W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2021 – 04:19 PM EDT

PDB ID	:	1GUQ
Title	:	STRUCTURE OF NUCLEOTIDYLTRANSFERASE COMPLEXED WITH UDP-GLUCOSE
Authors	:	Thoden, J.B.; Rayment, I.; Holden, H.
Deposited on	:	1996-10-23
Resolution	:	1.80 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

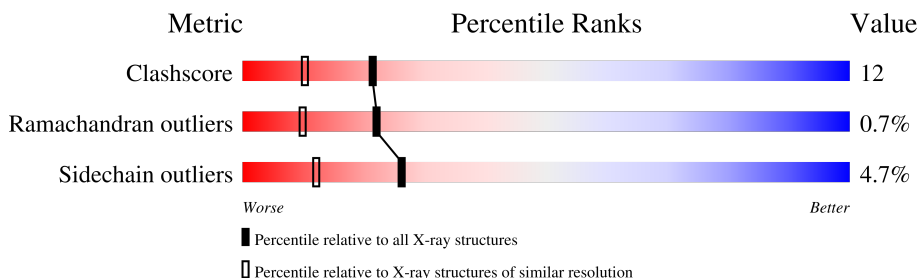|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 1.80 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	Percentile Ranks	Value
Clashscore		12
Ramachandran outliers		0.7%
Sidechain outliers		4.7%

Worse — Better

▮ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	348	70% ... 25% ... ·
1	B	348	67% ... 27% ... · ··
1	C	348	61% ... 32% ... 5% ··
1	D	348	54% ... 37% ... 7% ··

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 12597 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 GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	347	Total 2784	C 1763	N 492	O 516	S 13	0	0	0
1	B	344	Total 2766	C 1753	N 489	O 511	S 13	0	0	0
1	C	344	Total 2766	C 1753	N 489	O 511	S 13	0	0	0
1	D	344	Total 2766	C 1753	N 489	O 511	S 13	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	166	GLY	HIS	engineered mutation	UNP P09148
B	166	GLY	HIS	engineered mutation	UNP P09148
C	166	GLY	HIS	engineered mutation	UNP P09148
D	166	GLY	HIS	engineered mutation	UNP P09148

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

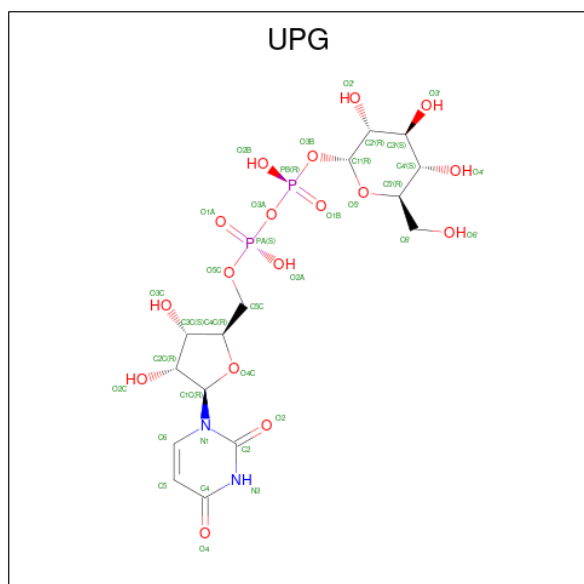| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 1	Zn 1	0	0
2	B	1	Total 1	Zn 1	0	0
2	C	1	Total 1	Zn 1	0	0
2	D	1	Total 1	Zn 1	0	0

- Molecule 3 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Fe 1	0	0
3	B	1	Total 1	Fe 1	0	0
3	C	1	Total 1	Fe 1	0	0
3	D	1	Total 1	Fe 1	0	0

- Molecule 4 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total 1	K 1	0	0
4	B	1	Total 1	K 1	0	0
4	C	1	Total 1	K 1	0	0
4	D	1	Total 1	K 1	0	0

- Molecule 5 is URIDINE-5'-DIPHOSPHATE-GLUCOSE (three-letter code: UPG) (formula: $C_{15}H_{24}N_2O_{17}P_2$).

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

Continued on next page...

Continued from previous page...

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

- Molecule 6 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	368	Total	O	0	0
			368	368		
6	B	363	Total	O	0	0
			363	363		
6	C	351	Total	O	0	0
			351	351		
6	D	277	Total	O	0	0
			277	277		

3 Residue-property plots ⓘ

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. 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. 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.
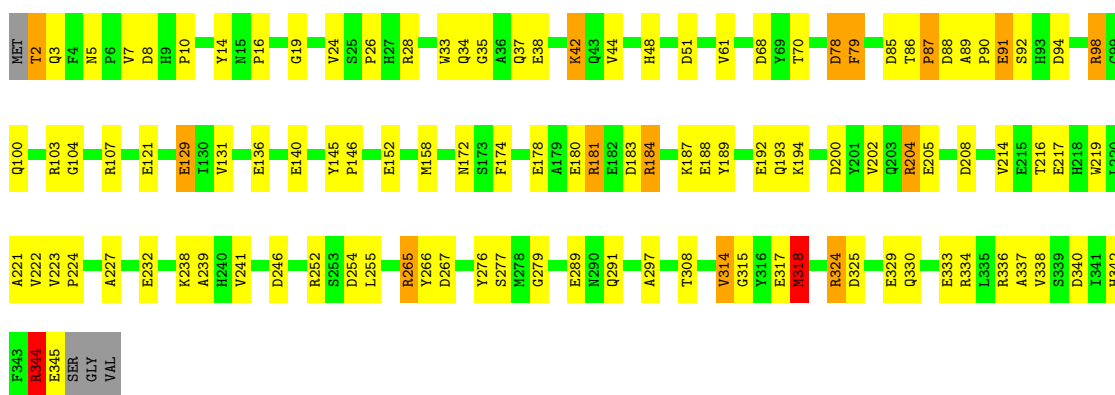
Note EDS was not executed.

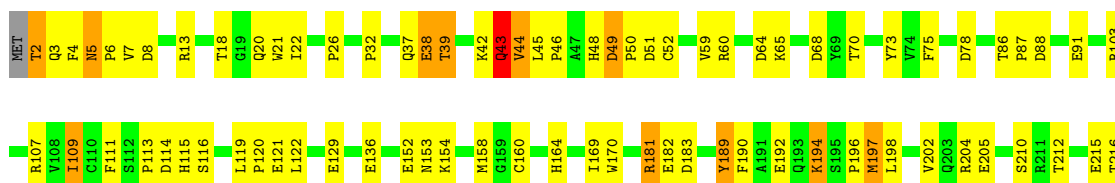• Molecule 1: GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE
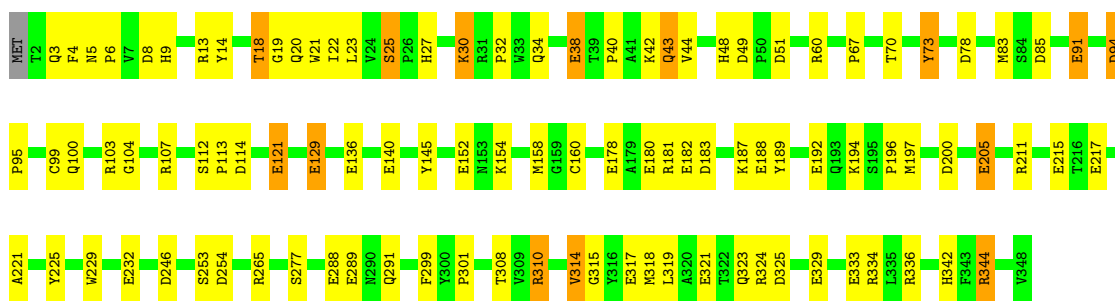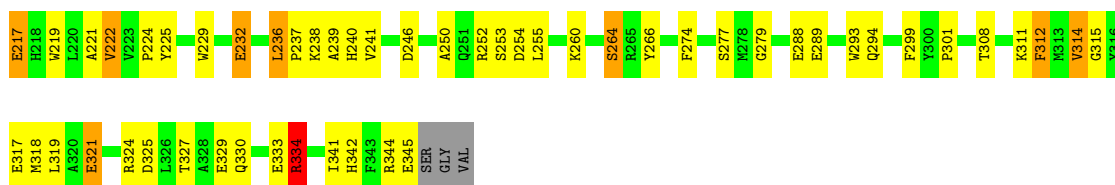

• Molecule 1: GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE
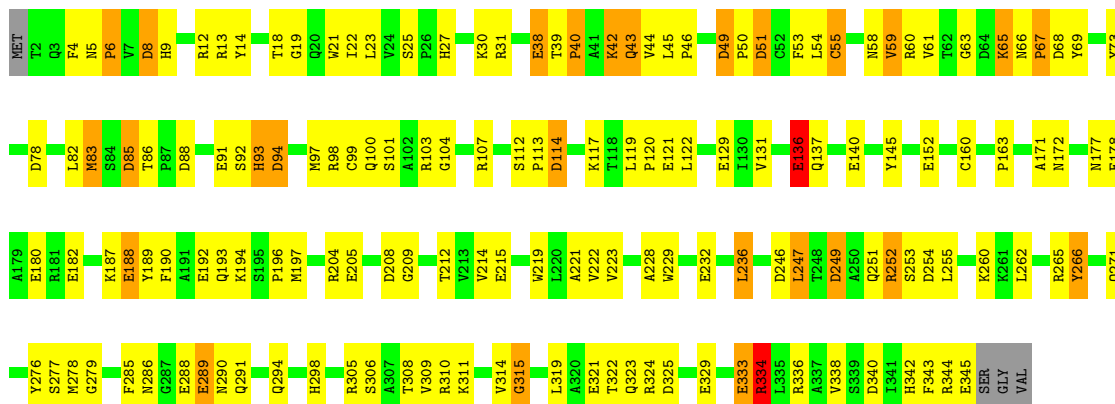

• Molecule 1: GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE

E217 H218 W219 L220 A221 V222 P224 Y225 W229 E232 L236 P237 K238 A239 H240 V241 D246 A250 Q251 R252 S253 D254 L255 K260 S264 R265 Y266 F274 S277 M278 G279 E288 E289 W293 Q294 F299 Y300 P301 T308 K311 F312 M313 V314 G315 Y316

E317 M318 L319 A320 E321 R324 D325 L326 T327 A328 E329 Q330 E333 R334 I341 H342 F343 R344 E345 SER GLY VAL

● Molecule 1: GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE

Chain D:

54%   37%   7%   ··

MET T2 Q3 F4 N5 P6 V7 D8 H9 R12 R13 Y14 T18 G19 Q20 W21 I22 L23 V24 S25 P26 H27 K30 R31 E38 T39 P40 A41 K42 Q43 V44 L45 P46 D49 P50 D51 C52 F53 L54 C55 N58 V59 R60 V61 T62 G63 D64 K65 N66 P67 D68 Y69 Y73

D78 L82 M83 S84 D85 T86 P87 D88 E91 S92 H93 D94 M97 R98 C99 Q100 S101 A102 R103 G104 R107 S112 P113 D114 K117 T118 L119 P120 E121 L122 E129 I130 V131 E136 Q137 E140 Y145 E152 C160 P163 A171 N172 N177 E178

A179 E180 R181 E182 K187 E188 Y189 F190 A191 E192 Q193 K194 S195 P196 M197 R204 E205 D208 G209 T212 V213 V214 E215 W219 L220 A221 V222 V223 A228 W229 E232 L236 D246 L247 T248 D249 A250 Q251 R252 S253 D254 L255 K260 K261 L262 R265 Y266 Q271

Y276 S277 M278 G279 F285 N286 G287 E288 E289 N290 Q291 Q294 H298 R305 S306 A307 T308 V309 R310 K311 V314 G315 L319 A320 E321 T322 Q323 R324 D325 E329 E333 R334 L335 R336 A337 V338 S339 D340 I341 H342 F343 R344 E345 SER GLY VAL

4 Data and refinement statistics ⓘ

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	68.40Å 57.50Å 188.90Å 90.00° 100.13° 90.00°	Depositor
Resolution (Å)	30.00 – 1.80	Depositor
% Data completeness (in resolution range)	(Not available) (30.00-1.80)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	TNT	Depositor
R, R_{free}	0.174 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	12597	wwPDB-VP
Average B, all atoms (Å^2)	25.0	wwPDB-VP

5　Model quality　ⓘ

5.1　Standard geometry　ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: FE, K, UPG, ZN

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	1.04	21/2869 (0.7%)	1.48	40/3912 (1.0%)
1	B	1.03	19/2851 (0.7%)	1.50	52/3889 (1.3%)
1	C	1.04	18/2851 (0.6%)	1.51	48/3889 (1.2%)
1	D	1.01	18/2851 (0.6%)	1.51	60/3889 (1.5%)
All	All	1.03	76/11422 (0.7%)	1.50	200/15579 (1.3%)

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	1	0

The worst 5 of 76 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	205	GLU	CD-OE2	9.70	1.36	1.25
1	C	329	GLU	CD-OE1	9.11	1.35	1.25
1	B	152	GLU	CD-OE1	8.44	1.34	1.25
1	C	232	GLU	CD-OE1	8.09	1.34	1.25
1	A	121	GLU	CD-OE1	8.05	1.34	1.25

The worst 5 of 200 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(º)	Ideal(º)
1	C	334	ARG	NE-CZ-NH1	11.25	125.92	120.30
1	B	336	ARG	NE-CZ-NH2	-11.07	114.77	120.30
1	C	107	ARG	NE-CZ-NH2	-10.98	114.81	120.30
1	B	324	ARG	NE-CZ-NH1	9.80	125.20	120.30

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	254	ASP	CB-CG-OD2	-9.71	109.56	118.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	C	44	VAL	CA

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	2784	0	2661	53	0
1	B	2766	0	2644	54	0
1	C	2766	0	2644	77	0
1	D	2766	0	2644	96	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	36	0	22	0	0
5	B	36	0	22	0	0
5	C	36	0	22	0	0
5	D	36	0	22	1	0
6	A	368	0	0	9	1
6	B	363	0	0	11	1
6	C	351	0	0	6	0
6	D	277	0	0	6	0
All	All	12597	0	10681	263	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 12.

The worst 5 of 263 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:289:GLU:HG2	1:A:291:GLN:NE2	1.47	1.26
1:A:289:GLU:CG	1:A:291:GLN:HE22	1.68	1.07
1:C:21:TRP:CH2	1:D:100:GLN:HG3	1.94	1.01
1:A:38:GLU:HB2	1:A:308:THR:HA	1.44	0.99
1:D:342:HIS:CE1	1:D:344:ARG:HB2	1.96	0.99

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)
6:A:471:HOH:O	6:B:675:HOH:O[1_655]	2.18	0.02

5.3   Torsion angles   ⓘ

5.3.1   Protein backbone   ⓘ

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	345/348 (99%)	327 (95%)	16 (5%)	2 (1%)	25	12
1	B	342/348 (98%)	324 (95%)	16 (5%)	2 (1%)	25	12
1	C	342/348 (98%)	319 (93%)	19 (6%)	4 (1%)	13	3
1	D	342/348 (98%)	326 (95%)	14 (4%)	2 (1%)	25	12
All	All	1371/1392 (98%)	1296 (94%)	65 (5%)	10 (1%)	22	10

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	44	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	315	GLY
1	C	43	GLN
1	D	314	VAL
1	A	314	VAL

5.3.2 Protein sidechains ⓘ

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	297/298 (100%)	287 (97%)	10 (3%)	37	22
1	B	295/298 (99%)	285 (97%)	10 (3%)	37	22
1	C	295/298 (99%)	280 (95%)	15 (5%)	24	10
1	D	295/298 (99%)	275 (93%)	20 (7%)	16	5
All	All	1182/1192 (99%)	1127 (95%)	55 (5%)	26	12

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	210	SER
1	D	8	ASP
1	D	334	ARG
1	D	249	ASP
1	C	217	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	27	HIS
1	C	48	HIS
1	D	100	GLN
1	C	37	GLN
1	C	100	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 16 ligands modelled in this entry, 12 are monoatomic - leaving 4 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
5	UPG	D	352	-	31,38,38	1.54	4 (12%)	41,58,58	1.19	2 (4%)
5	UPG	C	352	-	31,38,38	1.53	4 (12%)	41,58,58	1.21	5 (12%)
5	UPG	B	352	-	31,38,38	1.44	3 (9%)	41,58,58	1.40	7 (17%)
5	UPG	A	352	-	31,38,38	1.27	5 (16%)	41,58,58	1.53	6 (14%)

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	UPG	D	352	-	-	5/21/59/59	0/3/3/3
5	UPG	C	352	-	-	9/21/59/59	0/3/3/3
5	UPG	B	352	-	-	8/21/59/59	0/3/3/3
5	UPG	A	352	-	-	7/21/59/59	0/3/3/3

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	C	352	UPG	C4-N3	4.99	1.41	1.33
5	D	352	UPG	C4-N3	4.61	1.41	1.33
5	B	352	UPG	C4-N3	4.16	1.40	1.33
5	A	352	UPG	C4-N3	3.54	1.39	1.33
5	A	352	UPG	PB-O3B	3.48	1.69	1.60

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	352	UPG	O4C-C1C-C2C	-4.51	100.33	106.93
5	A	352	UPG	C5-C4-N3	-4.35	113.75	123.31
5	B	352	UPG	C5-C4-N3	-4.07	114.36	123.31
5	C	352	UPG	C5-C4-N3	-3.86	114.82	123.31
5	D	352	UPG	C5-C4-N3	-3.74	115.08	123.31

There are no chirality outliers.

5 of 29 torsion outliers are listed below:

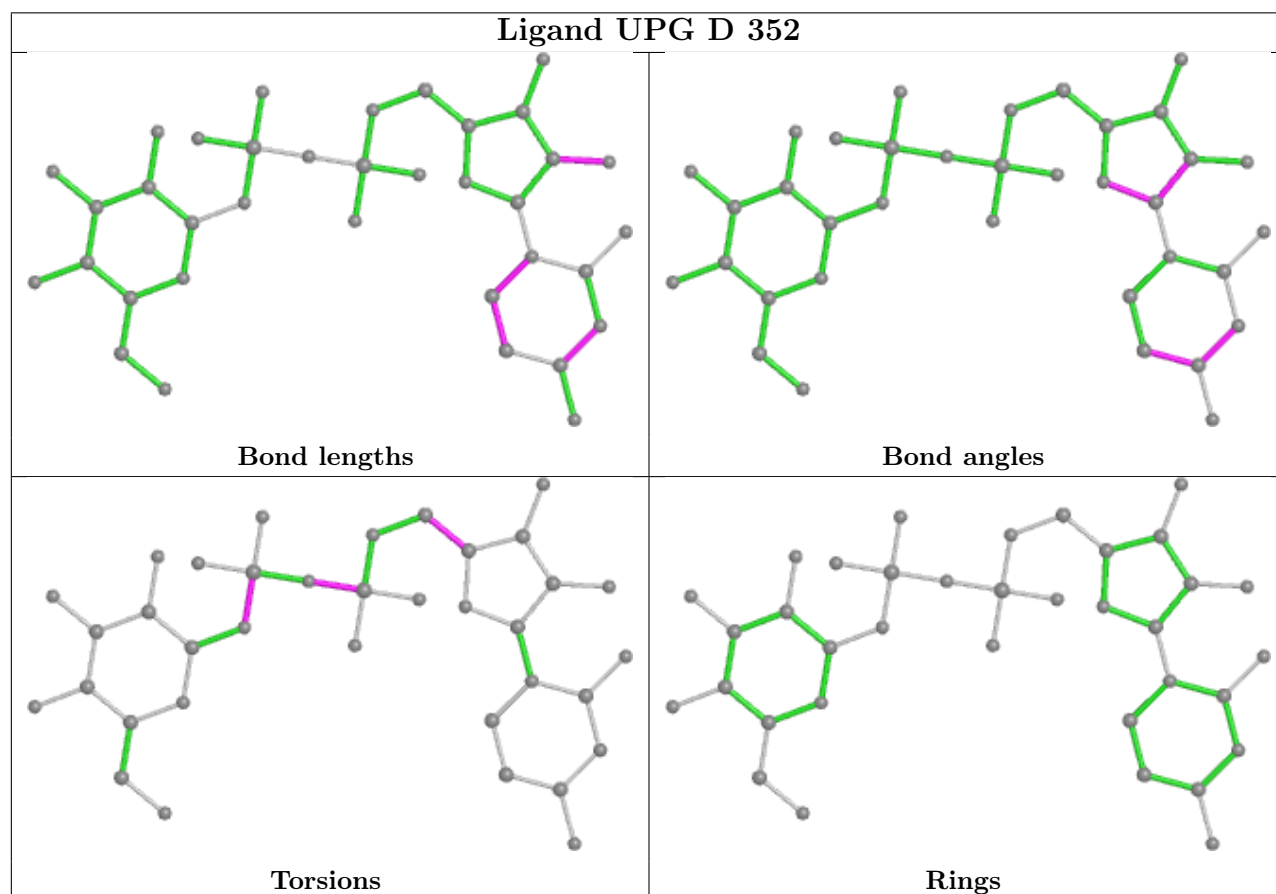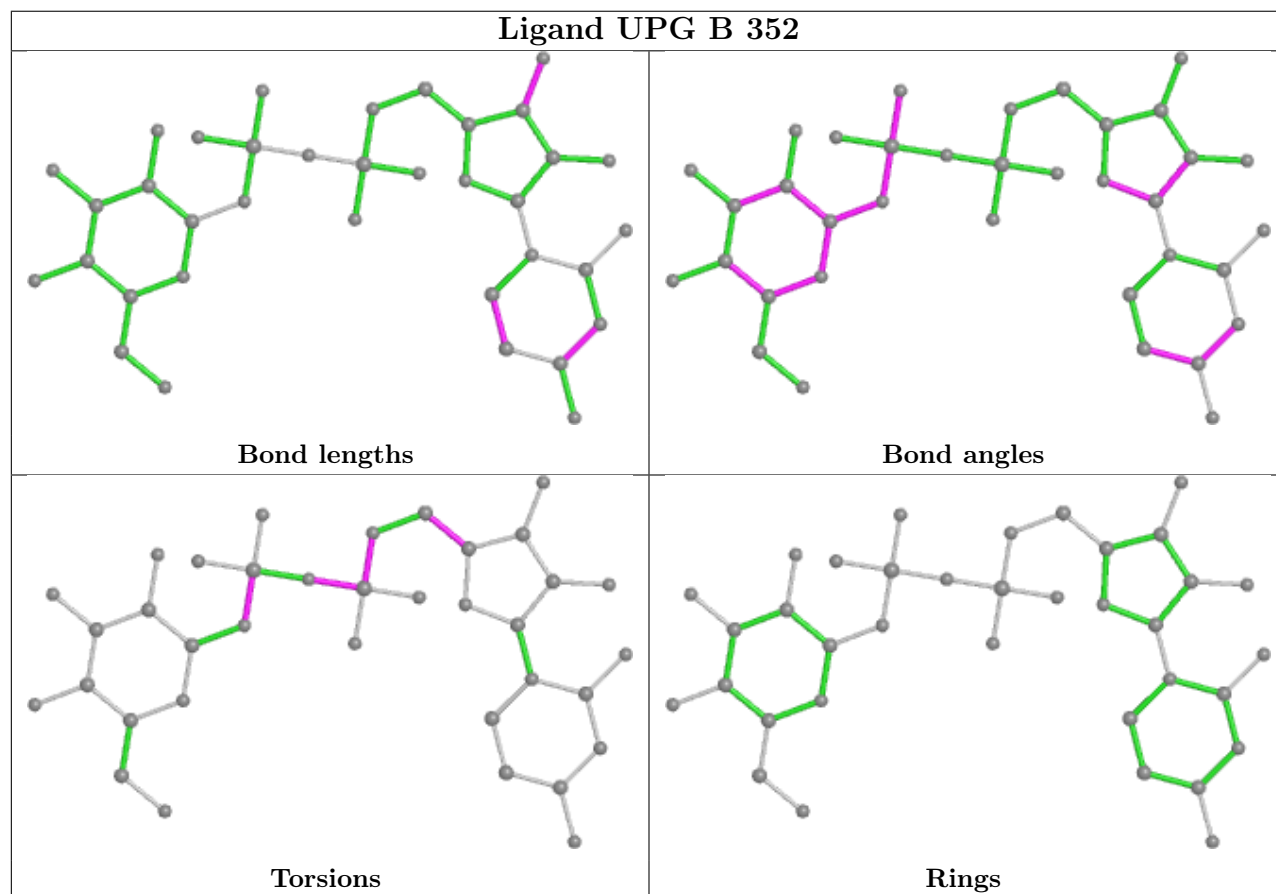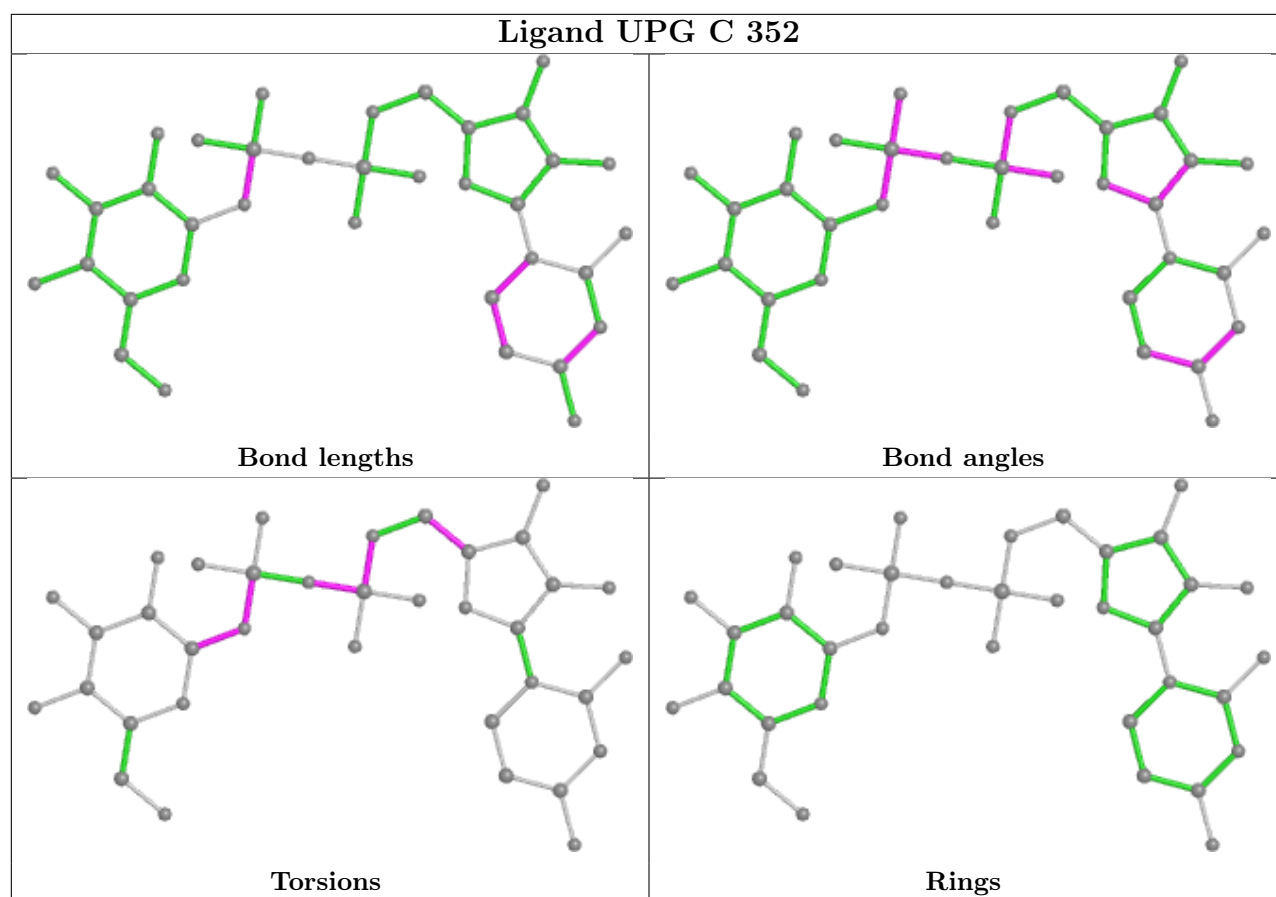| Mol	Chain	Res	Type	Atoms
5	A	352	UPG	C5C-O5C-PA-O1A
5	A	352	UPG	C1'-O3B-PB-O3A
5	B	352	UPG	C5C-O5C-PA-O1A
5	B	352	UPG	C1'-O3B-PB-O3A
5	C	352	UPG	C5C-O5C-PA-O1A

There are no ring outliers.
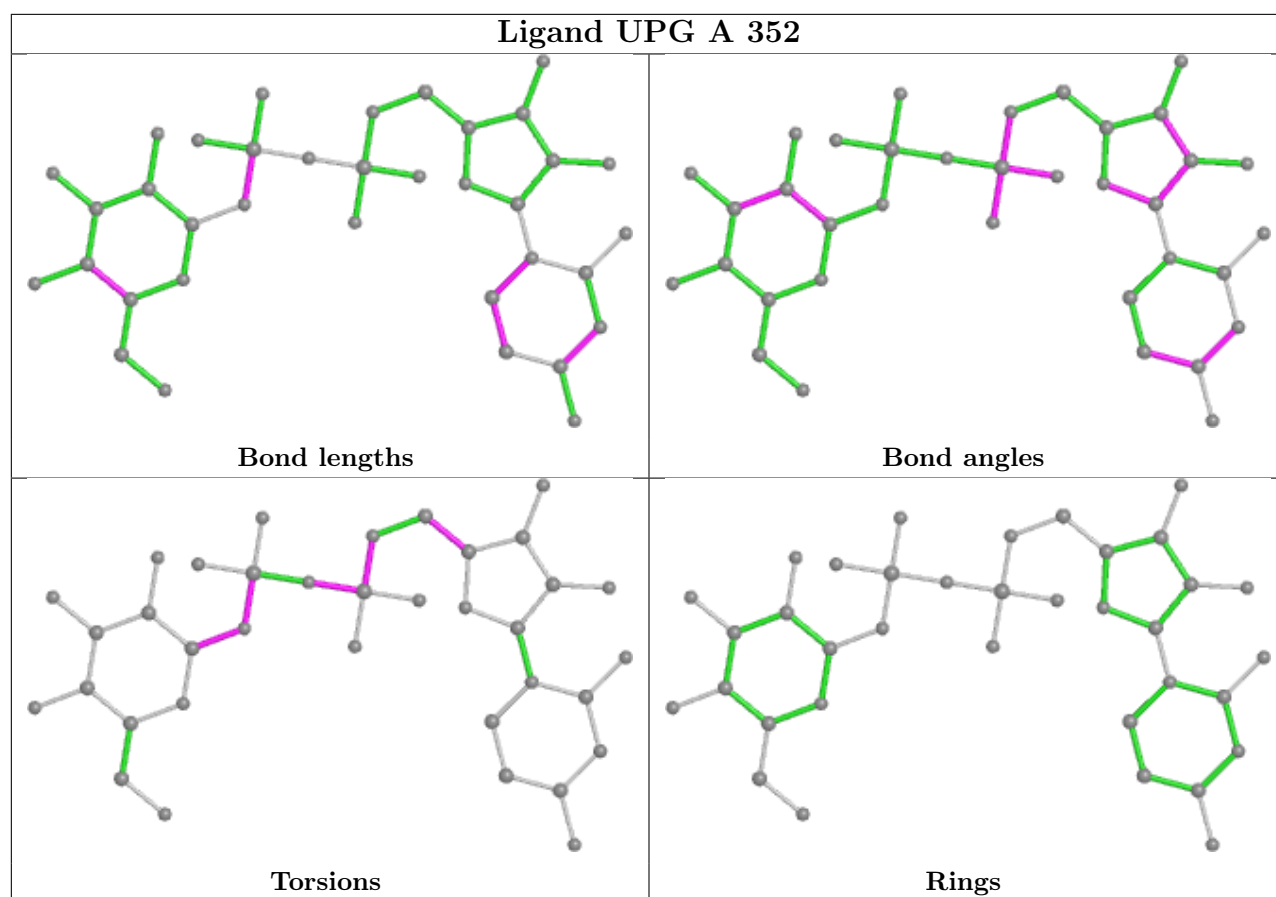
1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	D	352	UPG	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 UPG D 352

Bond lengths

Bond angles

Torsions

Rings

Ligand UPG C 352

Bond lengths

Bond angles

Torsions

Rings

Ligand UPG B 352

Bond lengths

Bond angles

Torsions

Rings

Ligand UPG A 352

Bond lengths

Bond angles

Torsions

Rings

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

6.4 Ligands ⓘ

EDS was not executed - this section is therefore empty.

6.5 Other polymers ⓘ

EDS was not executed - this section is therefore empty.