wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2023 – 05:45 pm GMT

PDB ID	:	4CJ8
Title	:	monoclinic crystal form of Bogt6a E192Q in complex with UDP-GalNAc, UDP and GalNAc
Authors	:	Pham, T.; Stinson, B.; Thiyagarajan, N.; Lizotte-Waniewski, M.; Brew, K.; Acharya, K.R.
Deposited on	:	2013-12-19
Resolution	:	3.50 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 (i)) were used in the production of this report:

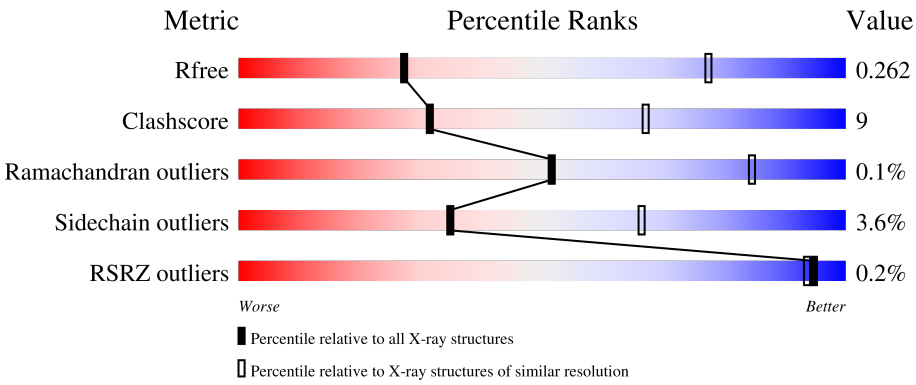
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.50 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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	Quality of chain
1	A	248	73% 22% • •
1	B	248	76% 19% • •
1	C	248	72% 23% •
1	D	248	75% 19% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	E	248	77% 21%
1	F	248	79% 18%
1	G	248	78% 19%
1	H	248	79% 21%
1	I	248	77% 17% 5%
1	J	248	72% 20% 5%
1	K	248	73% 21%
1	L	248	72% 21% 5%
1	M	248	77% 21%
1	N	248	75% 20%
1	O	248	2% 71% 23%
1	P	248	82% 15%

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
3	UDP	N	1244	-	-	X	-

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 32877 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 GLYCOSYLTRANSFERASE FAMILY 6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	238	1989	1300	324	358	7	10	0	0
1	B	238	1989	1300	324	358	7	10	0	0
1	C	238	1989	1300	324	358	7	10	0	0
1	D	237	1983	1297	323	356	7	10	0	0
1	E	246	2057	1342	341	367	7	10	0	0
1	F	246	2057	1342	341	367	7	10	0	0
1	G	246	2057	1342	341	367	7	10	0	0
1	H	247	2063	1345	342	369	7	10	0	0
1	I	236	1973	1291	320	355	7	10	0	0
1	J	236	1973	1291	320	355	7	10	0	0
1	K	237	1983	1297	323	356	7	10	0	0
1	L	236	1973	1291	320	355	7	10	0	0
1	M	245	2047	1336	338	366	7	10	0	0
1	N	243	2027	1324	332	364	7	10	0	0
1	O	240	2000	1306	326	361	7	10	0	0
1	P	243	2027	1324	332	364	7	10	0	0

There are 48 discrepancies between the modelled and reference sequences:

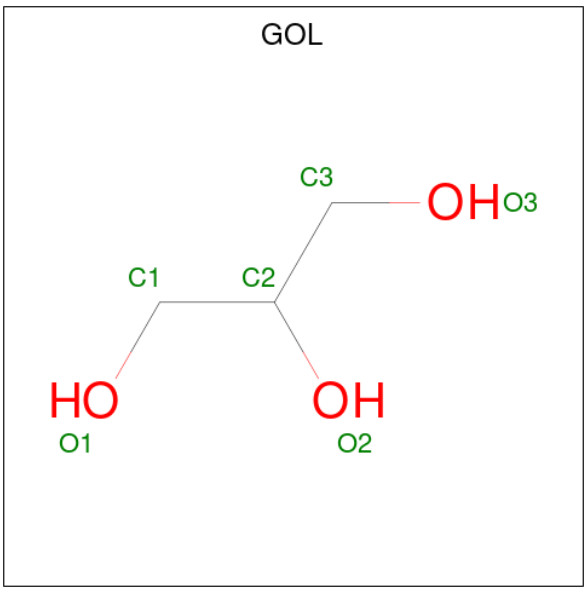
Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	SER	-	expression tag	UNP A7LVT2
A	0	HIS	-	expression tag	UNP A7LVT2
A	192	GLN	GLU	engineered mutation	UNP A7LVT2
B	-1	SER	-	expression tag	UNP A7LVT2
B	0	HIS	-	expression tag	UNP A7LVT2
B	192	GLN	GLU	engineered mutation	UNP A7LVT2
C	-1	SER	-	expression tag	UNP A7LVT2
C	0	HIS	-	expression tag	UNP A7LVT2
C	192	GLN	GLU	engineered mutation	UNP A7LVT2
D	-1	SER	-	expression tag	UNP A7LVT2
D	0	HIS	-	expression tag	UNP A7LVT2
D	192	GLN	GLU	engineered mutation	UNP A7LVT2
E	-1	SER	-	expression tag	UNP A7LVT2
E	0	HIS	-	expression tag	UNP A7LVT2
E	192	GLN	GLU	engineered mutation	UNP A7LVT2
F	-1	SER	-	expression tag	UNP A7LVT2
F	0	HIS	-	expression tag	UNP A7LVT2
F	192	GLN	GLU	engineered mutation	UNP A7LVT2
G	-1	SER	-	expression tag	UNP A7LVT2
G	0	HIS	-	expression tag	UNP A7LVT2
G	192	GLN	GLU	engineered mutation	UNP A7LVT2
H	-1	SER	-	expression tag	UNP A7LVT2
H	0	HIS	-	expression tag	UNP A7LVT2
H	192	GLN	GLU	engineered mutation	UNP A7LVT2
I	-1	SER	-	expression tag	UNP A7LVT2
I	0	HIS	-	expression tag	UNP A7LVT2
I	192	GLN	GLU	engineered mutation	UNP A7LVT2
J	-1	SER	-	expression tag	UNP A7LVT2
J	0	HIS	-	expression tag	UNP A7LVT2
J	192	GLN	GLU	engineered mutation	UNP A7LVT2
K	-1	SER	-	expression tag	UNP A7LVT2
K	0	HIS	-	expression tag	UNP A7LVT2
K	192	GLN	GLU	engineered mutation	UNP A7LVT2
L	-1	SER	-	expression tag	UNP A7LVT2
L	0	HIS	-	expression tag	UNP A7LVT2
L	192	GLN	GLU	engineered mutation	UNP A7LVT2
M	-1	SER	-	expression tag	UNP A7LVT2
M	0	HIS	-	expression tag	UNP A7LVT2
M	192	GLN	GLU	engineered mutation	UNP A7LVT2
N	-1	SER	-	expression tag	UNP A7LVT2
N	0	HIS	-	expression tag	UNP A7LVT2
N	192	GLN	GLU	engineered mutation	UNP A7LVT2

Continued on next page...

Continued from previous page...

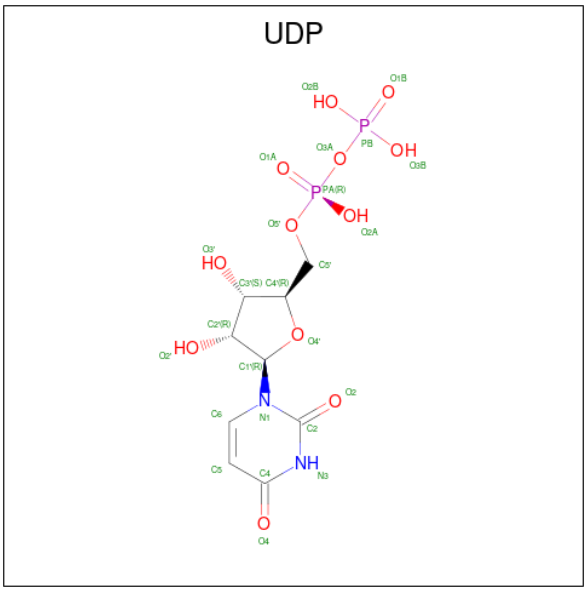
Chain	Residue	Modelled	Actual	Comment	Reference
O	-1	SER	-	expression tag	UNP A7LVT2
O	0	HIS	-	expression tag	UNP A7LVT2
O	192	GLN	GLU	engineered mutation	UNP A7LVT2
P	-1	SER	-	expression tag	UNP A7LVT2
P	0	HIS	-	expression tag	UNP A7LVT2
P	192	GLN	GLU	engineered mutation	UNP A7LVT2

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



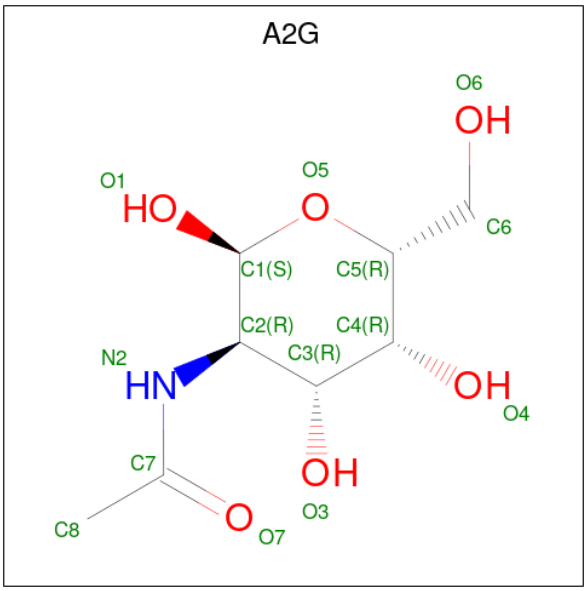
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O			
2	A	1	6	3	3		0	0
			Total	C	O			
2	A	1	6	3	3		0	0
			Total	C	O			
2	D	1	6	3	3		0	0
			Total	C	O			
2	G	1	6	3	3		0	0
			Total	C	O			
2	G	1	6	3	3		0	0

- Molecule 3 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula: $C_9H_{14}N_2O_{12}P_2$).



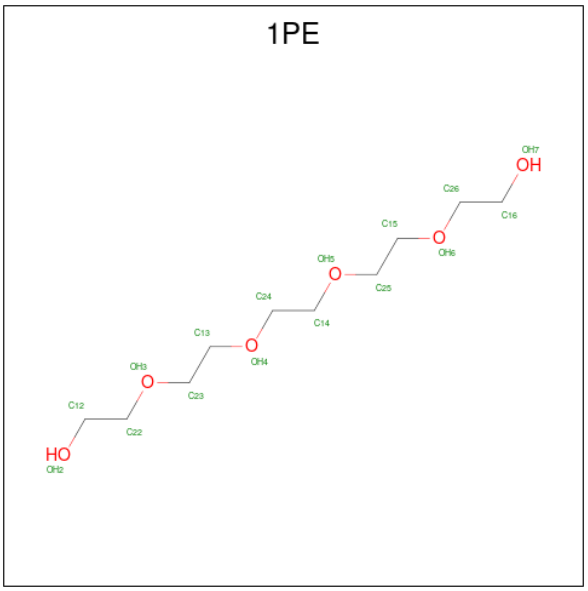
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	B	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	C	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	D	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	I	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	J	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	K	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	L	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	M	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	N	1	Total	C	N	O	P	0	0
			25	9	2	12	2		

- Molecule 4 is 2-acetamido-2-deoxy-alpha-D-galactopyranose (three-letter code: A2G) (formula: $C_8H_{15}NO_6$).



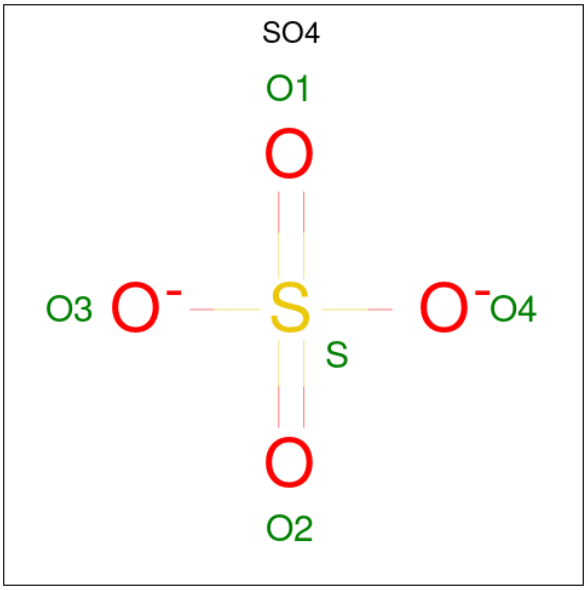
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
4	A	1	15	8	1	6	0	0
4	B	1	15	8	1	6	0	0
4	C	1	15	8	1	6	0	0
4	D	1	15	8	1	6	0	0
4	I	1	15	8	1	6	0	0
4	J	1	15	8	1	6	0	0
4	K	1	15	8	1	6	0	0
4	L	1	15	8	1	6	0	0

- Molecule 5 is PENTAETHYLENE GLYCOL (three-letter code: 1PE) (formula: $C_{10}H_{22}O_6$).



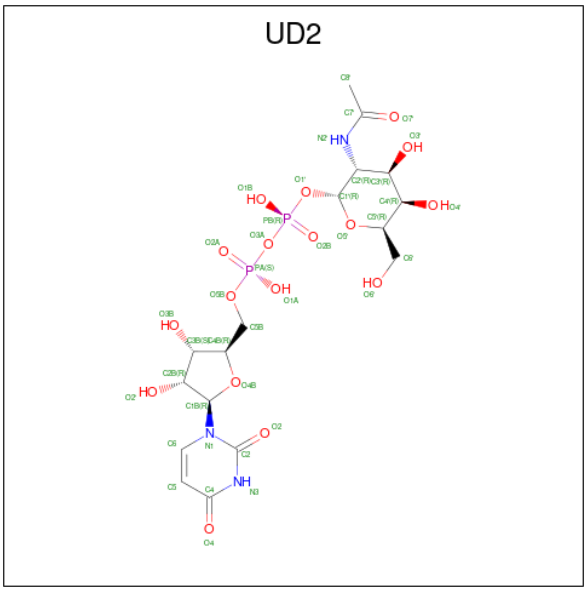
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
5	A	1	16	10	6	0	0

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



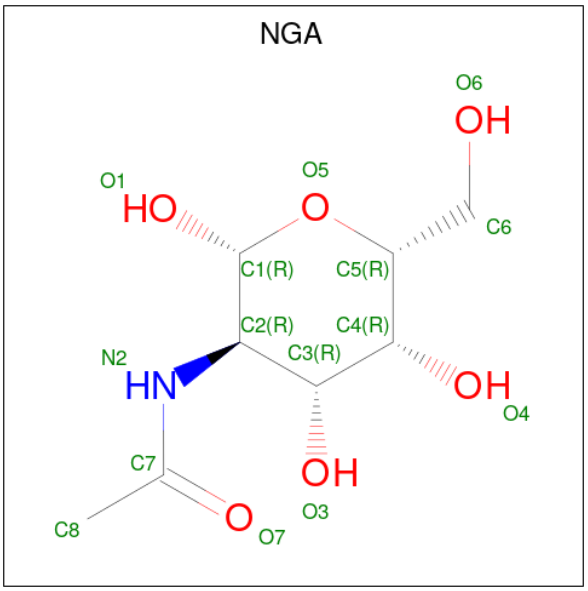
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
6	B	1	5	4	1	0	0

- Molecule 7 is URIDINE-DIPHOSPHATE-N-ACETYLGALACTOSAMINE (three-letter code: UD2) (formula: $C_{17}H_{27}N_3O_{17}P_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	E	1	Total	C	N	O	P	0	0
			39	17	3	17	2		
7	F	1	Total	C	N	O	P	0	0
			39	17	3	17	2		
7	G	1	Total	C	N	O	P	0	0
			39	17	3	17	2		
7	H	1	Total	C	N	O	P	0	0
			39	17	3	17	2		
7	O	1	Total	C	N	O	P	0	0
			39	17	3	17	2		
7	P	1	Total	C	N	O	P	0	0
			39	17	3	17	2		

- Molecule 8 is 2-acetamido-2-deoxy-beta-D-galactopyranose (three-letter code: NGA) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
8	M	1	14	8	1	5	0	0
			Total	C	N	O		
8	N	1	14	8	1	5	0	0

- Molecule 9 is water.

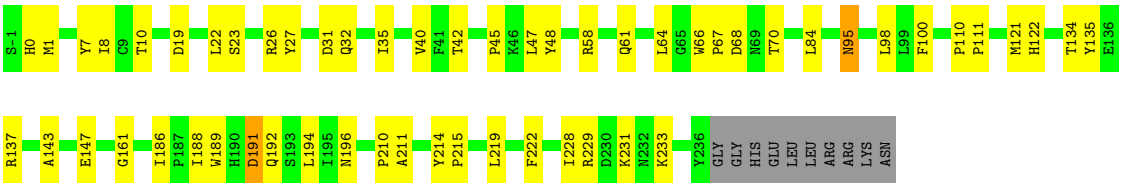
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
9	B	2	2	2	0	0
			Total	O		
9	C	1	1	1	0	0
			Total	O		
9	D	1	1	1	0	0
			Total	O		
9	E	1	1	1	0	0
			Total	O		
9	L	2	2	2	0	0

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.

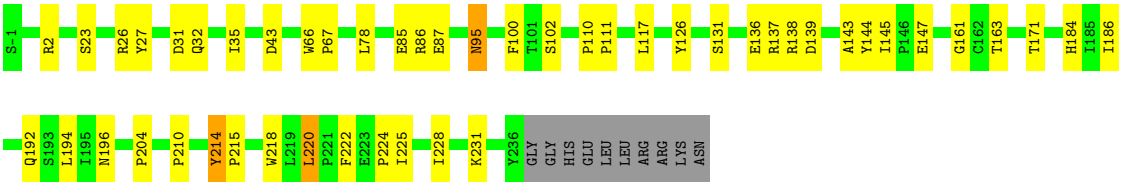
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain A: 73% 22%



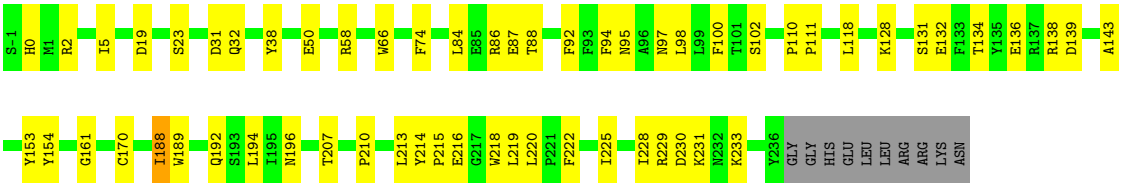
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain B: 76% 19%



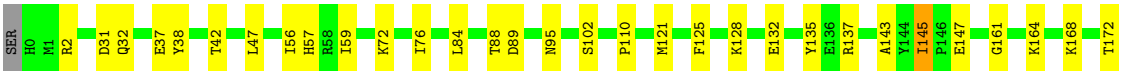
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

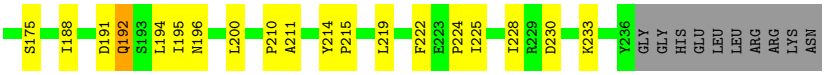
Chain C: 72% 23%



• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

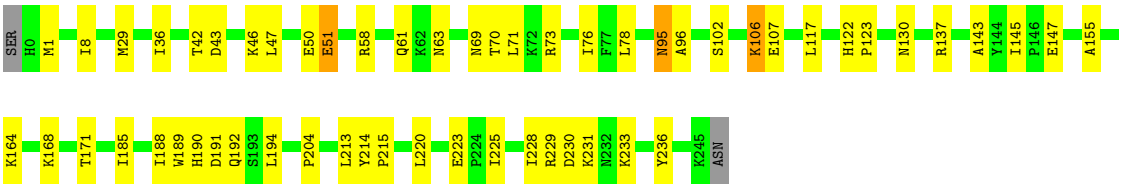
Chain D: 75% 19%





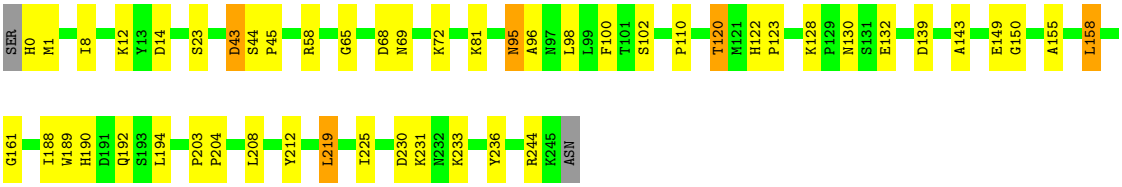
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain E:



• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain F:

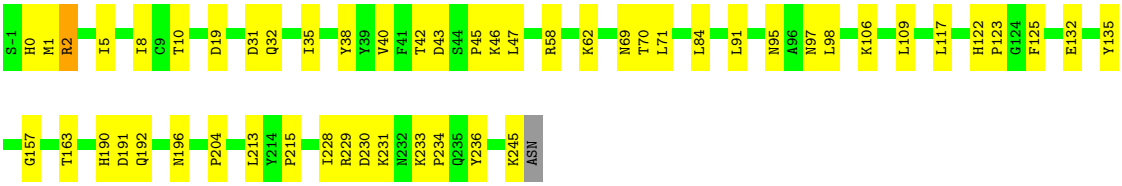


• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain G:

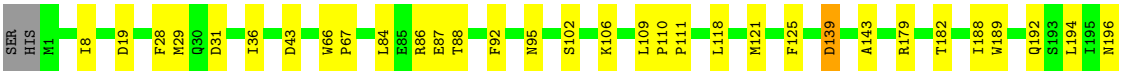
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

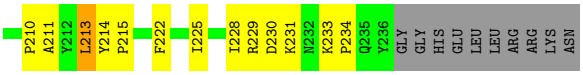
Chain H:



• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

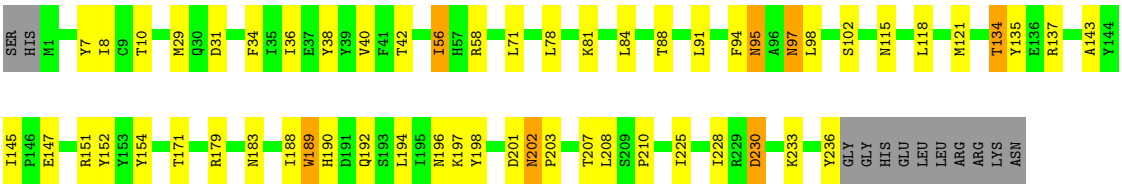
Chain I:





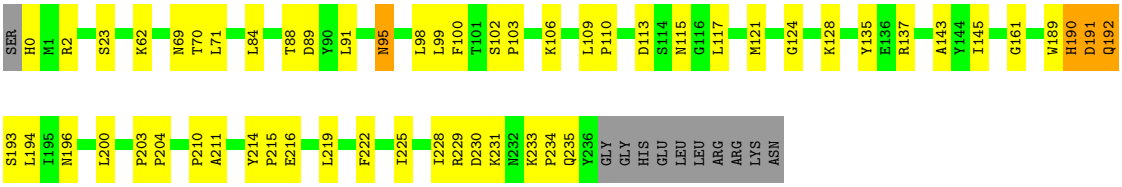
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain J: 72% 20% • 5%



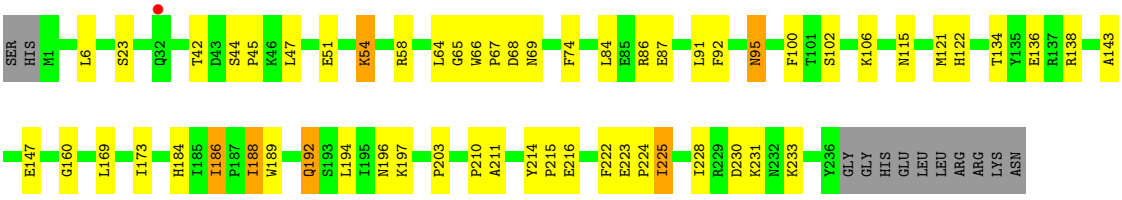
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain K: 73% 21% • •



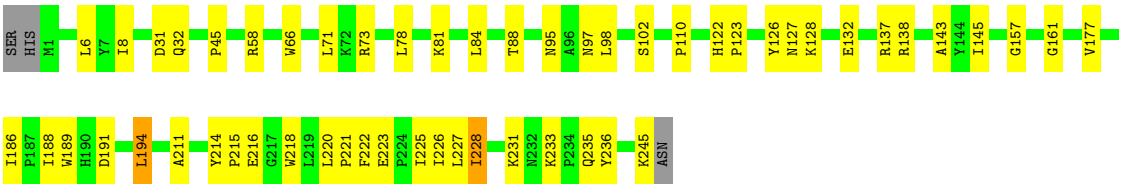
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain L: 72% 21% • 5%



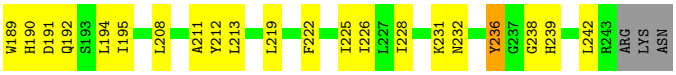
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain M: 77% 21% ••



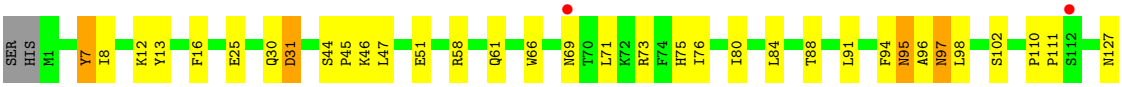
• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

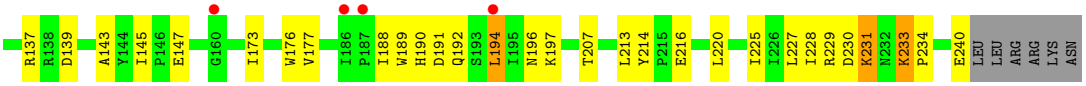
Chain N: 75% 20% • •



• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

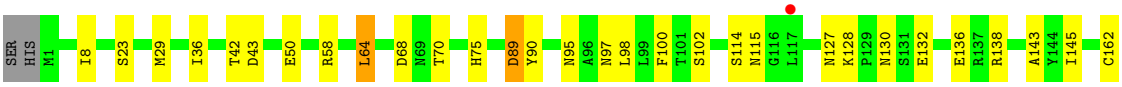
Chain O:

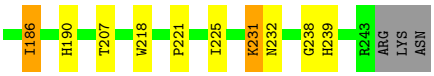




• Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

Chain P:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	176.69Å 79.74Å 179.08Å 90.00° 95.23° 90.00°	Depositor
Resolution (Å)	87.98 – 3.50 87.98 – 3.50	Depositor EDS
% Data completeness (in resolution range)	97.7 (87.98-3.50) 87.0 (87.98-3.50)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.01 (at 3.49Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.4_1496)	Depositor
R, R_{free}	0.225 , 0.249 0.236 , 0.262	Depositor DCC
R_{free} test set	3137 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	60.6	Xtriage
Anisotropy	0.241	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 18.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.086 for l,-k,h	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	32877	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 26.29 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.7089e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1]Intensities estimated from amplitudes.

[2]Theoretical 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.

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: NGA, SO4, 1PE, UDP, GOL, UD2, A2G

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	0.21	0/2055	0.35	0/2791
1	B	0.21	0/2055	0.35	0/2791
1	C	0.21	0/2055	0.35	0/2791
1	D	0.21	0/2049	0.36	0/2783
1	E	0.21	0/2124	0.36	0/2881
1	F	0.21	0/2124	0.37	0/2881
1	G	0.21	0/2124	0.36	0/2881
1	H	0.21	0/2130	0.35	0/2889
1	I	0.21	0/2038	0.36	0/2768
1	J	0.21	0/2038	0.36	0/2768
1	K	0.21	0/2049	0.35	0/2783
1	L	0.21	0/2038	0.35	0/2768
1	M	0.21	0/2113	0.36	0/2866
1	N	0.21	0/2093	0.35	0/2841
1	O	0.25	0/2066	0.37	0/2805
1	P	0.21	0/2093	0.36	0/2841
All	All	0.21	0/33244	0.36	0/45128

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	1989	0	1926	36	0
1	B	1989	0	1926	34	0
1	C	1989	0	1926	33	0
1	D	1983	0	1921	28	0
1	E	2057	0	2001	31	0
1	F	2057	0	2001	29	0
1	G	2057	0	2001	37	0
1	H	2063	0	2006	32	0
1	I	1973	0	1914	28	0
1	J	1973	0	1914	38	0
1	K	1983	0	1921	40	0
1	L	1973	0	1914	35	1
1	M	2047	0	1992	33	0
1	N	2027	0	1966	44	0
1	O	2000	0	1933	64	0
1	P	2027	0	1968	34	0
2	A	12	0	16	0	0
2	D	6	0	8	0	0
2	G	12	0	16	1	0
3	A	25	0	11	4	0
3	B	25	0	11	5	0
3	C	25	0	11	3	1
3	D	25	0	11	6	0
3	I	25	0	11	2	0
3	J	25	0	11	2	0
3	K	25	0	11	4	0
3	L	25	0	11	2	0
3	M	25	0	11	4	0
3	N	25	0	11	18	0
4	A	15	0	12	1	0
4	B	15	0	12	0	0
4	C	15	0	12	0	0
4	D	15	0	12	0	0
4	I	15	0	12	0	0
4	J	15	0	12	0	0
4	K	15	0	12	2	0
4	L	15	0	12	5	0
5	A	16	0	22	1	0
6	B	5	0	0	0	0
7	E	39	0	25	5	0
7	F	39	0	25	5	0
7	G	39	0	25	7	0
7	H	39	0	25	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	O	39	0	25	20	0
7	P	39	0	25	3	0
8	M	14	0	13	2	0
8	N	14	0	13	2	0
9	B	2	0	0	0	0
9	C	1	0	0	0	0
9	D	1	0	0	0	0
9	E	1	0	0	0	0
9	L	2	0	0	0	0
All	All	32877	0	31674	575	1

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 9.

The worst 5 of 575 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:N:95:ASN:ND2	3:N:1244:UDP:H5'2	1.74	1.02
1:D:191:ASP:OD1	1:D:192:GLN:HG3	1.63	0.98
1:C:231:LYS:NZ	3:C:1237:UDP:O2A	1.96	0.97
1:G:95:ASN:ND2	1:G:96:ALA:H	1.63	0.96
1:F:95:ASN:HD22	1:F:96:ALA:H	0.99	0.93

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-2	Interatomic distance (Å)	Clash overlap (Å)
1:L:184:HIS:CE1	3:C:1237:UDP:O3B[1_654]	2.14	0.06

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	236/248 (95%)	219 (93%)	17 (7%)	0	100	100
1	B	236/248 (95%)	221 (94%)	15 (6%)	0	100	100
1	C	236/248 (95%)	219 (93%)	17 (7%)	0	100	100
1	D	235/248 (95%)	220 (94%)	14 (6%)	1 (0%)	34	72
1	E	244/248 (98%)	231 (95%)	13 (5%)	0	100	100
1	F	244/248 (98%)	227 (93%)	17 (7%)	0	100	100
1	G	244/248 (98%)	231 (95%)	12 (5%)	1 (0%)	34	72
1	H	245/248 (99%)	230 (94%)	15 (6%)	0	100	100
1	I	234/248 (94%)	221 (94%)	13 (6%)	0	100	100
1	J	234/248 (94%)	218 (93%)	16 (7%)	0	100	100
1	K	235/248 (95%)	216 (92%)	19 (8%)	0	100	100
1	L	234/248 (94%)	218 (93%)	16 (7%)	0	100	100
1	M	243/248 (98%)	233 (96%)	10 (4%)	0	100	100
1	N	241/248 (97%)	227 (94%)	13 (5%)	1 (0%)	34	72
1	O	238/248 (96%)	222 (93%)	15 (6%)	1 (0%)	34	72
1	P	241/248 (97%)	224 (93%)	17 (7%)	0	100	100
All	All	3820/3968 (96%)	3577 (94%)	239 (6%)	4 (0%)	51	84

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	237	GLY
1	N	146	PRO
1	D	224	PRO
1	O	45	PRO

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	215/223 (96%)	209 (97%)	6 (3%)	43	72

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	215/223 (96%)	211 (98%)	4 (2%)	57	80
1	C	215/223 (96%)	209 (97%)	6 (3%)	43	72
1	D	214/223 (96%)	209 (98%)	5 (2%)	50	77
1	E	221/223 (99%)	212 (96%)	9 (4%)	30	63
1	F	221/223 (99%)	209 (95%)	12 (5%)	22	55
1	G	221/223 (99%)	213 (96%)	8 (4%)	35	66
1	H	222/223 (100%)	217 (98%)	5 (2%)	50	77
1	I	213/223 (96%)	208 (98%)	5 (2%)	50	77
1	J	213/223 (96%)	205 (96%)	8 (4%)	33	65
1	K	214/223 (96%)	206 (96%)	8 (4%)	34	65
1	L	213/223 (96%)	203 (95%)	10 (5%)	26	60
1	M	220/223 (99%)	211 (96%)	9 (4%)	30	63
1	N	218/223 (98%)	208 (95%)	10 (5%)	27	61
1	O	215/223 (96%)	205 (95%)	10 (5%)	26	60
1	P	218/223 (98%)	209 (96%)	9 (4%)	30	63
All	All	3468/3568 (97%)	3344 (96%)	124 (4%)	35	66

5 of 124 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	125	PHE
1	O	97	ASN
1	K	95	ASN
1	O	95	ASN
1	P	89	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	J	95	ASN
1	K	192	GLN
1	K	95	ASN
1	L	95	ASN
1	E	239	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

33 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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	A2G	A	1238	-	15,15,15	0.40	0	21,21,21	0.82	1 (4%)
4	A2G	K	1238	-	15,15,15	0.41	0	21,21,21	0.46	0
3	UDP	J	1237	-	24,26,26	1.11	2 (8%)	37,40,40	1.63	7 (18%)
8	NGA	M	1247	1	14,14,15	0.55	0	17,19,21	0.65	0
2	GOL	A	1239	-	5,5,5	0.37	0	5,5,5	0.27	0
4	A2G	I	1238	-	15,15,15	0.41	0	21,21,21	0.55	0
2	GOL	G	1248	-	5,5,5	0.37	0	5,5,5	0.26	0
3	UDP	D	1237	-	24,26,26	1.11	2 (8%)	37,40,40	1.72	7 (18%)
6	SO4	B	1239	-	4,4,4	0.14	0	6,6,6	0.05	0
7	UD2	F	1246	-	38,41,41	0.88	2 (5%)	57,62,62	1.37	7 (12%)
7	UD2	H	1246	-	38,41,41	0.88	2 (5%)	57,62,62	1.39	8 (14%)
2	GOL	A	1236	-	5,5,5	0.38	0	5,5,5	0.24	0
3	UDP	B	1237	-	24,26,26	1.12	2 (8%)	37,40,40	1.66	7 (18%)
4	A2G	J	1238	-	15,15,15	0.41	0	21,21,21	0.47	0
5	1PE	A	1240	-	15,15,15	0.53	0	14,14,14	0.46	0
3	UDP	N	1244	-	24,26,26	1.12	2 (8%)	37,40,40	1.72	7 (18%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	GOL	D	1239	-	5,5,5	0.37	0	5,5,5	0.27	0
4	A2G	B	1238	-	15,15,15	0.40	0	21,21,21	0.87	1 (4%)
3	UDP	A	1237	-	24,26,26	1.11	2 (8%)	37,40,40	1.71	6 (16%)
3	UDP	L	1237	-	24,26,26	1.11	2 (8%)	37,40,40	1.62	7 (18%)
3	UDP	I	1237	-	24,26,26	1.11	2 (8%)	37,40,40	1.65	7 (18%)
3	UDP	M	1246	-	24,26,26	1.12	3 (12%)	37,40,40	1.68	7 (18%)
7	UD2	O	1241	-	38,41,41	0.89	2 (5%)	57,62,62	1.40	7 (12%)
7	UD2	G	1246	-	38,41,41	0.88	2 (5%)	57,62,62	1.41	8 (14%)
3	UDP	C	1237	-	24,26,26	1.12	2 (8%)	37,40,40	1.71	7 (18%)
4	A2G	C	1238	-	15,15,15	0.41	0	21,21,21	0.85	1 (4%)
7	UD2	P	1244	-	38,41,41	0.88	2 (5%)	57,62,62	1.35	7 (12%)
7	UD2	E	1246	-	38,41,41	0.89	2 (5%)	57,62,62	1.36	6 (10%)
3	UDP	K	1237	-	24,26,26	1.11	2 (8%)	37,40,40	1.71	7 (18%)
4	A2G	D	1238	-	15,15,15	0.41	0	21,21,21	0.47	0
2	GOL	G	1247	-	5,5,5	0.37	0	5,5,5	0.23	0
8	NGA	N	1245	1	14,14,15	0.54	0	17,19,21	0.65	0
4	A2G	L	1238	-	15,15,15	0.42	0	21,21,21	0.48	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
4	A2G	A	1238	-	-	4/6/26/26	0/1/1/1
4	A2G	K	1238	-	-	3/6/26/26	0/1/1/1
3	UDP	J	1237	-	-	3/16/32/32	0/2/2/2
8	NGA	M	1247	1	-	3/6/23/26	0/1/1/1
2	GOL	A	1239	-	-	2/4/4/4	-
4	A2G	I	1238	-	-	5/6/26/26	0/1/1/1
2	GOL	G	1248	-	-	2/4/4/4	-
3	UDP	D	1237	-	-	4/16/32/32	0/2/2/2
7	UD2	F	1246	-	-	12/26/63/63	0/3/3/3
7	UD2	H	1246	-	-	8/26/63/63	0/3/3/3
2	GOL	A	1236	-	-	2/4/4/4	-
3	UDP	B	1237	-	-	0/16/32/32	0/2/2/2
4	A2G	J	1238	-	-	5/6/26/26	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	1PE	A	1240	-	-	3/13/13/13	-
3	UDP	N	1244	-	-	6/16/32/32	0/2/2/2
2	GOL	D	1239	-	-	2/4/4/4	-
4	A2G	B	1238	-	-	3/6/26/26	0/1/1/1
3	UDP	A	1237	-	-	7/16/32/32	0/2/2/2
3	UDP	L	1237	-	-	5/16/32/32	0/2/2/2
3	UDP	I	1237	-	-	0/16/32/32	0/2/2/2
3	UDP	M	1246	-	-	5/16/32/32	0/2/2/2
7	UD2	O	1241	-	-	10/26/63/63	0/3/3/3
7	UD2	G	1246	-	-	6/26/63/63	0/3/3/3
3	UDP	C	1237	-	-	1/16/32/32	0/2/2/2
4	A2G	C	1238	-	-	3/6/26/26	0/1/1/1
7	UD2	P	1244	-	-	11/26/63/63	0/3/3/3
7	UD2	E	1246	-	-	9/26/63/63	0/3/3/3
3	UDP	K	1237	-	-	0/16/32/32	0/2/2/2
4	A2G	D	1238	-	-	2/6/26/26	0/1/1/1
2	GOL	G	1247	-	-	2/4/4/4	-
8	NGA	N	1245	1	-	3/6/23/26	0/1/1/1
4	A2G	L	1238	-	-	4/6/26/26	0/1/1/1

The worst 5 of 33 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	L	1237	UDP	C4-N3	-2.43	1.34	1.38
3	N	1244	UDP	C4-N3	-2.43	1.34	1.38
3	D	1237	UDP	C4-N3	-2.42	1.34	1.38
3	K	1237	UDP	C4-N3	-2.42	1.34	1.38
3	A	1237	UDP	C4-N3	-2.42	1.34	1.38

The worst 5 of 115 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	O	1241	UD2	C4-N3-C2	-4.52	120.62	126.58
3	D	1237	UDP	C4-N3-C2	-4.51	120.62	126.58
3	L	1237	UDP	C4-N3-C2	-4.50	120.65	126.58
3	N	1244	UDP	C4-N3-C2	-4.50	120.65	126.58
3	B	1237	UDP	C4-N3-C2	-4.50	120.65	126.58

There are no chirality outliers.

5 of 135 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1236	GOL	O1-C1-C2-C3
2	D	1239	GOL	O1-C1-C2-C3
2	G	1247	GOL	O1-C1-C2-C3
2	G	1248	GOL	O1-C1-C2-C3
3	A	1237	UDP	C5'-O5'-PA-O1A

There are no ring outliers.

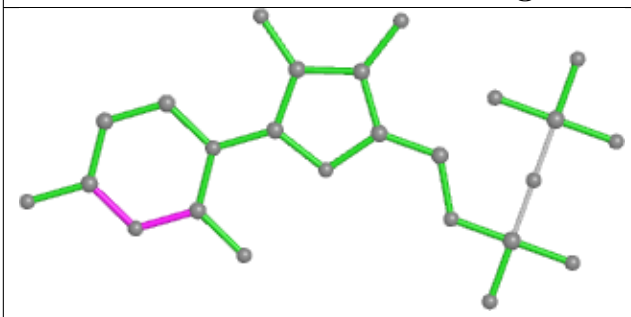
23 monomers are involved in 105 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1238	A2G	1	0
4	K	1238	A2G	2	0
3	J	1237	UDP	2	0
8	M	1247	NGA	2	0
3	D	1237	UDP	6	0
7	F	1246	UD2	5	0
7	H	1246	UD2	4	0
3	B	1237	UDP	5	0
5	A	1240	1PE	1	0
3	N	1244	UDP	18	0
3	A	1237	UDP	4	0
3	L	1237	UDP	2	0
3	I	1237	UDP	2	0
3	M	1246	UDP	4	0
7	O	1241	UD2	20	0
7	G	1246	UD2	7	0
3	C	1237	UDP	3	1
7	P	1244	UD2	3	0
7	E	1246	UD2	5	0
3	K	1237	UDP	4	0
2	G	1247	GOL	1	0
8	N	1245	NGA	2	0
4	L	1238	A2G	5	0

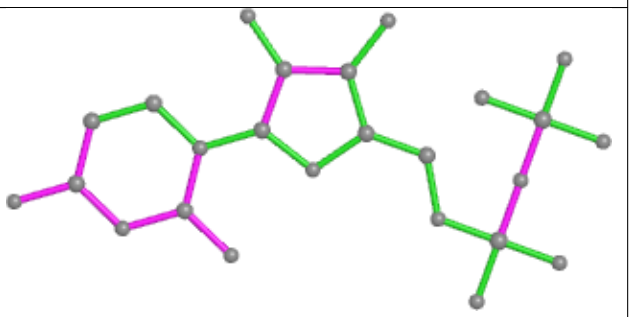
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included.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

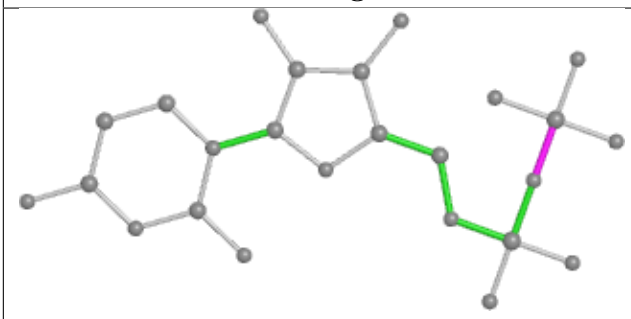
Ligand UDP J 1237



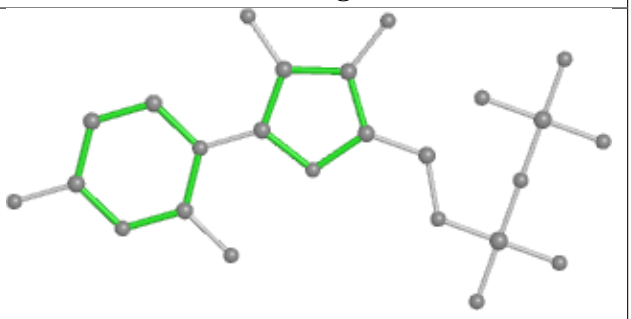
Bond lengths



Bond angles

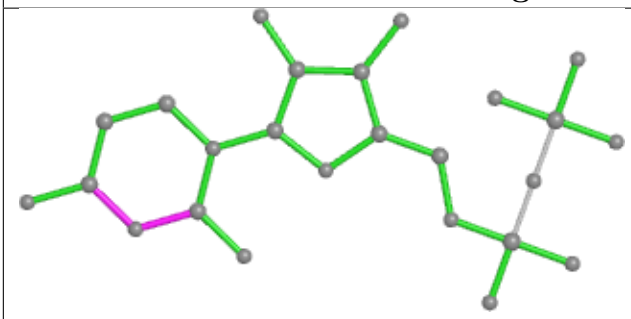


Torsions

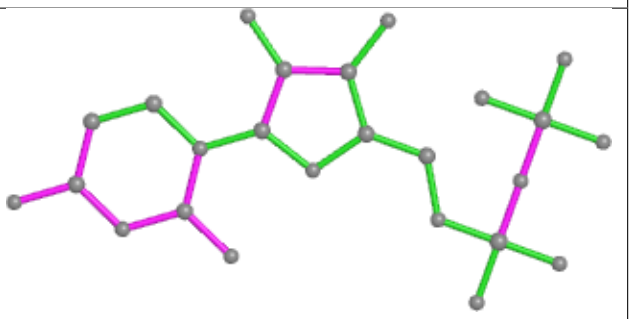


Rings

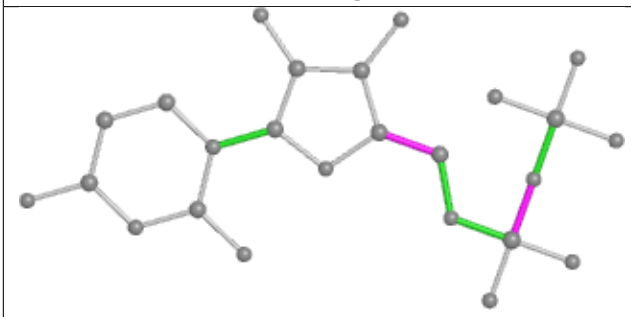
Ligand UDP D 1237



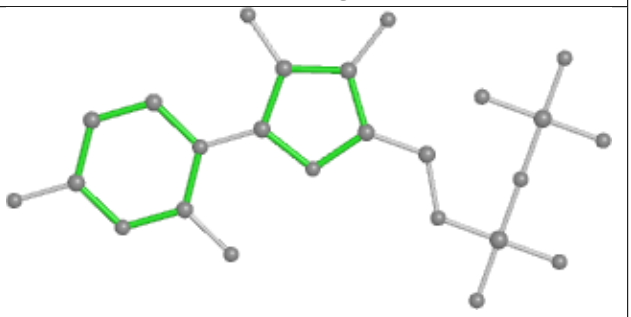
Bond lengths



Bond angles

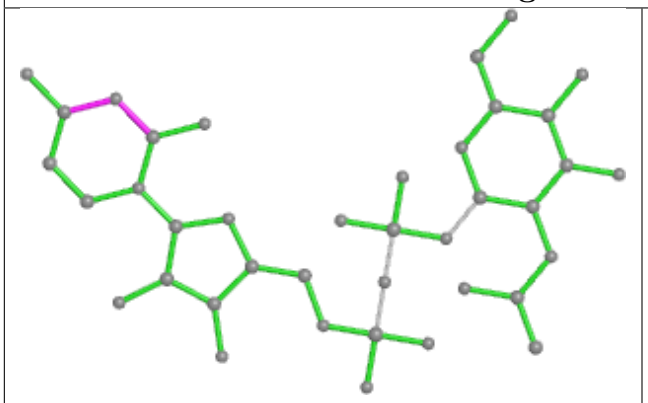


Torsions

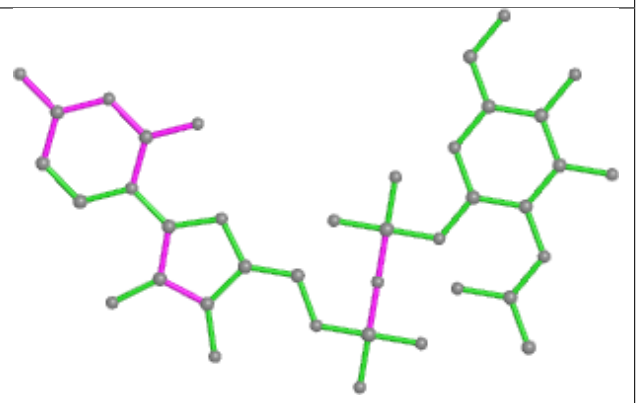


Rings

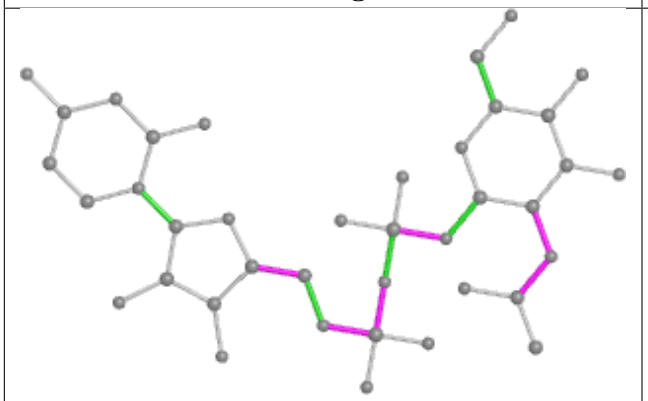
Ligand UD2 F 1246



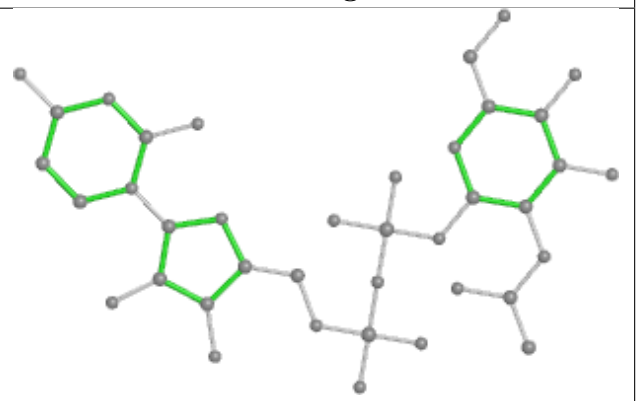
Bond lengths



Bond angles

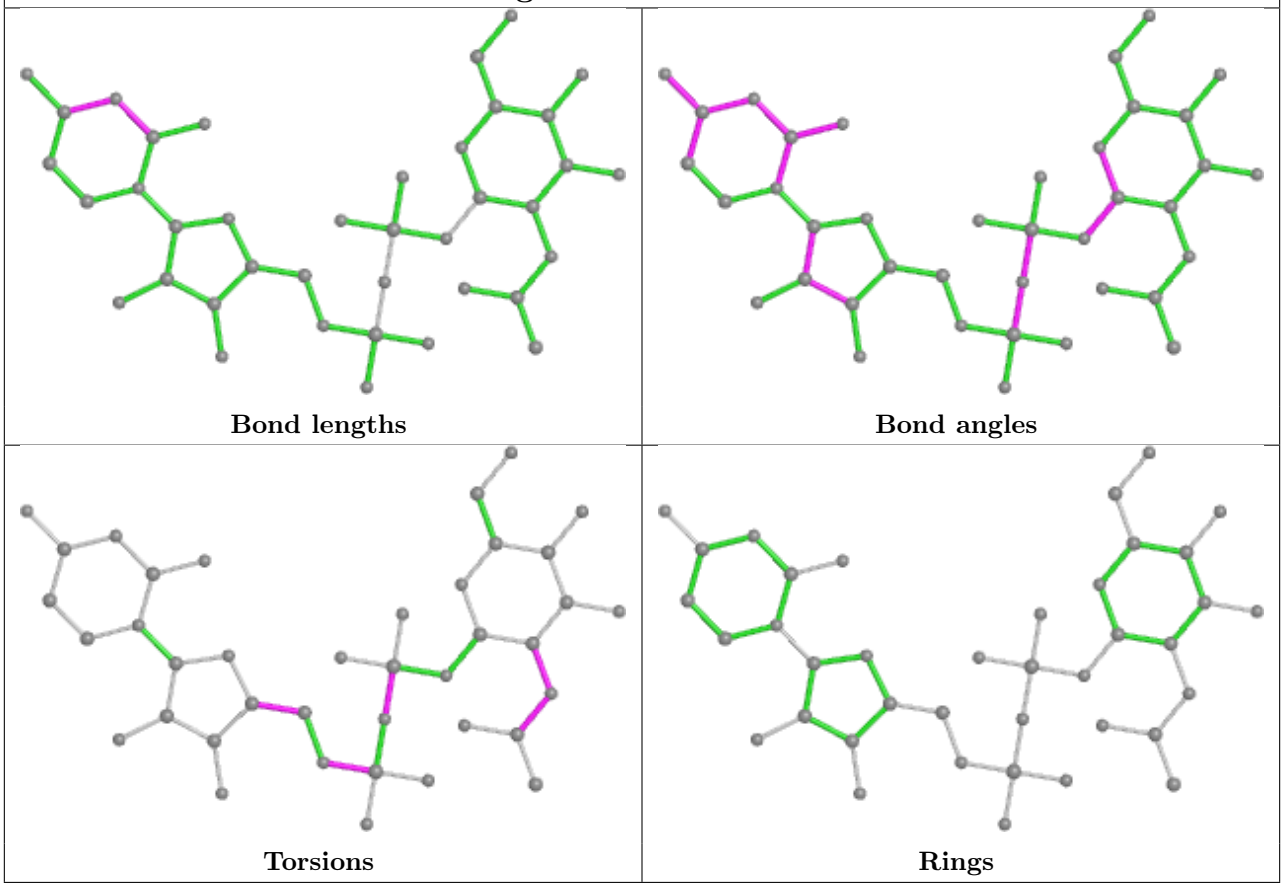


Torsions

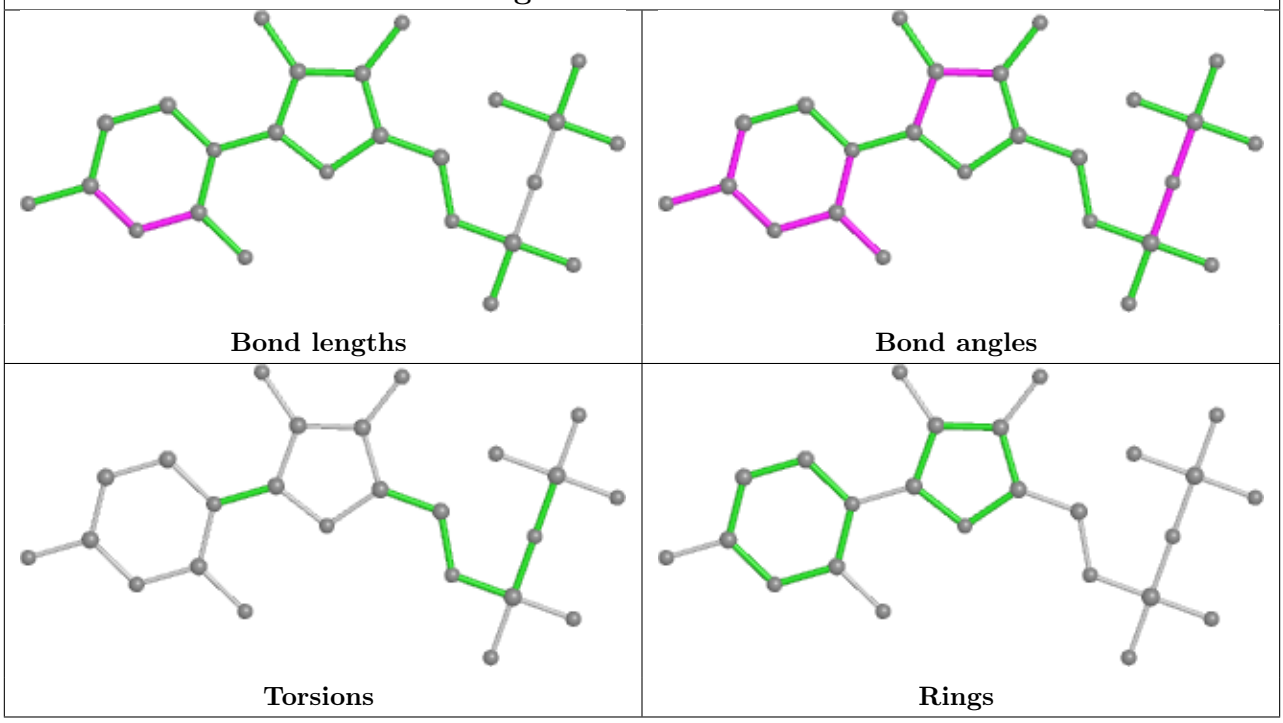


Rings

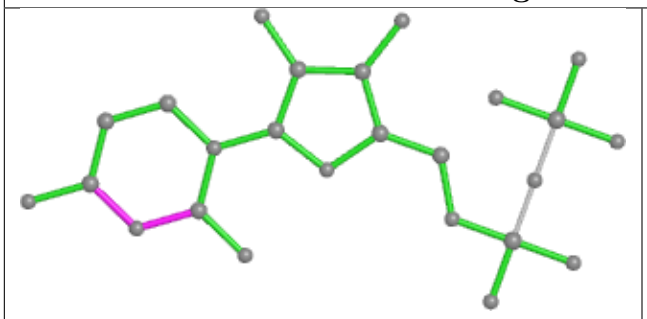
Ligand UD2 H 1246



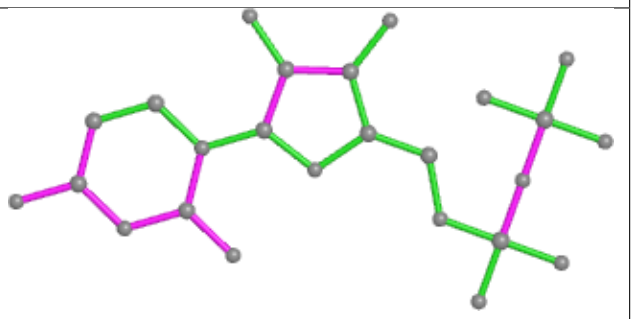
Ligand UDP B 1237



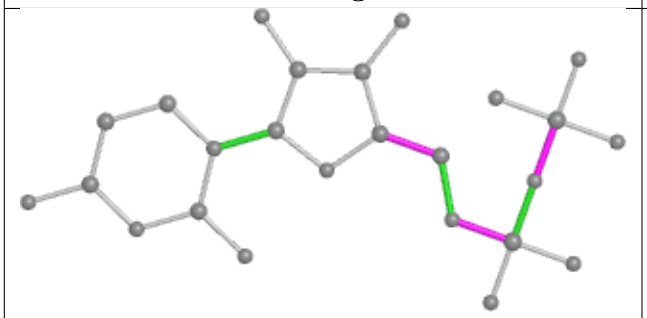
Ligand UDP N 1244



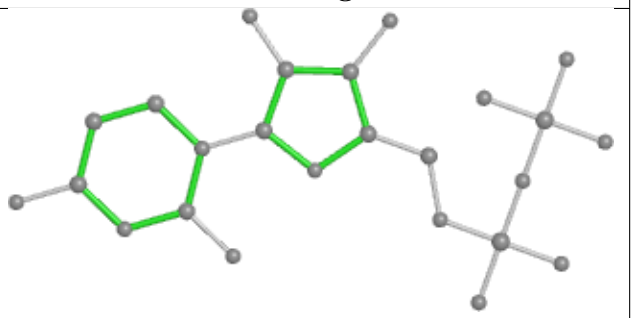
Bond lengths



Bond angles

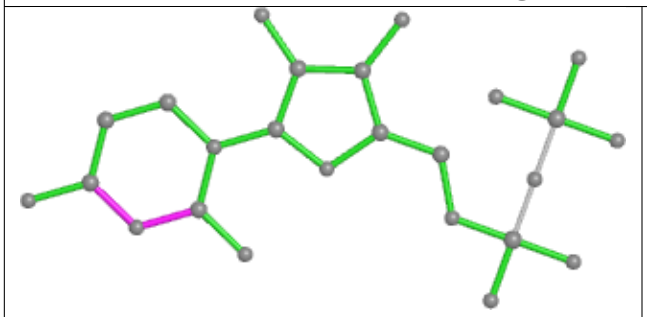


Torsions

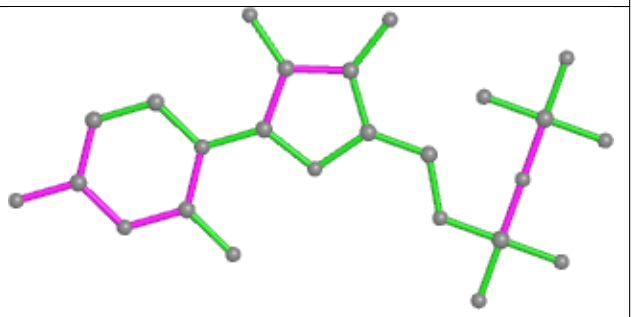


Rings

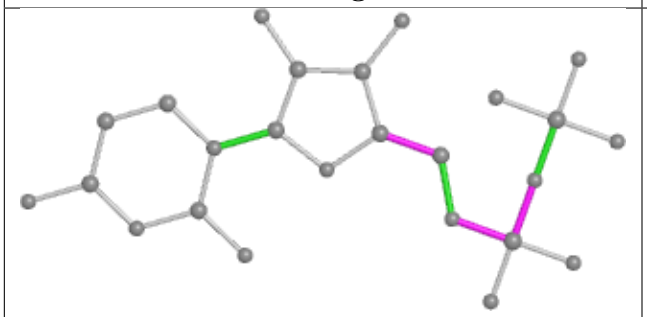
Ligand UDP A 1237



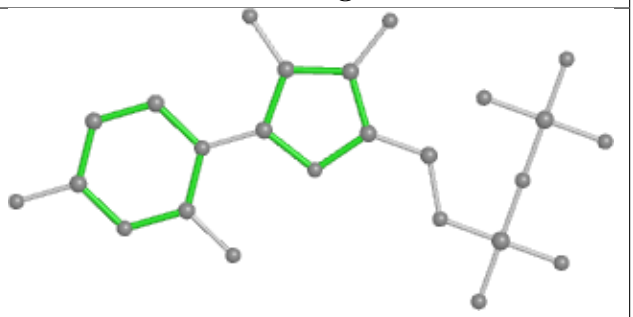
Bond lengths



Bond angles

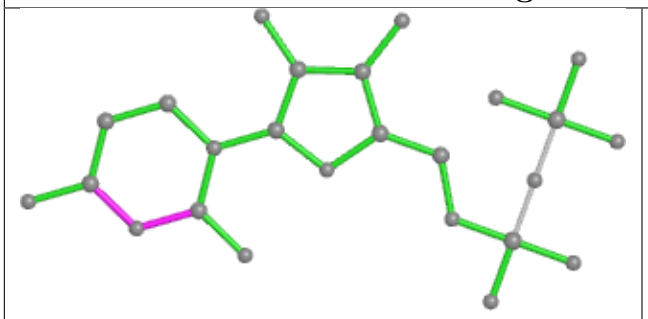


Torsions

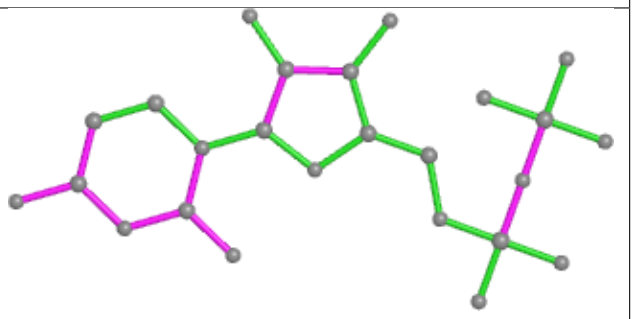


Rings

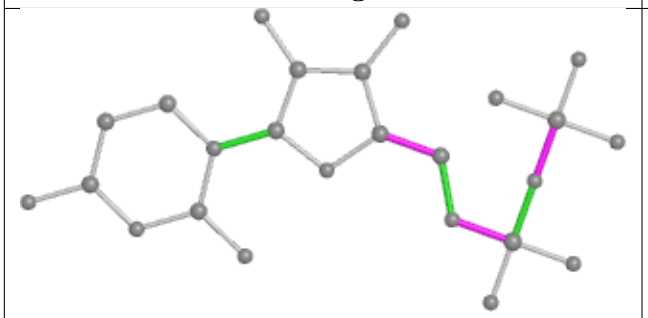
Ligand UDP L 1237



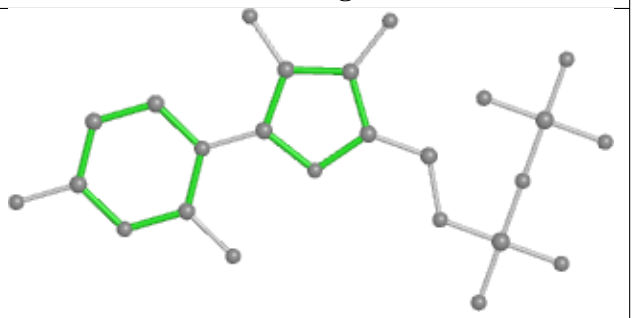
Bond lengths



Bond angles

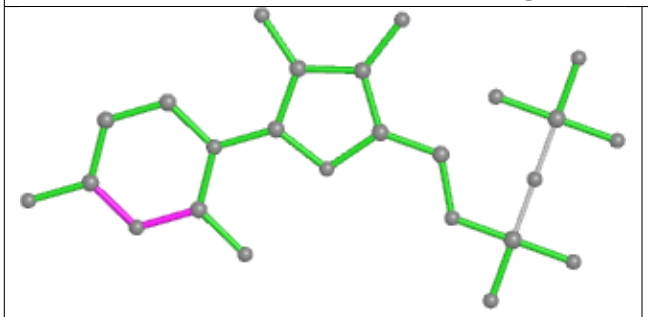


Torsions

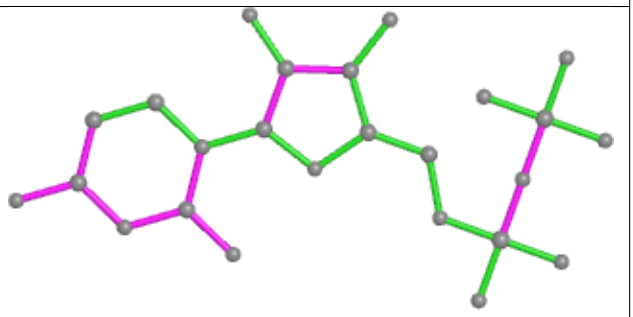


Rings

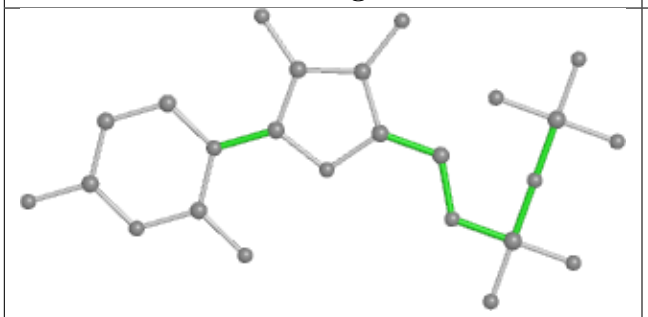
Ligand UDP I 1237



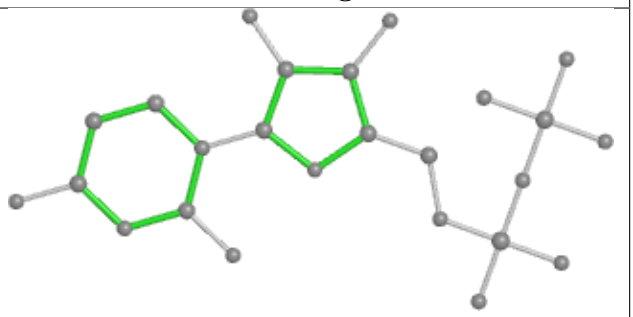
Bond lengths



Bond angles

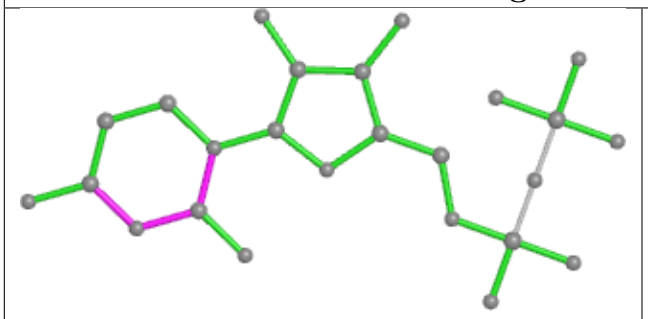


Torsions

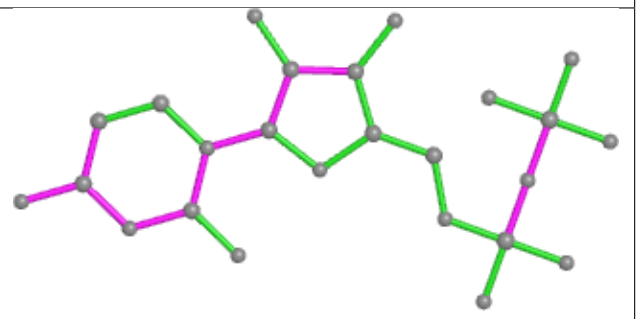


Rings

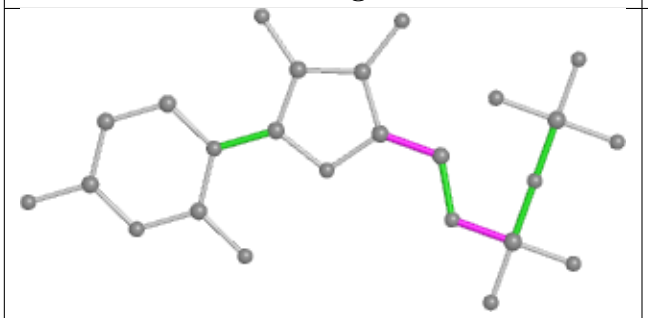
Ligand UDP M 1246



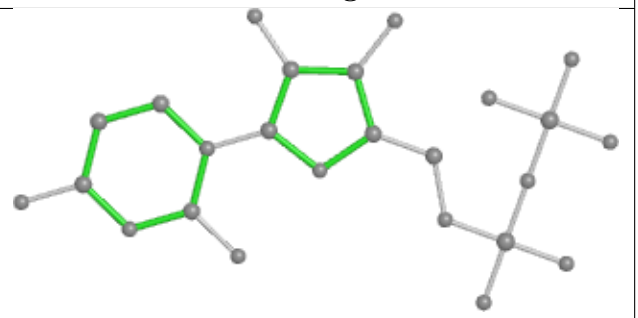
Bond lengths



Bond angles

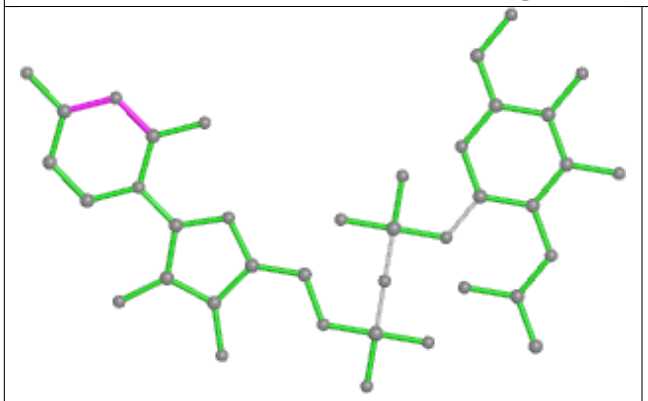


Torsions

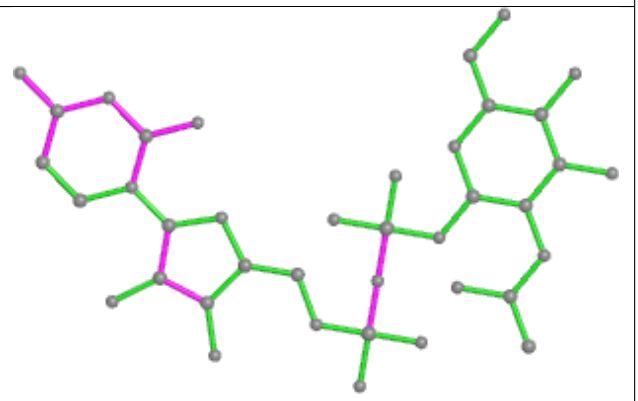


Rings

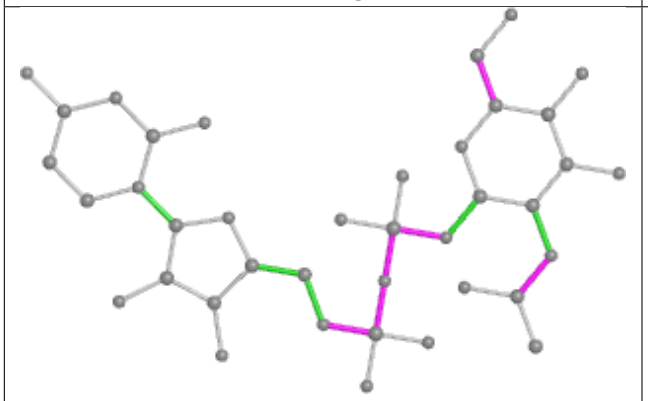
Ligand UD2 O 1241



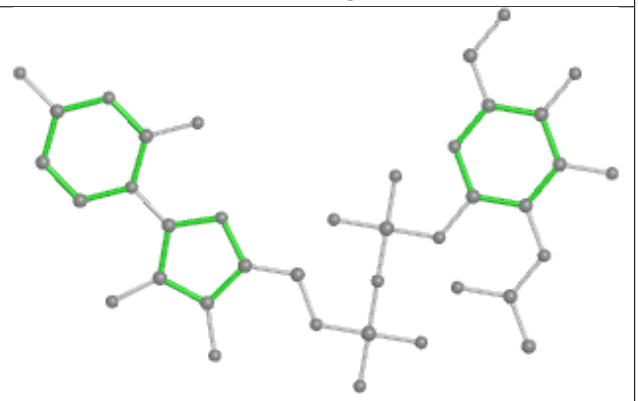
Bond lengths



Bond angles

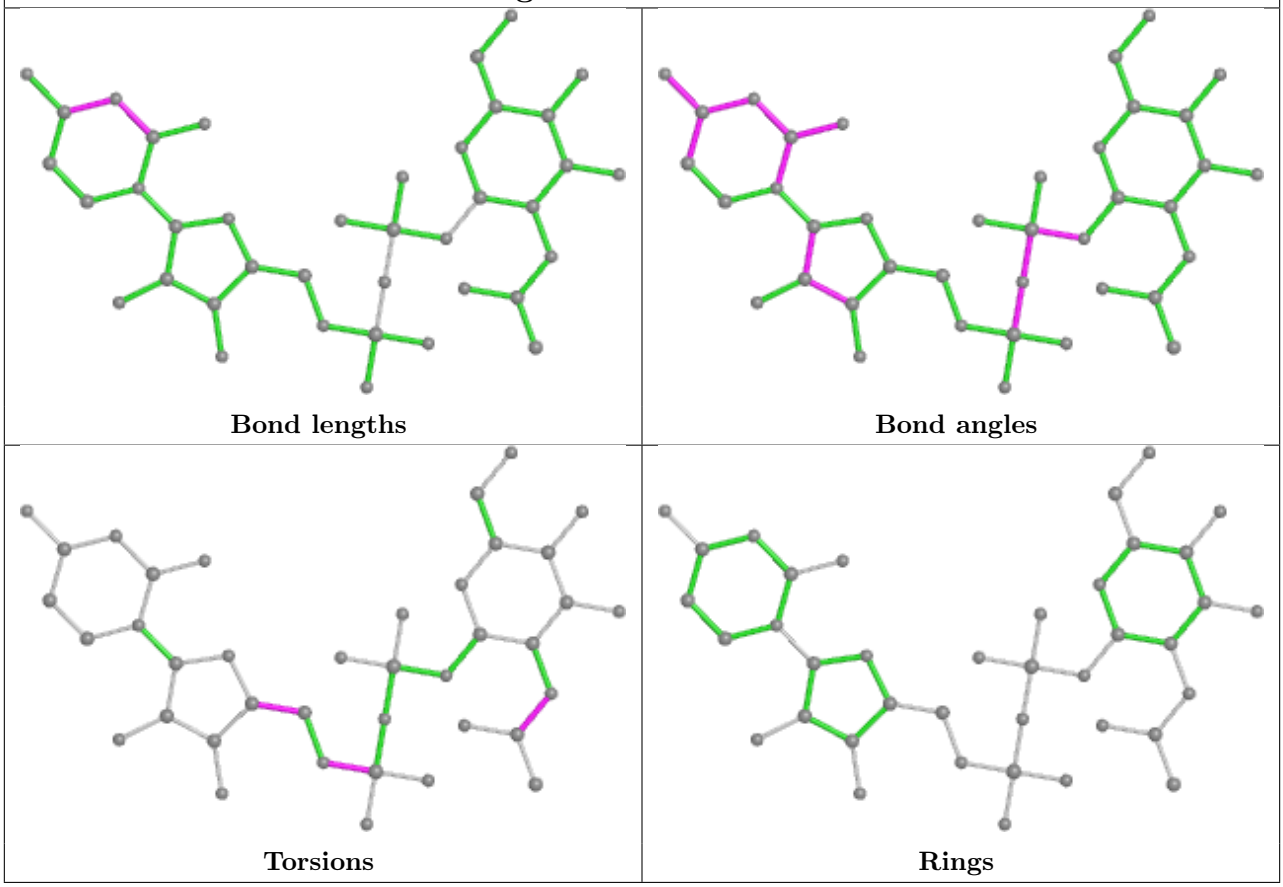


Torsions

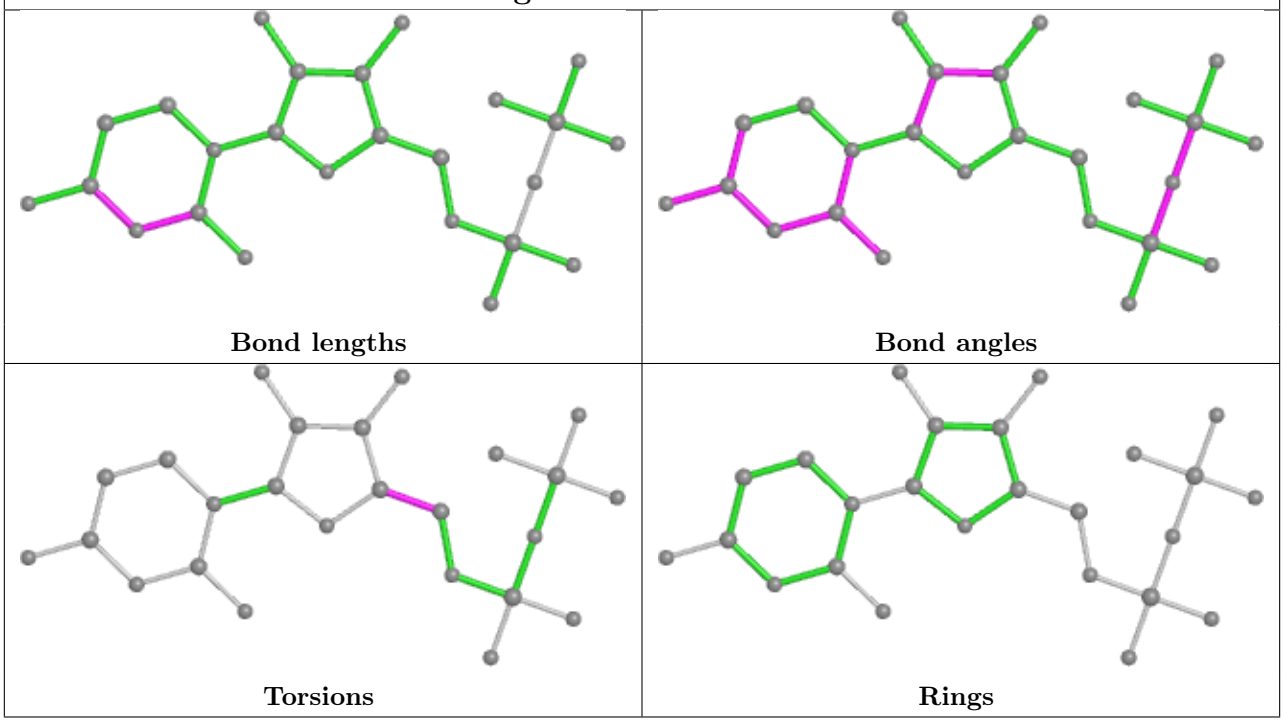


Rings

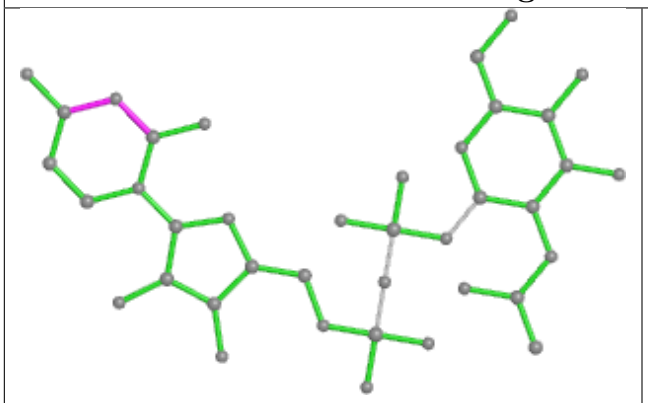
Ligand UD2 G 1246



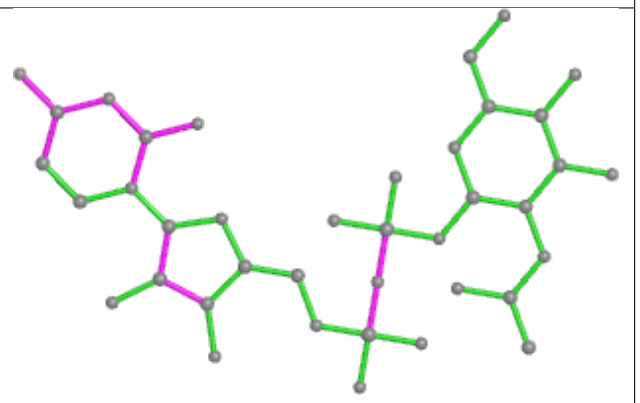
Ligand UDP C 1237



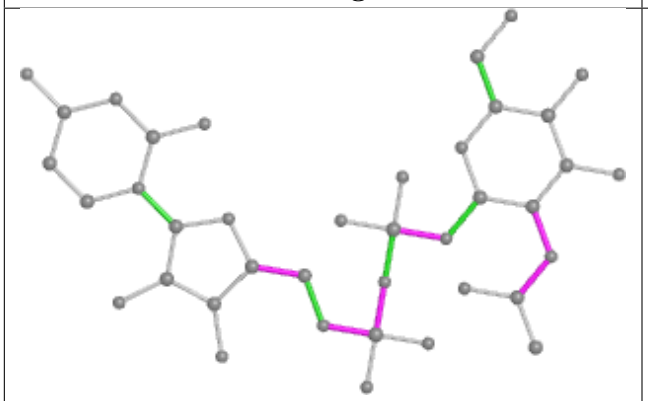
Ligand UD2 P 1244



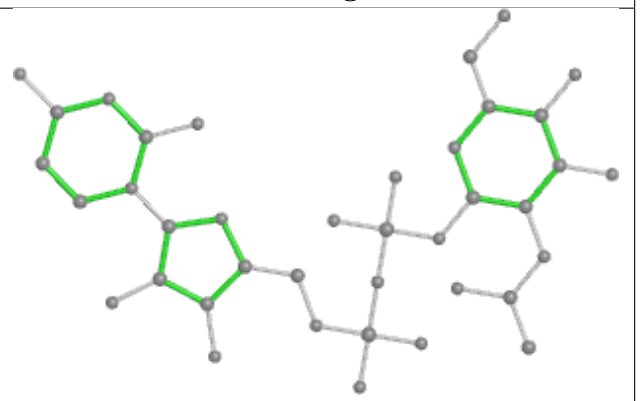
Bond lengths



Bond angles

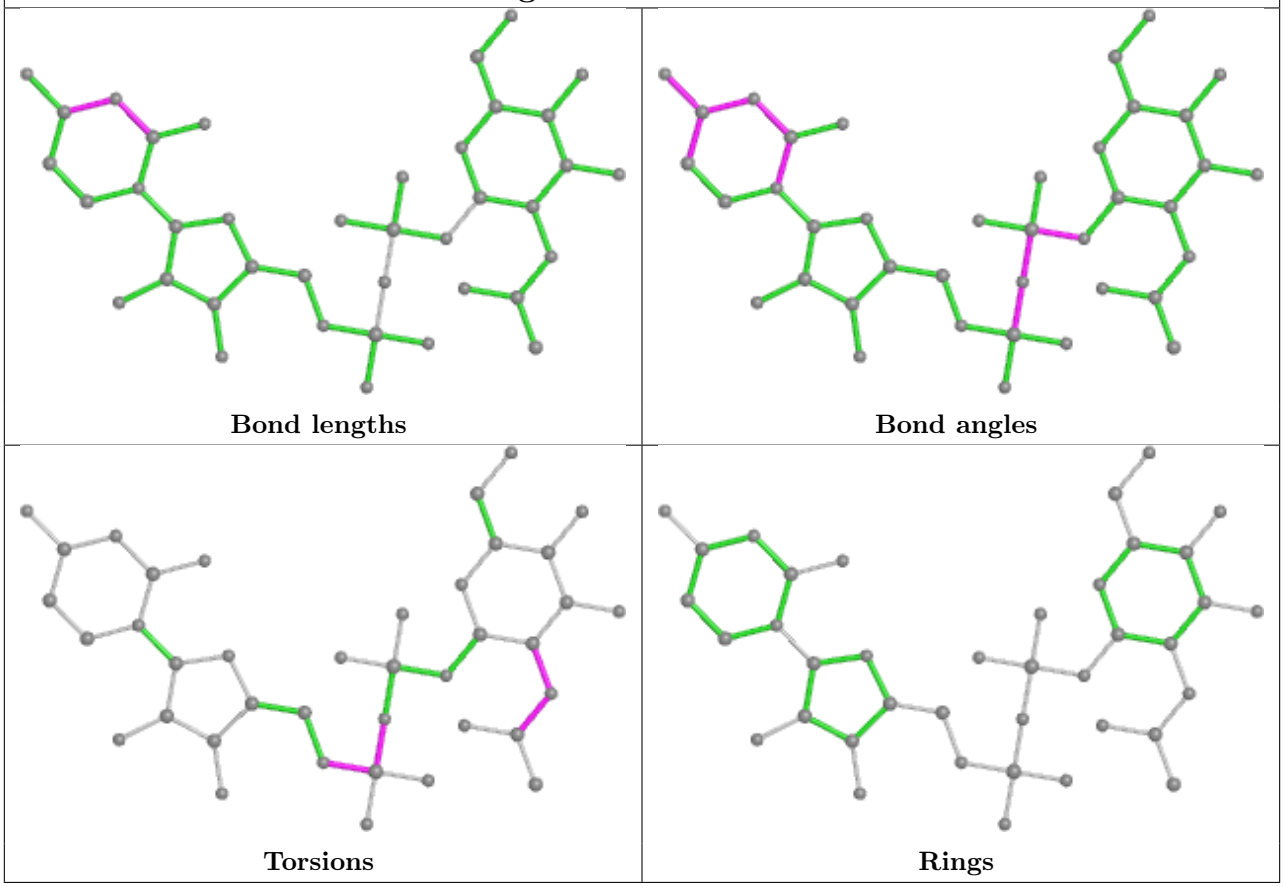


Torsions

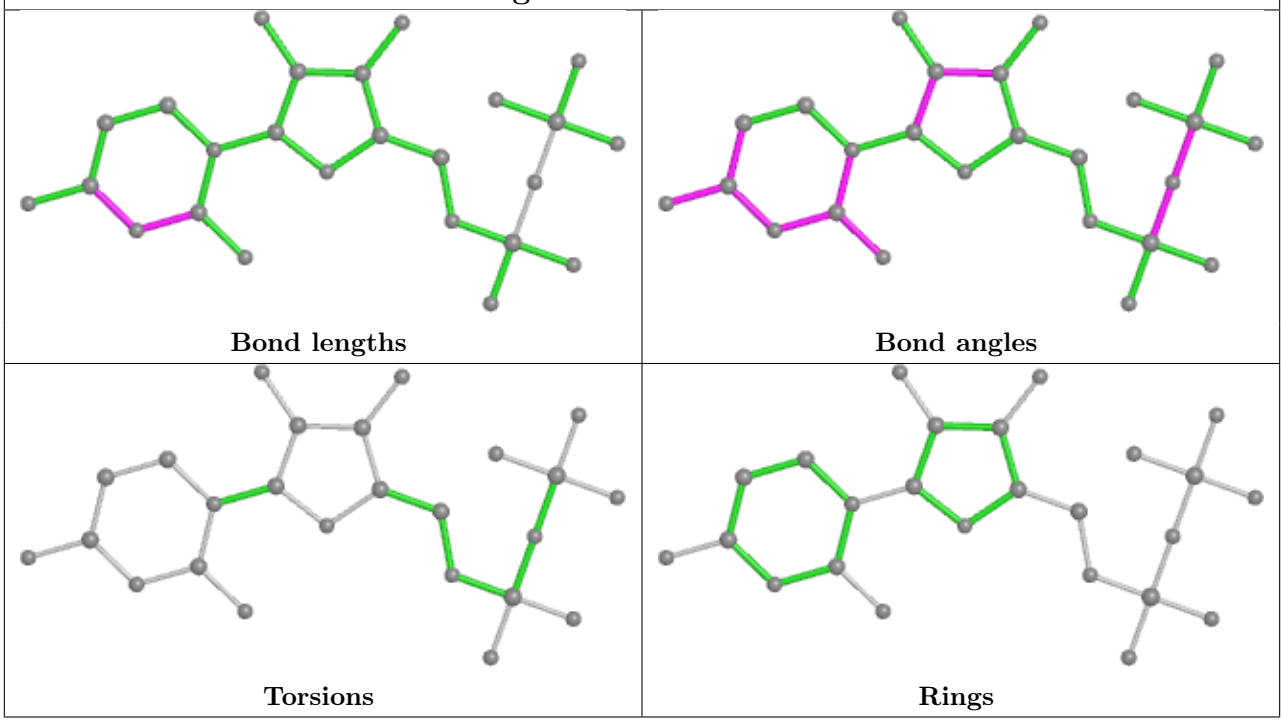


Rings

Ligand UD2 E 1246



Ligand UDP K 1237



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, 95th 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^2)	Q<0.9
1	A	238/248 (95%)	-0.44	0	100	100	35, 52, 66, 92	3 (1%)
1	B	238/248 (95%)	-0.39	0	100	100	41, 63, 81, 102	3 (1%)
1	C	238/248 (95%)	-0.35	0	100	100	46, 63, 82, 100	3 (1%)
1	D	237/248 (95%)	-0.33	0	100	100	46, 64, 82, 89	3 (1%)
1	E	246/248 (99%)	-0.42	0	100	100	38, 56, 84, 94	3 (1%)
1	F	246/248 (99%)	-0.37	0	100	100	47, 62, 88, 99	3 (1%)
1	G	246/248 (99%)	-0.42	0	100	100	43, 59, 79, 92	3 (1%)
1	H	247/248 (99%)	-0.32	0	100	100	43, 66, 94, 104	3 (1%)
1	I	236/248 (95%)	-0.27	0	100	100	41, 66, 88, 109	3 (1%)
1	J	236/248 (95%)	-0.13	0	100	100	59, 84, 99, 109	3 (1%)
1	K	237/248 (95%)	-0.30	0	100	100	41, 70, 85, 110	3 (1%)
1	L	236/248 (95%)	-0.22	1 (0%)	92	90	51, 79, 102, 111	3 (1%)
1	M	245/248 (98%)	-0.24	0	100	100	42, 66, 85, 102	3 (1%)
1	N	243/248 (97%)	-0.01	0	100	100	53, 94, 110, 120	3 (1%)
1	O	240/248 (96%)	0.27	6 (2%)	57	51	66, 106, 126, 140	3 (1%)
1	P	243/248 (97%)	-0.09	1 (0%)	92	90	50, 81, 98, 113	3 (1%)
All	All	3852/3968 (97%)	-0.25	8 (0%)	95	93	35, 68, 105, 140	48 (1%)

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	O	160	GLY	5.7
1	O	186	ILE	4.1
1	O	187	PRO	3.1
1	L	32	GLN	2.6
1	P	117	LEU	2.6

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, 95th 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^2)	Q<0.9
6	SO4	B	1239	5/5	0.76	0.24	98,98,105,119	0
4	A2G	I	1238	15/15	0.82	0.25	56,76,86,94	0
5	1PE	A	1240	16/16	0.83	0.27	68,77,84,86	0
4	A2G	K	1238	15/15	0.85	0.24	69,81,85,89	0
7	UD2	O	1241	39/39	0.85	0.25	87,114,125,131	0
4	A2G	C	1238	15/15	0.86	0.22	52,68,77,78	0
3	UDP	J	1237	25/25	0.86	0.20	86,102,113,122	0
4	A2G	A	1238	15/15	0.86	0.24	44,53,68,73	0
4	A2G	B	1238	15/15	0.88	0.22	42,55,65,66	0
3	UDP	N	1244	25/25	0.88	0.21	94,104,109,113	0
4	A2G	L	1238	15/15	0.88	0.29	60,67,74,75	0
3	UDP	B	1237	25/25	0.89	0.26	64,78,111,133	0
3	UDP	C	1237	25/25	0.89	0.26	61,78,107,131	0
4	A2G	J	1238	15/15	0.89	0.20	50,70,76,78	0
3	UDP	A	1237	25/25	0.89	0.23	51,57,90,100	0
2	GOL	D	1239	6/6	0.90	0.19	55,65,68,76	0
2	GOL	A	1236	6/6	0.90	0.20	55,65,67,68	0
8	NGA	M	1247	14/15	0.90	0.27	84,93,100,104	0
4	A2G	D	1238	15/15	0.91	0.20	69,75,81,81	0
2	GOL	G	1248	6/6	0.91	0.25	54,67,71,75	0
3	UDP	L	1237	25/25	0.91	0.20	73,86,101,108	0
8	NGA	N	1245	14/15	0.91	0.25	84,93,100,104	0
2	GOL	A	1239	6/6	0.92	0.15	53,57,59,61	0
7	UD2	P	1244	39/39	0.92	0.25	70,80,93,94	0
7	UD2	H	1246	39/39	0.93	0.24	52,61,72,84	0
3	UDP	I	1237	25/25	0.93	0.18	63,70,91,112	0
2	GOL	G	1247	6/6	0.93	0.16	66,71,72,81	0

Continued on next page...

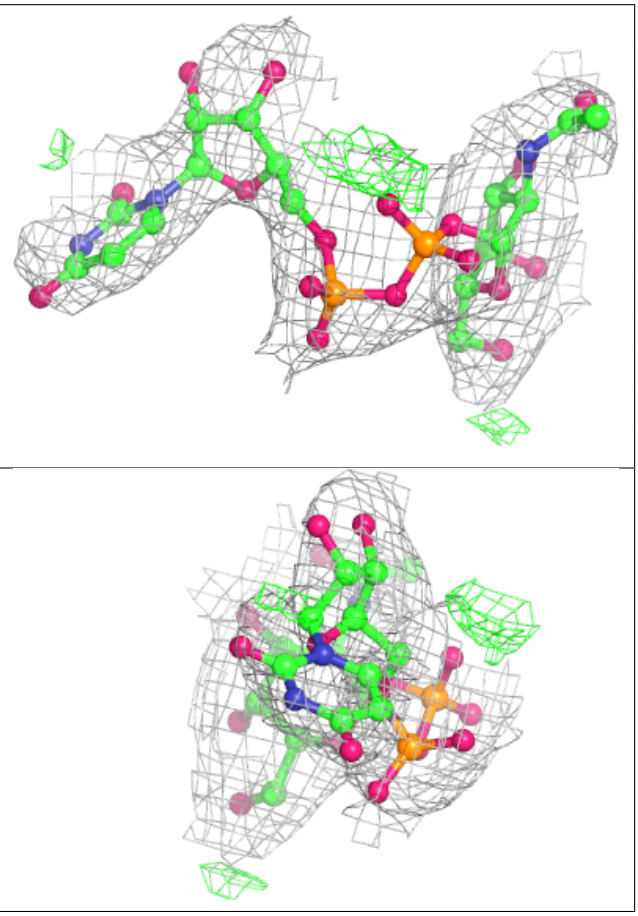
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
7	UD2	E	1246	39/39	0.93	0.20	48,59,75,79	0
7	UD2	F	1246	39/39	0.93	0.23	48,61,77,79	0
3	UDP	M	1246	25/25	0.94	0.21	66,71,77,82	0
3	UDP	K	1237	25/25	0.94	0.16	56,74,91,101	0
7	UD2	G	1246	39/39	0.94	0.21	50,58,70,78	0
3	UDP	D	1237	25/25	0.94	0.23	66,73,101,104	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.

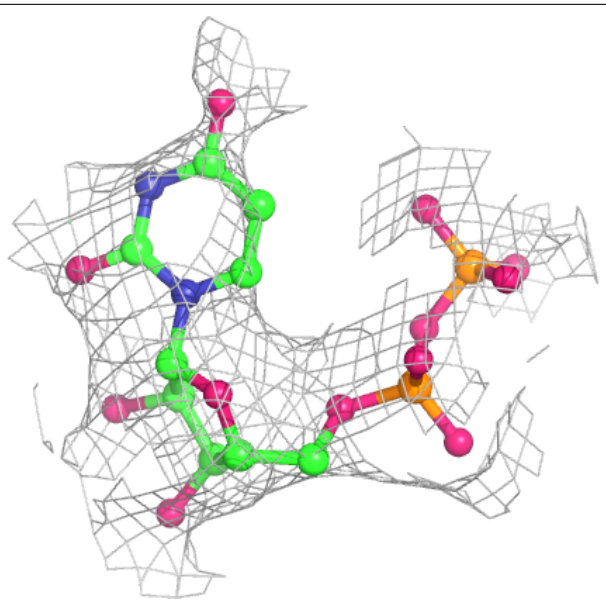
Electron density around UD2 O 1241:

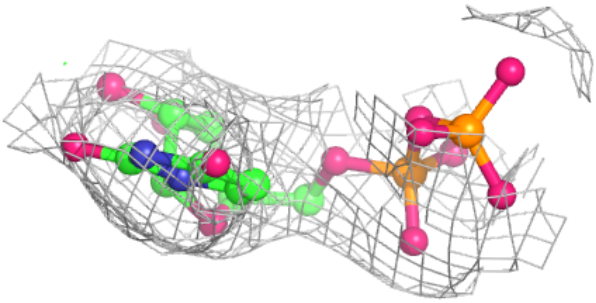
2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)

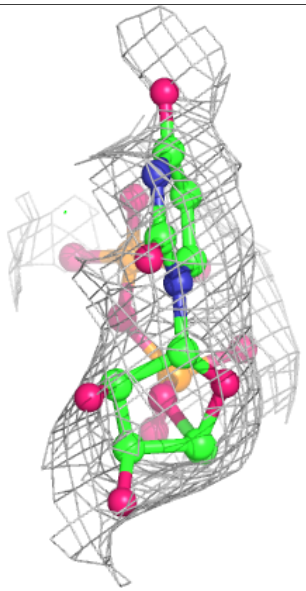


Electron density around UDP J 1237:

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

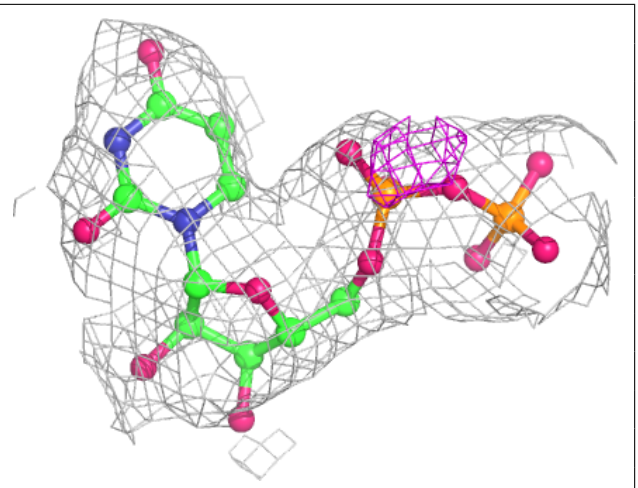


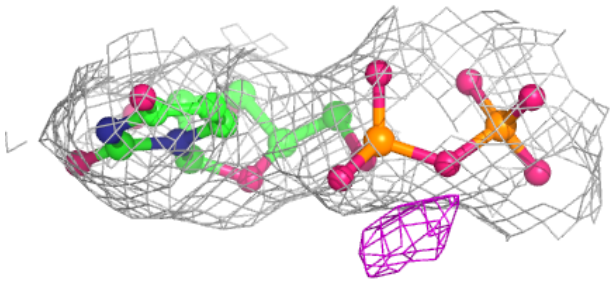


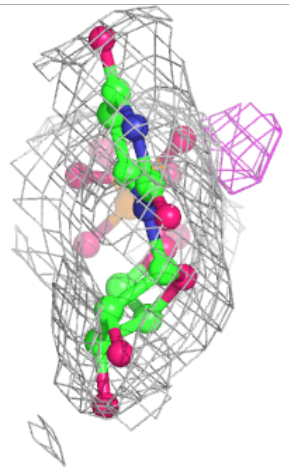


Electron density around UDP N 1244:

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

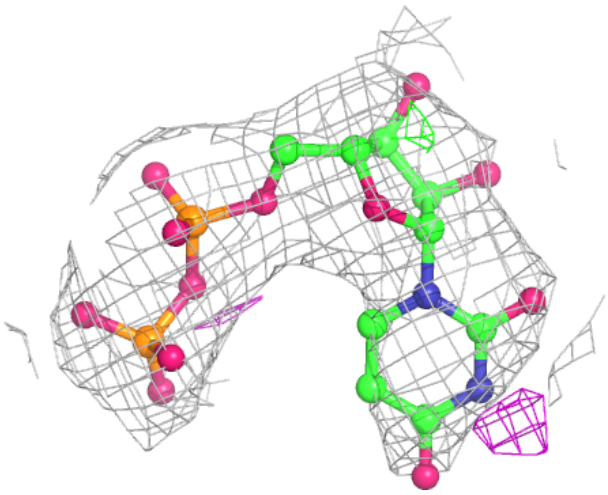


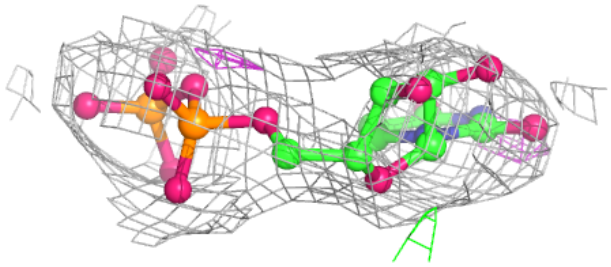


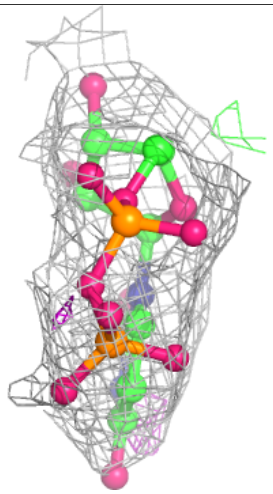


Electron density around UDP B 1237:

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

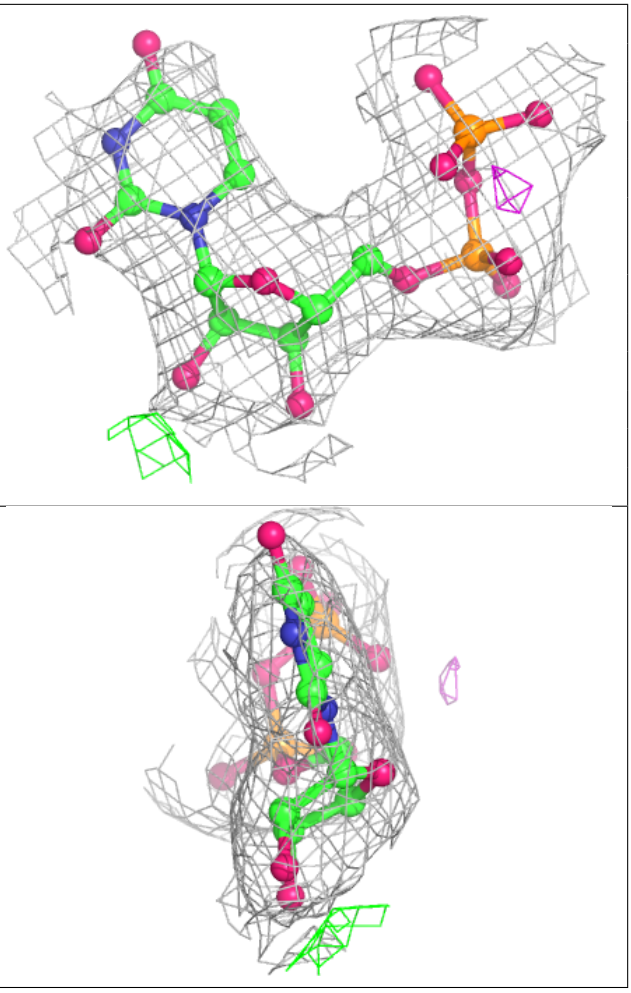






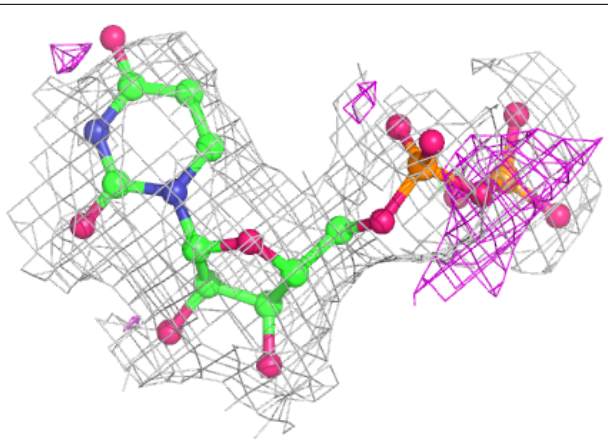
Electron density around UDP C 1237:

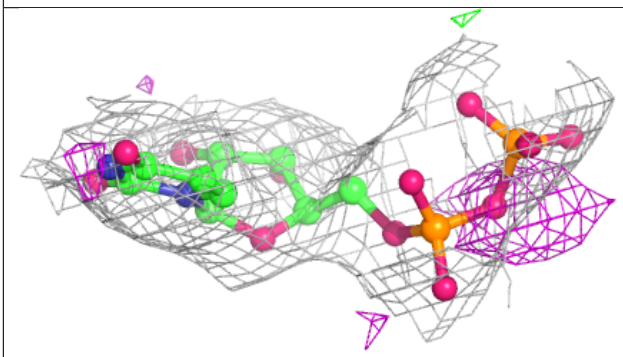
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

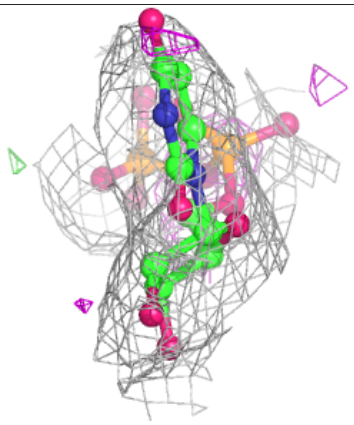


Electron density around UDP A 1237:

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

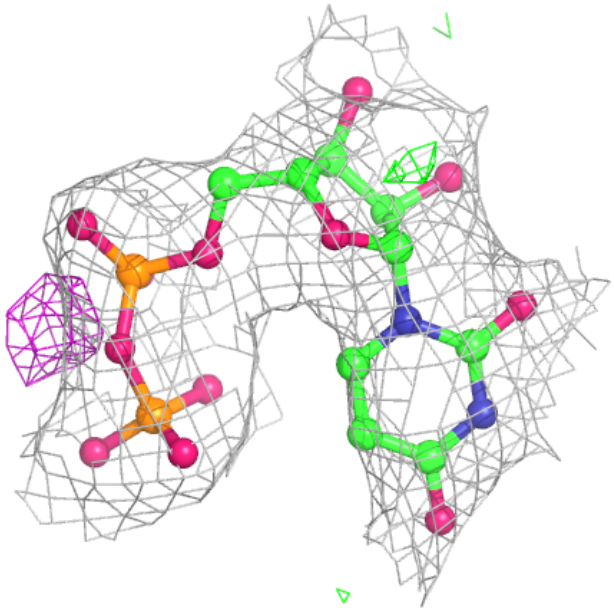


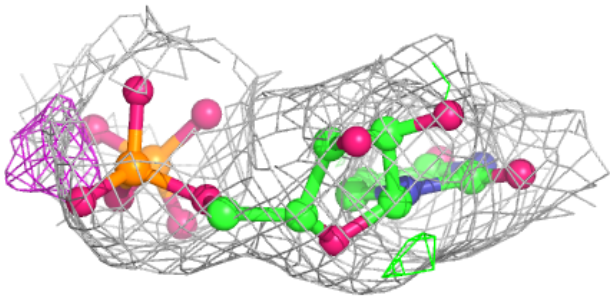


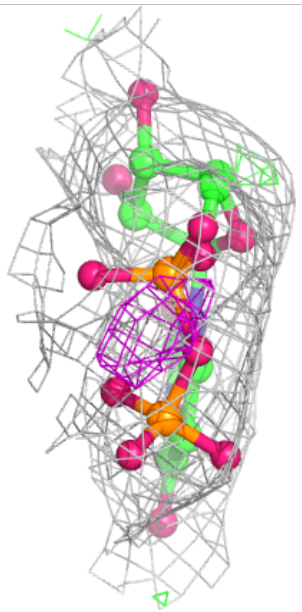


Electron density around UDP L 1237:

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

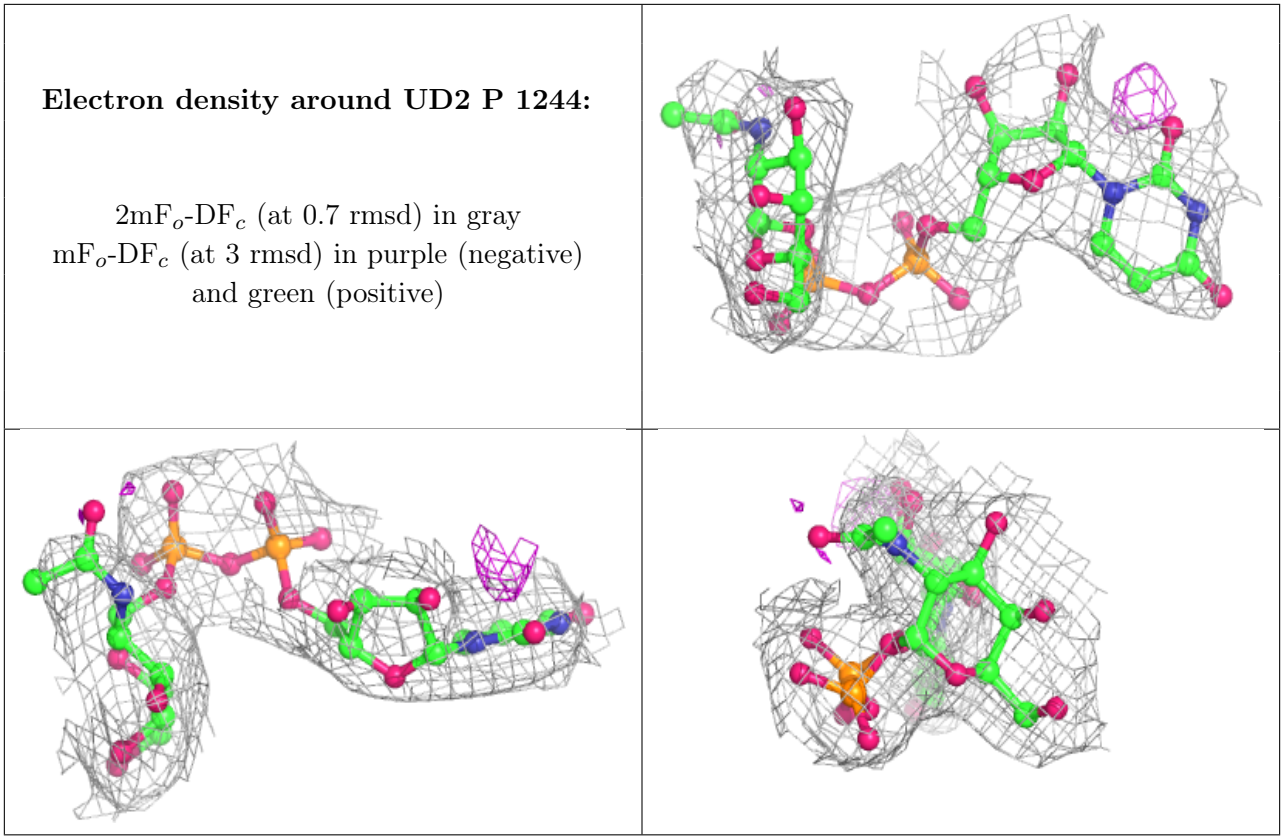






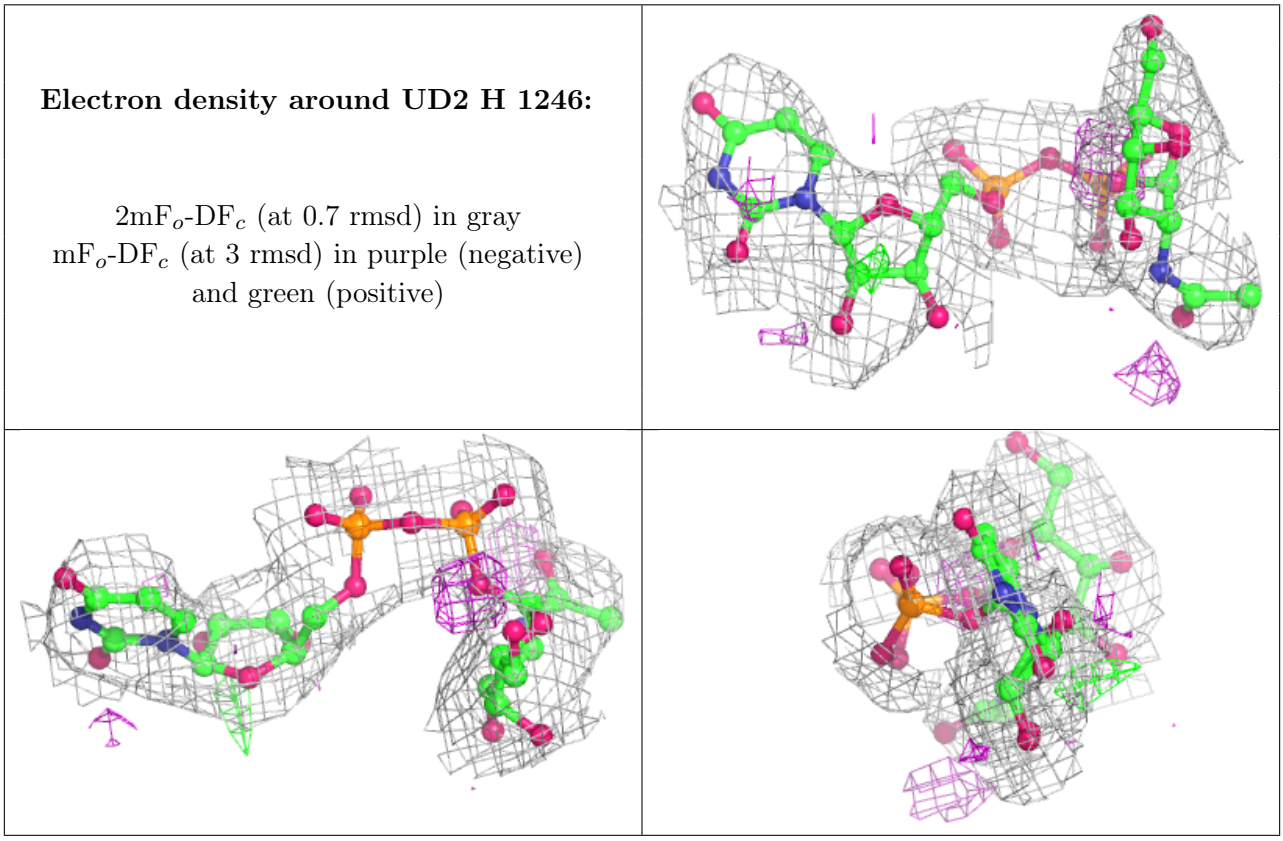
Electron density around UD2 P 1244:

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



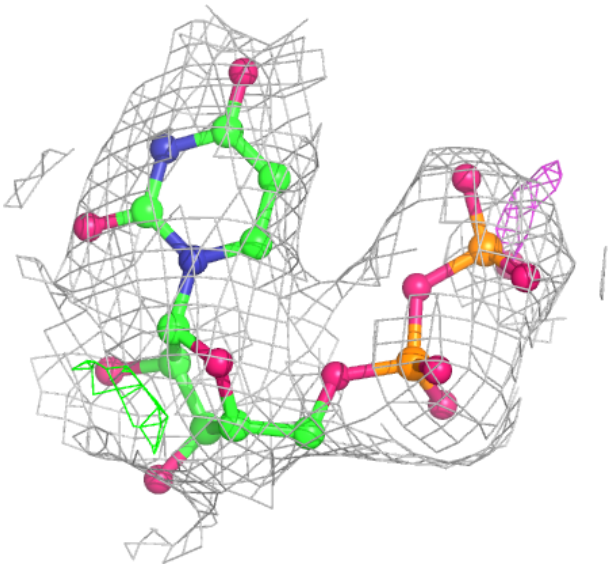
Electron density around UD2 H 1246:

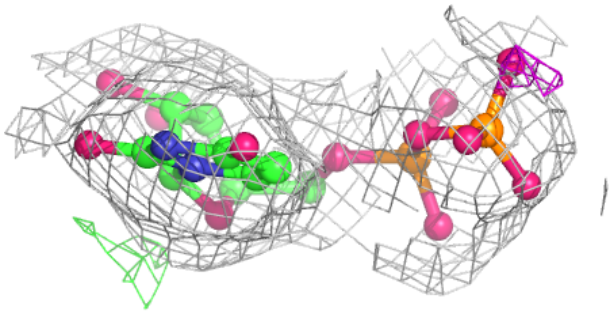
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

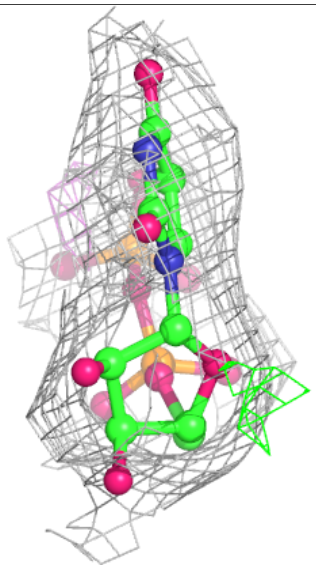


Electron density around UDP I 1237:

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

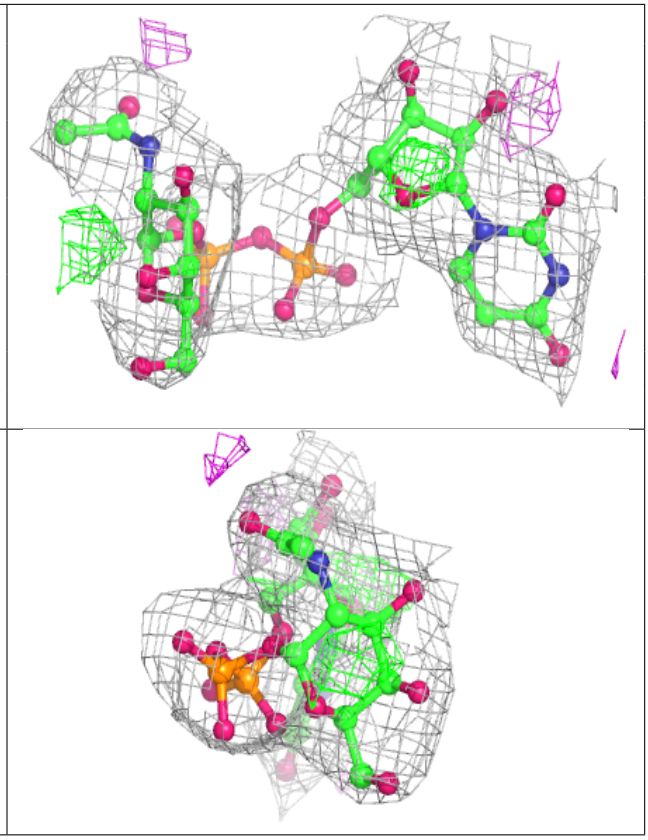






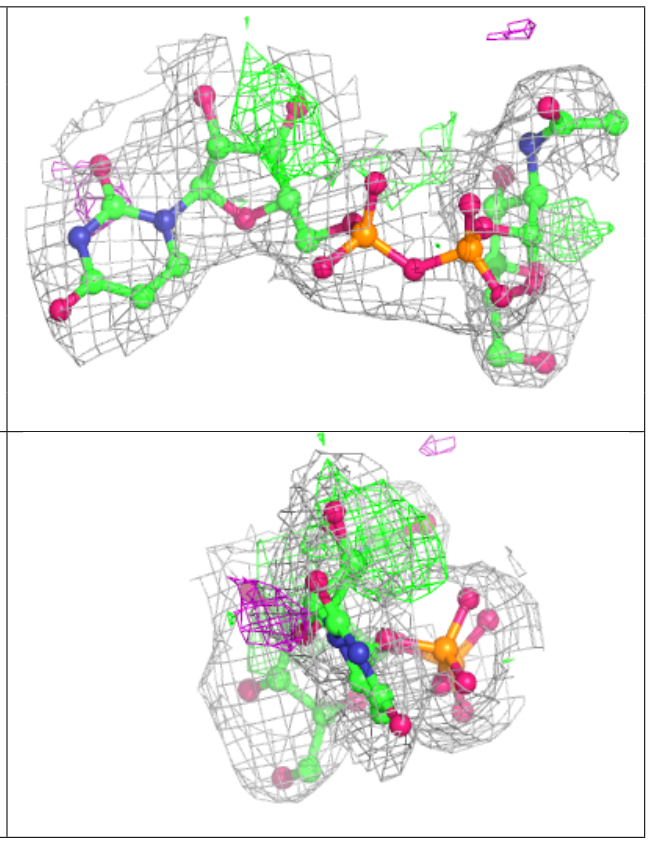
Electron density around UD2 E 1246:

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



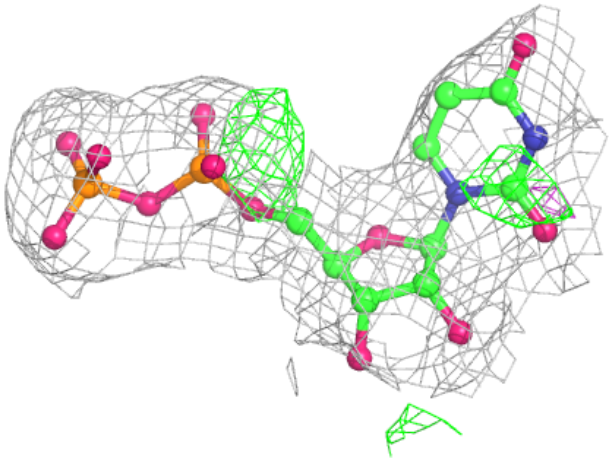
Electron density around UD2 F 1246:

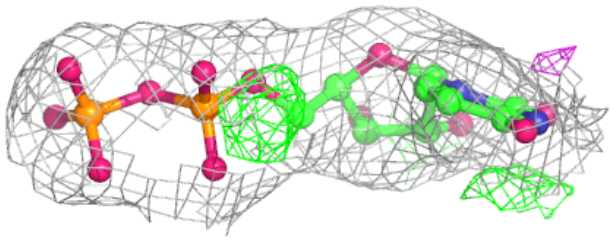
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

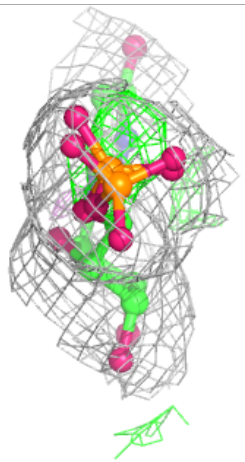


Electron density around UDP M 1246:

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

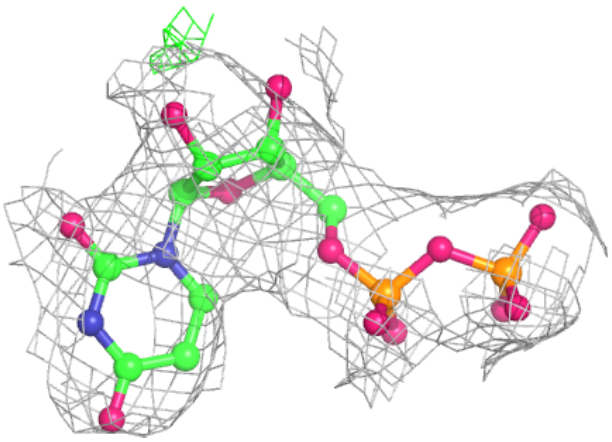


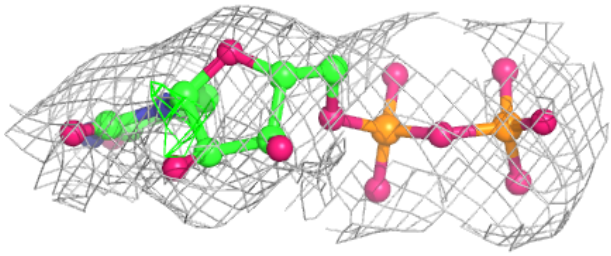


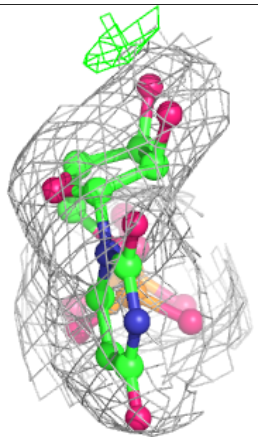


Electron density around UDP K 1237:

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

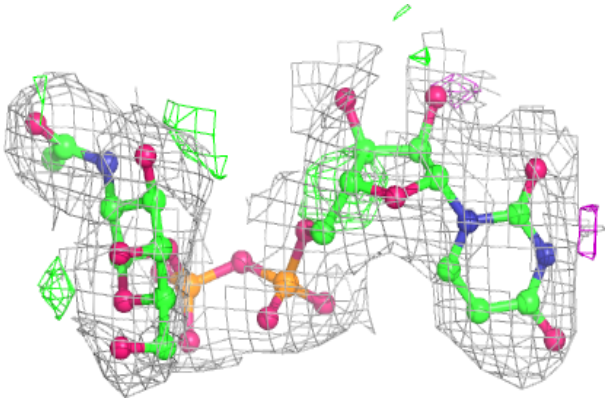


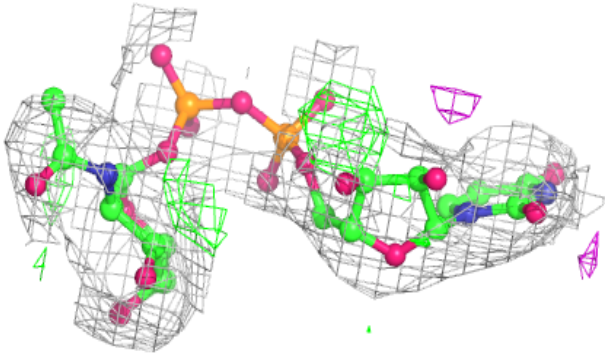


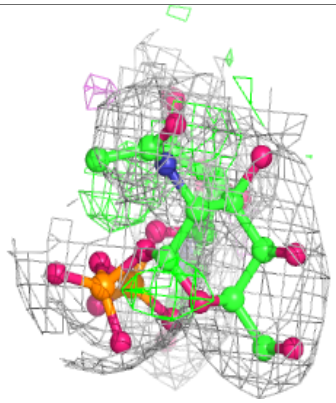


Electron density around UD2 G 1246:

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

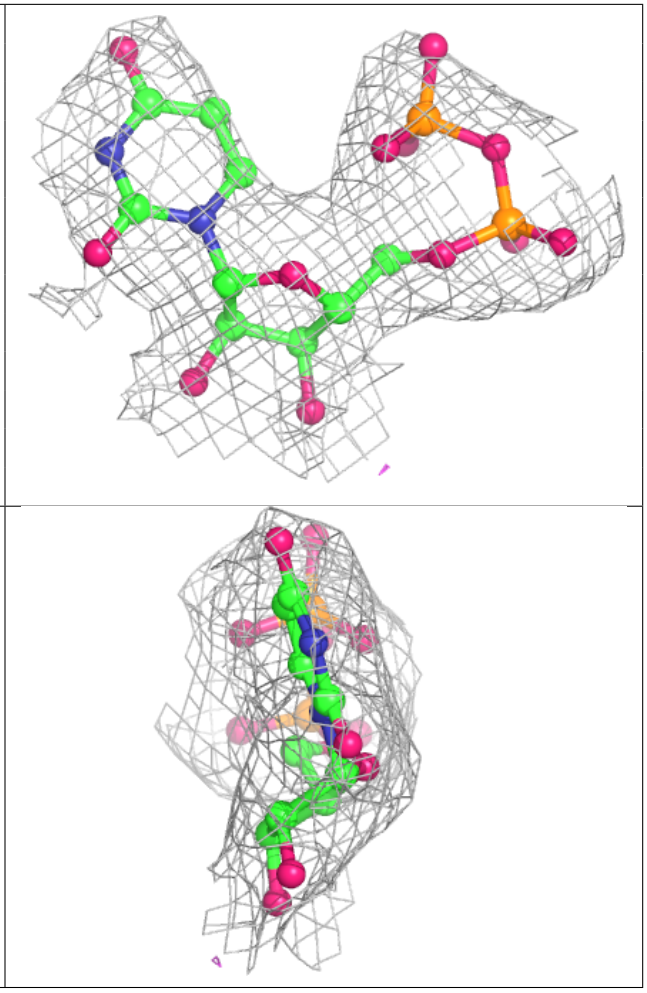






Electron density around UDP D 1237:

$2mF_o$-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 (i)

There are no such residues in this entry.