wwPDB X-ray Structure Validation Summary Report (i)

Oct 16, 2023 – 06:43 PM EDT

PDB ID	:	1YA6
Title	:	alpha-glucosyltransferase in complex with UDP and a 13-mer DNA containing a central A:G mismatch
Authors	:	Lariviere, L.; Sommer, N.; Morera, S.
Deposited on	:	2004-12-17
Resolution	:	2.40 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:

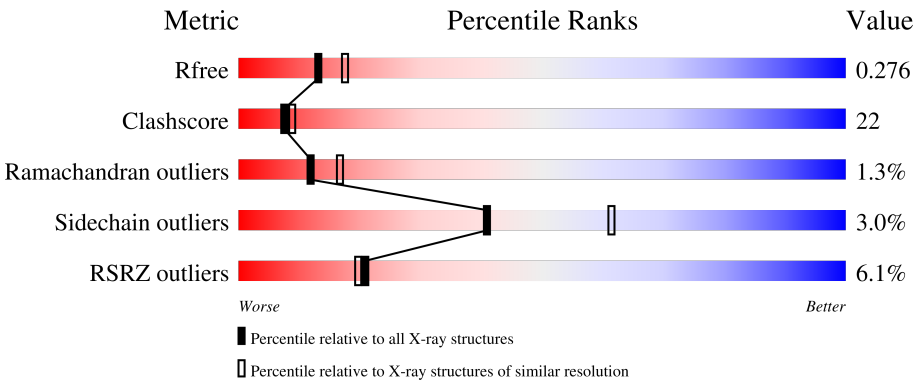
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (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 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	C	12	50% 42% 8%
2	D	12	75% 25%
3	A	403	6% 60% 32% • 6%
3	B	403	5% 51% 40% • 6%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 6977 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 DNA chain called 5'-D(*AP*TP*AP*CP*TP*AP*AP*GP*AP*TP*AP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	12	246	119	49	67	11	0	0	0

- Molecule 2 is a DNA chain called 5'-D(*CP*TP*AP*TP*CP*TP*GP*AP*GP*TP*AP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	12	242	118	41	72	11	0	0	0

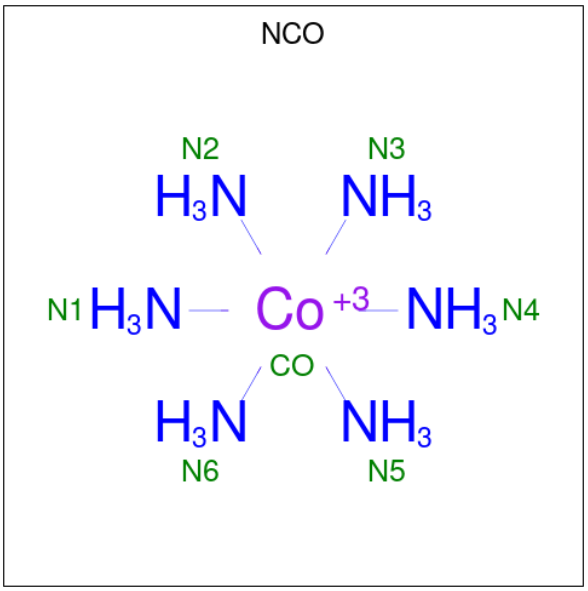
- Molecule 3 is a protein called DNA alpha-glucosyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	380	3118	1986	532	584	16	0	0	0
3	B	377	3091	1970	527	577	17	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

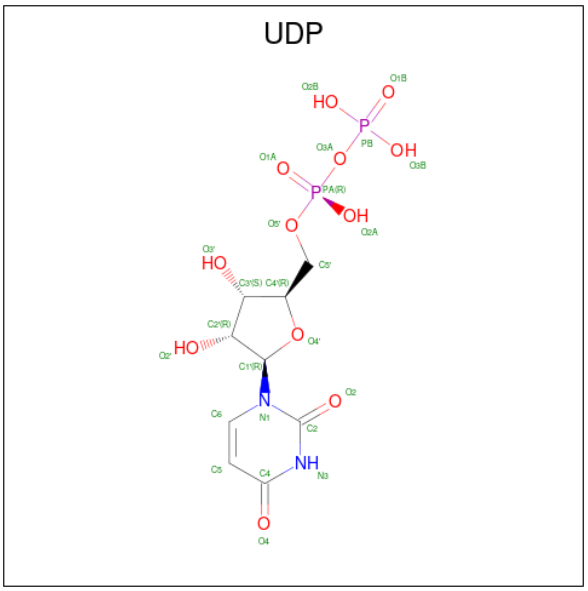
Chain	Residue	Modelled	Actual	Comment	Reference
A	998	MET	-	cloning artifact	UNP P04519
A	999	GLY	-	cloning artifact	UNP P04519
A	1000	SER	-	cloning artifact	UNP P04519
B	998	MET	-	cloning artifact	UNP P04519
B	999	GLY	-	cloning artifact	UNP P04519
B	1000	SER	-	cloning artifact	UNP P04519

- Molecule 4 is COBALT HEXAMMINE(III) (three-letter code: NCO) (formula: $CoH_{18}N_6$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	Co	N		
4	D	1	7	1	6	0	0
4	A	1	7	1	6	0	0
4	A	1	7	1	6	0	0
4	A	1	7	1	6	0	0
4	B	1	7	1	6	0	0
4	B	1	7	1	6	0	0
4	B	1	7	1	6	0	0

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



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

- Molecule 6 is water.

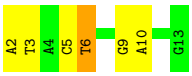
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	C	5	5	5	0	0
6	D	6	6	6	0	0
6	A	89	89	89	0	0
6	B	81	81	81	0	0

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.

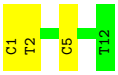
• Molecule 1: 5'-D(*AP*TP*AP*CP*TP*AP*AP*GP*AP*TP*AP*G)-3'

Chain C:



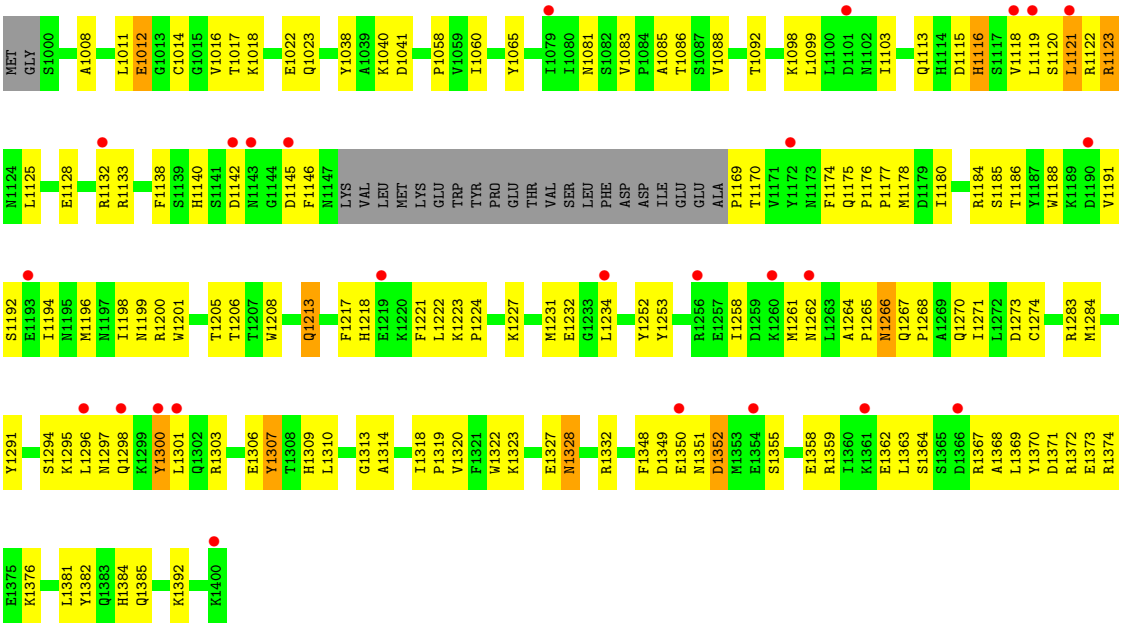
• Molecule 2: 5'-D(*CP*TP*AP*TP*CP*TP*GP*AP*GP*TP*AP*T)-3'

Chain D:



• Molecule 3: DNA alpha-glucosyltransferase

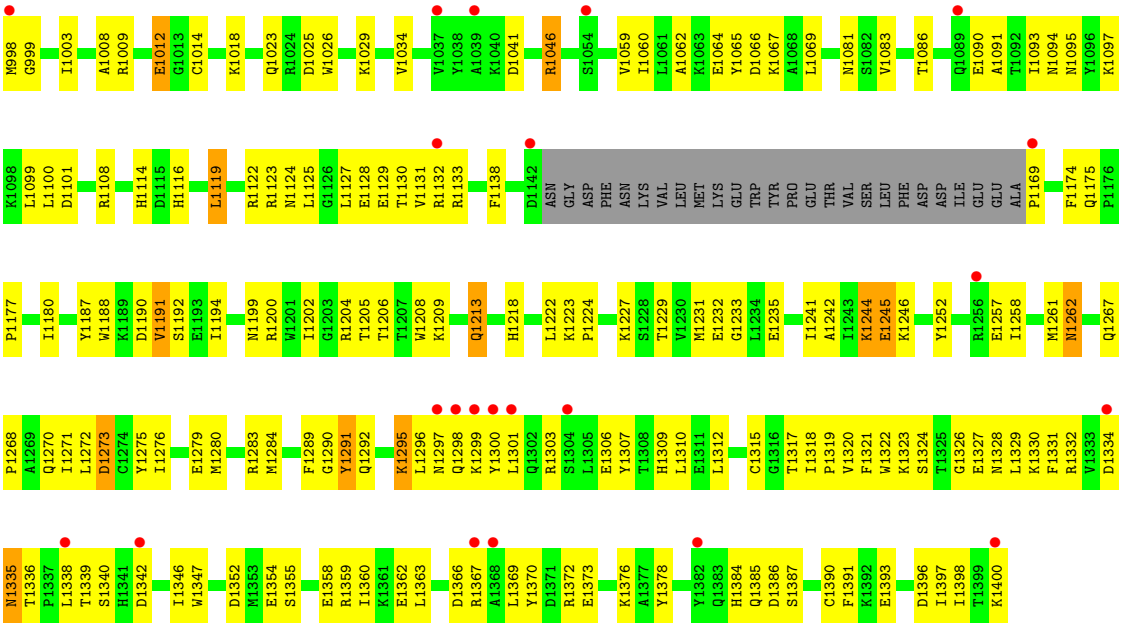
Chain A:



• Molecule 3: DNA alpha-glucosyltransferase

Chain B:

5%
51%
40%
6%


M998 G999 I1003 A1008 R1009 E1012 G1013 C1014 K1018 Q1023 R1024 D1025 W1026 K1029 V1034 V1037 Y1038 A1039 K1040 D1041 R1046 S1054 V1059 I1060 L1061 A1062 K1063 E1064 Y1065 D1066 K1067 A1068 L1069 N1081 S1082 V1083 T1086 Q1089 E1090 A1091 T1092 I1093 N1094 N1095 Y1096 K1097
K1098 L1099 L1100 D1101 R1108 H1114 D1115 H1116 L1119 R1122 R1123 N1124 L1125 G1126 L1127 E1128 E1129 T1130 V1131 R1132 R1133 F1138 D1142 ASN GLY ASP PHE ASN LYS VAL LEU MET LYS GLU TRP TYR PRO GLU THR VAL SER LEU PHE ASP ASP ILE GLU GLU ALA P1169 F1174 Q1175 P1176
P1177 I1180 Y1187 W1188 K1189 D1190 V1191 S1192 E1193 I1194 N1199 R1200 W1201 I1202 G1203 R1204 T1205 T1206 T1207 W1208 K1209 Q1213 H1218 L1222 K1223 P1224 K1227 S1228 T1229 V1230 M1231 E1232 G1233 L1234 E1235 I1241 A1242 I1243 K1244 E1245 K1246 Y1252 R1256 E1257 I1258 M1261 N1262 Q1267
P1268 A1269 Q1270 I1271 L1272 D1273 C1274 Y1275 I1276 E1279 M1280 R1283 M1284 F1289 G1290 Y1291 Q1292 K1296 L1296 N1297 Q1298 K1299 Y1300 L1301 Q1302 R1303 S1304 L1305 E1306 Y1307 T1308 H1309 L1310 E1311 L1312 C1315 G1316 T1317 I1318 P1319 V1320 F1321 W1322 K1323 S1324 T1325 G1326 E1327 N1328 L1329 K1330 F1331 R1332 V1333 D1334
N1335 T1336 P1337 L1338 T1339 S1340 H1341 D1342 I1346 W1347 D1352 M1353 E1354 S1355 E1358 R1359 I1360 K1361 E1362 L1363 D1366 R1367 A1368 L1369 Y1370 D1371 R1372 E1373 K1376 A1377 Y1378 Y1382 Q1383 H1384 Q1385 D1386 S1387 C1390 F1391 K1392 E1393 D1396 I1397 I1398 T1399 K1400

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	45.85Å 119.32Å 86.81Å 90.00° 92.87° 90.00°	Depositor
Resolution (Å)	20.00 – 2.40 19.98 – 2.40	Depositor EDS
% Data completeness (in resolution range)	(Not available) (20.00-2.40) 99.9 (19.98-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	2.30 (at 2.41Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.230 , 0.280 0.223 , 0.276	Depositor DCC
R_{free} test set	1795 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	50.4	Xtriage
Anisotropy	0.718	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 50.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.030 for h,-k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6977	wwPDB-VP
Average B, all atoms (Å^2)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.49% of the height of the origin peak. No significant pseudotranslation is detected.*

[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: UDP, NCO

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	C	0.48	0/277	0.76	0/426
2	D	0.44	0/270	0.71	0/415
3	A	0.41	0/3190	0.63	0/4301
3	B	0.39	0/3162	0.62	0/4262
All	All	0.41	0/6899	0.64	0/9404

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	C	0	1
3	A	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	1307	TYR	Sidechain
1	C	6	DT	Sidechain

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	C	246	0	137	4	0
2	D	242	0	139	6	0
3	A	3118	0	3060	143	0
3	B	3091	0	3044	141	0
4	A	21	0	0	3	0
4	B	21	0	0	4	0
4	D	7	0	0	0	0
5	A	25	0	11	0	0
5	B	25	0	11	1	0
6	A	89	0	0	3	0
6	B	81	0	0	3	0
6	C	5	0	0	0	0
6	D	6	0	0	0	0
All	All	6977	0	6402	290	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 22.

The worst 5 of 290 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)
3:A:1213:GLN:H	3:A:1213:GLN:NE2	1.68	0.92
3:A:1118:VAL:HG12	3:A:1145:ASP:HB2	1.55	0.87
3:A:1113:GLN:HE22	3:A:1115:ASP:HB2	1.40	0.85
3:A:1116:HIS:CE1	4:A:2002:NCO:N3	2.45	0.85
3:A:1318:ILE:HG12	3:A:1370:TYR:CE1	2.15	0.82

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	
3	A	376/403 (93%)	336 (89%)	36 (10%)	4 (1%)	14	20
3	B	373/403 (93%)	336 (90%)	31 (8%)	6 (2%)	9	13
All	All	749/806 (93%)	672 (90%)	67 (9%)	10 (1%)	12	17

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	1352	ASP
3	B	1191	VAL
3	B	1335	ASN
3	A	1185	SER
3	B	1273	ASP

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	
3	A	347/368 (94%)	338 (97%)	9 (3%)	46	66
3	B	344/368 (94%)	332 (96%)	12 (4%)	36	55
All	All	691/736 (94%)	670 (97%)	21 (3%)	41	61

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	1245	GLU
3	B	1291	TYR
3	B	1340	SER
3	B	1292	GLN
3	B	1262	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37

such sidechains are listed below:

Mol	Chain	Res	Type
3	B	1255	ASN
3	B	1384	HIS
3	B	1267	GLN
3	B	1335	ASN
3	A	1328	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

9 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	NCO	D	1003	-	6,6,6	1.43	0	-		
4	NCO	A	2004	-	6,6,6	1.32	0	-		
4	NCO	B	3002	-	6,6,6	1.25	0	-		
4	NCO	A	2002	-	6,6,6	1.17	0	-		
4	NCO	B	3004	-	6,6,6	1.34	0	-		
4	NCO	A	2003	-	6,6,6	1.39	0	-		
5	UDP	A	2001	-	24,26,26	0.57	0	37,40,40	0.58	1 (2%)
5	UDP	B	3001	-	24,26,26	0.58	0	37,40,40	0.62	1 (2%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	NCO	B	3003	-	6,6,6	1.33	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
5	UDP	B	3001	-	-	5/16/32/32	0/2/2/2
5	UDP	A	2001	-	-	3/16/32/32	0/2/2/2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	2001	UDP	O2B-PB-O1B	2.24	119.46	110.68
5	B	3001	UDP	O2B-PB-O1B	2.18	119.22	110.68

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	2001	UDP	PA-O3A-PB-O1B
5	A	2001	UDP	PB-O3A-PA-O5'
5	B	3001	UDP	PB-O3A-PA-O5'
5	B	3001	UDP	O4'-C4'-C5'-O5'
5	A	2001	UDP	O4'-C4'-C5'-O5'

There are no ring outliers.

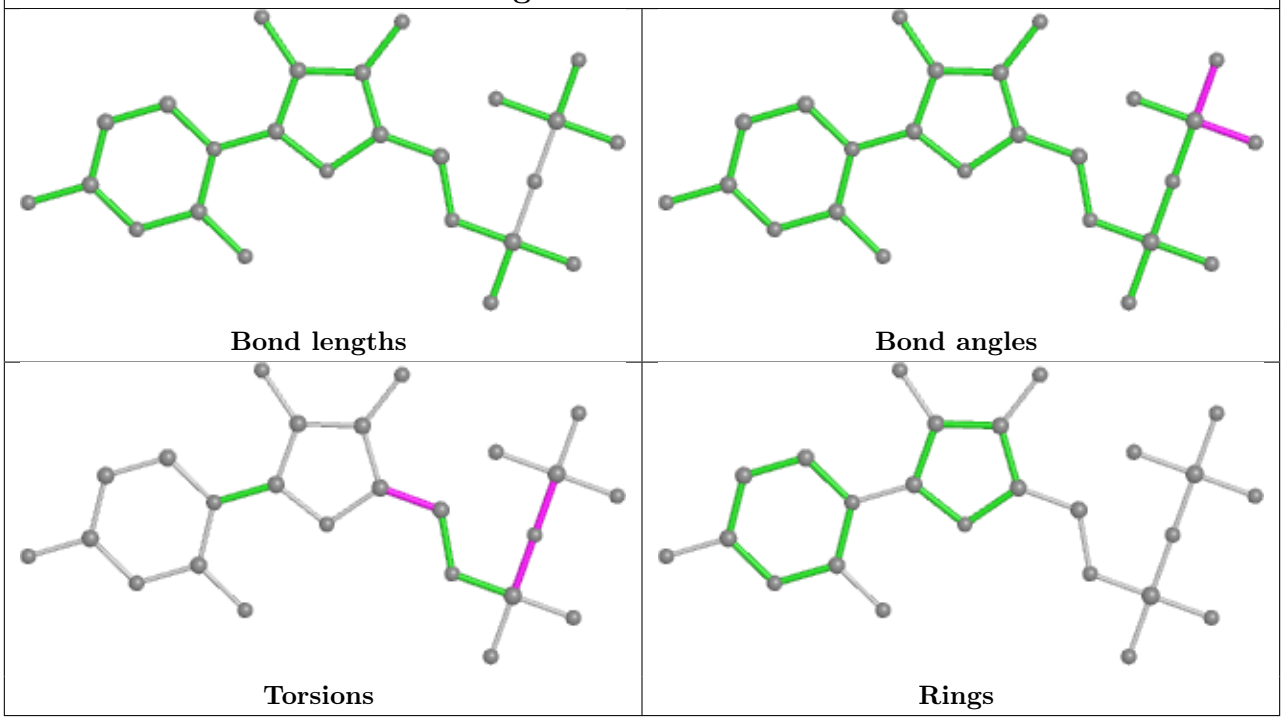
3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	3002	NCO	4	0
4	A	2002	NCO	3	0
5	B	3001	UDP	1	0

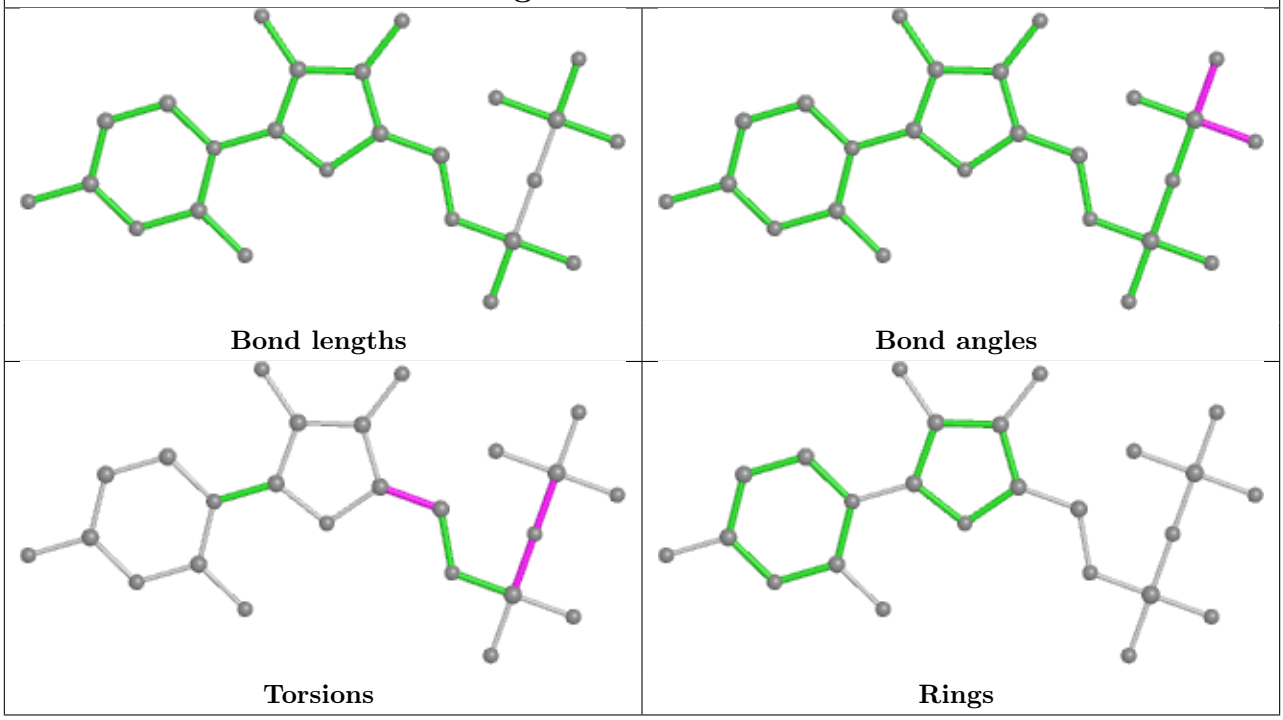
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. 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 A 2001



Ligand UDP B 3001



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	C	12/12 (100%)	0.14	0	100	100	38, 62, 71, 72	0
2	D	12/12 (100%)	-0.24	0	100	100	50, 60, 74, 74	0
3	A	380/403 (94%)	0.28	26 (6%)	17	15	34, 60, 87, 100	0
3	B	377/403 (93%)	0.33	22 (5%)	23	22	39, 65, 84, 97	0
All	All	781/830 (94%)	0.29	48 (6%)	21	20	34, 62, 86, 100	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	1143	ASN	7.1
3	A	1145	ASP	5.3
3	B	1300	TYR	4.3
3	A	1296	LEU	3.8
3	B	1132	ARG	3.4

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)

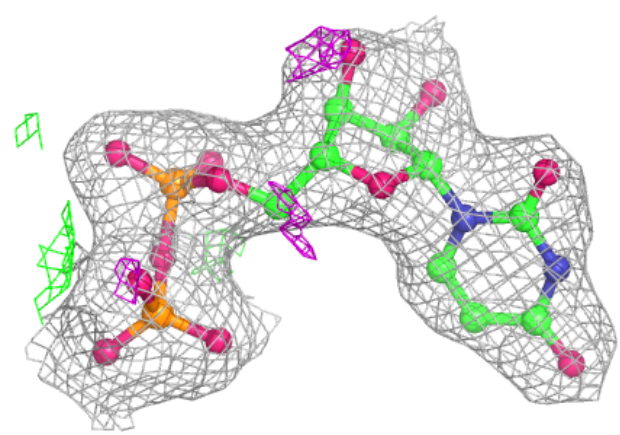
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

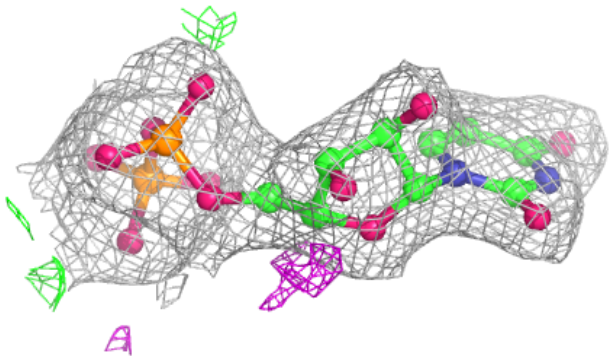
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	NCO	B	3004	7/7	0.92	0.20	71,71,71,73	7
4	NCO	A	2004	7/7	0.93	0.19	92,92,93,93	7
4	NCO	B	3003	7/7	0.94	0.17	98,99,99,99	0
4	NCO	B	3002	7/7	0.94	0.22	78,79,79,79	0
4	NCO	A	2002	7/7	0.96	0.19	63,63,65,66	0
4	NCO	A	2003	7/7	0.96	0.15	81,82,84,84	0
4	NCO	D	1003	7/7	0.97	0.14	88,89,89,91	0
5	UDP	A	2001	25/25	0.98	0.08	38,44,46,48	0
5	UDP	B	3001	25/25	0.98	0.10	37,43,47,49	0

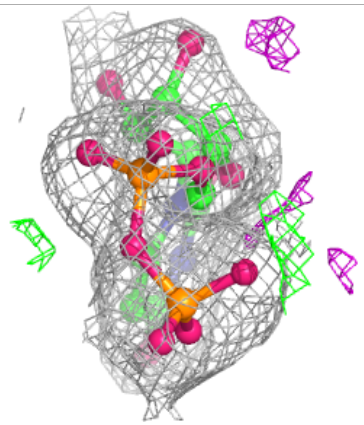
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around UDP A 2001:

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

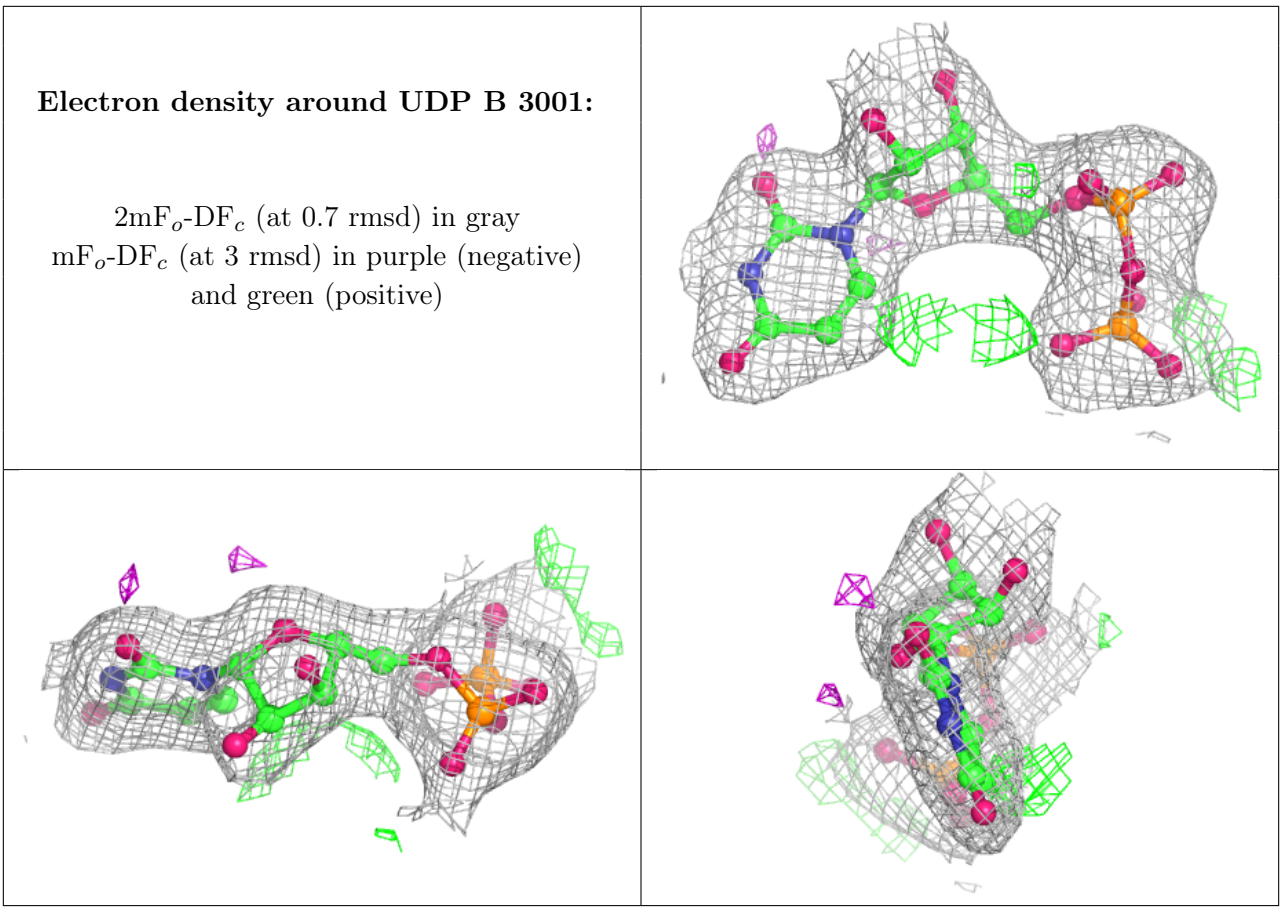






Electron density around UDP B 3001:

$2mF_o$-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.