# W O R L D W I D E
# wwPDB
## P R O T E I N   D A T A   B A N K

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

Jun 24, 2024 – 11:16 PM EDT

|  |  |  |
|---|---|---|
| PDB ID | : | 6IWR |
| Title | : | Crystal structure of GalNAc-T7 with UDP, GalNAc and Mn2+ |
| Authors | : | Yu, C.; Yin, Y.X. |
| Deposited on | : | 2018-12-06 |
| Resolution | : | 2.60 Å(reported) |

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

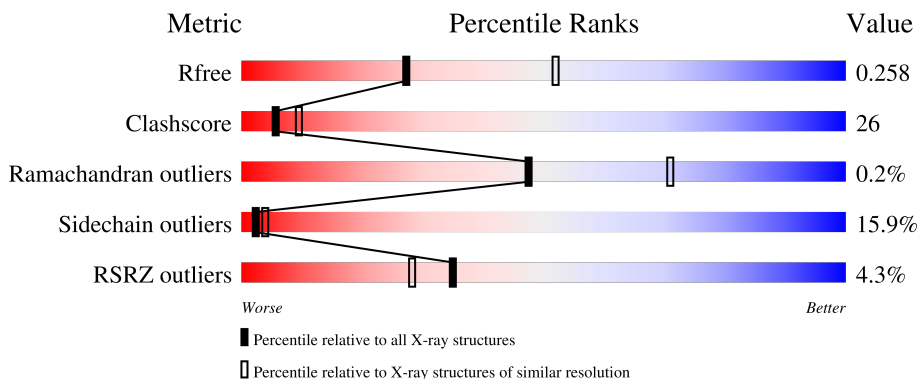|  |  |  |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.37.1 |
| buster-report | : | 1.1.7 (2018) |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 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) | : | 2.37.1 |

# 1    Overall quality at a glance ⓘ

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

The reported resolution of this entry is 2.60 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 |
|--------|------------------|-------|
| Rfree | | 0.258 |
| Clashscore | | 26 |
| Ramachandran outliers | | 0.2% |
| Sidechain outliers | | 15.9% |
| RSRZ outliers | | 4.3% |

*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)) |
|--------|--------------------------|------------------------------------------------------|
| R$_{free}$ | 130704 | 3163 (2.60-2.60) |
| Clashscore | 141614 | 3518 (2.60-2.60) |
| Ramachandran outliers | 138981 | 3455 (2.60-2.60) |
| Sidechain outliers | 138945 | 3455 (2.60-2.60) |
| RSRZ outliers | 127900 | 3104 (2.60-2.60) |

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 | 597 | 3% / 51% 32% 7% 9% |
| 1 | B | 597 | 2% / 50% 32% 9% 9% |
| 1 | C | 597 | 5% / 48% 34% 9% 9% |
| 1 | D | 597 | 2% / 52% 32% 6% · 10% |
| 1 | E | 597 | 5% / 53% 31% 7% 9% |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Length | Quality of chain |
|-----|-------|--------|------------------|
| 1 | F | 597 | |

7%

53% 32% 7% 9%

# 2   Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 27098 atoms, of which 26 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 N-acetylgalactosaminyltransferase 7.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 546 | Total 4440 | C 2855 | N 757 | O 805 | S 23 | 0 | 0 | 0 |
| 1 | B | 546 | Total 4440 | C 2855 | N 757 | O 805 | S 23 | 0 | 0 | 0 |
| 1 | C | 546 | Total 4440 | C 2855 | N 757 | O 805 | S 23 | 0 | 0 | 0 |
| 1 | D | 538 | Total 4380 | C 2814 | N 748 | O 795 | S 23 | 0 | 0 | 0 |
| 1 | E | 546 | Total 4440 | C 2855 | N 757 | O 805 | S 23 | 0 | 0 | 0 |
| 1 | F | 546 | Total 4440 | C 2855 | N 757 | O 805 | S 23 | 0 | 0 | 0 |

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| 2 | A | 1 | Total 1 | Mn 1 | 0 | 0 |
| 2 | B | 1 | Total 1 | Mn 1 | 0 | 0 |
| 2 | C | 1 | Total 1 | Mn 1 | 0 | 0 |
| 2 | D | 1 | Total 1 | Mn 1 | 0 | 0 |
| 2 | E | 1 | Total 1 | Mn 1 | 0 | 0 |
| 2 | F | 1 | Total 1 | Mn 1 | 0 | 0 |

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

NAG



| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---|---|---|---------|---------|
| 3 | D | 1 | Total | C | H | N | O | 0 | 0 |
| | | | 30 | 8 | 15 | 1 | 6 | | |

- Molecule 4 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula: $C_9H_{14}N_2O_{12}P_2$).

UDP
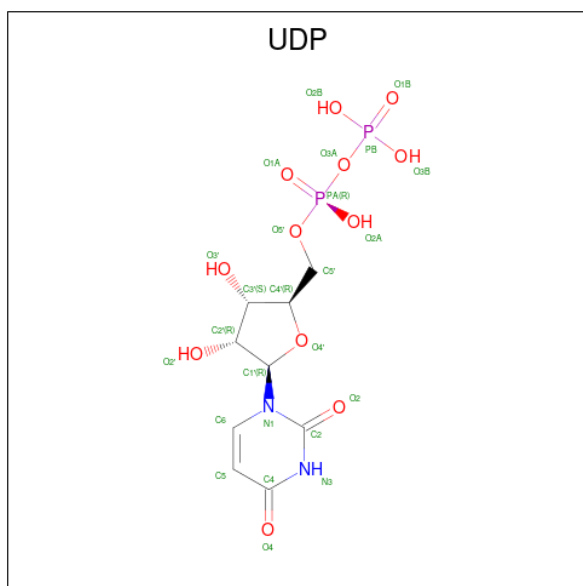


| Mol | Chain | Residues | Atoms | | | | | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---|---|---|---|---------|---------|
| 4 | D | 1 | Total | C | H | N | O | P | 0 | 0 |
| | | | 36 | 9 | 11 | 2 | 12 | 2 | | |

- Molecule 5 is URIDINE-DIPHOSPHATE-N-ACETYLGALACTOSAMINE (three-letter code: UD2) (formula: $C_{17}H_{27}N_3O_{17}P_2$).

UD2
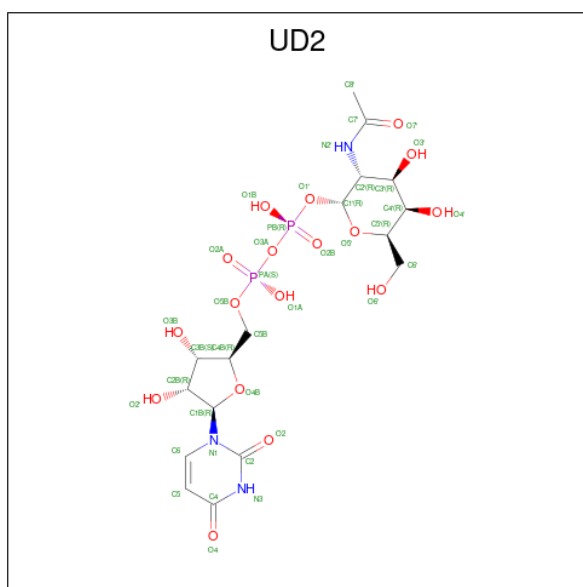
| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---|---|---|---------|---------|
| 5 | D | 1 | Total | C | N | O | P | 0 | 0 |
| | | | 39 | 17 | 3 | 17 | 2 | | |

- Molecule 6 is water.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---------|---------|
| 6 | A | 100 | Total | O | 0 | 0 |
| | | | 100 | 100 | | |
| 6 | B | 77 | Total | O | 0 | 0 |
| | | | 77 | 77 | | |
| 6 | C | 60 | Total | O | 0 | 0 |
| | | | 60 | 60 | | |
| 6 | D | 55 | Total | O | 0 | 0 |
| | | | 55 | 55 | | |
| 6 | E | 68 | Total | O | 0 | 0 |
| | | | 68 | 68 | | |
| 6 | F | 47 | Total | O | 0 | 0 |
| | | | 47 | 47 | | |

# 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 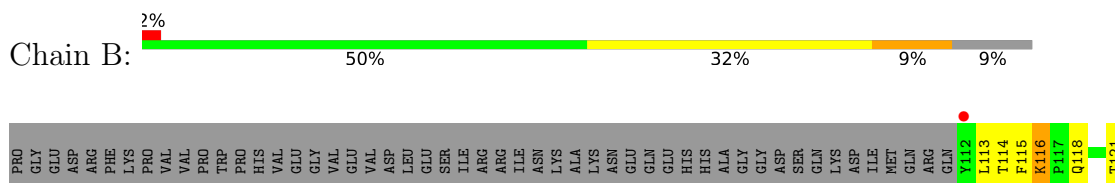the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: 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: N-acetylgalactosaminyltransferase 7

Chain A:



• Molecule 1: N-acetylgalactosaminyltransferase 7

Chain B:

• Molecule 1: N-acetylgalactosaminyltransferase 7

Chain C:
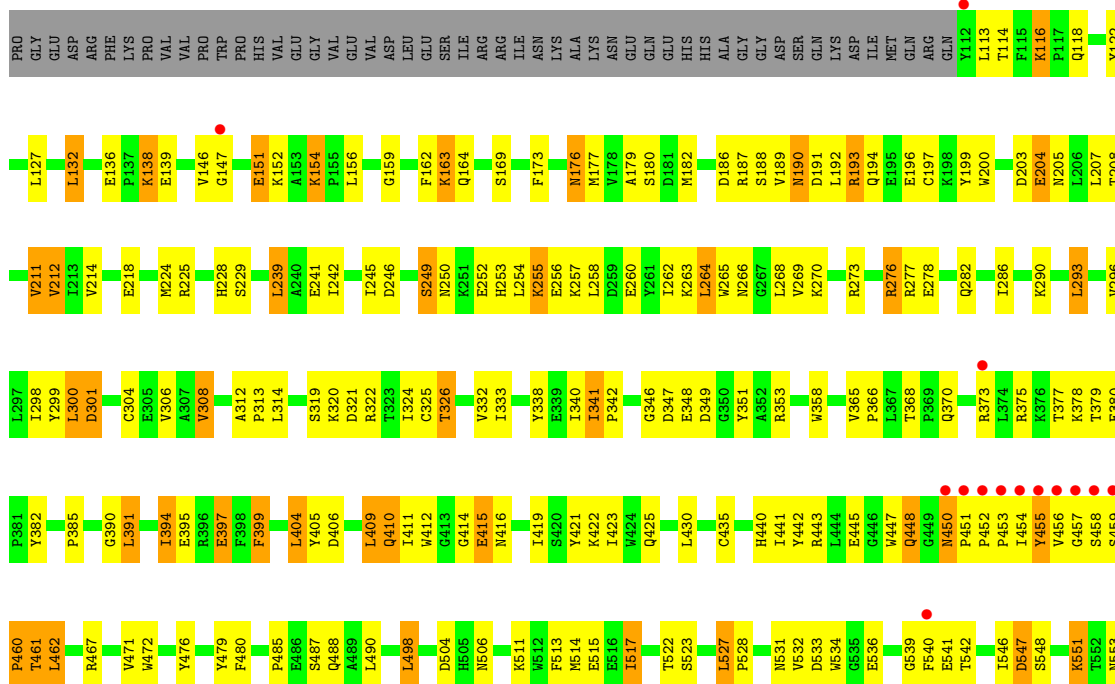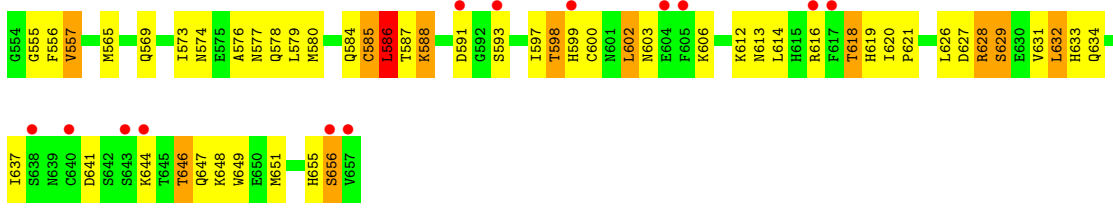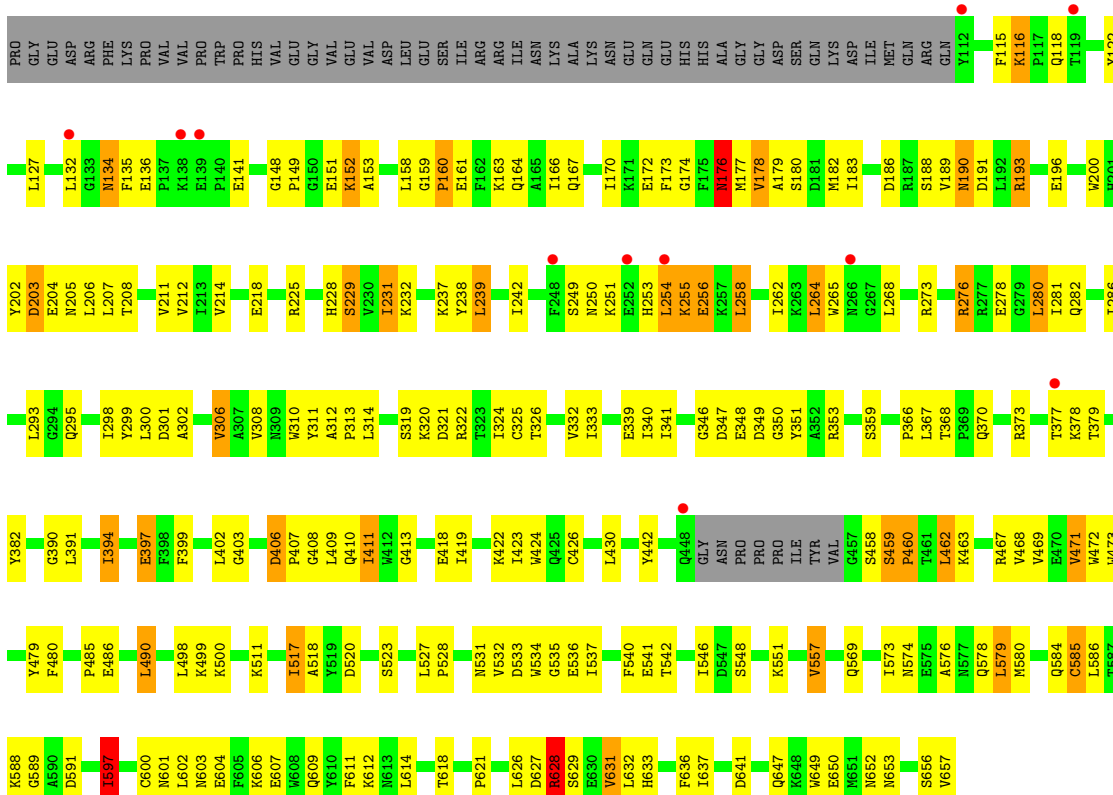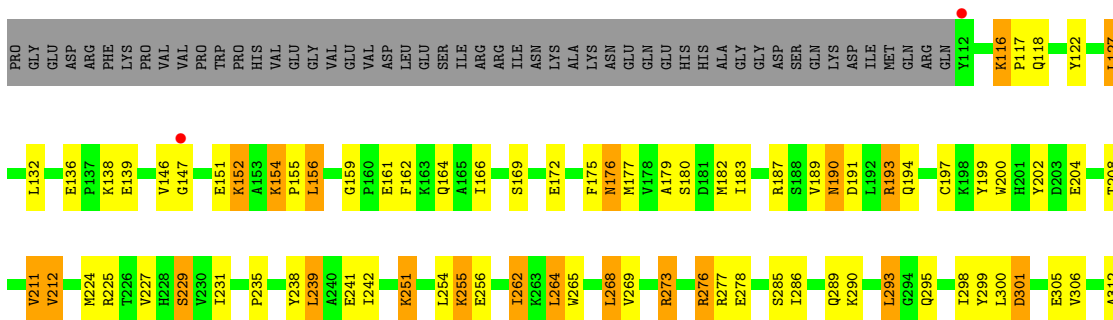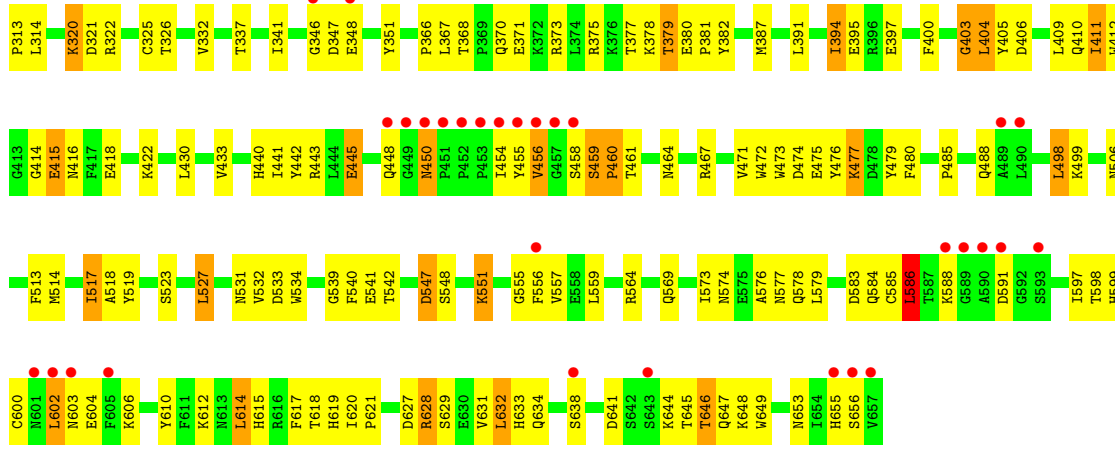
G554 G555 F556 V557 M565 Q569 I573 N574 E575 A576 N577 Q578 L579 M580 Q584 C585 L586 T587 K588 D591 G592 S593 I597 T598 H599 C600 N601 L602 N603 E604 F605 K606 K612 N613 L614 H615 R616 F617 T618 H619 I620 P621 L626 D627 R628 S629 E630 V631 L632 H633 Q634

I637 S638 N639 C640 D641 S642 S643 K644 T645 T646 Q647 K648 W649 E650 M651 H655 S656 V657

• Molecule 1: N-acetylgalactosaminyltransferase 7

Chain D:

2%

52%          32%          6% •  10%

PRO GLY GLU ASP ARG PHE LYS PRO VAL VAL PRO TRP PRO HIS VAL GLU GLY VAL GLU VAL ASP GLU SER ILE ARG ARG ILE ASN LYS ALA LYS ASN GLU GLN GLN HIS HIS ALA GLY GLY ASP SER GLN LYS ASP ILE MET GLN ARG GLN Y112 F115 K116 P117 Q118 T119 Y122

L127 L132 G133 N134 F135 E136 P137 K138 E139 P140 E141 G148 P149 G150 E151 K152 A153 L158 G159 P160 E161 F162 K163 Q164 A165 I166 Q167 I170 K171 E172 F173 G174 F175 M176 M177 V178 A179 S180 D181 M182 I183 D186 R187 S188 V189 N190 D191 L192 R193 E196 W200 H201

Y202 D203 E204 N205 L206 L207 T208 V211 V212 I213 V214 E218 R225 H228 S229 V230 I231 K232 K237 Y238 L239 I242 F248 S249 N250 K251 E252 H253 L254 K255 E256 K257 L258 I262 K263 L264 W265 N266 G267 L268 R273 R276 R277 E278 G279 L280 I281 Q282 I286

L293 G294 Q295 I298 Y299 L300 D301 A302 V306 A307 V308 N309 W310 Y311 A312 P313 L314 S319 K320 D321 R322 T323 I324 C325 T326 V332 I333 E339 I340 I341 G346 D347 E348 D349 G350 Y351 A352 R353 S359 P366 L367 T368 P369 Q370 R373 T377 K378 T379

Y382 G390 L391 I394 E397 F398 F399 L402 G403 D406 P407 G408 L409 Q410 I411 W412 G413 E418 I419 K422 I423 W424 Q425 C426 L430 Y442 Q448 GLY ASN PRO PRO PRO ILE TYR VAL G457 S458 S459 P460 T461 L462 K463 R467 V468 V469 E470 V471 W472 W473

Y479 F480 P485 E486 L490 L498 K499 K500 K511 I517 A518 Y519 D520 S523 L527 P528 N531 V532 D533 W534 G535 E536 I537 F540 E541 T542 I546 D547 S548 K551 V557 Q569 I573 N574 E575 A576 N577 Q578 L579 M580 Q584 C585 L586 T587

K588 G589 A590 D591 I597 C600 N601 L602 N603 E604 F605 K606 E607 W608 Q609 Y610 F611 K612 N613 L614 T618 P621 L626 D627 R628 S629 E630 V631 L632 H633 F636 I637 D641 Q647 K648 W649 E650 M651 N652 N653 S656 V657

• Molecule 1: N-acetylgalactosaminyltransferase 7

Chain E:

5%

53%          31%          7%   9%

PRO GLY GLU ASP ARG PHE LYS PRO VAL VAL PRO TRP PRO HIS VAL GLU GLY VAL GLU VAL ASP GLU SER ILE ARG ARG ILE ASN LYS ALA LYS ASN GLU GLN GLN HIS HIS ALA GLY GLY ASP SER GLN LYS ASP ILE MET GLN ARG GLN Y112 K116 P117 Q118 Y122 L127

L132 E136 P137 K138 E139 V146 G147 E151 K152 A153 K154 P155 L156 G159 P160 E161 F162 K163 Q164 A165 I166 S169 E172 F175 M176 M177 V178 A179 S180 D181 M182 I183 R187 S188 V189 N190 D191 L192 R193 Q194 C197 K198 Y199 W200 H201 Y202 D203 E204 T208

V211 V212 M224 R225 T226 V227 H228 S229 V230 I231 P235 Y238 L239 A240 E241 I242 K251 L254 K255 E256 I262 K263 L264 W265 L268 V269 R273 R276 R277 E278 S285 I286 Q289 K290 L293 G294 Q295 I298 Y299 L300 D301 E305 V306 A312

• Molecule 1: N-acetylgalactosaminyltransferase 7

Chain F:

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 21 21 21 | Depositor |
| Cell constants a, b, c, $\alpha$, $\beta$, $\gamma$ | 137.44Å 158.26Å 251.48Å<br>90.00° 90.00° 90.00° | Depositor |
| Resolution (Å) | 49.25 – 2.60 | Depositor |
| | 49.25 – 2.60 | EDS |
| % Data completeness (in resolution range) | 95.7 (49.25-2.60) | Depositor |
| | 95.7 (49.25-2.60) | EDS |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 2.84 (at 2.61Å) | Xtriage |
| Refinement program | PHENIX (dev_2400: ???) | Depositor |
| R, $R_{free}$ | 0.220 , 0.258 | Depositor |
| | 0.220 , 0.258 | DCC |
| $R_{free}$ test set | 8025 reflections (5.00%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 43.2 | Xtriage |
| Anisotropy | 0.130 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.33 , 43.6 | EDS |
| L-test for twinning[2] | $< |L| > = 0.48, < L^2 > = 0.31$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| $F_o,F_c$ correlation | 0.92 | EDS |
| Total number of atoms | 27098 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 47.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.65% 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 ⓘ

## 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: UD2, UDP, MN, NAG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

| Mol | Chain | Bond lengths | | Bond angles | |
|-----|-------|------|------------|------|------------|
| | | RMSZ | #$|Z| > 5$ | RMSZ | #$|Z| > 5$ |
| 1 | A | 0.56 | 2/4569 (0.0%) | 0.71 | 2/6192 (0.0%) |
| 1 | B | 0.55 | 2/4569 (0.0%) | 0.70 | 4/6192 (0.1%) |
| 1 | C | 0.54 | 3/4569 (0.1%) | 0.68 | 4/6192 (0.1%) |
| 1 | D | 0.54 | 4/4504 (0.1%) | 0.70 | 6/6098 (0.1%) |
| 1 | E | 0.51 | 3/4569 (0.1%) | 0.66 | 3/6192 (0.0%) |
| 1 | F | 0.47 | 0/4569 | 0.62 | 2/6192 (0.0%) |
| All | All | 0.53 | 14/27349 (0.1%) | 0.68 | 21/37058 (0.1%) |

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 | A | 0 | 1 |
| 1 | B | 0 | 2 |
| 1 | E | 0 | 1 |
| 1 | F | 0 | 1 |
| All | All | 0 | 5 |

The worst 5 of 14 bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|-----|-------|-----|------|-------|------|-------------|----------|
| 1 | A | 585 | CYS | CB-SG | -7.43 | 1.69 | 1.82 |
| 1 | B | 585 | CYS | CB-SG | -6.53 | 1.71 | 1.82 |
| 1 | D | 585 | CYS | CB-SG | -5.51 | 1.72 | 1.81 |
| 1 | E | 585 | CYS | CB-SG | -5.38 | 1.73 | 1.81 |
| 1 | A | 320 | LYS | CD-CE | 5.38 | 1.64 | 1.51 |

The worst 5 of 21 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|-----|-------|-----|------|-------|-----|------------|---------|
| 1 | D | 403 | GLY | N-CA-C | 8.30 | 133.84 | 113.10 |
| 1 | B | 148 | GLY | C-N-CD | -7.82 | 103.39 | 120.60 |
| 1 | A | 320 | LYS | CD-CE-NZ | 6.68 | 127.06 | 111.70 |
| 1 | E | 403 | GLY | N-CA-C | 6.63 | 129.67 | 113.10 |
| 1 | B | 547 | ASP | CB-CG-OD1 | 6.62 | 124.26 | 118.30 |

There are no chirality outliers.

All (5) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group |
|-----|-------|-----|------|-------|
| 1 | A | 410 | GLN | Peptide |
| 1 | B | 402 | LEU | Peptide |
| 1 | B | 413 | GLY | Peptide |
| 1 | E | 403 | GLY | Peptide |
| 1 | F | 402 | LEU | Peptide |

## 5.2 Too-close contacts ⓘ

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 | 4440 | 0 | 4325 | 212 | 0 |
| 1 | B | 4440 | 0 | 4325 | 254 | 0 |
| 1 | C | 4440 | 0 | 4325 | 244 | 0 |
| 1 | D | 4380 | 0 | 4265 | 228 | 0 |
| 1 | E | 4440 | 0 | 4325 | 200 | 0 |
| 1 | F | 4440 | 0 | 4325 | 227 | 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 |
| 2 | E | 1 | 0 | 0 | 0 | 0 |
| 2 | F | 1 | 0 | 0 | 0 | 0 |
| 3 | D | 15 | 15 | 15 | 1 | 0 |
| 4 | D | 25 | 11 | 11 | 2 | 0 |
| 5 | D | 39 | 0 | 25 | 3 | 0 |
| 6 | A | 100 | 0 | 0 | 8 | 0 |
| 6 | B | 77 | 0 | 0 | 6 | 0 |
| 6 | C | 60 | 0 | 0 | 1 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|-----|-------|-------|----------|----------|---------|--------------|
| 6 | D | 55 | 0 | 0 | 10 | 0 |
| 6 | E | 68 | 0 | 0 | 1 | 0 |
| 6 | F | 47 | 0 | 0 | 4 | 0 |
| All | All | 27072 | 26 | 25941 | 1346 | 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 26.

The worst 5 of 1346 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|--------|--------|--------------------------|-------------------|
| 1:B:456:VAL:HG23 | 1:B:458:SER:CA | 1.41 | 1.47 |
| 1:F:453:PRO:HB2 | 1:F:455:TYR:CZ | 1.56 | 1.37 |
| 1:F:453:PRO:HB2 | 1:F:455:TYR:CE2 | 1.62 | 1.34 |
| 1:B:454:ILE:HG22 | 1:B:457:GLY:CA | 1.62 | 1.27 |
| 1:C:454:ILE:O | 1:C:458:SER:HA | 1.39 | 1.22 |

There are no symmetry-related clashes.

## 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 | 544/597 (91%) | 516 (95%) | 26 (5%) | 2 (0%) | 34 | 57 |
| 1 | B | 544/597 (91%) | 520 (96%) | 24 (4%) | 0 | 100 | 100 |
| 1 | C | 544/597 (91%) | 515 (95%) | 27 (5%) | 2 (0%) | 34 | 57 |
| 1 | D | 534/597 (89%) | 513 (96%) | 20 (4%) | 1 (0%) | 47 | 71 |
| 1 | E | 544/597 (91%) | 518 (95%) | 26 (5%) | 0 | 100 | 100 |
| 1 | F | 544/597 (91%) | 509 (94%) | 32 (6%) | 3 (1%) | 25 | 47 |
| All | All | 3254/3582 (91%) | 3091 (95%) | 155 (5%) | 8 (0%) | 47 | 71 |

5 of 8 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | A | 459 | SER |
| 1 | F | 411 | ILE |
| 1 | F | 413 | GLY |
| 1 | D | 176 | ASN |
| 1 | C | 453 | PRO |

### 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 | 480/525 (91%) | 400 (83%) | 80 (17%) | 2 | 3 |
| 1 | B | 480/525 (91%) | 395 (82%) | 85 (18%) | 2 | 3 |
| 1 | C | 480/525 (91%) | 397 (83%) | 83 (17%) | 2 | 3 |
| 1 | D | 473/525 (90%) | 418 (88%) | 55 (12%) | 5 | 10 |
| 1 | E | 480/525 (91%) | 401 (84%) | 79 (16%) | 2 | 3 |
| 1 | F | 480/525 (91%) | 406 (85%) | 74 (15%) | 2 | 4 |
| All | All | 2873/3150 (91%) | 2417 (84%) | 456 (16%) | 2 | 4 |

5 of 456 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | C | 586 | LEU |
| 1 | F | 598 | THR |
| 1 | D | 490 | LEU |
| 1 | F | 579 | LEU |
| 1 | F | 239 | LEU |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 88 such sidechains are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | D | 647 | GLN |
| 1 | E | 603 | ASN |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | D | 655 | HIS |
| 1 | E | 450 | ASN |
| 1 | F | 176 | ASN |

### 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 9 ligands modelled in this entry, 6 are monoatomic - leaving 3 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 | $\#|Z| > 2$ |
| 3 | NAG | D | 701 | - | 15,15,15 | 1.56 | 2 (13%) | 21,21,21 | 1.04 | 1 (4%) |
| 5 | UD2 | D | 704 | - | 38,41,41 | 3.14 | 20 (52%) | 57,62,62 | 2.42 | 16 (28%) |
| 4 | UDP | D | 702 | 2 | 24,26,26 | 1.31 | 4 (16%) | 37,40,40 | 1.03 | 1 (2%) |

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 |
|-----|------|-------|-----|------|---------|----------|-------|
| 3 | NAG | D | 701 | - | - | 1/6/26/26 | 0/1/1/1 |
| 5 | UD2 | D | 704 | - | - | 2/26/63/63 | 0/3/3/3 |
| 4 | UDP | D | 702 | 2 | - | 8/16/32/32 | 0/2/2/2 |

The worst 5 of 26 bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|-----|-------|-----|------|-------|---|-------------|----------|
| 5 | D | 704 | UD2 | PB-O1' | 10.34 | 1.87 | 1.60 |
| 5 | D | 704 | UD2 | O4B-C4B | 5.90 | 1.58 | 1.45 |
| 5 | D | 704 | UD2 | C2-N3 | 5.29 | 1.47 | 1.38 |
| 5 | D | 704 | UD2 | C4-N3 | 4.87 | 1.47 | 1.38 |
| 5 | D | 704 | UD2 | O5'-C1' | 4.78 | 1.54 | 1.41 |

The worst 5 of 18 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(°) | Ideal(°) |
|-----|-------|-----|------|-------|---|-------------|----------|
| 5 | D | 704 | UD2 | O3A-PB-O1' | -8.79 | 84.76 | 102.48 |
| 5 | D | 704 | UD2 | O1'-PB-O2B | -7.91 | 79.79 | 109.47 |
| 5 | D | 704 | UD2 | PB-O3A-PA | -6.05 | 112.08 | 132.83 |
| 4 | D | 702 | UDP | PA-O3A-PB | -5.25 | 114.81 | 132.83 |
| 5 | D | 704 | UD2 | O1B-PB-O1' | -4.31 | 89.79 | 106.78 |

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|-----|-------|-----|------|-------|
| 4 | D | 702 | UDP | C5'-O5'-PA-O1A |
| 5 | D | 704 | UD2 | C5B-O5B-PA-O3A |
| 3 | D | 701 | NAG | O5-C5-C6-O6 |
| 4 | D | 702 | UDP | C2'-C1'-N1-C6 |
| 4 | D | 702 | UDP | C5'-O5'-PA-O3A |

There are no ring outliers.

3 monomers are involved in 6 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|-----|-------|-----|------|---------|--------------|
| 3 | D | 701 | NAG | 1 | 0 |
| 5 | D | 704 | UD2 | 3 | 0 |
| 4 | D | 702 | UDP | 2 | 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 UD2 D 704**



Bond lengths

Bond angles

Torsions

Rings

**Ligand UDP D 702**



Bond lengths



Bond angles



Torsions



Rings
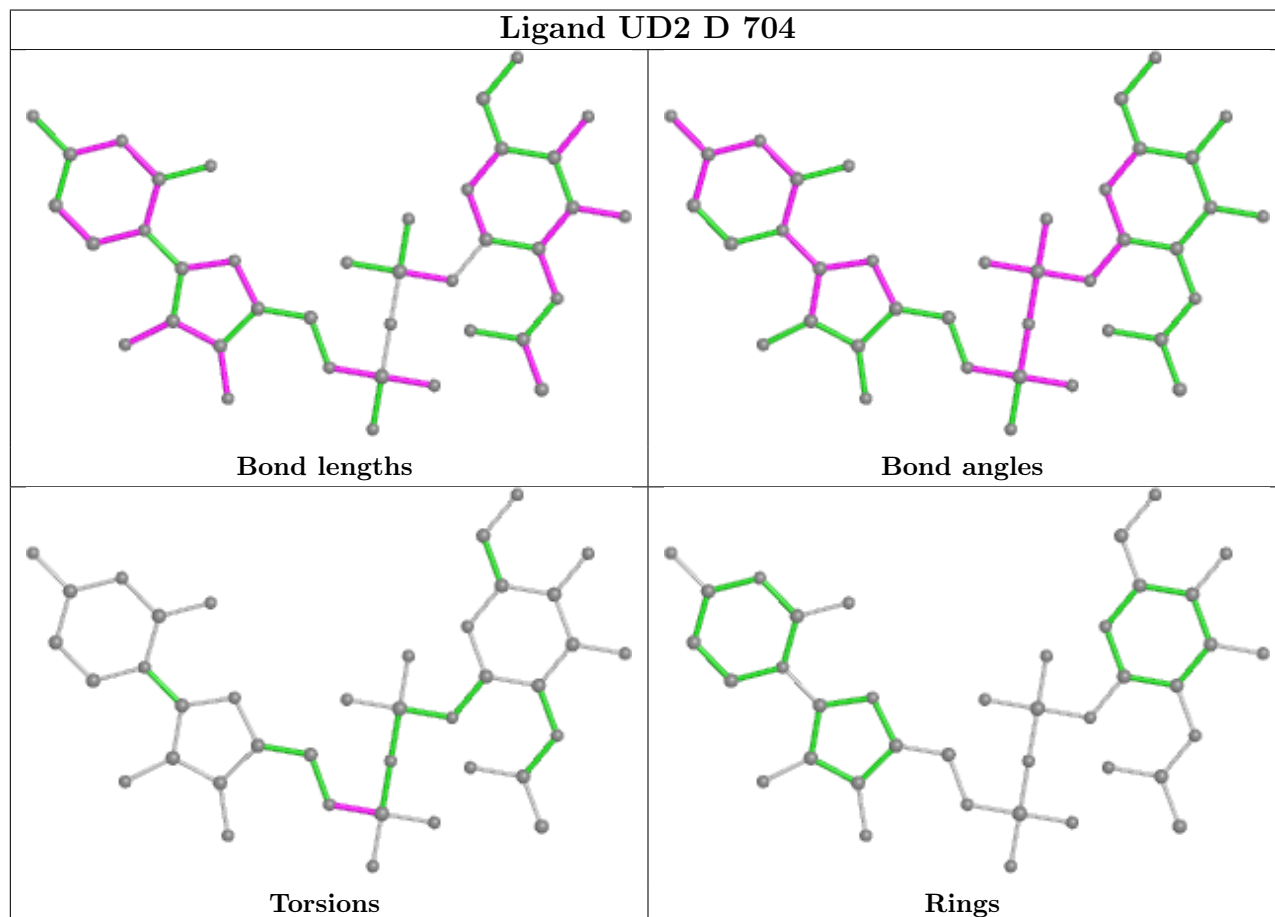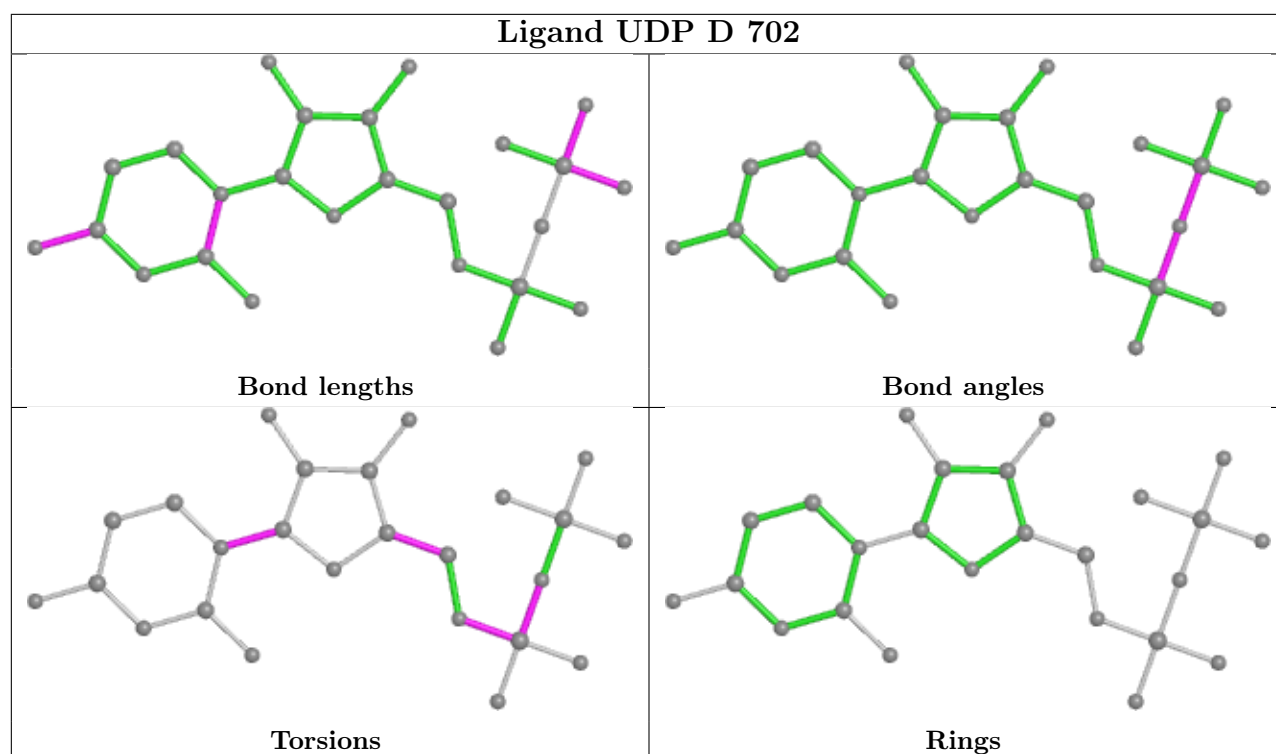
## 5.7 Other polymers ⓘ

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 ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, $95^{th}$ percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

| Mol | Chain | Analysed | <RSRZ> | #RSRZ>2 | | | OWAB(Å²) | Q<0.9 |
|-----|-------|----------|--------|---------|-----|-----|----------|-------|
| 1 | A | 546/597 (91%) | -0.10 | 15 (2%) | 54 | 48 | 23, 33, 59, 111 | 0 |
| 1 | B | 546/597 (91%) | -0.00 | 14 (2%) | 56 | 50 | 26, 37, 64, 116 | 0 |
| 1 | C | 546/597 (91%) | 0.22 | 27 (4%) | 29 | 23 | 24, 42, 82, 138 | 0 |
| 1 | D | 538/597 (90%) | 0.07 | 11 (2%) | 65 | 60 | 29, 47, 75, 92 | 0 |
| 1 | E | 546/597 (91%) | 0.19 | 32 (5%) | 22 | 17 | 26, 46, 83, 154 | 0 |
| 1 | F | 546/597 (91%) | 0.46 | 42 (7%) | 13 | 10 | 34, 59, 86, 136 | 0 |
| All | All | 3268/3582 (91%) | 0.14 | 141 (4%) | 35 | 28 | 23, 43, 79, 154 | 0 |

The worst 5 of 141 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|-----|-------|-----|------|------|
| 1 | A | 458 | SER | 14.3 |
| 1 | F | 455 | TYR | 11.4 |
| 1 | C | 454 | ILE | 10.7 |
| 1 | E | 657 | VAL | 10.0 |
| 1 | C | 456 | VAL | 9.8 |

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

There are no non-standard protein/DNA/RNA residues in this entry.
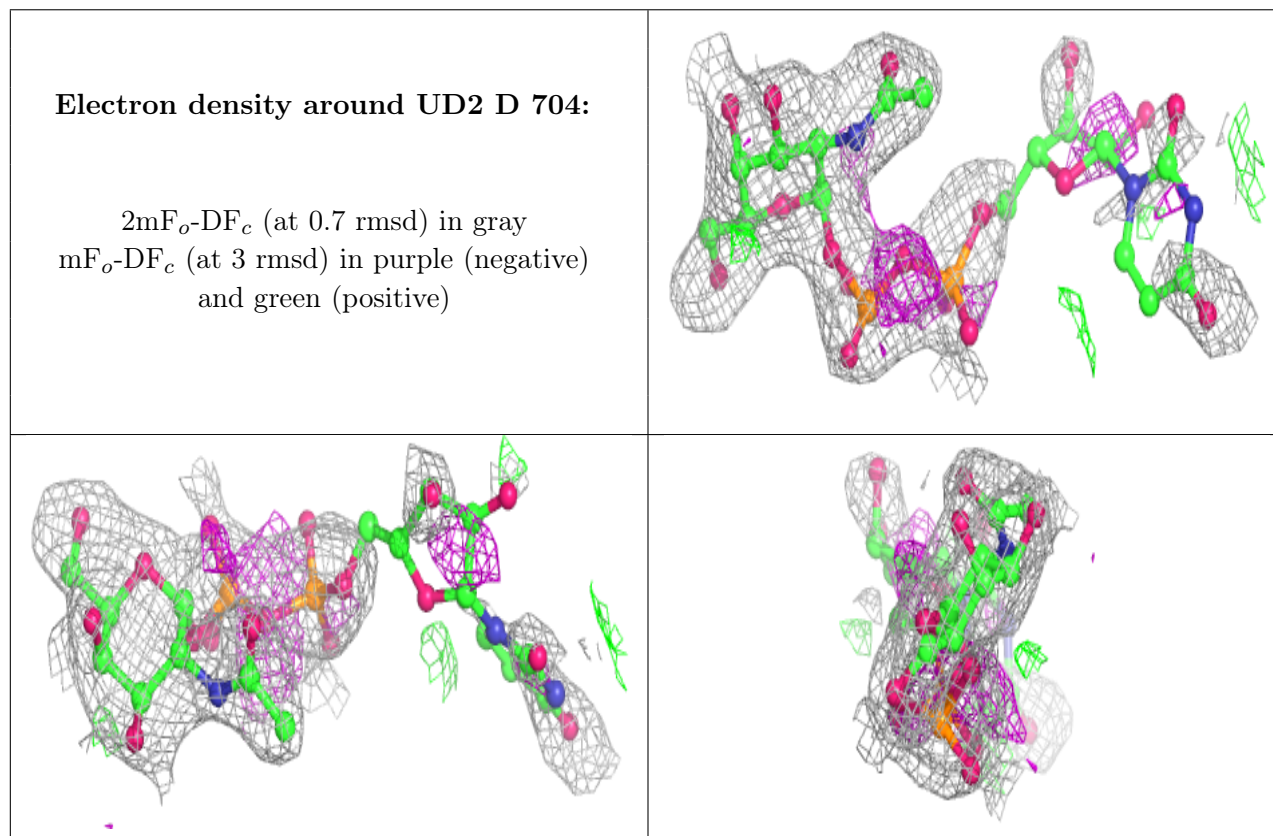
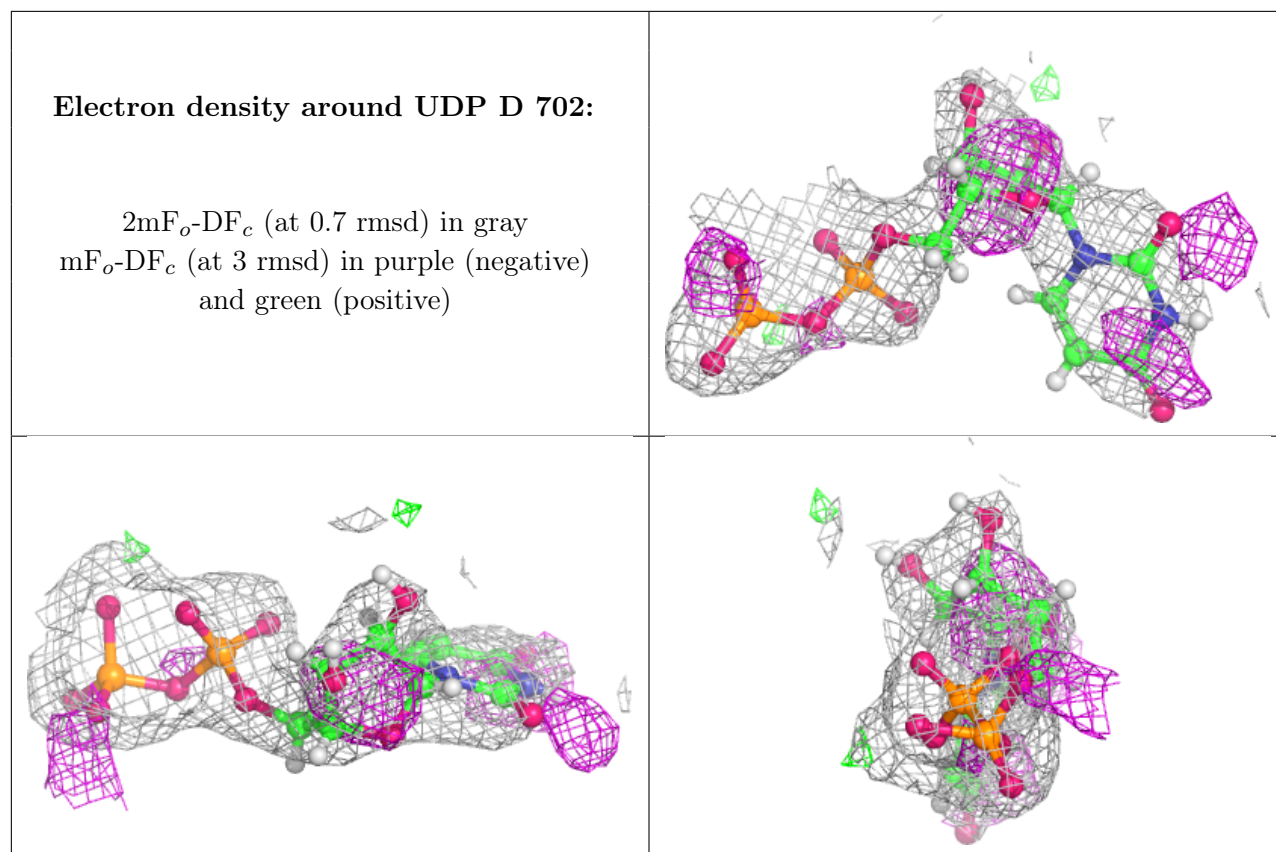## 6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 6.4 Ligands (i)

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

| Mol | Type | Chain | Res | Atoms | RSCC | RSR | B-factors($\text{Å}^2$) | Q<0.9 |
|---|---|---|---|---|---|---|---|---|
| 2 | MN | A | 701 | 1/1 | 0.22 | 0.32 | 100,100,100,100 | 0 |
| 2 | MN | F | 701 | 1/1 | 0.66 | 0.32 | 91,91,91,91 | 0 |
| 2 | MN | C | 701 | 1/1 | 0.70 | 0.27 | 93,93,93,93 | 0 |
| 2 | MN | E | 701 | 1/1 | 0.78 | 0.19 | 82,82,82,82 | 0 |
| 3 | NAG | D | 701 | 15/15 | 0.83 | 0.33 | 52,68,79,82 | 0 |
| 2 | MN | B | 701 | 1/1 | 0.85 | 0.22 | 87,87,87,87 | 0 |
| 5 | UD2 | D | 704 | 39/39 | 0.85 | 0.33 | 36,81,113,113 | 0 |
| 4 | UDP | D | 702 | 25/25 | 0.86 | 0.28 | 55,69,81,82 | 0 |
| 2 | MN | D | 703 | 1/1 | 0.97 | 0.14 | 56,56,56,56 | 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 D 704:**

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

**Electron density around UDP D 702:**

$2mF_o$-$DF_c$ (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.