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

Oct 5, 2023 – 11:08 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 6UU1 |
| Title | : | E. coli sigma-S transcription initiation complex with a 4-nt RNA and a CTP ("Fresh" crystal soaked with CTP, GTP, and ddTTP for 30 minutes) |
| Authors | : | Zuo, Y.; De, S.; Steitz, T.A. |
| Deposited on | : | 2019-10-30 |
| Resolution | : | 4.10 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:

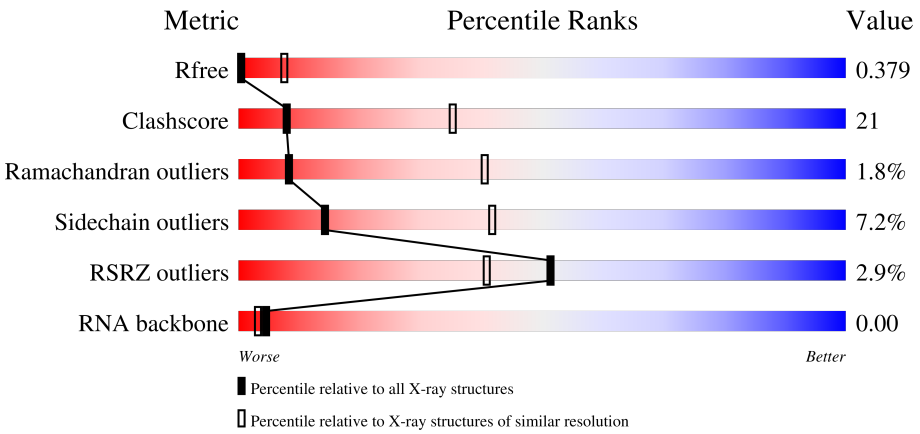
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.35.1 |
| buster-report | : | 1.1.7 (2018) |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 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) | : | 2.35.1 |

# 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 4.10 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 | 1193 (4.50-3.70) |
| Clashscore | 141614 | 1003 (4.44-3.76) |
| Ramachandran outliers | 138981 | 1005 (4.48-3.72) |
| Sidechain outliers | 138945 | 1199 (4.50-3.70) |
| RSRZ outliers | 127900 | 1034 (4.50-3.70) |
| RNA backbone | 3102 | 1049 (5.04-3.00) |

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 | AAA | 242 | 5% 62% 29% • 5% |
| 1 | BBB | 242 | 5% 62% 29% • 6% |
| 2 | CCC | 1342 | 2% 64% 33% • |
| 3 | DDD | 1407 | 2% 58% 35% • • |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 4 | EEE | 90 | %, 60%, 26%, 12% |
| 5 | FFF | 336 | 6%, 50%, 30%, 18% |
| 6 | 111 | 50 | 6%, 30%, 30%, 40% |
| 7 | 222 | 50 | 8%, 22%, 50%, 28% |
| 8 | 333 | 4 | 50%, 50% |

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 |
|---|---|---|---|---|---|---|---|
| 11 | CTP | 333 | 201 | - | - | X | - |
| 9 | ZN | DDD | 1502 | - | - | X | - |

# 2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 29064 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 DNA-directed RNA polymerase subunit alpha.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 1 | AAA | 230 | 1787 | 1112 | 317 | 352 | 6 | 0 | 0 | 0 |
| 1 | BBB | 228 | 1767 | 1100 | 312 | 349 | 6 | 0 | 0 | 0 |

There are 14 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| AAA | -6 | ALA | - | expression tag | UNP A0A377D9Q8 |
| AAA | -5 | HIS | - | expression tag | UNP A0A377D9Q8 |
| AAA | -4 | HIS | - | expression tag | UNP A0A377D9Q8 |
| AAA | -3 | HIS | - | expression tag | UNP A0A377D9Q8 |
| AAA | -2 | HIS | - | expression tag | UNP A0A377D9Q8 |
| AAA | -1 | HIS | - | expression tag | UNP A0A377D9Q8 |
| AAA | 0 | HIS | - | expression tag | UNP A0A377D9Q8 |
| BBB | -6 | ALA | - | expression tag | UNP A0A377D9Q8 |
| BBB | -5 | HIS | - | expression tag | UNP A0A377D9Q8 |
| BBB | -4 | HIS | - | expression tag | UNP A0A377D9Q8 |
| BBB | -3 | HIS | - | expression tag | UNP A0A377D9Q8 |
| BBB | -2 | HIS | - | expression tag | UNP A0A377D9Q8 |
| BBB | -1 | HIS | - | expression tag | UNP A0A377D9Q8 |
| BBB | 0 | HIS | - | expression tag | UNP A0A377D9Q8 |

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 2 | CCC | 1341 | 10576 | 6636 | 1842 | 2055 | 43 | 0 | 0 | 0 |

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 3 | DDD | 1362 | 10568 | 6633 | 1887 | 1998 | 50 | 0 | 0 | 0 |

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 4 | EEE | 79 | 627 | 382 | 118 | 126 | 1 | 0 | 0 | 0 |

- Molecule 5 is a protein called RNA polymerase sigma factor RpoS.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 5 | FFF | 277 | 2253 | 1411 | 415 | 423 | 4 | 0 | 0 | 0 |

There are 10 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| FFF | 2 | GLY | SER | conflict | UNP P13445 |
| FFF | 33 | GLU | GLN | conflict | UNP P13445 |
| FFF | 329 | LEU | - | expression tag | UNP P13445 |
| FFF | 330 | GLU | - | expression tag | UNP P13445 |
| FFF | 331 | HIS | - | expression tag | UNP P13445 |
| FFF | 332 | HIS | - | expression tag | UNP P13445 |
| FFF | 333 | HIS | - | expression tag | UNP P13445 |
| FFF | 334 | HIS | - | expression tag | UNP P13445 |
| FFF | 335 | HIS | - | expression tag | UNP P13445 |
| FFF | 336 | HIS | - | expression tag | UNP P13445 |

- Molecule 6 is a DNA chain called Synthetic DNA 50-MER (promoter non-template strand).

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | P | | | |
| 6 | 111 | 30 | 619 | 294 | 114 | 181 | 30 | 0 | 0 | 0 |

- Molecule 7 is a DNA chain called Synthetic DNA 50-MER (promoter template strand).

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | P | | | |
| 7 | 222 | 36 | 738 | 352 | 137 | 214 | 35 | 0 | 0 | 0 |

- Molecule 8 is a RNA chain called RNA 4-mer.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | P | | | |
| 8 | 333 | 4 | 96 | 40 | 17 | 33 | 6 | 0 | 0 | 0 |

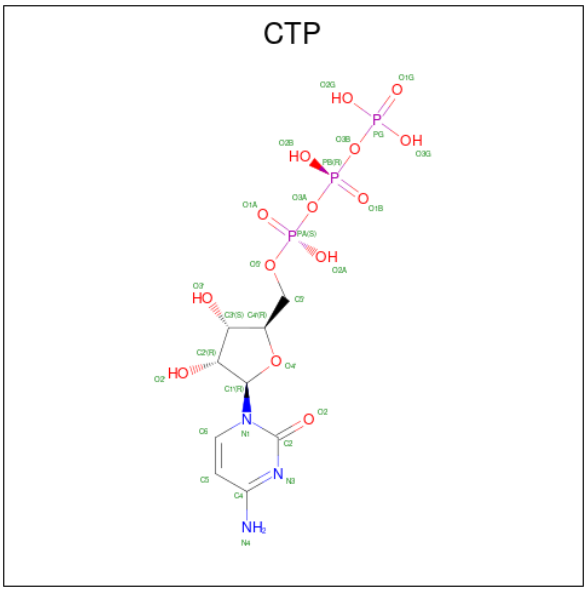
- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | Zn | | |
| 9 | DDD | 2 | 2 | 2 | 0 | 0 |

- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | Mg | | |
| 10 | DDD | 2 | 2 | 2 | 0 | 0 |

- Molecule 11 is CYTIDINE-5'-TRIPHOSPHATE (three-letter code: CTP) (formula: $C_9H_{16}N_3O_{14}P_3$).



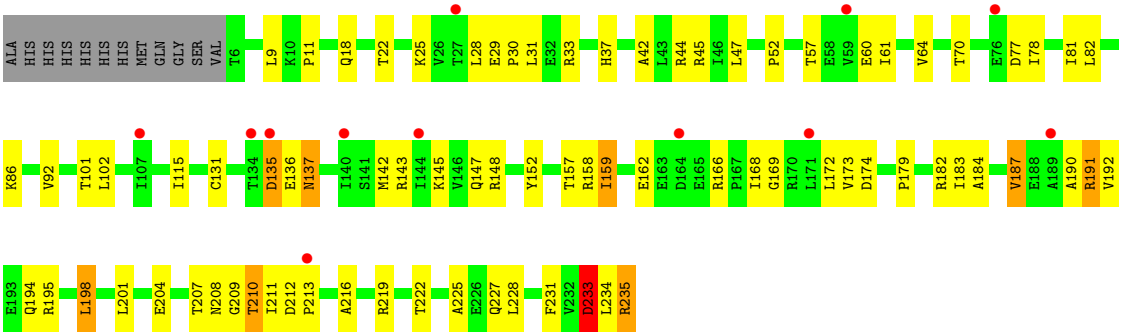

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | P | | |
| 11 | 333 | 1 | 29 | 9 | 3 | 14 | 3 | 0 | 0 |

# 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.

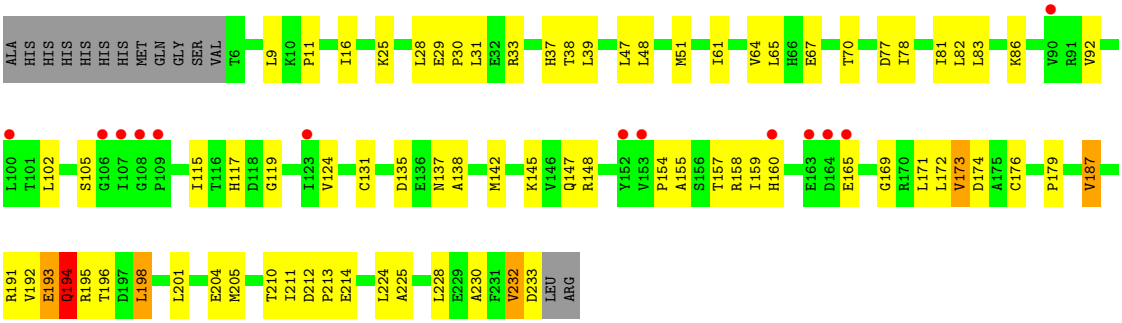
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain AAA:



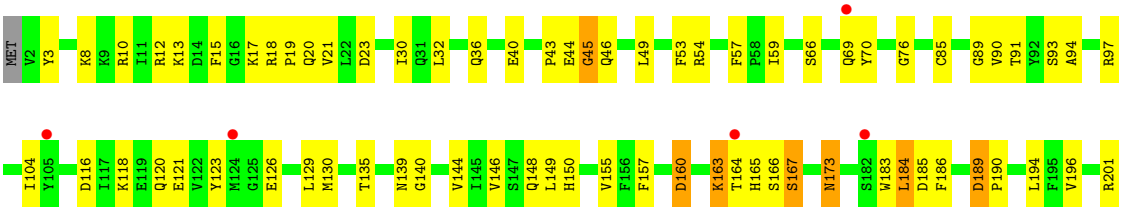

• Molecule 1: DNA-directed RNA polymerase subunit alpha

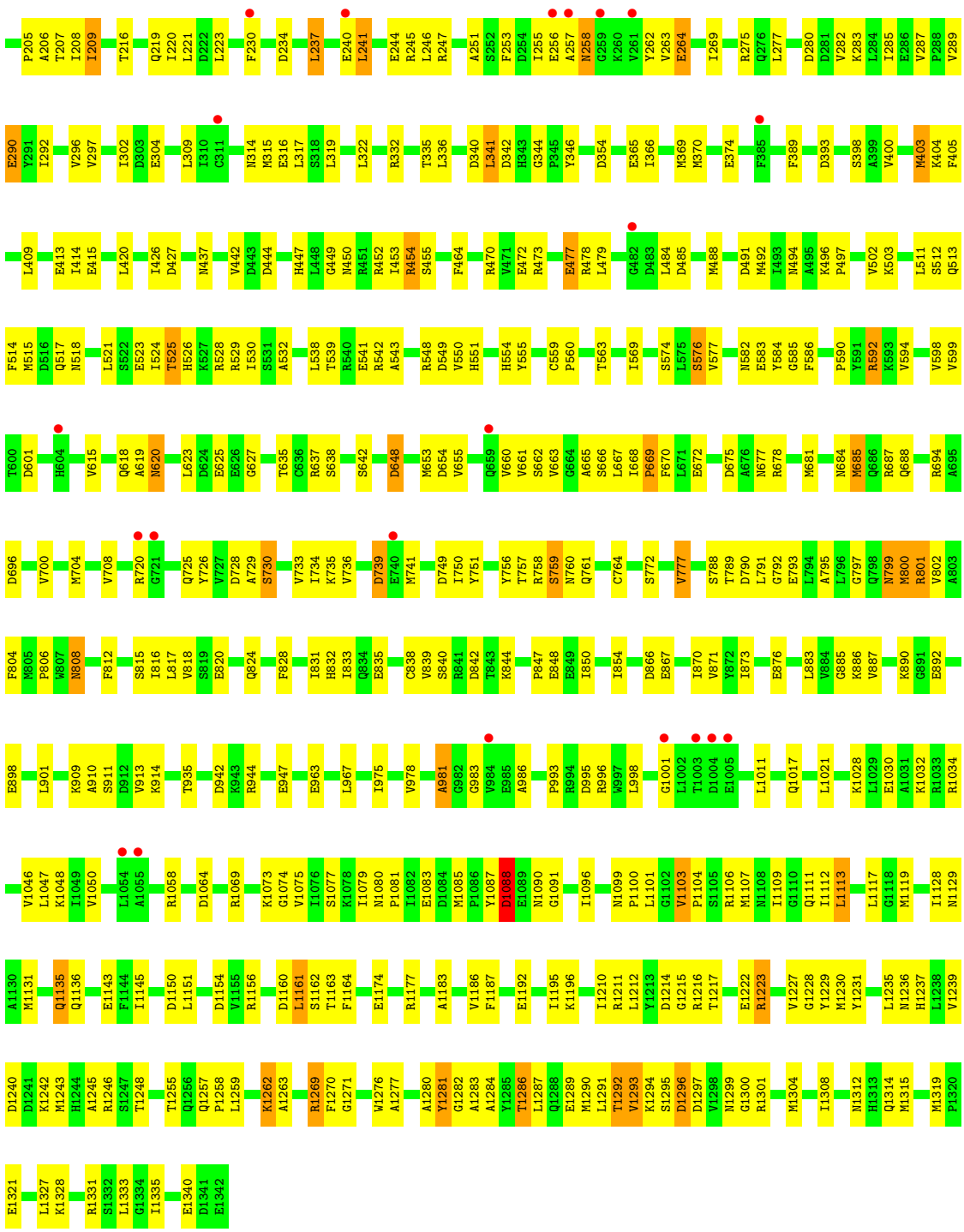
Chain BBB:




• Molecule 2: DNA-directed RNA polymerase subunit beta

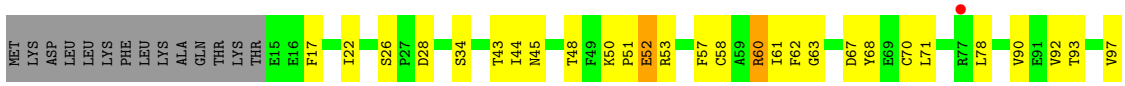
Chain CCC:

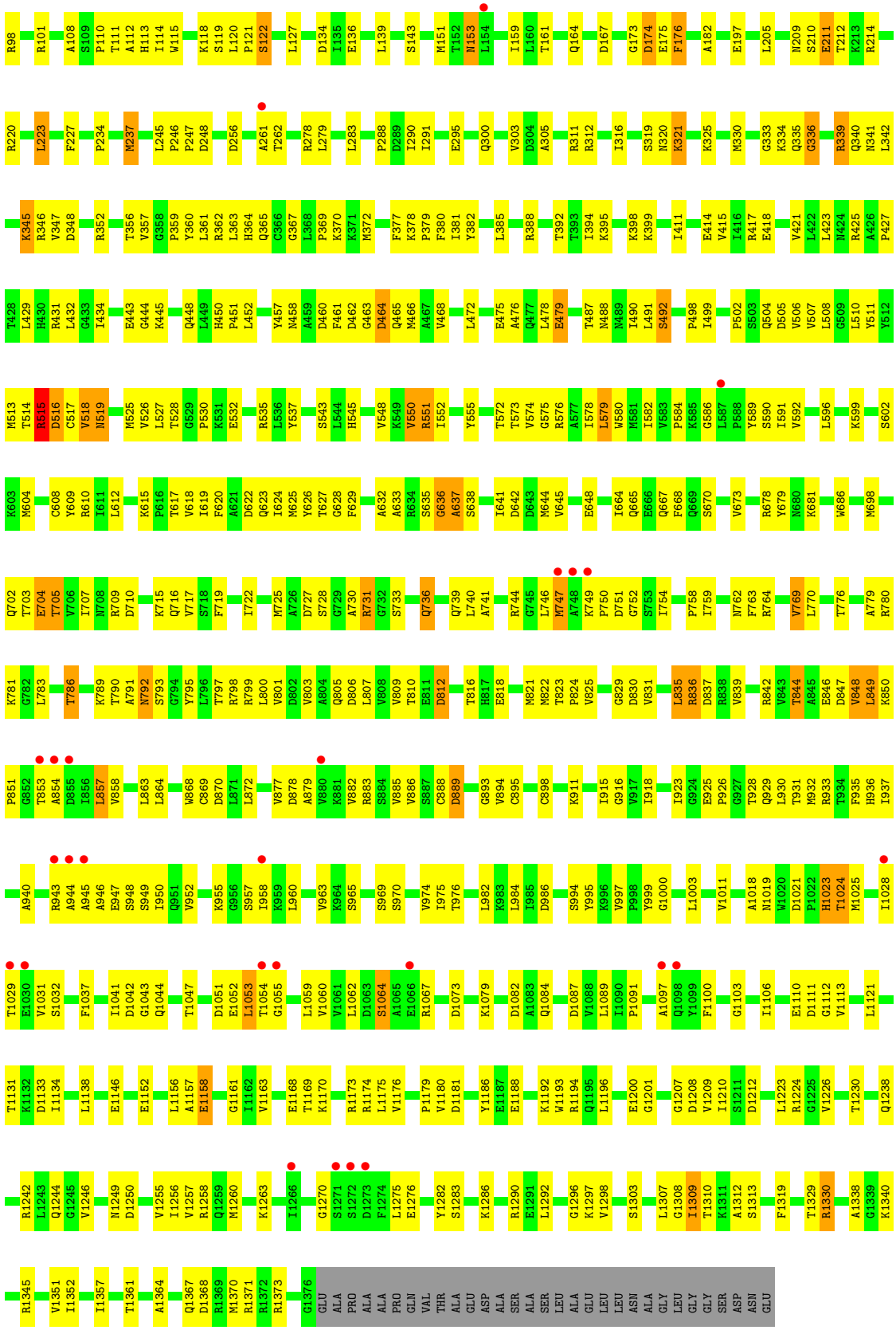





- Molecule 3: DNA-directed RNA polymerase subunit beta'

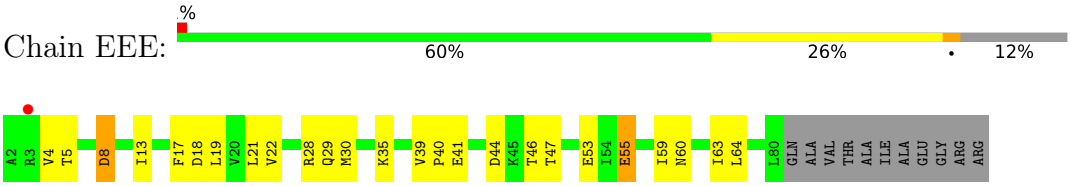
Chain DDD:





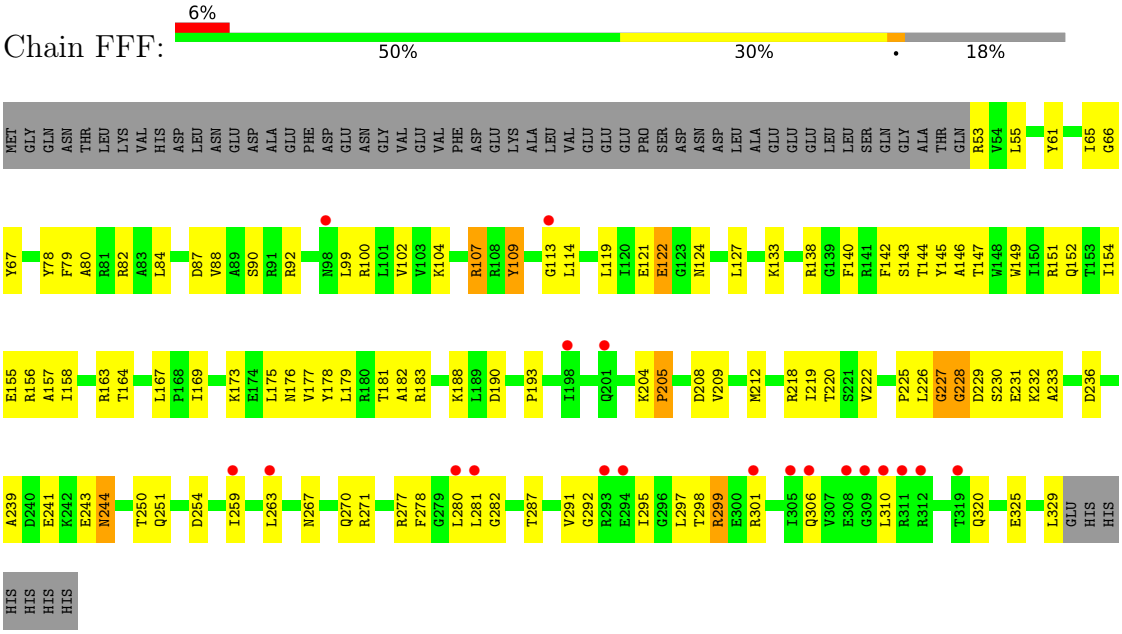

- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain EEE:



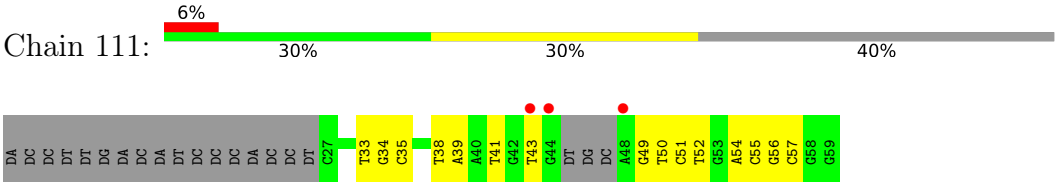

• Molecule 5: RNA polymerase sigma factor RpoS

Chain FFF:



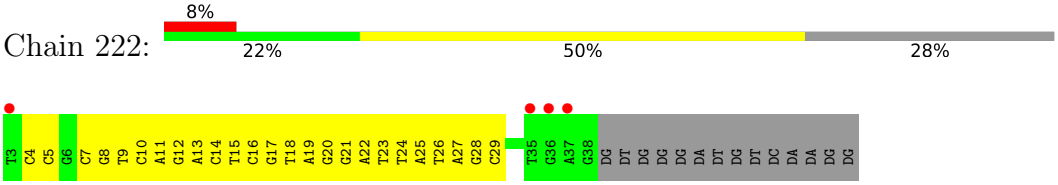

• Molecule 6: Synthetic DNA 50-MER (promoter non-template strand)

Chain 111:




• Molecule 7: Synthetic DNA 50-MER (promoter template strand)

Chain 222:




• Molecule 8: RNA 4-mer

Chain 333:

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 21 21 21 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 132.85Å 154.25Å 230.73Å<br>90.00° 90.00° 90.00° | Depositor |
| Resolution (Å) | 49.18 – 4.10<br>49.18 – 4.10 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 98.0 (49.18-4.10)<br>98.2 (49.18-4.10) | Depositor<br>EDS |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 1.07 (at 4.14Å) | Xtriage |
| Refinement program | REFMAC 5.8.0257 | Depositor |
| R, $R_{free}$ | 0.332 , 0.396<br>0.319 , 0.379 | Depositor<br>DCC |
| $R_{free}$ test set | 1774 reflections (4.76%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 187.4 | Xtriage |
| Anisotropy | 0.501 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.24 , 211.1 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.34$, $< L^2 > = 0.17$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| $F_o$,$F_c$ correlation | 0.90 | EDS |
| Total number of atoms | 29064 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 288.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.37% 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: 2DT, MG, CTP, GTP, 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 | AAA | 0.63 | 0/1809 | 0.74 | 0/2450 |
| 1 | BBB | 0.64 | 0/1789 | 0.73 | 0/2425 |
| 2 | CCC | 0.62 | 0/10745 | 0.81 | 3/14499 (0.0%) |
| 3 | DDD | 0.63 | 0/10729 | 0.80 | 5/14487 (0.0%) |
| 4 | EEE | 0.62 | 0/629 | 0.80 | 0/847 |
| 5 | FFF | 0.65 | 0/2282 | 0.66 | 0/3076 |
| 6 | 111 | 0.29 | 0/693 | 0.61 | 0/1066 |
| 7 | 222 | 0.27 | 0/828 | 0.70 | 0/1277 |
| 8 | 333 | 0.19 | 0/50 | 0.56 | 0/76 |
| All | All | 0.62 | 0/29554 | 0.78 | 8/40203 (0.0%) |

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 |
|---|---|---|---|
| 2 | CCC | 0 | 1 |
| 3 | DDD | 0 | 2 |
| All | All | 0 | 3 |

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 3 | DDD | 636 | GLY | C-N-CA | -7.74 | 102.34 | 121.70 |
| 3 | DDD | 636 | GLY | N-CA-C | -7.16 | 95.20 | 113.10 |
| 3 | DDD | 636 | GLY | CA-C-N | -6.29 | 103.37 | 117.20 |
| 2 | CCC | 1048 | LYS | CB-CA-C | -5.97 | 98.46 | 110.40 |
| 3 | DDD | 637 | ALA | CB-CA-C | 5.75 | 118.72 | 110.10 |

There are no chirality outliers.

All (3) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group |
|---|---|---|---|---|
| 2 | CCC | 1282 | GLY | Peptide |
| 3 | DDD | 636 | GLY | Peptide,Mainchain |

## 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 | AAA | 1787 | 0 | 1813 | 61 | 0 |
| 1 | BBB | 1767 | 0 | 1789 | 67 | 0 |
| 2 | CCC | 10576 | 0 | 10591 | 454 | 0 |
| 3 | DDD | 10568 | 0 | 10782 | 638 | 0 |
| 4 | EEE | 627 | 0 | 634 | 17 | 0 |
| 5 | FFF | 2253 | 0 | 2298 | 109 | 1 |
| 6 | 111 | 619 | 0 | 340 | 22 | 0 |
| 7 | 222 | 738 | 0 | 407 | 74 | 0 |
| 8 | 333 | 96 | 0 | 46 | 13 | 0 |
| 9 | DDD | 2 | 0 | 0 | 3 | 0 |
| 10 | DDD | 2 | 0 | 0 | 0 | 0 |
| 11 | 333 | 29 | 0 | 12 | 11 | 0 |
| All | All | 29064 | 0 | 28712 | 1238 | 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 21.

The worst 5 of 1238 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:DDD:504:GLN:NE2 | 3:DDD:731:ARG:HH21 | 1.12 | 1.39 |
| 3:DDD:750:PRO:HA | 3:DDD:781:LYS:CG | 1.57 | 1.35 |
| 7:222:21:DG:H4' | 7:222:22:DA:C5' | 1.54 | 1.34 |
| 3:DDD:750:PRO:CA | 3:DDD:781:LYS:HG2 | 1.61 | 1.28 |
| 7:222:21:DG:C4' | 7:222:22:DA:H5' | 1.68 | 1.24 |

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|---|---|---|---|
| 5:FFF:67:TYR:O | 5:FFF:299:ARG:NH2[3_644] | 1.95 | 0.25 |

## 5.3 Torsion angles (i)

### 5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

| Mol | Chain | Analysed | Favoured | Allowed | Outliers | Percentiles | |
|---|---|---|---|---|---|---|---|
| 1 | AAA | 228/242 (94%) | 210 (92%) | 12 (5%) | 6 (3%) | 5 | 34 |
| 1 | BBB | 226/242 (93%) | 207 (92%) | 14 (6%) | 5 (2%) | 6 | 37 |
| 2 | CCC | 1339/1342 (100%) | 1241 (93%) | 72 (5%) | 26 (2%) | 8 | 39 |
| 3 | DDD | 1360/1407 (97%) | 1246 (92%) | 92 (7%) | 22 (2%) | 9 | 43 |
| 4 | EEE | 77/90 (86%) | 73 (95%) | 4 (5%) | 0 | 100 | 100 |
| 5 | FFF | 275/336 (82%) | 254 (92%) | 16 (6%) | 5 (2%) | 8 | 40 |
| All | All | 3505/3659 (96%) | 3231 (92%) | 210 (6%) | 64 (2%) | 8 | 40 |

5 of 64 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | BBB | 117 | HIS |
| 1 | BBB | 119 | GLY |
| 1 | BBB | 193 | GLU |
| 1 | BBB | 194 | GLN |
| 2 | CCC | 46 | GLN |

### 5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was

analysed, and the total number of residues.

| Mol | Chain | Analysed | Rotameric | Outliers | Percentiles | |
|---|---|---|---|---|---|---|
| 1 | AAA | 198/208 (95%) | 180 (91%) | 18 (9%) | 9 | 33 |
| 1 | BBB | 196/208 (94%) | 182 (93%) | 14 (7%) | 14 | 42 |
| 2 | CCC | 1156/1157 (100%) | 1065 (92%) | 91 (8%) | 12 | 39 |
| 3 | DDD | 1135/1168 (97%) | 1059 (93%) | 76 (7%) | 16 | 44 |
| 4 | EEE | 67/74 (90%) | 64 (96%) | 3 (4%) | 27 | 54 |
| 5 | FFF | 240/292 (82%) | 226 (94%) | 14 (6%) | 20 | 48 |
| All | All | 2992/3107 (96%) | 2776 (93%) | 216 (7%) | 14 | 42 |

5 of 216 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 2 | CCC | 1262 | LYS |
| 3 | DDD | 515 | ARG |
| 4 | EEE | 8 | ASP |
| 2 | CCC | 1293 | VAL |
| 3 | DDD | 176 | PHE |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA (i)

| Mol | Chain | Analysed | Backbone Outliers | Pucker Outliers |
|---|---|---|---|---|
| 8 | 333 | 1/4 (25%) | 1 (100%) | 0 |

All (1) RNA backbone outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 8 | 333 | 16 | G |

There are no RNA pucker outliers to report.

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

1 non-standard protein/DNA/RNA residue is 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 | $\#\|Z\| > 2$ |
| 8 | 2DT | 333 | 101 | 7 | 17,20,21 | 1.51 | 4 (23%) | 22,28,31 | 2.70 | 8 (36%) |

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 |
|---|---|---|---|---|---|---|---|
| 8 | 2DT | 333 | 101 | 7 | - | 1/7/18/19 | 0/2/2/2 |

All (4) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 8 | 333 | 101 | 2DT | C2-N1 | 2.76 | 1.42 | 1.38 |
| 8 | 333 | 101 | 2DT | C6-C5 | 2.70 | 1.39 | 1.34 |
| 8 | 333 | 101 | 2DT | C4-C5 | 2.28 | 1.48 | 1.44 |
| 8 | 333 | 101 | 2DT | C4-N3 | -2.14 | 1.34 | 1.38 |

The worst 5 of 8 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 8 | 333 | 101 | 2DT | C4-N3-C2 | -5.22 | 120.59 | 127.35 |
| 8 | 333 | 101 | 2DT | C5-C4-N3 | 4.80 | 119.41 | 115.31 |
| 8 | 333 | 101 | 2DT | O4-C4-C5 | -4.80 | 119.34 | 124.90 |
| 8 | 333 | 101 | 2DT | C3'-C2'-C1' | 4.73 | 108.24 | 102.78 |
| 8 | 333 | 101 | 2DT | N3-C2-N1 | 4.50 | 120.86 | 114.89 |

There are no chirality outliers.

All (1) torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|---|---|---|---|---|
| 8 | 333 | 101 | 2DT | C3'-C4'-C5'-O5' |

There are no ring outliers.

1 monomer is involved in 6 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 8 | 333 | 101 | 2DT | 6 | 0 |

## 5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 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 | $\#\|Z\| > 2$ |
| 11 | CTP | 333 | 201 | 10 | 26,30,30 | 0.54 | 0 | 39,47,47 | 0.52 | 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 |
|---|---|---|---|---|---|---|---|
| 11 | CTP | 333 | 201 | 10 | - | 5/22/38/38 | 0/2/2/2 |

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|---|---|---|---|---|
| 11 | 333 | 201 | CTP | PB-O3A-PA-O1A |
| 11 | 333 | 201 | CTP | PB-O3B-PG-O1G |
| 11 | 333 | 201 | CTP | PB-O3B-PG-O2G |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type | Atoms |
|---|---|---|---|---|
| 11 | 333 | 201 | CTP | PB-O3B-PG-O3G |
| 11 | 333 | 201 | CTP | PB-O3A-PA-O2A |

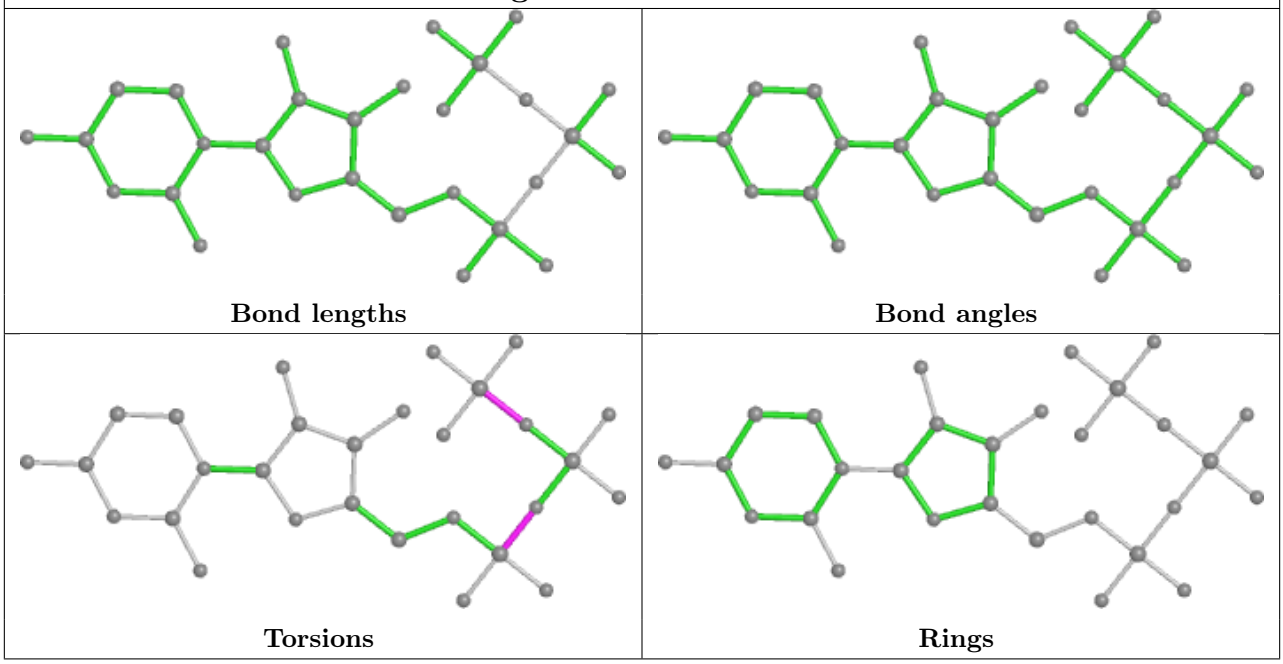
There are no ring outliers.

1 monomer is involved in 11 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 11 | 333 | 201 | CTP | 11 | 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 CTP 333 201**



## 5.7 Other polymers (i)

There are no such residues in this entry.

## 5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

| Mol | Chain | Number of breaks |
|---|---|---|
| 8 | 333 | 1 |

All chain breaks are listed below:

| Model | Chain | Residue-1 | Atom-1 | Residue-2 | Atom-2 | Distance (Å) |
|---|---|---|---|---|---|---|
| 1 | 333 | 16:G | O3' | 101:2DT | P | 1.78 |

# 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, $95^{th}$ 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(Å$^2$) | Q<0.9 |
|---|---|---|---|---|---|---|---|---|
| 1 | AAA | 230/242 (95%) | 0.12 | 12 (5%) | 27 | 23 | 232, 309, 389, 488 | 0 |
| 1 | BBB | 228/242 (94%) | 0.02 | 13 (5%) | 23 | 20 | 215, 331, 426, 456 | 0 |
| 2 | CCC | 1341/1342 (99%) | -0.11 | 26 (1%) | 66 | 58 | 149, 259, 404, 485 | 0 |
| 3 | DDD | 1362/1407 (96%) | -0.17 | 27 (1%) | 65 | 56 | 142, 269, 383, 504 | 0 |
| 4 | EEE | 79/90 (87%) | -0.29 | 1 (1%) | 77 | 68 | 249, 322, 473, 567 | 0 |
| 5 | FFF | 277/336 (82%) | 0.15 | 19 (6%) | 16 | 13 | 215, 334, 439, 467 | 0 |
| 6 | 111 | 30/50 (60%) | -0.16 | 3 (10%) | 7 | 7 | 272, 329, 405, 531 | 0 |
| 7 | 222 | 36/50 (72%) | 0.40 | 4 (11%) | 5 | 6 | 230, 334, 489, 507 | 0 |
| 8 | 333 | 2/4 (50%) | 0.92 | 0 | 100 | 100 | 231, 231, 231, 302 | 0 |
| All | All | 3585/3763 (95%) | -0.09 | 105 (2%) | 51 | 41 | 142, 281, 411, 567 | 0 |

The worst 5 of 105 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 3 | DDD | 853 | THR | 5.5 |
| 7 | 222 | 36 | DG | 5.3 |
| 3 | DDD | 1272 | SER | 4.9 |
| 1 | AAA | 144 | ILE | 4.7 |
| 1 | BBB | 108 | GLY | 4.5 |

## 6.2 Non-standard residues in protein, DNA, RNA chains (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(Å$^2$) | Q<0.9 |
|---|---|---|---|---|---|---|---|---|
| 8 | 2DT | 333 | 101 | 19/20 | 0.89 | 0.27 | 223,258,283,290 | 0 |

## 6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

## 6.4 Ligands (i)

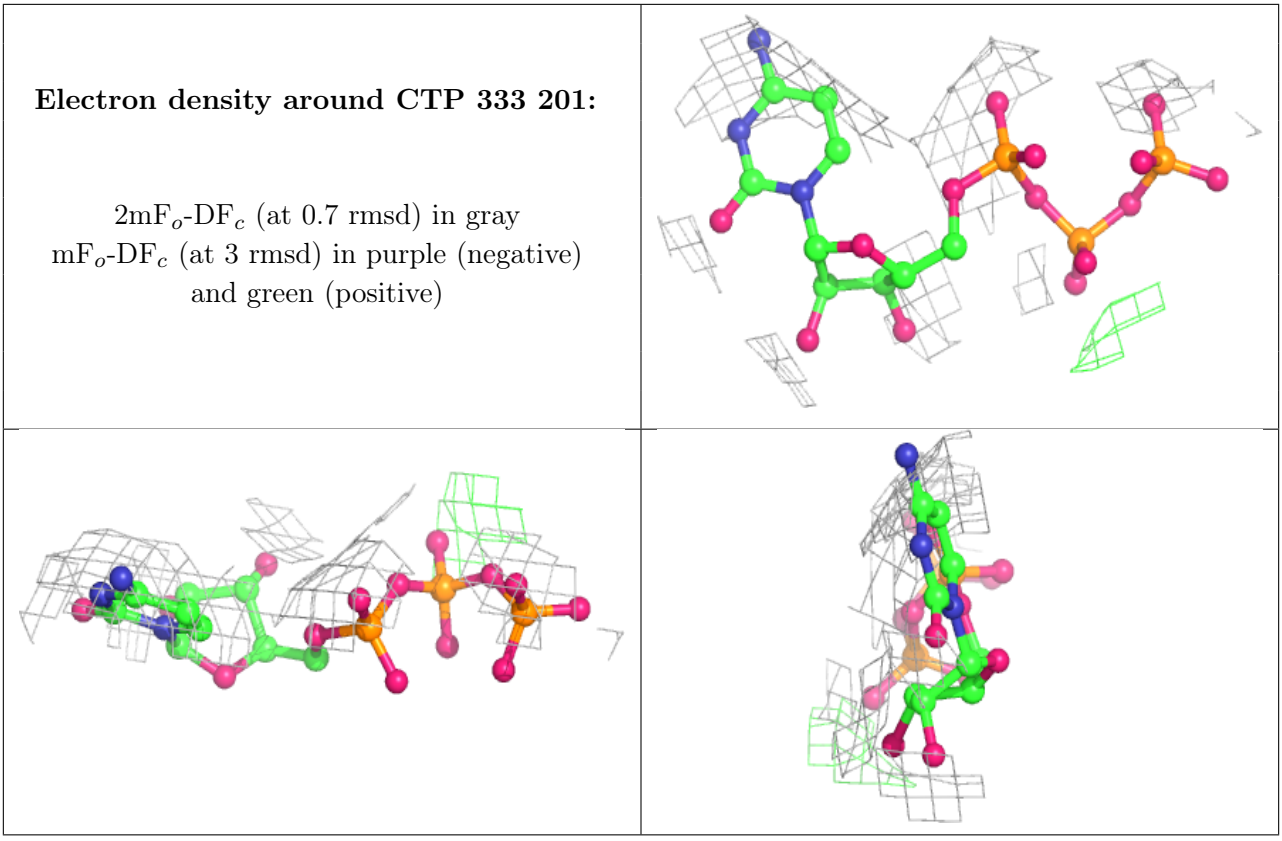
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95$^{th}$ percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

| Mol | Type | Chain | Res | Atoms | RSCC | RSR | B-factors(Å$^2$) | Q<0.9 |
|---|---|---|---|---|---|---|---|---|
| 9 | ZN | DDD | 1502 | 1/1 | 0.94 | 0.17 | 293,293,293,293 | 0 |
| 11 | CTP | 333 | 201 | 29/29 | 0.94 | 0.29 | 172,203,288,301 | 0 |
| 9 | ZN | DDD | 1501 | 1/1 | 0.96 | 0.19 | 421,421,421,421 | 0 |
| 10 | MG | DDD | 1504 | 1/1 | 0.97 | 0.20 | 205,205,205,205 | 0 |
| 10 | MG | DDD | 1503 | 1/1 | 0.99 | 0.30 | 133,133,133,133 | 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 CTP 333 201:**

$2mF_o$-$DF_c$ (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.