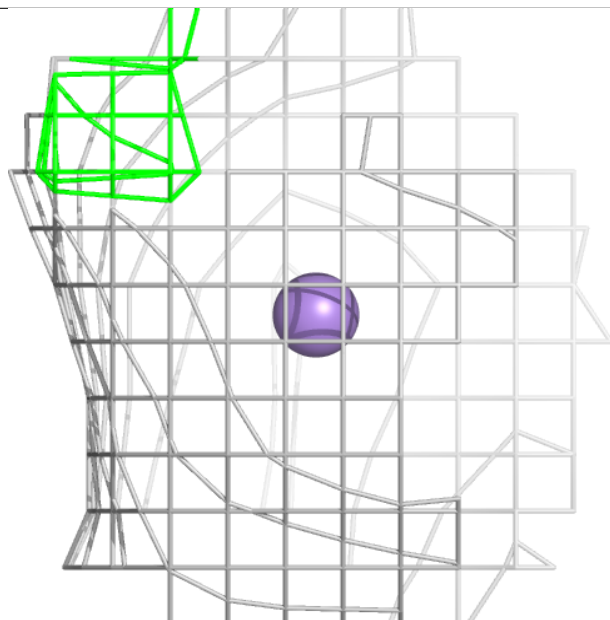
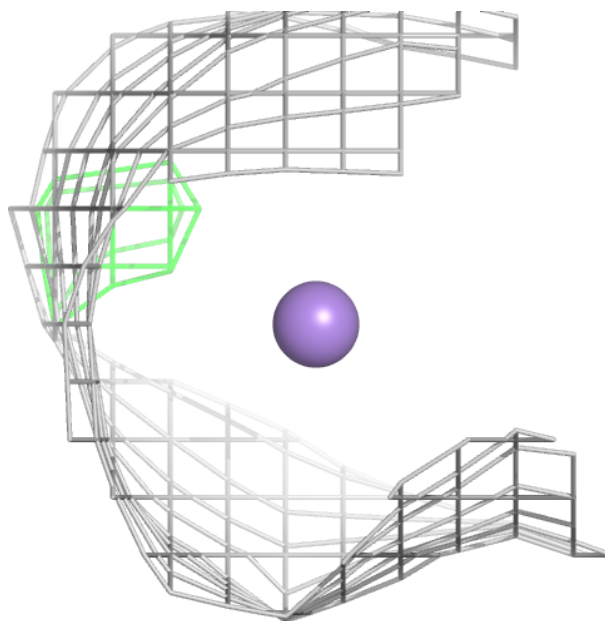
**Electron density around CA B 706:**

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

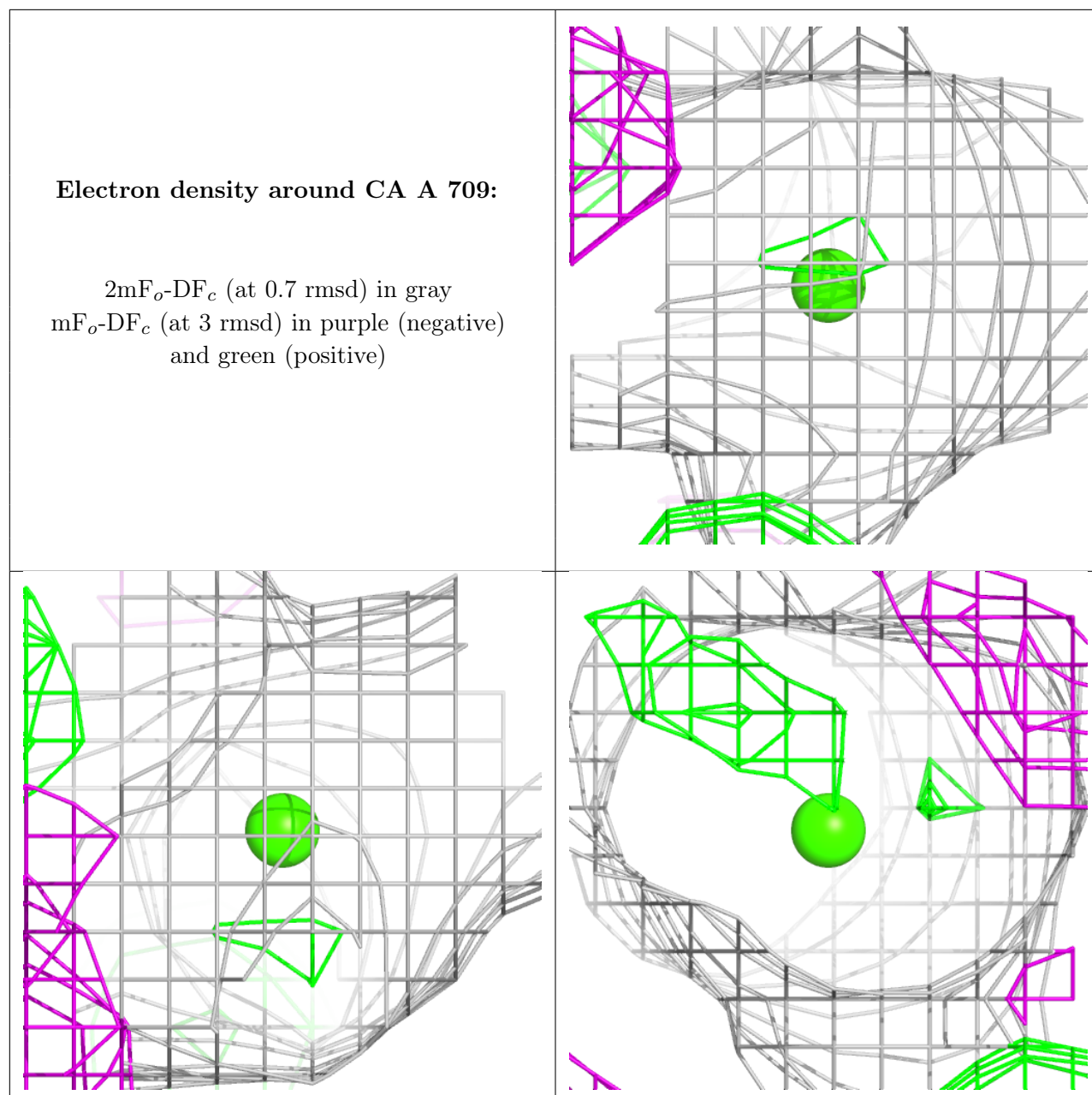


**Electron density around MN A 703:**

$2mF_o-DF_c$  (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.