# W O R L D W I D E
# wwPDB
## PROTEIN DATA BANK

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

May 23, 2020 – 08:15 pm BST

| | | |
|---|---|---|
| PDB ID | : | 1Y1Y |
| Title | : | RNA Polymerase II-TFIIS-DNA/RNA complex |
| Authors | : | Cramer, P.; Kettenberger, H.; Armache, K.-J. |
| Deposited on | : | 2004-11-19 |
| Resolution | : | 4.00 Å(reported) |

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

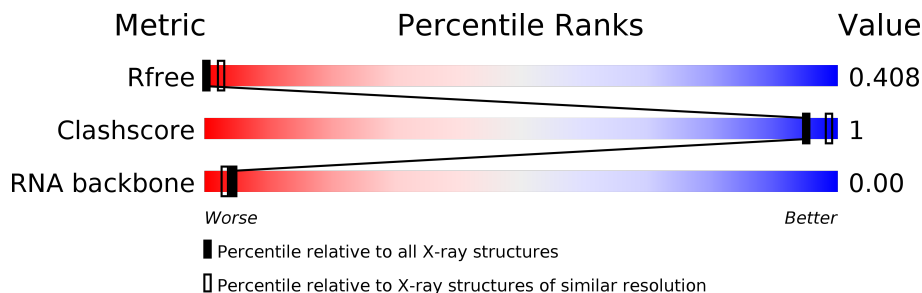| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.11 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (Gargrove) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.11 |

# 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 4.00 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.
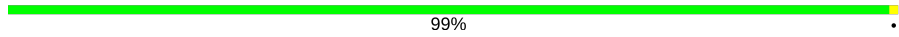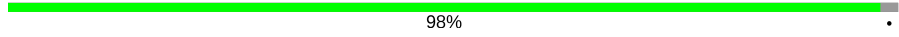
| Metric | Percentile Ranks | Value |
|---|---|---|
| Rfree | | 0.408 |
| Clashscore | | 1 |
| RNA backbone | | 0.00 |

*Worse* *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
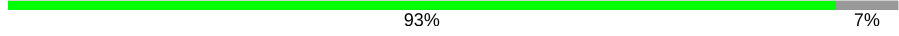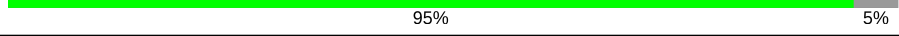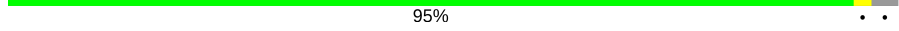
| Metric | Whole archive (#Entries) | Similar resolution (#Entries, resolution range(Å)) |
|---|---|---|
| $R_{free}$ | 130704 | 1087 (4.30-3.70) |
| Clashscore | 141614 | 1148 (4.30-3.70) |
| RNA backbone | 3102 | 1048 (5.00-3.00) |

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 on 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%

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 1 | T | 7 | 100% |
| 2 | P | 4 | 100% |
| 3 | A | 1733 | 82% / 18% |
| 4 | B | 1224 | 91% / 9% |
| 5 | C | 318 | 84% / 16% |
| 6 | D | 221 | 80% / 20% |
| 7 | E | 215 | 99% / . |
| 8 | F | 155 | 54% / 46% |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Length | Quality of chain |
|-----|-------|--------|------------------|
| 9 | G | 171 | 99% |
| 10 | H | 146 | 91% 9% |
| 11 | I | 122 | 98% |
| 12 | J | 70 | 93% 7% |
| 13 | K | 120 | 95% 5% |
| 14 | L | 70 | 66% 34% |
| 15 | S | 179 | 95% |

# 2 Entry composition (i)

There are 15 unique types of molecules in this entry. The entry contains 4112 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 5'-D(P*TP*AP*CP*GP*CP*CP*T)-3'.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---|---------|---------|-------|
| 1 | T | 7 | Total 7 | P 7 | 0 | 0 | 7 |

- Molecule 2 is a RNA chain called 5'-R(P*AP*GP*GP*C)-3'.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---|---------|---------|-------|
| 2 | P | 4 | Total 4 | P 4 | 0 | 0 | 4 |

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

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---|---------|---------|-------|
| 3 | A | 1426 | Total 1426 | C 1426 | 0 | 0 | 1426 |

- Molecule 4 is a protein called DNA-directed RNA polymerase II 140 kDa polypeptide.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---|---------|---------|-------|
| 4 | B | 1112 | Total 1112 | C 1112 | 8 | 0 | 1112 |

- Molecule 5 is a protein called DNA-directed RNA polymerase II 45 kDa polypeptide.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---|---------|---------|-------|
| 5 | C | 266 | Total 266 | C 266 | 0 | 0 | 266 |

- Molecule 6 is a protein called DNA-directed RNA polymerase II 32 kDa polypeptide.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---|---------|---------|-------|
| 6 | D | 177 | Total 177 | C 177 | 0 | 0 | 177 |

- Molecule 7 is a protein called DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|
| 7 | E | 214 | Total<br>214 | C<br>214 | 0 | 0 | 214 |

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|
| 8 | F | 84 | Total<br>84 | C<br>84 | 0 | 0 | 84 |

- Molecule 9 is a protein called DNA-directed RNA polymerase II 19 kDa polypeptide.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|
| 9 | G | 171 | Total<br>171 | C<br>171 | 0 | 0 | 171 |

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|
| 10 | H | 133 | Total<br>133 | C<br>133 | 0 | 0 | 133 |

- Molecule 11 is a protein called DNA-directed RNA polymerase II subunit 9.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|
| 11 | I | 119 | Total<br>119 | C<br>119 | 0 | 0 | 119 |

- Molecule 12 is a protein called DNA-directed RNA polymerases I/II/III subunit 10.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|
| 12 | J | 65 | Total<br>65 | C<br>65 | 0 | 0 | 65 |

- Molecule 13 is a protein called DNA-directed RNA polymerase II 13.6 kDa polypeptide.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|
| 13 | K | 114 | Total<br>114 | C<br>114 | 0 | 0 | 114 |

- Molecule 14 is a protein called DNA-directed RNA polymerases I, II, and III 7.7 kDa polypeptide.

| Mol | Chain | Residues | Atoms | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---------|---------|-------|
| 14 | L | 46 | Total C<br>46 46 | 0 | 0 | 46 |

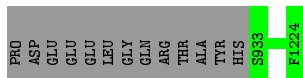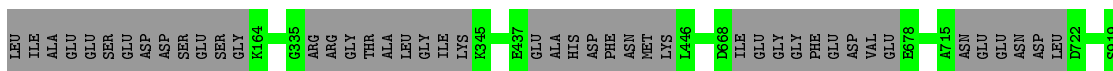- Molecule 15 is a protein called Transcription elongation factor S-II.

| Mol | Chain | Residues | Atoms | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---------|---------|-------|
| 15 | S | 174 | Total C<br>174 174 | 0 | 0 | 174 |

# 3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: 5'-D(P*TP*AP*CP*GP*CP*CP*T)-3'

Chain T:



There are no outlier residues recorded for this chain.

- Molecule 2: 5'-R(P*AP*GP*GP*C)-3'

Chain P:



There are no outlier residues recorded for this chain.

- Molecule 3: DNA-directed RNA polymerase II largest subunit

Chain A:



- Molecule 4: DNA-directed RNA polymerase II 140 kDa polypeptide

Chain B:

LEU ILE ALA GLU GLU SER GLU ASP SER GLU SER GLY K164 G335 ARG ARG GLY THR ALA LEU GLY ILE LYS K345 E437 GLU ALA HIS ASP PHE ASN MET LYS L446 D68 ILE GLU GLY PHE GLU ASP VAL GLU E678 A715 ASN GLU GLU ASN ASP LEU D722 S919

PRO ASP GLU GLU GLU LEU GLY GLN ARG THR ALA TYR HIS S933 F1224

• Molecule 5: DNA-directed RNA polymerase II 45 kDa polypeptide

Chain C:

84% 16%

MET SER E3 D268 LYS VAL ASN PHE ALA SER GLY ASP ASN ASN THR ALA SER ASN MET LEU GLY SER ASN GLU ASP VAL MET THR GLY ALA GLU GLU ASP PRO TYR SER ASN ALA SER GLN MET GLY ASN THR GLY SER GLY GLY TYR ASP ASN ALA TRP
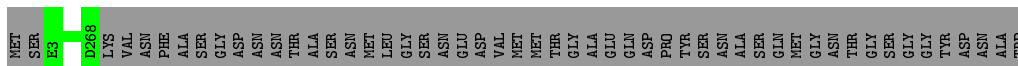
• Molecule 6: DNA-directed RNA polymerase II 32 kDa polypeptide

Chain D:

80% 20%

MET ASN VAL S4 K76 HIS LYS LYS LYS HIS LEU LYS HIS GLU ASN ALA ASN ASP GLU THR ALA VAL GLU ASP GLU ASP ASP LEU ASP GLU ASP VAL ASN ALA ASP ASP ASP PHE MET HIS SER GLU T118 Y221

• Molecule 7: DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide

Chain E:

99% .

MET D2 P128 P129 M215

• Molecule 8: DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide

Chain F:

54% 46%

MET SER ASP TYR GLU GLU ALA PHE ASN ASP GLY ASN ASN GLU ASN PHE GLU ASP PHE ASP VAL GLU HIS PHE SER GLU GLU THR TYR GLU GLU LYS PRO GLN PHE LYS ASP GLY GLU THR THR ASP ALA ASN LYS THR ILE VAL THR GLY GLY ASN GLY PRO GLU ASP PHE GLN GLN

HIS GLU GLN ILE ARG ARG LYS THR LEU LYS GLU K72 L155
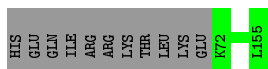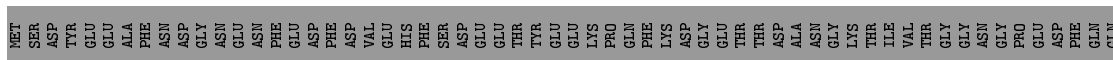
• Molecule 9: DNA-directed RNA polymerase II 19 kDa polypeptide

Chain G:

99% .

M1 N126 P127 I171

• Molecule 10: DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide

Chain H:

91% 9%

• Molecule 11: DNA-directed RNA polymerase II subunit 9

Chain I:





• Molecule 12: DNA-directed RNA polymerases I/II/III subunit 10

Chain J:





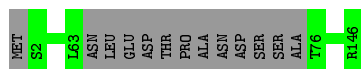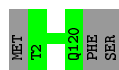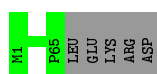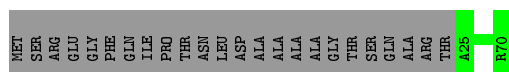• Molecule 13: DNA-directed RNA polymerase II 13.6 kDa polypeptide

Chain K:





• Molecule 14: DNA-directed RNA polymerases I, II, and III 7.7 kDa polypeptide

Chain L:





• Molecule 15: Transcription elongation factor S-II

Chain S:

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|----------|-------|--------|
| Space group | C 2 2 21 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 220.20Å   395.70Å   282.10Å<br>90.00°    90.00°    90.00° | Depositor |
| Resolution (Å) | 50.00 – 4.00<br>39.19 – 3.98 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | (Not available) (50.00-4.00)<br>98.1 (39.19-3.98) | Depositor<br>EDS |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | 0.07 | Depositor |
| $< I/\sigma(I) >$ [1] | 4.11 (at 4.00Å) | Xtriage |
| Refinement program | unknown | Depositor |
| R, $R_{free}$ | 0.281 , (Not available)<br>0.413 , 0.408 | Depositor<br>DCC |
| $R_{free}$ test set | 2038 reflections (1.97%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 106.8 | Xtriage |
| Anisotropy | 0.380 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.34 , 98.7 | EDS |
| L-test for twinning[2] | $< |L| > = 0.48, < L^2 > = 0.31$ | Xtriage |
| Estimated twinning fraction | 0.004 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l<br>0.015 for 1/2*h+1/2*k,3/2*h-1/2*k,-l | Xtriage |
| $F_o,F_c$ correlation | 0.65 | EDS |
| Total number of atoms | 4112 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 87.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.11% 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 ⓘ

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).

There are no protein, RNA or DNA chains available to summarize Z scores of covalent bonds and angles.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

## 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 | 7 | 0 | 0 | 0 | 0 |
| 2 | P | 4 | 0 | 0 | 0 | 0 |
| 3 | A | 1426 | 0 | 0 | 1 | 0 |
| 4 | B | 1112 | 0 | 0 | 0 | 0 |
| 5 | C | 266 | 0 | 0 | 0 | 0 |
| 6 | D | 177 | 0 | 0 | 0 | 0 |
| 7 | E | 214 | 0 | 0 | 1 | 0 |
| 8 | F | 84 | 0 | 0 | 0 | 0 |
| 9 | G | 171 | 0 | 0 | 1 | 0 |
| 10 | H | 133 | 0 | 0 | 0 | 0 |
| 11 | I | 119 | 0 | 0 | 0 | 0 |
| 12 | J | 65 | 0 | 0 | 0 | 0 |
| 13 | K | 114 | 0 | 0 | 0 | 0 |
| 14 | L | 46 | 0 | 0 | 0 | 0 |
| 15 | S | 174 | 0 | 0 | 2 | 0 |
| All | All | 4112 | 0 | 0 | 5 | 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 1.

All (5) 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) |
|---|---|---|---|
| 15:S:218:ILE:CA | 15:S:219:ALA:CA | 2.73 | 0.66 |
| 9:G:126:ASN:CA | 9:G:127:PRO:CA | 2.78 | 0.62 |
| 3:A:447:GLN:CA | 3:A:448:PRO:CA | 2.83 | 0.56 |
| 15:S:231:CYS:CA | 15:S:232:ASP:CA | 2.86 | 0.53 |
| 7:E:128:PRO:CA | 7:E:129:PRO:CA | 2.90 | 0.49 |

There are no symmetry-related clashes.

## 5.3 Torsion angles (i)

### 5.3.1 Protein backbone (i)

There are no protein backbone outliers to report in this entry.

### 5.3.2 Protein sidechains (i)

There are no protein residues with a non-rotameric sidechain to report in this entry.

### 5.3.3 RNA (i)

| Mol | Chain | Analysed | Backbone Outliers | Pucker Outliers |
|---|---|---|---|---|
| 2 | P | 0/4 | - | - |

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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

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

## 5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry (i)

There are no ligands in this entry.

## 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 (i)

## 6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

## 6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

## 6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

## 6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.