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

Sep 17, 2023 – 05:52 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 4Y7N |
| Title | : | The Structure Insight into 5-Carboxycytosine Recognition by RNA Polymerase II during Transcription Elongation. |
| Authors | : | Wang, L.; Chong, J.; Wang, D. |
| Deposited on | : | 2015-02-15 |
| Resolution | : | 3.30 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:

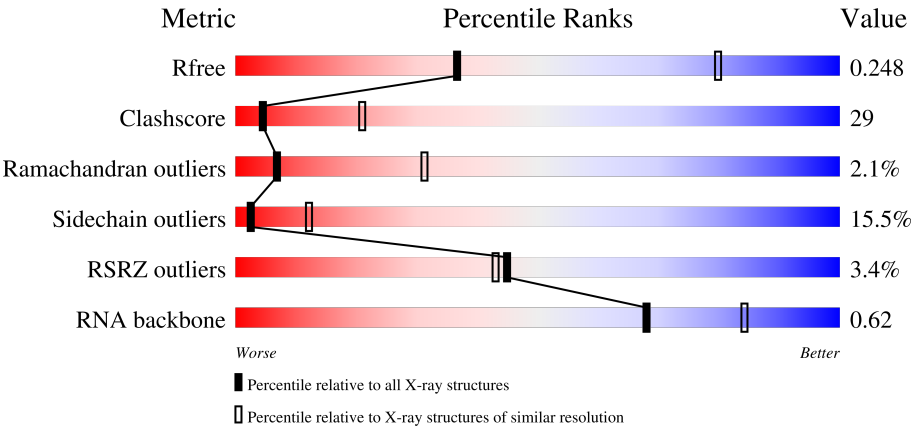
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.35.1 |
| buster-report | : | 1.1.7 (2018) |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 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 3.30 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.



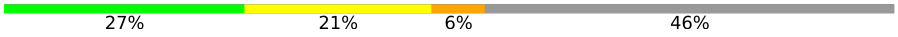
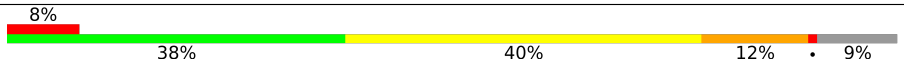
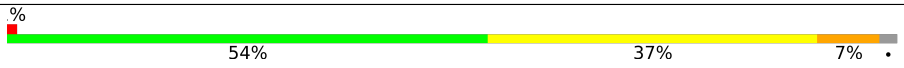
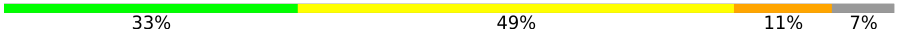
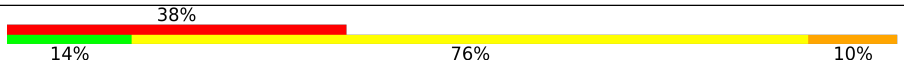
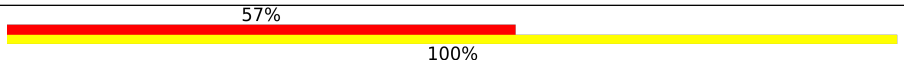
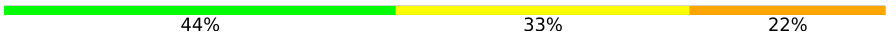

| Metric | Whole archive (#Entries) | Similar resolution (#Entries, resolution range(Å)) |
|---|---|---|
| $R_{free}$ | 130704 | 1149 (3.34-3.26) |
| Clashscore | 141614 | 1205 (3.34-3.26) |
| Ramachandran outliers | 138981 | 1183 (3.34-3.26) |
| Sidechain outliers | 138945 | 1182 (3.34-3.26) |
| RSRZ outliers | 127900 | 1115 (3.34-3.26) |
| RNA backbone | 3102 | 1117 (3.70-2.90) |

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 | 1733 | 3% 43% 30% 7% 20% |
| 2 | B | 1224 | 2% 48% 35% 6% 10% |
| 3 | C | 318 | 44% 34% 6% 16% |
| 4 | E | 215 | 5% 59% 35% 6% |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 5 | F | 155 |   |
| 6 | H | 146 |   |
| 7 | I | 122 |   |
| 8 | J | 70 |   |
| 9 | K | 120 |   |
| 10 | L | 70 |   |
| 11 | T | 29 |   |
| 12 | N | 14 |   |
| 13 | R | 9 |   |

## 2 Entry composition (i)

There are 16 unique types of molecules in this entry. The entry contains 29193 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 II subunit RPB1.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 1393 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 10953 | 6908 | 1921 | 2063 | 61 | | | |

- Molecule 2 is a protein called DNA-directed RNA polymerase II subunit RPB2,DNA-directed RNA polymerase II subunit RPB2.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | B | 1103 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 8762 | 5549 | 1532 | 1626 | 55 | | | |

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | C | 266 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 2095 | 1317 | 348 | 417 | 13 | | | |

- Molecule 4 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | E | 214 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 1752 | 1111 | 309 | 321 | 11 | | | |

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | F | 84 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 679 | 434 | 115 | 127 | 3 | | | |

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 6 | H | 133 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 1068 | 673 | 180 | 211 | 4 | | | |

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB9.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 7 | I | 119 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 971 | 596 | 179 | 186 | 10 | | | |

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | J | 65 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 532 | 339 | 93 | 94 | 6 | | | |

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB11.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 9 | K | 114 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 919 | 590 | 156 | 171 | 2 | | | |

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | L | 44 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 351 | 217 | 70 | 60 | 4 | | | |

- Molecule 11 is a DNA chain called DNA (29-MER).

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 11 | T | 29 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 587 | 280 | 108 | 171 | 28 | | | |

- Molecule 12 is a DNA chain called DNA (5'-D(*CP*TP*GP*CP*TP*TP*AP*TP*CP*GP *GP*TP*AP*G)-3').

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 12 | N | 14 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 284 | 137 | 49 | 85 | 13 | | | |

- Molecule 13 is a RNA chain called RNA (5'-D(*AP*UP*GP*GP*AP*GP*AP*GP*G)-3').

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 13 | R | 9 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 198 | 89 | 42 | 59 | 8 | | | |

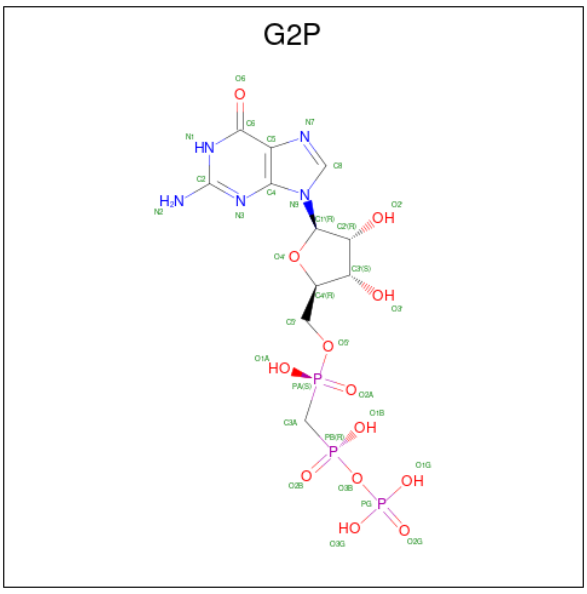
- Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| 14 | A | 2 | Total | Zn | 0 | 0 |
| | | | 2 | 2 | | |
| 14 | B | 1 | Total | Zn | 0 | 0 |
| | | | 1 | 1 | | |
| 14 | C | 1 | Total | Zn | 0 | 0 |
| | | | 1 | 1 | | |
| 14 | I | 2 | Total | Zn | 0 | 0 |
| | | | 2 | 2 | | |
| 14 | J | 1 | Total | Zn | 0 | 0 |
| | | | 1 | 1 | | |
| 14 | L | 1 | Total | Zn | 0 | 0 |
| | | | 1 | 1 | | |

- Molecule 15 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| 15 | A | 2 | Total | Mg | 0 | 0 |
| | | | 2 | 2 | | |

- Molecule 16 is PHOSPHOMETHYLPHOSPHONIC ACID GUANYLATE ESTER (three-letter code: G2P) (formula: $C_{11}H_{18}N_5O_{13}P_3$).



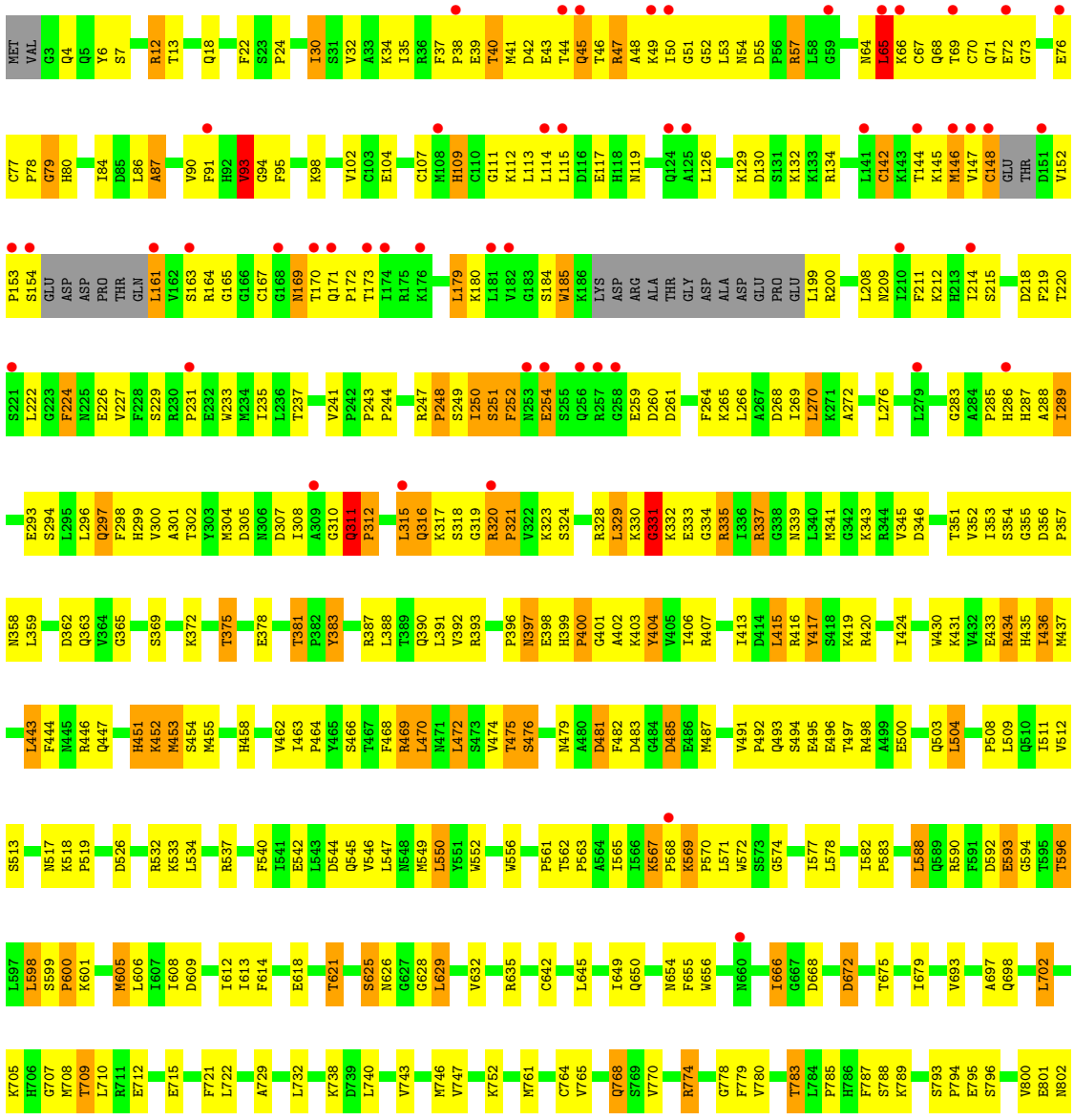
| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|---|
| 16 | T | 1 | Total | C | N | O | P | 0 | 0 |
| | | | 32 | 11 | 5 | 13 | 3 | | |

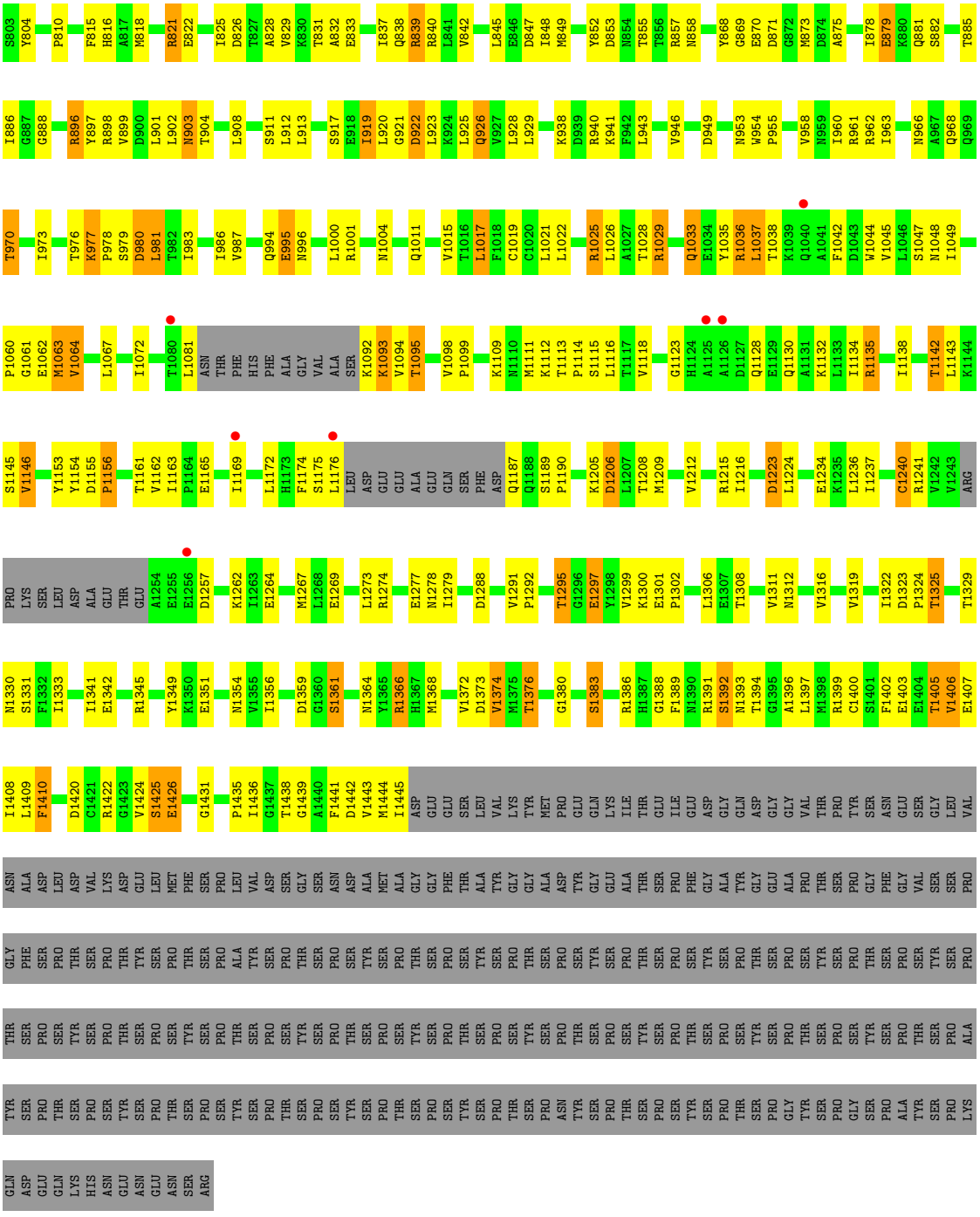
# 3 Residue-property plots (i)

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 II subunit RPB1

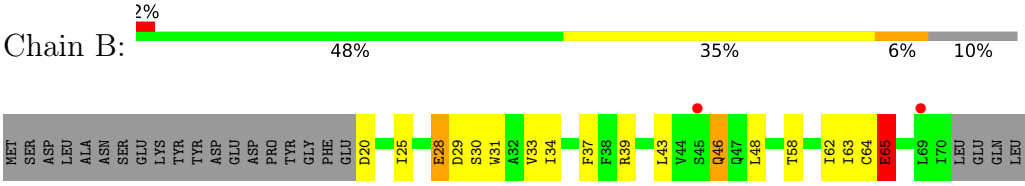
Chain A:

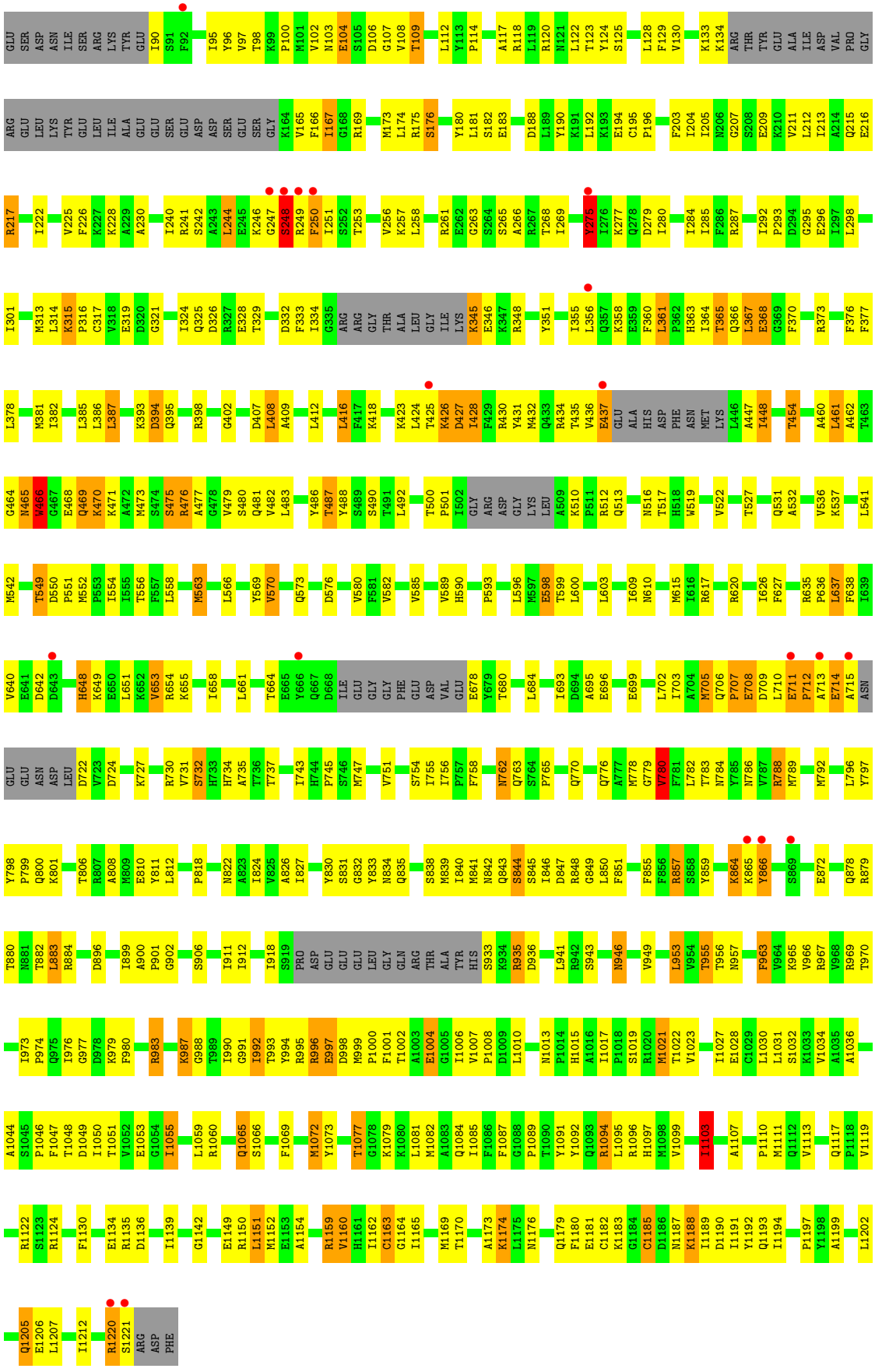





• Molecule 2: DNA-directed RNA polymerase II subunit RPB2,DNA-directed RNA polymerase II subunit RPB2

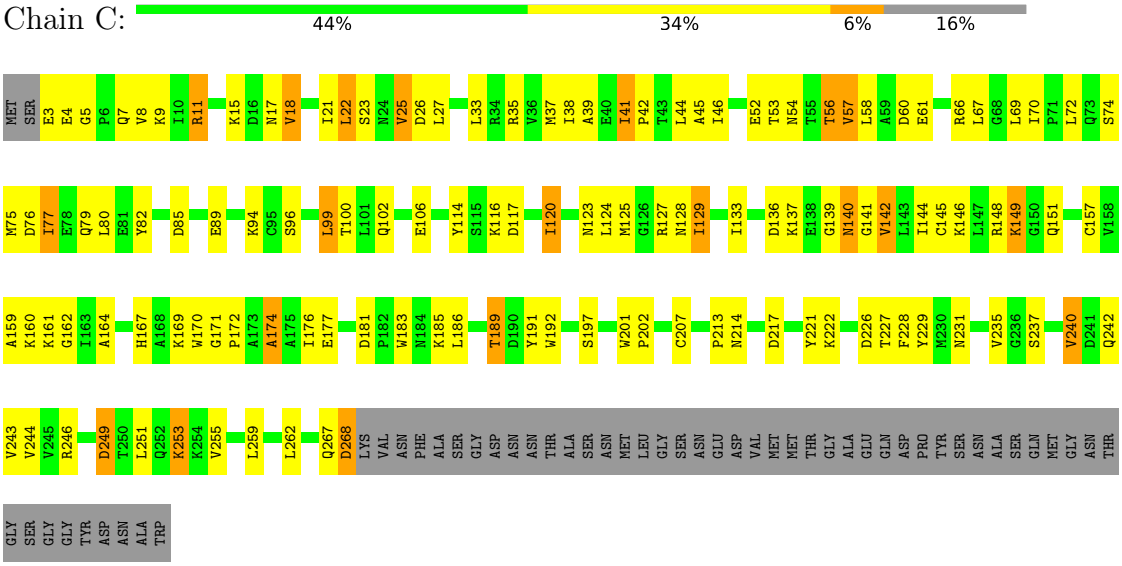
Chain B:




GLU I90 S91 F92 I95 Y96 V97 T98 K99 P100 M101 V102 N103 E104 S105 D106 G107 V108 T109 L112 Y113 P114 A117 R118 L119 R120 N121 L122 T123 Y124 S125 L128 F129 V130 K133 K134
ARG GLU LEU LYS TYR GLU LEU ILE ALA GLU GLU SER GLU ASP ASP SER GLU SER GLY K164 V165 F166 I167 G168 R169 M173 L174 R175 S176 Y180 L181 S182 E183 D188 L189 Y190 K191 L192 K193 E194 C195 P196 F203 I204 I205 N206 G207 S208 E209 K210 V211 L212 I213 A214 Q215 E216
R217 I222 V225 F226 K227 K228 A229 A230 I240 R241 S242 A243 L244 E245 K246 G247 S248 R249 F250 I251 S252 T253 V256 K257 L258 R261 E262 G263 S264 S265 A266 R267 T268 I269 Y276 I276 K277 Q278 D279 I280 I284 I285 P286 R287 I292 P293 D294 G295 E296 I297 L298
I301 M313 L314 K315 P316 C317 V318 E319 D320 G321 I324 Q325 D326 R327 E328 T329 D332 F333 I334 G335 ARG ARG GLY THR ALA LEU GLY ILE LYS K345 E346 K347 R348 Y351 I355 L356 Q357 K358 E359 F360 L361 P362 H363 I364 T365 Q366 L367 E368 G369 F370 R373 F376 F377
L378 M381 I382 L385 L386 L387 K393 D394 Q395 R398 G402 D407 L408 A409 L412 L416 F417 K418 K423 L424 T425 K426 D427 I428 F429 R430 Y431 M432 Q433 R434 T435 V436 E437 GLU ALA HIS ASP PHE ASN MET LYS L446 A447 I448 T454 A460 L461 A462 T463
G464 N465 K466 G467 E468 Q469 K470 K471 A472 M473 S474 S475 R476 A477 G478 V479 S480 Q481 V482 L483 Y486 T487 Y488 S489 S490 T491 L492 T500 P501 I502 GLY ARG ASP GLY LYS LEU A509 K510 P511 R512 Q513 N516 T517 H518 W519 V522 T527 Q531 A532 V536 K537 L541
M542 T549 D550 P551 M552 P553 I554 L555 T556 F557 L558 M563 L566 Y569 V570 Q573 D576 V580 F581 V582 V585 V589 H590 P593 L596 H597 E598 T599 L600 L603 I609 M610 M615 T616 R617 R620 I626 F627 R635 P636 L637 F638 I639
V640 E641 D642 D643 H648 K649 E650 L651 K652 V653 R654 K655 I658 L661 T664 E665 Y666 Q667 D668 ILE GLU GLY GLY PHE GLU ASP VAL GLU E678 Y679 T680 L684 I693 D694 A695 E696 E699 L702 I703 A704 M705 Q706 P707 E708 D709 L710 E711 P712 A713 E714 A715 ASN
GLU GLU ASN ASP LEU D722 V723 D724 K727 R730 V731 S732 H733 H734 A735 T736 T737 I743 H744 P745 S746 M747 V751 S754 I755 I756 P757 F758 N762 Q763 S764 P765 Q770 Q776 A777 M778 G779 V780 F781 L782 T783 N784 Y785 N786 V787 R788 M789 M792 L796 Y797
Y798 P799 Q800 K801 T806 R807 A808 M809 E810 Y811 L812 P818 N822 A823 I824 V825 A826 I827 Y830 S831 G832 Y833 N834 Q835 S838 M839 I840 M841 N842 Q843 S844 S845 I846 D847 R848 G849 L850 F851 F855 F856 R857 S858 Y859 K864 K865 Y866 S869 E872 Q878 R879
T880 N881 T882 L883 R884 D896 I899 A900 P901 G902 S906 I911 I912 I918 S919 PRO ASP GLU GLU GLU LEU GLY GLN ARG THR ALA TYR HIS S933 K934 R935 D936 L941 R942 S943 N946 V949 L953 V954 Y955 T956 N957 F963 V964 K965 V966 R967 V968 R969 T970
I973 P974 Q976 I976 G977 D978 K979 F980 R983 K987 G988 T989 I990 G991 I992 T993 Y994 R995 R996 E997 D998 M999 P1000 F1001 T1002 A1003 E1004 G1005 I1006 V1007 P1008 D1009 L1010 N1013 P1014 H1015 A1016 I1017 P1018 S1019 R1020 M1021 T1022 V1023 I1027 E1028 C1029 L1030 L1031 S1032 K1033 V1034 A1035 A1036
A1044 S1045 P1046 F1047 T1048 D1049 I1050 T1051 V1052 E1053 G1054 I1055 L1059 R1060 Q1066 S1066 F1069 M1072 Y1073 T1077 G1078 K1079 K1080 L1081 M1082 A1083 Q1084 I1085 F1086 F1087 G1088 P1089 T1090 Y1091 Y1092 Q1093 R1094 L1095 R1096 H1097 M1098 V1099 I1103 A1107 P1110 M1111 Q1112 V1113 Q1117 P1118 V1119
R1122 S1123 R1124 F1130 E1134 R1135 D1136 I1139 G1142 E1149 R1150 L1151 M1152 E1153 A1154 R1159 V1160 H1161 I1162 C1163 G1164 I1165 M1169 T1170 A1173 K1174 L1175 N1176 Q1179 F1180 E1181 C1182 K1183 G1184 C1185 D1186 N1187 K1188 I1189 D1190 I1191 Y1192 Q1193 I1194 P1197 Y1198 A1199 L1202
Q1205 E1206 L1207 I1212 R1220 S1221 ARG ASP PHE

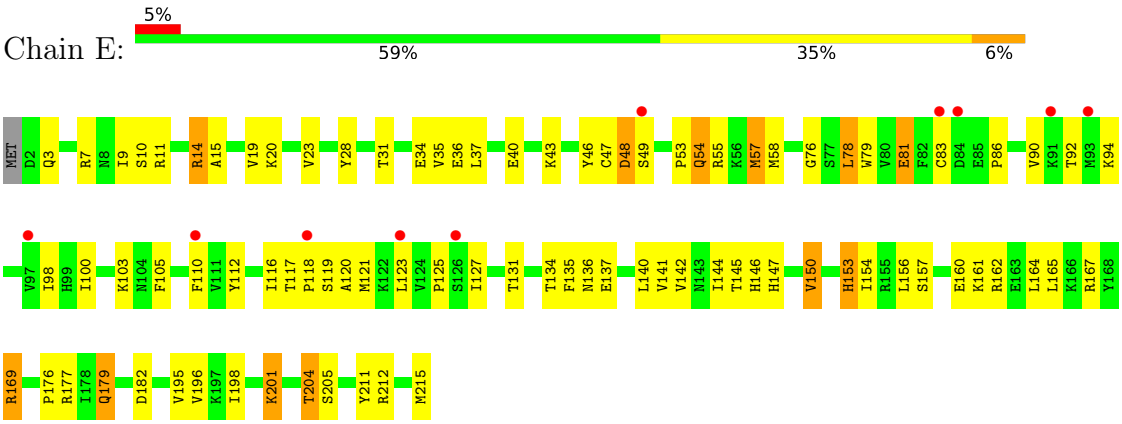
• Molecule 3: DNA-directed RNA polymerase II subunit RPB3

Chain C:



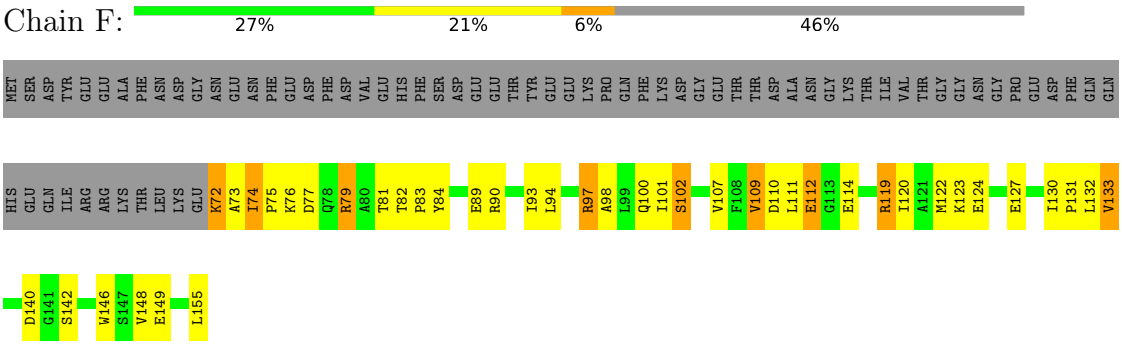

• Molecule 4: DNA-directed RNA polymerases I, II, and III subunit RPABC1

Chain E:



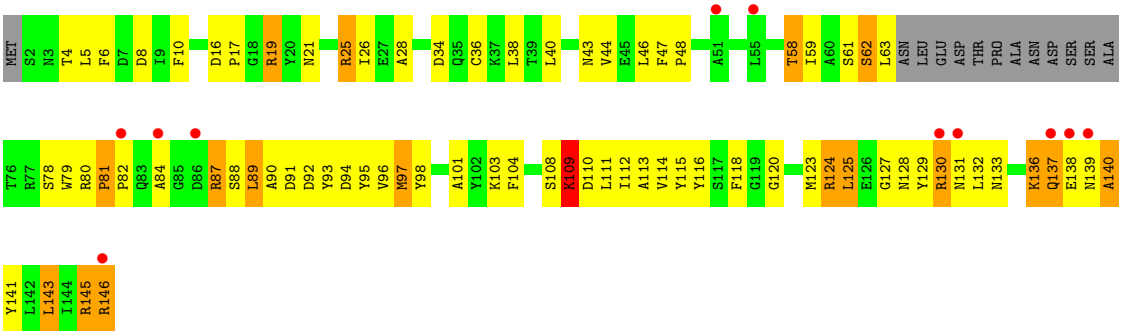

• Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:

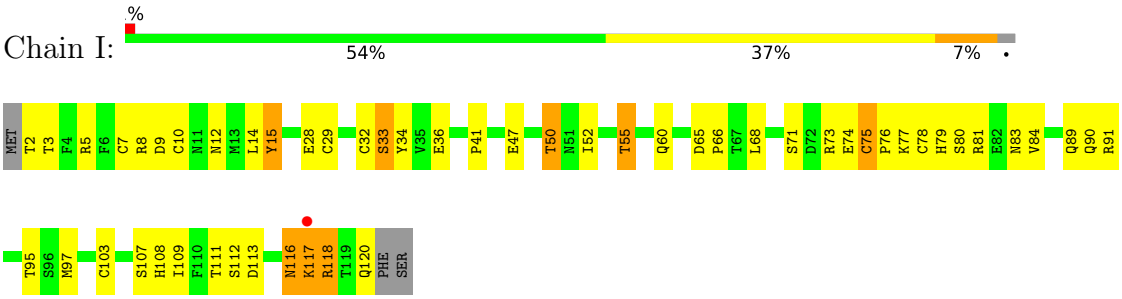



• Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC3

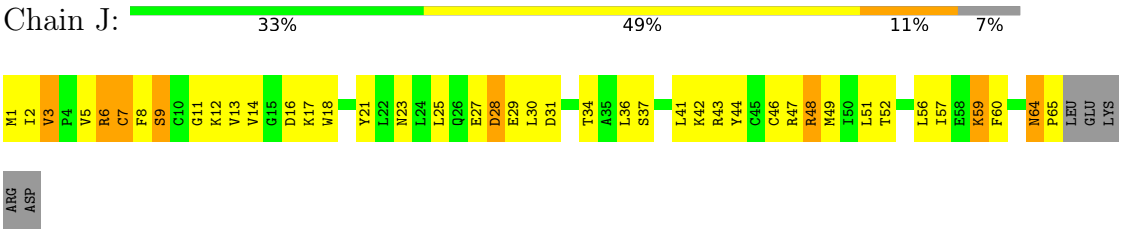
Chain H:



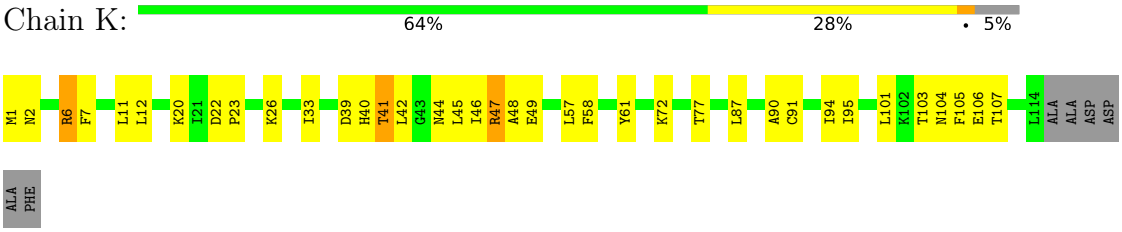

• Molecule 7: DNA-directed RNA polymerase II subunit RPB9



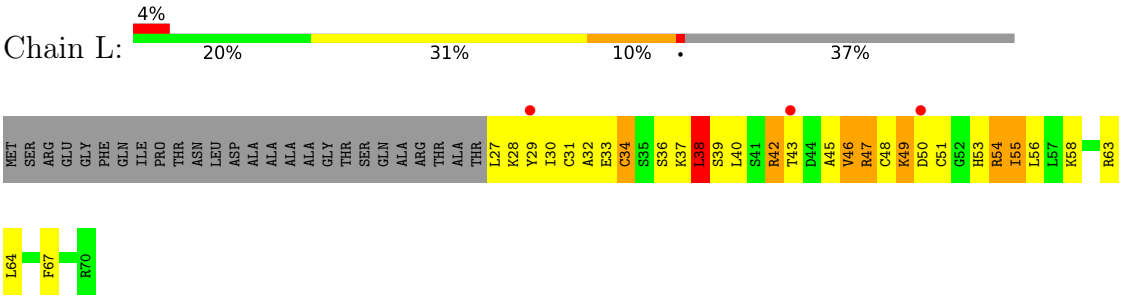

• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC5




• Molecule 9: DNA-directed RNA polymerase II subunit RPB11

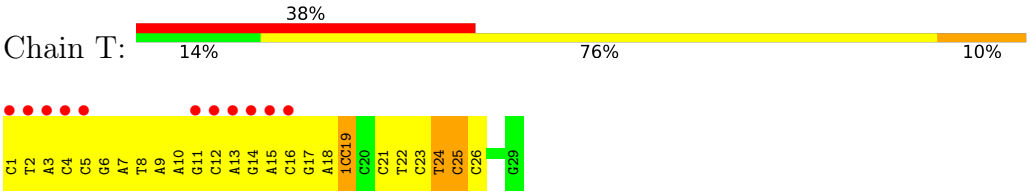



• Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC4



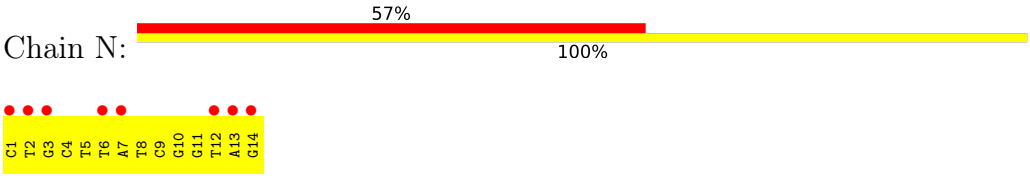
• Molecule 11: DNA (29-MER)

Chain T:



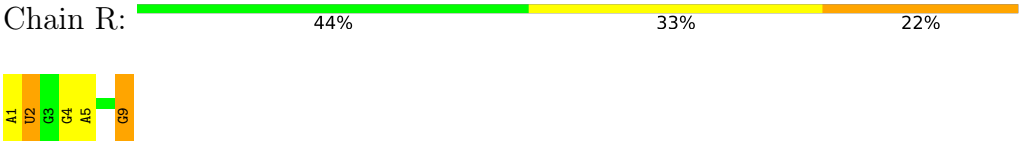

• Molecule 12: DNA (5'-D(*CP*TP*GP*CP*TP*TP*AP*TP*CP*GP*GP*TP*AP*G)-3')

Chain N:




• Molecule 13: RNA (5'-D(*AP*UP*GP*GP*AP*GP*AP*GP*G)-3')

Chain R:



# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | C 1 2 1 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 168.18Å 222.61Å 192.83Å<br>90.00° 101.59° 90.00° | Depositor |
| Resolution (Å) | 48.88 – 3.30<br>48.88 – 3.26 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 98.0 (48.88-3.30)<br>98.1 (48.88-3.26) | Depositor<br>EDS |
| $R_{merge}$ | 0.15 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 1.64 (at 3.25Å) | Xtriage |
| Refinement program | PHENIX | Depositor |
| R, $R_{free}$ | 0.207 , 0.250<br>0.205 , 0.248 | Depositor<br>DCC |
| $R_{free}$ test set | 5268 reflections (4.98%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 83.9 | Xtriage |
| Anisotropy | 0.481 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.27 , 52.7 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.47$, $< L^2 > = 0.30$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| $F_o$,$F_c$ correlation | 0.93 | EDS |
| Total number of atoms | 29193 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 102.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.70% 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: 1CC, ZN, MG, G2P

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 | 0.43 | 1/11146 (0.0%) | 0.65 | 5/15066 (0.0%) |
| 2 | B | 0.47 | 1/8932 (0.0%) | 0.67 | 3/12045 (0.0%) |
| 3 | C | 0.47 | 0/2133 | 0.71 | 1/2891 (0.0%) |
| 4 | E | 0.38 | 0/1788 | 0.55 | 0/2406 |
| 5 | F | 0.40 | 0/691 | 0.61 | 0/933 |
| 6 | H | 0.38 | 0/1086 | 0.56 | 0/1470 |
| 7 | I | 0.40 | 0/989 | 0.64 | 1/1331 (0.1%) |
| 8 | J | 0.49 | 0/541 | 0.69 | 0/727 |
| 9 | K | 0.46 | 0/937 | 0.61 | 0/1265 |
| 10 | L | 0.32 | 0/353 | 0.63 | 0/468 |
| 11 | T | 0.61 | 0/632 | 1.17 | 7/969 (0.7%) |
| 12 | N | 0.18 | 0/317 | 0.56 | 0/488 |
| 13 | R | 0.56 | 0/223 | 1.01 | 0/348 |
| All | All | 0.44 | 2/29768 (0.0%) | 0.67 | 17/40407 (0.0%) |

All (2) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 2 | B | 707 | PRO | N-CD | 5.28 | 1.55 | 1.47 |
| 1 | A | 400 | PRO | N-CD | 5.19 | 1.55 | 1.47 |

The worst 5 of 17 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 1 | A | 331 | GLY | N-CA-C | 7.95 | 132.97 | 113.10 |
| 11 | T | 24 | DT | N3-C4-O4 | 7.17 | 124.20 | 119.90 |
| 11 | T | 24 | DT | P-O5'-C5' | -6.69 | 110.20 | 120.90 |
| 1 | A | 919 | ILE | CB-CA-C | -6.33 | 98.94 | 111.60 |
| 11 | T | 25 | DC | O4'-C4'-C3' | -6.09 | 102.06 | 104.50 |

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|---|---|---|---|---|---|---|
| 1 | A | 10953 | 0 | 11053 | 681 | 0 |
| 2 | B | 8762 | 0 | 8797 | 459 | 0 |
| 3 | C | 2095 | 0 | 2051 | 104 | 0 |
| 4 | E | 1752 | 0 | 1776 | 60 | 0 |
| 5 | F | 679 | 0 | 701 | 31 | 0 |
| 6 | H | 1068 | 0 | 1040 | 88 | 0 |
| 7 | I | 971 | 0 | 927 | 36 | 0 |
| 8 | J | 532 | 0 | 542 | 66 | 0 |
| 9 | K | 919 | 0 | 929 | 29 | 0 |
| 10 | L | 351 | 0 | 374 | 65 | 0 |
| 11 | T | 587 | 0 | 326 | 123 | 0 |
| 12 | N | 284 | 0 | 161 | 49 | 0 |
| 13 | R | 198 | 0 | 99 | 9 | 0 |
| 14 | A | 2 | 0 | 0 | 0 | 0 |
| 14 | B | 1 | 0 | 0 | 0 | 0 |
| 14 | C | 1 | 0 | 0 | 0 | 0 |
| 14 | I | 2 | 0 | 0 | 0 | 0 |
| 14 | J | 1 | 0 | 0 | 0 | 0 |
| 14 | L | 1 | 0 | 0 | 0 | 0 |
| 15 | A | 2 | 0 | 0 | 0 | 0 |
| 16 | T | 32 | 0 | 14 | 2 | 0 |
| All | All | 29193 | 0 | 28790 | 1651 | 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 29.

The worst 5 of 1651 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|---|---|---|---|
| 1:A:317:LYS:CG | 1:A:318:SER:HB2 | 1.25 | 1.56 |
| 2:B:714:GLU:HA | 2:B:715:ALA:CB | 1.40 | 1.45 |
| 1:A:317:LYS:HG3 | 1:A:318:SER:CB | 1.46 | 1.42 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|---|---|---|---|
| 1:A:317:LYS:CG | 1:A:318:SER:CB | 1.98 | 1.42 |
| 10:L:38:LEU:HD12 | 10:L:40:LEU:CG | 1.52 | 1.36 |

There are no symmetry-related clashes.

## 5.3 Torsion angles (i)

### 5.3.1 Protein backbone (i)

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

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

| Mol | Chain | Analysed | Favoured | Allowed | Outliers | Percentiles |
|---|---|---|---|---|---|---|
| 1 | A | 1379/1733 (80%) | 1169 (85%) | 177 (13%) | 33 (2%) | 6 28 |
| 2 | B | 1085/1224 (89%) | 927 (85%) | 135 (12%) | 23 (2%) | 7 31 |
| 3 | C | 264/318 (83%) | 228 (86%) | 35 (13%) | 1 (0%) | 34 66 |
| 4 | E | 212/215 (99%) | 183 (86%) | 25 (12%) | 4 (2%) | 8 34 |
| 5 | F | 82/155 (53%) | 72 (88%) | 10 (12%) | 0 | 100 100 |
| 6 | H | 129/146 (88%) | 106 (82%) | 16 (12%) | 7 (5%) | 2 12 |
| 7 | I | 117/122 (96%) | 98 (84%) | 17 (14%) | 2 (2%) | 9 35 |
| 8 | J | 63/70 (90%) | 56 (89%) | 6 (10%) | 1 (2%) | 9 36 |
| 9 | K | 112/120 (93%) | 106 (95%) | 6 (5%) | 0 | 100 100 |
| 10 | L | 42/70 (60%) | 28 (67%) | 11 (26%) | 3 (7%) | 1 7 |
| All | All | 3485/4173 (84%) | 2973 (85%) | 438 (13%) | 74 (2%) | 7 31 |

5 of 74 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 286 | HIS |
| 1 | A | 321 | PRO |
| 1 | A | 593 | GLU |
| 1 | A | 1393 | ASN |
| 2 | B | 230 | ALA |

### 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 | A | 1216/1520 (80%) | 1030 (85%) | 186 (15%) | 2 12 |
| 2 | B | 957/1061 (90%) | 813 (85%) | 144 (15%) | 3 13 |
| 3 | C | 234/274 (85%) | 198 (85%) | 36 (15%) | 2 12 |
| 4 | E | 196/197 (100%) | 167 (85%) | 29 (15%) | 3 13 |
| 5 | F | 74/137 (54%) | 59 (80%) | 15 (20%) | 1 5 |
| 6 | H | 117/128 (91%) | 99 (85%) | 18 (15%) | 2 12 |
| 7 | I | 113/116 (97%) | 91 (80%) | 22 (20%) | 1 5 |
| 8 | J | 60/65 (92%) | 50 (83%) | 10 (17%) | 2 10 |
| 9 | K | 99/102 (97%) | 89 (90%) | 10 (10%) | 7 27 |
| 10 | L | 39/57 (68%) | 29 (74%) | 10 (26%) | 0 2 |
| All | All | 3105/3657 (85%) | 2625 (84%) | 480 (16%) | 2 12 |

5 of 480 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 2 | B | 427 | ASP |
| 7 | I | 90 | GLN |
| 2 | B | 946 | ASN |
| 7 | I | 60 | GLN |
| 10 | L | 46 | VAL |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 3 | C | 195 | GLN |
| 3 | C | 231 | ASN |
| 6 | H | 137 | GLN |
| 2 | B | 465 | ASN |
| 2 | B | 395 | GLN |

### 5.3.3 RNA (i)

| Mol | Chain | Analysed | Backbone Outliers | Pucker Outliers |
|---|---|---|---|---|
| 13 | R | 8/9 (88%) | 2 (25%) | 0 |

All (2) RNA backbone outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 13 | R | 2 | U |
| 13 | R | 9 | 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 | $\#\|Z\| > 2$ |
| 11 | 1CC | T | 19 | 11 | 20,23,24 | 1.23 | 3 (15%) | 26,33,36 | 1.39 | 3 (11%) |

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 | 1CC | T | 19 | 11 | - | 0/11/25/26 | 0/2/2/2 |

All (3) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 11 | T | 19 | 1CC | O5'-C5' | -2.51 | 1.38 | 1.44 |
| 11 | T | 19 | 1CC | C5-C4 | 2.23 | 1.47 | 1.43 |
| 11 | T | 19 | 1CC | C2-N1 | -2.20 | 1.35 | 1.40 |

All (3) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 11 | T | 19 | 1CC | C5-C6-N1 | -2.87 | 119.28 | 123.10 |
| 11 | T | 19 | 1CC | C2'-C1'-N1 | -2.65 | 107.67 | 113.77 |
| 11 | T | 19 | 1CC | O2-C2-N3 | -2.57 | 118.15 | 122.33 |

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 11 | T | 19 | 1CC | 3 | 0 |

## 5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry (i)

Of 11 ligands modelled in this entry, 10 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 | $\#\|Z\| > 2$ |
| 16 | G2P | T | 101 | 15 | 27,34,34 | 2.36 | 10 (37%) | 33,54,54 | 1.94 | 5 (15%) |

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 |
|---|---|---|---|---|---|---|---|
| 16 | G2P | T | 101 | 15 | - | 4/15/38/38 | 0/3/3/3 |

The worst 5 of 10 bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 16 | T | 101 | G2P | C6-N1 | 4.89 | 1.41 | 1.33 |
| 16 | T | 101 | G2P | PB-O2B | 4.31 | 1.61 | 1.51 |
| 16 | T | 101 | G2P | C2-N1 | 4.19 | 1.42 | 1.35 |
| 16 | T | 101 | G2P | PA-O2A | 4.15 | 1.61 | 1.51 |
| 16 | T | 101 | G2P | O4'-C1' | 4.10 | 1.46 | 1.41 |

All (5) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 16 | T | 101 | G2P | C5-C6-N1 | -7.52 | 113.15 | 123.43 |
| 16 | T | 101 | G2P | C2-N1-C6 | 3.29 | 121.15 | 115.93 |
| 16 | T | 101 | G2P | O1G-PG-O3B | 3.03 | 114.81 | 104.64 |
| 16 | T | 101 | G2P | C1'-N9-C4 | -2.26 | 122.67 | 126.64 |
| 16 | T | 101 | G2P | O2B-PB-C3A | -2.13 | 103.44 | 109.07 |

There are no chirality outliers.

All (4) torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|---|---|---|---|---|
| 16 | T | 101 | G2P | O4'-C4'-C5'-O5' |
| 16 | T | 101 | G2P | C3'-C4'-C5'-O5' |
| 16 | T | 101 | G2P | C5'-O5'-PA-O1A |
| 16 | T | 101 | G2P | C4'-C5'-O5'-PA |

There are no ring outliers.

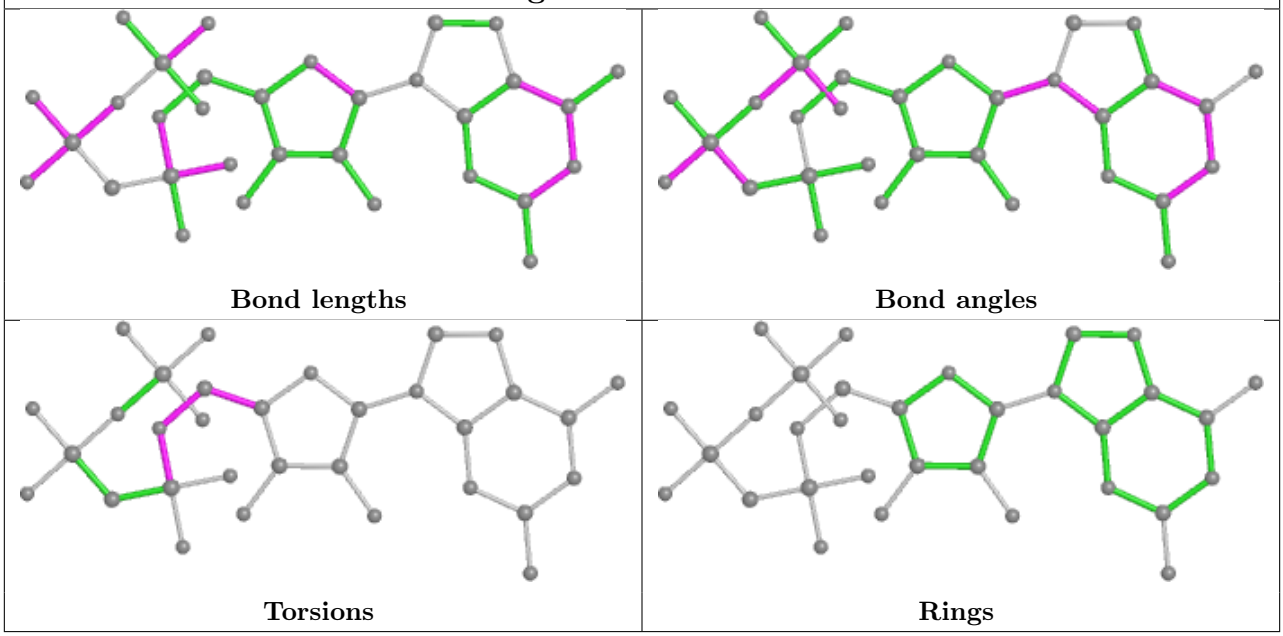
1 monomer is involved in 2 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 16 | T | 101 | G2P | 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 G2P T 101**




## 5.7 Other polymers (i)

There are no such residues in this entry.

## 5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

# 6 Fit of model and data

## 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(Å$^2$) | Q<0.9 |
|---|---|---|---|---|---|---|
| 1 | A | 1393/1733 (80%) | 0.05 | 58 (4%) 36 34 | 45, 88, 191, 305 | 0 |
| 2 | B | 1103/1224 (90%) | -0.08 | 21 (1%) 66 65 | 42, 79, 153, 277 | 0 |
| 3 | C | 266/318 (83%) | -0.31 | 0 100 100 | 50, 79, 117, 170 | 0 |
| 4 | E | 214/215 (99%) | 0.04 | 10 (4%) 31 29 | 63, 114, 201, 236 | 0 |
| 5 | F | 84/155 (54%) | -0.24 | 0 100 100 | 65, 93, 138, 182 | 0 |
| 6 | H | 133/146 (91%) | 0.46 | 11 (8%) 11 11 | 75, 116, 207, 288 | 0 |
| 7 | I | 119/122 (97%) | -0.18 | 1 (0%) 86 86 | 58, 97, 139, 186 | 0 |
| 8 | J | 65/70 (92%) | -0.23 | 0 100 100 | 56, 74, 113, 131 | 0 |
| 9 | K | 114/120 (95%) | -0.13 | 0 100 100 | 54, 86, 115, 164 | 0 |
| 10 | L | 44/70 (62%) | 0.69 | 3 (6%) 17 17 | 73, 158, 247, 292 | 0 |
| 11 | T | 28/29 (96%) | 1.43 | 11 (39%) 0 0 | 60, 289, 368, 372 | 0 |
| 12 | N | 14/14 (100%) | 1.82 | 8 (57%) 0 0 | 249, 290, 330, 348 | 0 |
| 13 | R | 9/9 (100%) | -0.47 | 0 100 100 | 54, 66, 111, 116 | 0 |
| All | All | 3586/4225 (84%) | -0.00 | 123 (3%) 45 43 | 42, 87, 187, 372 | 0 |

The worst 5 of 123 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 1 | A | 257 | ARG | 6.9 |
| 6 | H | 86 | ASP | 6.5 |
| 2 | B | 1221 | SER | 6.1 |
| 2 | B | 250 | PHE | 5.8 |
| 10 | L | 50 | ASP | 5.7 |

## 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 |
|---|---|---|---|---|---|---|---|---|
| 11 | 1CC | T | 19 | 22/23 | 0.95 | 0.13 | 71,87,132,144 | 0 |

## 6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

## 6.4 Ligands (i)

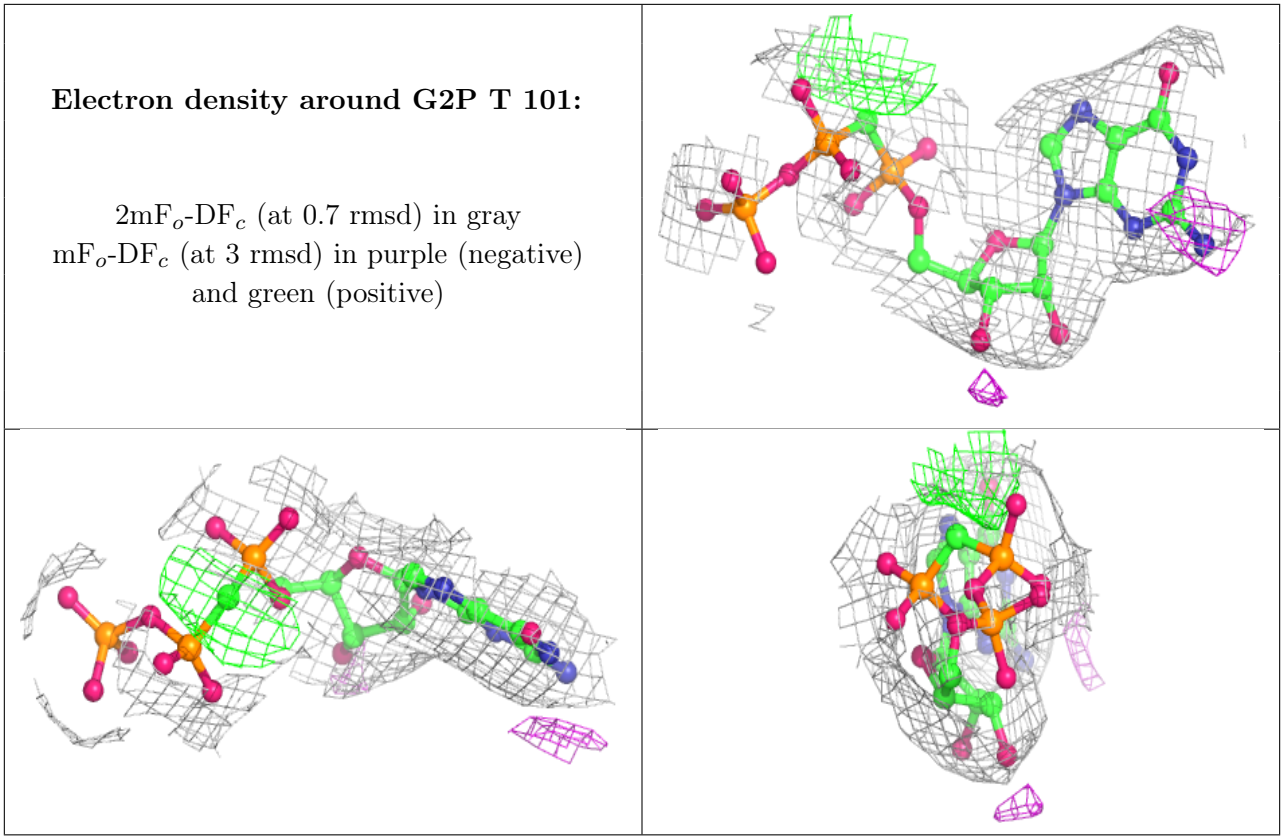
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, $95^{th}$ percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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 |
|---|---|---|---|---|---|---|---|---|
| 14 | ZN | A | 1801 | 1/1 | 0.70 | 0.08 | 266,266,266,266 | 0 |
| 15 | MG | A | 1804 | 1/1 | 0.83 | 0.20 | 87,87,87,87 | 0 |
| 14 | ZN | L | 101 | 1/1 | 0.87 | 0.06 | 172,172,172,172 | 0 |
| 16 | G2P | T | 101 | 32/32 | 0.91 | 0.18 | 59,73,127,131 | 0 |
| 14 | ZN | A | 1802 | 1/1 | 0.94 | 0.08 | 136,136,136,136 | 0 |
| 14 | ZN | B | 1301 | 1/1 | 0.97 | 0.12 | 123,123,123,123 | 0 |
| 15 | MG | A | 1803 | 1/1 | 0.97 | 0.17 | 42,42,42,42 | 0 |
| 14 | ZN | I | 202 | 1/1 | 0.97 | 0.13 | 93,93,93,93 | 0 |
| 14 | ZN | J | 101 | 1/1 | 0.97 | 0.19 | 92,92,92,92 | 0 |
| 14 | ZN | I | 201 | 1/1 | 0.98 | 0.09 | 104,104,104,104 | 0 |
| 14 | ZN | C | 401 | 1/1 | 0.99 | 0.12 | 107,107,107,107 | 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 G2P T 101:**

$2mF_o$-$DF_c$ (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.