# WORLDWIDE PDB PROTEIN DATA BANK

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

Aug 15, 2023 – 04:09 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 1T05 |
| Title | : | HIV-1 reverse transcriptase crosslinked to template-primer with tenofovir-diphosphate bound as the incoming nucleotide substrate |
| Authors | : | Tuske, S.; Sarafianos, S.G.; Ding, J.; Arnold, E. |
| Deposited on | : | 2004-04-07 |
| Resolution | : | 3.00 Å(reported) |

---

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

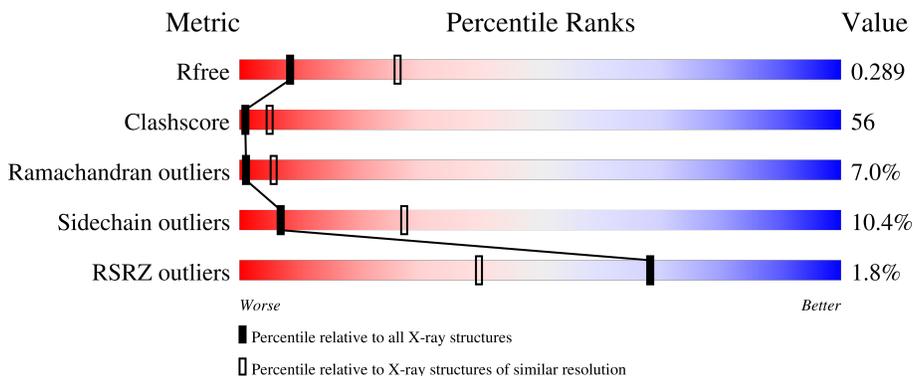| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.35 |
| buster-report | : | 1.1.7 (2018) |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 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) | : | 2.35 |

# 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 3.00 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 | | 0.289 |
| Clashscore | | 56 |
| Ramachandran outliers | | 7.0% |
| Sidechain outliers | | 10.4% |
| RSRZ outliers | | 1.8% |

*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 | 2092 (3.00-3.00) |
| Clashscore | 141614 | 2416 (3.00-3.00) |
| Ramachandran outliers | 138981 | 2333 (3.00-3.00) |
| Sidechain outliers | 138945 | 2336 (3.00-3.00) |
| RSRZ outliers | 127900 | 1990 (3.00-3.00) |

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 | T | 27 | 11%   78%   11% |
| 2 | P | 21 | 10%   57%   29%   5% |
| 3 | A | 558 | %   28%   59%   11%   •• |
| 4 | B | 437 | 3%   24%   57%   13%   • 5% |

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 |
|-----|------|-------|-----|-----------|----------|---------|------------------|
| 7 | GOL | A | 824 | - | - | X | - |

# 2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 8898 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 oligonucleotide template.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|-----|-----|-----|-----|---------|---------|-------|
| 1 | T | 24 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 494 | 233 | 97 | 141 | 23 | | | |

- Molecule 2 is a DNA chain called oligonucleotide primer.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|-----|-----|-----|-----|---------|---------|-------|
| 2 | P | 20 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 403 | 192 | 72 | 120 | 19 | | | |

- Molecule 3 is a protein called POL polyprotein.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|------|-----|-----|----|---------|---------|-------|
| 3 | A | 554 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 4506 | 2917 | 749 | 832 | 8 | | | |

There are 2 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|---------------------|-------------|
| A | 258 | CYS | GLN | engineered mutation | UNP P03366 |
| A | 280 | SER | CYS | engineered mutation | UNP P03366 |

- Molecule 4 is a protein called POL polyprotein.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|------|-----|-----|----|---------|---------|-------|
| 4 | B | 414 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 3415 | 2221 | 567 | 621 | 6 | | | |

There are 9 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|---------------------|-------------|
| B | 280 | SER | CYS | engineered mutation | UNP P04585 |
| B | 430 | GLY | - | cloning artifact | UNP P04585 |

*Continued on next page...*

*Continued from previous page...*

| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|---------|-----------|
| B | 431 | GLY | - | cloning artifact | UNP P04585 |
| B | 432 | HIS | - | expression tag | UNP P04585 |
| B | 433 | HIS | - | expression tag | UNP P04585 |
| B | 434 | HIS | - | expression tag | UNP P04585 |
| B | 435 | HIS | - | expression tag | UNP P04585 |
| B | 436 | HIS | - | expression tag | UNP P04585 |
| B | 437 | HIS | - | expression tag | UNP P04585 |

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---------|---------|
| | | | Total | Mg | | |
| 5 | A | 2 | 2 | 2 | 0 | 0 |

- Molecule 6 is [2-(6-AMINO-9H-PURIN-9-YL)-1-METHYLETHOXY]METHYL-TRIPHOSPHATE (three-letter code: TNV) (formula: $C_9H_{16}N_5O_{10}P_3$).



| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---|---|---|---------|---------|
| | | | Total | C | N | O | P | | |
| 6 | A | 1 | 27 | 9 | 5 | 10 | 3 | 0 | 0 |

- Molecule 7 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

```
                    GOL


                C3
                        OH O3

        C1      C2


        HO              OH
        O1              O2
```
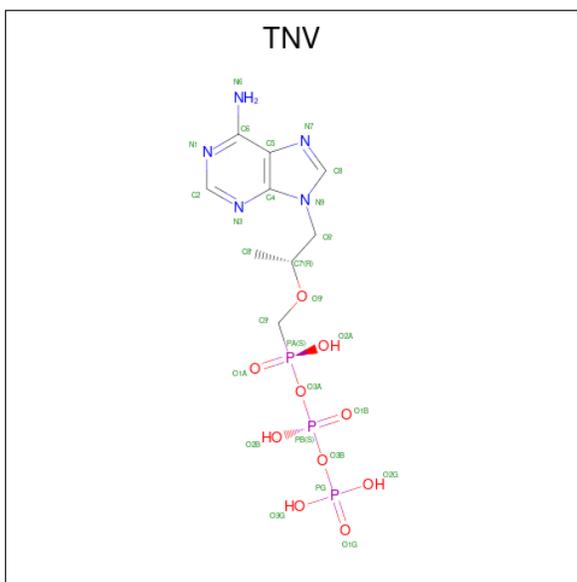
| Mol | Chain | Residues | Atoms | | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---|---------|---------|
| 7 | A | 1 | Total<br>6 | C<br>3 | O<br>3 | 0 | 0 |
| 7 | A | 1 | Total<br>6 | C<br>3 | O<br>3 | 0 | 0 |
| 7 | A | 1 | Total<br>6 | C<br>3 | O<br>3 | 0 | 0 |
| 7 | B | 1 | Total<br>6 | C<br>3 | O<br>3 | 0 | 0 |
| 7 | B | 1 | Total<br>6 | C<br>3 | O<br>3 | 0 | 0 |

- Molecule 8 is water.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---------|---------|
| 8 | T | 2 | Total<br>2 | O<br>2 | 0 | 0 |
| 8 | P | 3 | Total<br>3 | O<br>3 | 0 | 0 |
| 8 | A | 10 | Total<br>10 | O<br>10 | 0 | 0 |
| 8 | B | 6 | Total<br>6 | O<br>6 | 0 | 0 |

# 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: oligonucleotide template

Chain T:

- Molecule 2: oligonucleotide primer

Chain P:

- Molecule 3: POL polyprotein

Chain A:

G453 K454 Y457 V458 T459 N460 K461 G462 R463 Q464 K465 V466 L469 T470 N471 T472 T473 N474 Q475 K476 E477 E478 L479 Q480 A481 I482 Y483 L484 A485 L486 Q487 D488 S489 V493 N494 I495 V496 T497 D498 S499 Q500 Y501 A502 L503 G504 I505 Q509 P510 D511 K512 S513 E514 S515 E516 L517 V518

N519 Q520 I521 I522 E523 Q524 L525 I526 E529 K530 L533 A534 W535 V536 P537 A538 H539 K540 G541 I542 G543 G544 N545 E546 Q547 V548 D549 K550 L551 V552 S553 A554 GLY ILE ARG LYS

• Molecule 4: POL polyprotein

3%

Chain B:

24%     57%     13%    • 5%

PRO ILE S3 P4 I5 E6 T7 V8 K13 M16 D17 G18 P19 K20 V21 K22 Q23 W24 P25 L26 T27 E28 E29 K30 A33 L34 C38 M41 E42 K43 E44 G45 K46 I47 I50 G51 P52 E53 N54 P55 Y56 N57 T58 P59 V60 F61 A62 I63 K64 K65 K66 D67 S68

W71 R72 K73 L74 V75 D76 F77 R78 E79 L80 N81 K82 R83 T84 Q85 D86 F87 W88 E89 V90 Q91 L92 G93 I94 P95 H96 P97 A98 G99 L100 K101 K102 K103 K104 S105 V106 T107 V108 L109 D110 V111 G112 D113 A114 Y115 F116 S117 V118 P119 L120 D121 E122 D123 F124 R125 K126 Y127 T128 A129 F130

T131 I132 P133 S134 I135 N136 N137 E138 T139 P140 G141 I142 R143 Y144 Q145 Y146 N147 V148 L149 P150 Q151 G152 W153 K154 G155 S156 P157 A158 Q161 S162 S163 M164 T165 K166 I167 L168 E169 P170 F171 R172 K173 Q174 N175 P176 I180 Y181 Q182 Y183 M184 D185 D186 L187 Y188 V189 G190 S191 D192 L193

E194 I195 G196 Q197 H198 R199 T200 K201 I202 E203 E204 L205 R206 Q207 H208 L209 L210 R211 W212 G213 L214 T215 T216 P217 ASP LYS LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP MET G231 Y232 E233 L234 H235 P236 W299 T240 V241 Q242 P243 I244 V245 L246 K249 W252 T253 V254 N255 D256

I257 Q258 K259 L260 V261 G262 K263 L264 N265 W266 A267 I270 Y271 P272 G273 I274 K275 V276 R277 Q278 L279 S280 K281 L282 L283 R284 G285 T286 K287 A288 L289 T290 E291 V292 I293 P294 L295 T296 E297 E298 A299 E302 L303 A304 E305 N306 R307 E308 I309 L310 K311 E312 P313 V314 H315 G316 V317 Y318

Y319 D320 K323 D324 L325 I326 A327 E328 I329 Q330 K331 Q332 G333 Q334 G335 Q336 W337 T338 Y339 Q340 I341 Y342 Q343 E344 P345 F346 K347 N348 L349 K350 T351 G352 K353 R356 M357 R358 G359 A360 H361 T362 N363 D364 V365 K366 K367 L368 T369 E370 A371 V372 T376 T377 E378 S379 I380 V381 I382

W383 T386 P387 K388 F389 P392 I393 Q394 K395 E396 T397 W398 E399 T400 W401 W402 T403 E404 Y405 W406 Q407 A408 T409 W410 I411 P412 E413 W414 V417 N418 L422 V423 K424 L425 W426 Y427 Q428 L429 GLY GLY HIS HIS HIS HIS HIS HIS

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 31 1 2 | Depositor |
| Cell constants <br> a, b, c, $\alpha$, $\beta$, $\gamma$ | 171.57Å    171.57Å    156.27Å <br> 90.00°    90.00°    120.00° | Depositor |
| Resolution (Å) | 20.00   –   3.00 <br> 39.85   –   3.00 | Depositor <br> EDS |
| % Data completeness <br> (in resolution range) | 81.8 (20.00-3.00) <br> 91.3 (39.85-3.00) | Depositor <br> EDS |
| $R_{merge}$ | 0.10 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 1.35 (at 3.01Å) | Xtriage |
| Refinement program | CNS | Depositor |
| R, $R_{free}$ | 0.253   ,   0.292 <br> 0.255   ,   0.289 | Depositor <br> DCC |
| $R_{free}$ test set | 1926 reflections (4.01%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 70.0 | Xtriage |
| Anisotropy | 0.278 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.27 , 66.0 | EDS |
| L-test for twinning[2] | $< |L| > = 0.43, < L^2 > = 0.26$ | Xtriage |
| Estimated twinning fraction | 0.063 for -h,-k,l | Xtriage |
| $F_o$,$F_c$ correlation | 0.90 | EDS |
| Total number of atoms | 8898 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 84.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.97% 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: MRG, TNV, GOL, DDG, MG

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 | T | 0.59 | 0/555 | 1.13 | 2/856 (0.2%) |
| 2 | P | 0.59 | 0/400 | 1.24 | 5/612 (0.8%) |
| 3 | A | 0.55 | 1/4624 (0.0%) | 0.80 | 3/6282 (0.0%) |
| 4 | B | 0.48 | 0/3512 | 0.78 | 4/4774 (0.1%) |
| All | All | 0.53 | 1/9091 (0.0%) | 0.84 | 14/12524 (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 |
|---|---|---|---|
| 3 | A | 0 | 1 |

All (1) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 3 | A | 478 | GLU | CD-OE2 | 7.31 | 1.33 | 1.25 |

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

| Mol | Chain | Res | Type | Atoms | Z | Observed(°) | Ideal(°) |
|---|---|---|---|---|---|---|---|
| 4 | B | 3 | SER | C-N-CD | -20.07 | 76.44 | 120.60 |
| 4 | B | 295 | LEU | N-CA-C | 6.77 | 129.27 | 111.00 |
| 1 | T | 718 | DA | N9-C1'-C2' | 6.35 | 124.66 | 112.60 |
| 3 | A | 100 | LEU | CA-CB-CG | 6.22 | 129.60 | 115.30 |
| 2 | P | 815 | DG | N9-C1'-C2' | 6.15 | 124.28 | 112.60 |

There are no chirality outliers.

All (1) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group |
|-----|-------|-----|------|-------|
| 3 | A | 160 | PHE | Sidechain |

## 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 | T | 494 | 0 | 269 | 24 | 0 |
| 2 | P | 403 | 0 | 224 | 30 | 0 |
| 3 | A | 4506 | 0 | 4567 | 517 | 0 |
| 4 | B | 3415 | 0 | 3448 | 443 | 0 |
| 5 | A | 2 | 0 | 0 | 0 | 0 |
| 6 | A | 27 | 0 | 12 | 7 | 0 |
| 7 | A | 18 | 0 | 21 | 5 | 0 |
| 7 | B | 12 | 0 | 14 | 1 | 0 |
| 8 | A | 10 | 0 | 0 | 1 | 0 |
| 8 | B | 6 | 0 | 0 | 1 | 0 |
| 8 | P | 3 | 0 | 0 | 0 | 0 |
| 8 | T | 2 | 0 | 0 | 0 | 0 |
| All | All | 8898 | 0 | 8555 | 981 | 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 56.

The worst 5 of 981 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:498:ASP:HB2 | 3:A:538:ALA:HB2 | 1.31 | 1.08 |
| 3:A:103:LYS:H | 7:A:824:GOL:H32 | 1.13 | 1.08 |
| 3:A:205:LEU:HD22 | 3:A:209:LEU:HD11 | 1.35 | 1.07 |
| 3:A:408:ALA:HB1 | 4:B:364:ASP:HB3 | 1.37 | 1.07 |
| 3:A:72:ARG:HH12 | 6:A:823:TNV:H8'3 | 1.12 | 1.07 |

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 | |
|-----|-------|----------|----------|---------|----------|------|------|
| 3 | A | 552/558 (99%) | 410 (74%) | 112 (20%) | 30 (5%) | 2 | 11 |
| 4 | B | 410/437 (94%) | 278 (68%) | 95 (23%) | 37 (9%) | 1 | 3 |
| All | All | 962/995 (97%) | 688 (72%) | 207 (22%) | 67 (7%) | 1 | 6 |

5 of 67 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | A | 77 | PHE |
| 3 | A | 154 | LYS |
| 3 | A | 268 | SER |
| 3 | A | 345 | PRO |
| 3 | A | 438 | GLU |

### 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 | |
|-----|-------|----------|-----------|----------|------|------|
| 3 | A | 495/498 (99%) | 437 (88%) | 58 (12%) | 5 | 22 |
| 4 | B | 376/397 (95%) | 343 (91%) | 33 (9%) | 10 | 36 |
| All | All | 871/895 (97%) | 780 (90%) | 91 (10%) | 7 | 27 |

5 of 91 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | A | 547 | GLN |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 4 | B | 200 | THR |
| 4 | B | 34 | LEU |
| 4 | B | 133 | PRO |
| 4 | B | 239 | TRP |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | A | 494 | ASN |
| 4 | B | 340 | GLN |
| 3 | A | 519 | ASN |
| 4 | B | 407 | GLN |
| 4 | B | 242 | GLN |

### 5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues 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 | $\#|Z| > 2$ |
| 2 | MRG | P | 817 | 2,1 | 18,24,29 | 1.19 | 3 (16%) | 19,35,42 | 0.91 | 1 (5%) |
| 2 | DDG | P | 822 | 2,1 | 17,23,24 | 1.17 | 3 (17%) | 15,33,36 | 0.77 | 1 (6%) |

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 |
|-----|------|-------|-----|------|---------|----------|-------|
| 2 | MRG | P | 817 | 2,1 | - | 0/3/21/27 | 0/3/3/3 |
| 2 | DDG | P | 822 | 2,1 | - | 0/3/18/19 | 0/3/3/3 |

The worst 5 of 6 bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|-----|-------|-----|------|-------|---|-------------|----------|
| 2 | P | 822 | DDG | C5-C6 | -2.85 | 1.41 | 1.47 |
| 2 | P | 817 | MRG | C5-C6 | -2.49 | 1.42 | 1.47 |
| 2 | P | 817 | MRG | C8-N7 | -2.47 | 1.30 | 1.35 |
| 2 | P | 822 | DDG | C8-N7 | -2.41 | 1.30 | 1.35 |
| 2 | P | 822 | DDG | C5-C4 | -2.09 | 1.37 | 1.43 |

All (2) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|-----|-------|-----|------|-------|---|----------------|-------------|
| 2 | P | 817 | MRG | O3'-C3'-C2' | -2.97 | 100.28 | 110.90 |
| 2 | P | 822 | DDG | O6-C6-C5 | 2.19 | 128.64 | 124.37 |

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 5 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|-----|-------|-----|------|---------|--------------|
| 2 | P | 817 | MRG | 3 | 0 |
| 2 | P | 822 | DDG | 2 | 0 |

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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 | $\#\|Z\| > 2$ |
| 7 | GOL | A | 826 | - | 5,5,5 | 1.02 | 0 | 5,5,5 | 1.96 | 2 (40%) |
| 7 | GOL | B | 439 | - | 5,5,5 | 0.97 | 0 | 5,5,5 | 1.95 | 2 (40%) |
| 6 | TNV | A | 823 | 5 | 22,28,28 | 1.61 | 5 (22%) | 20,43,43 | 1.56 | 3 (15%) |
| 7 | GOL | B | 438 | - | 5,5,5 | 0.79 | 0 | 5,5,5 | 1.99 | 2 (40%) |
| 7 | GOL | A | 825 | - | 5,5,5 | 1.16 | 0 | 5,5,5 | 1.97 | 2 (40%) |
| 7 | GOL | A | 824 | - | 5,5,5 | 1.22 | 0 | 5,5,5 | 2.00 | 2 (40%) |

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 |
|-----|------|-------|-----|------|---------|----------|-------|
| 7 | GOL | A | 826 | - | - | 3/4/4/4 | - |
| 7 | GOL | B | 439 | - | - | 2/4/4/4 | - |
| 6 | TNV | A | 823 | 5 | - | 8/17/22/22 | 0/2/2/2 |
| 7 | GOL | B | 438 | - | - | 2/4/4/4 | - |
| 7 | GOL | A | 825 | - | - | 2/4/4/4 | - |
| 7 | GOL | A | 824 | - | - | 3/4/4/4 | - |

All (5) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|-----|-------|-----|------|-------|------|-------------|----------|
| 6 | A | 823 | TNV | O9'-C9' | -3.64 | 1.37 | 1.42 |
| 6 | A | 823 | TNV | PA-O3A | -3.02 | 1.55 | 1.58 |
| 6 | A | 823 | TNV | C6'-N9 | -2.96 | 1.45 | 1.48 |
| 6 | A | 823 | TNV | C2-N3 | 2.52 | 1.36 | 1.32 |
| 6 | A | 823 | TNV | C2-N1 | 2.03 | 1.37 | 1.33 |

The worst 5 of 13 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|-----|-------|-----|------|-------|------|----------------|-------------|
| 6 | A | 823 | TNV | PA-O3A-PB | -3.91 | 120.17 | 132.56 |
| 6 | A | 823 | TNV | PB-O3B-PG | -3.05 | 122.35 | 132.83 |
| 6 | A | 823 | TNV | C5-C6-N6 | 3.04 | 124.97 | 120.35 |
| 7 | B | 438 | GOL | C3-C2-C1 | 2.95 | 123.18 | 111.70 |
| 7 | A | 825 | GOL | O2-C2-C1 | 2.95 | 122.10 | 109.12 |

There are no chirality outliers.

5 of 20 torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|-----|-------|-----|------|-------|
| 6 | A | 823 | TNV | O9'-C9'-PA-O2A |
| 6 | A | 823 | TNV | O9'-C9'-PA-O3A |
| 6 | A | 823 | TNV | C6'-C7'-O9'-C9' |
| 6 | A | 823 | TNV | N9-C6'-C7'-O9' |
| 6 | A | 823 | TNV | N9-C6'-C7'-C8' |

There are no ring outliers.

5 monomers are involved in 13 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|-----|-------|-----|------|---------|--------------|
| 7 | B | 439 | GOL | 1 | 0 |
| 6 | A | 823 | TNV | 7 | 0 |
| 7 | B | 438 | GOL | 1 | 0 |
| 7 | A | 825 | GOL | 1 | 0 |
| 7 | A | 824 | GOL | 4 | 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 TNV A 823**



Bond lengths



Bond angles



Torsions



Rings
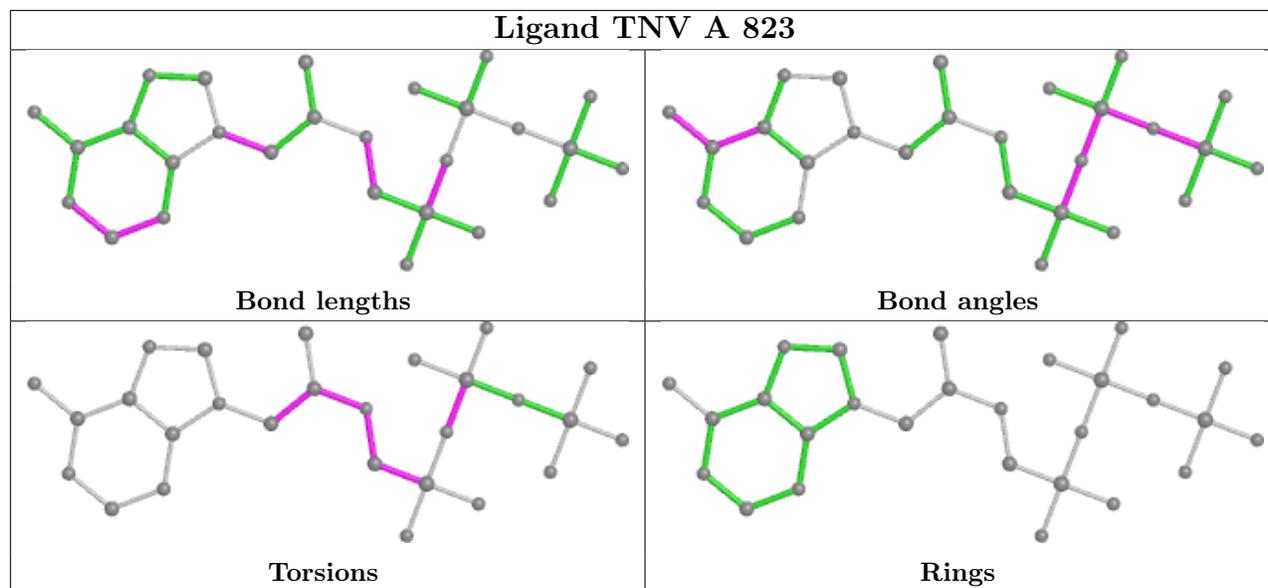
## 5.7 Other polymers ⓘ

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 | T | 24/27 (88%) | -0.26 | 0 | 100 | 100 | 42, 100, 120, 120 | 0 |
| 2 | P | 18/21 (85%) | -0.29 | 0 | 100 | 100 | 42, 104, 120, 120 | 0 |
| 3 | A | 554/558 (99%) | -0.27 | 3 (0%) | 91 | 75 | 13, 71, 119, 120 | 0 |
| 4 | B | 414/437 (94%) | -0.02 | 15 (3%) | 42 | 17 | 30, 100, 120, 120 | 0 |
| All | All | 1010/1043 (96%) | -0.17 | 18 (1%) | 68 | 40 | 13, 86, 120, 120 | 0 |

The worst 5 of 18 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|-----|-------|-----|------|------|
| 4 | B | 3 | SER | 9.3 |
| 4 | B | 4 | PRO | 5.9 |
| 4 | B | 5 | ILE | 4.5 |
| 4 | B | 279 | LEU | 3.5 |
| 4 | B | 358 | ARG | 3.4 |

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

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(Å²) | Q<0.9 |
|-----|------|-------|-----|-------|------|-----|---------------|-------|
| 2 | MRG | P | 817 | 22/27 | 0.94 | 0.12 | 37,60,89,98 | 0 |
| 2 | DDG | P | 822 | 21/22 | 0.96 | 0.19 | 39,53,60,74 | 0 |

## 6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

## 6.4 Ligands ⓘ

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 |
|-----|------|-------|-----|-------|------|-----|-------------------------|-------|
| 7 | GOL | A | 824 | 6/6 | 0.83 | 0.25 | 54,58,60,63 | 6 |
| 7 | GOL | A | 826 | 6/6 | 0.85 | 0.28 | 36,42,49,50 | 6 |
| 7 | GOL | A | 825 | 6/6 | 0.87 | 0.18 | 45,59,68,81 | 6 |
| 7 | GOL | B | 439 | 6/6 | 0.89 | 0.48 | 63,70,74,75 | 6 |
| 7 | GOL | B | 438 | 6/6 | 0.90 | 0.47 | 56,70,72,74 | 6 |
| 5 | MG | A | 600 | 1/1 | 0.90 | 0.21 | 16,16,16,16 | 0 |
| 5 | MG | A | 601 | 1/1 | 0.94 | 0.24 | 73,73,73,73 | 0 |
| 6 | TNV | A | 823 | 27/27 | 0.96 | 0.21 | 34,54,64,71 | 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 TNV A 823:**

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

## 6.5 Other polymers ⓘ

There are no such residues in this entry.