# W O R L D W I D E

# wwPDB

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

# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 09:11 am GMT

|  |  |  |
|---|---|---|
| PDB ID | : | 1CQT |
| Title | : | CRYSTAL STRUCTURE OF A TERNARY COMPLEX CONTAINING AN OCA-B PEPTIDE, THE OCT-1 POU DOMAIN, AND AN OCTAMER ELEMENT |
| Authors | : | Chasman, D.I.; Cepek, K.; Sharp, P.A.; Pabo, C.O. |
| Deposited on | : | 1999-08-11 |
| Resolution | : | 3.20 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
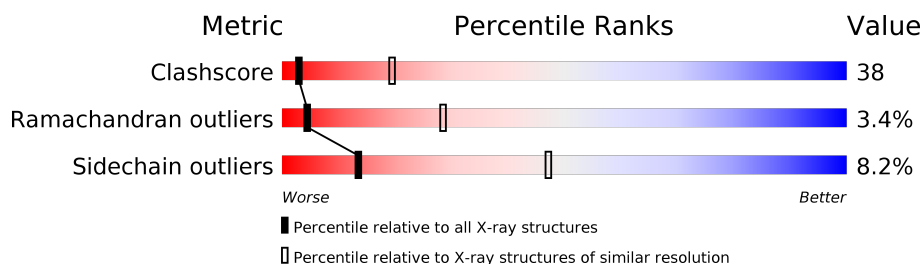with specific help available everywhere you see the ⓘ symbol.

# 1   Overall quality at a glance ⓘ

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

The reported resolution of this entry is 3.20 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 |
|---|---|---|
| Clashscore | | 38 |
| Ramachandran outliers | | 3.4% |
| Sidechain outliers | | 8.2% |

*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)) |
|---|---|---|
| Clashscore | 112137 | 1009 (3.20-3.20) |
| Ramachandran outliers | 110173 | 1118 (3.22-3.18) |
| Sidechain outliers | 110143 | 1117 (3.22-3.18) |

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. 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.

Note EDS was not executed.

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 1 | M | 15 | 27%    73% |
| 1 | O | 15 | 93%    7% |
| 2 | N | 15 | 100% |
| 2 | P | 15 | 7%    87%    7% |
| 3 | A | 163 | 54%    25%    • •    18% |
| 3 | B | 163 | 41%    32%    6%    21% |
| 4 | I | 44 | 36%    11%    •    50% |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Length | Quality of chain |
|-----|-------|--------|------------------|
| 4 | J | 44 | 34%    18%    48% |

# 2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 3649 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 DNA (5'-D(*TP*GP*TP*AP*TP*GP*CP*AP*AP*AP* TP*AP*AP*GP*G)-3').

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | M | 15 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 310 | 149 | 61 | 86 | 14 | | | |
| 1 | O | 14 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 293 | 139 | 59 | 81 | 14 | | | |

- Molecule 2 is a DNA chain called DNA (5'-D(*AP*CP*CP*TP*TP*AP*TP*TP*TP*GP* CP*AP*TP*AP*C)-3').

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | N | 15 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 299 | 146 | 49 | 90 | 14 | | | |
| 2 | P | 14 | Total | C | N | O | P | 0 | 0 | 0 |
| | | | 281 | 136 | 44 | 87 | 14 | | | |

- Molecule 3 is a protein called POU DOMAIN, CLASS 2, TRANSCRIPTION FACTOR 1.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | A | 134 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 1061 | 672 | 187 | 196 | 6 | | | |
| 3 | B | 129 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 1042 | 653 | 187 | 196 | 6 | | | |

There are 4 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | -2 | GLY | PRO | CONFLICT | UNP P14859 |
| A | 0 | ARG | LEU | CONFLICT | UNP P14859 |
| B | 498 | GLY | PRO | CONFLICT | UNP P14859 |
| B | 500 | ARG | LEU | CONFLICT | UNP P14859 |

- Molecule 4 is a protein called POU DOMAIN, CLASS 2, ASSOCIATING FACTOR 1.

| Mol | Chain | Residues | Atoms | | | | ZeroOcc | AltConf | Trace |
|:---:|:---:|:---:|:---|:---:|:---:|:---:|:---:|:---:|:---:|
| 4 | I | 22 | Total<br>179 | C<br>113 | N<br>38 | O<br>28 | 0 | 0 | 0 |
| 4 | J | 23 | Total<br>184 | C<br>116 | N<br>39 | O<br>29 | 0 | 0 | 0 |

# 3 Residue-property plots ⓘ

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 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.
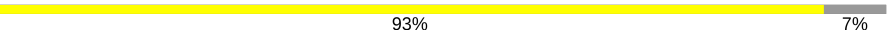
Note EDS was not executed.

• Molecule 1: DNA (5'-D(*TP*GP*TP*AP*TP*GP*CP*AP*AP*AP*TP*AP*AP*GP*G)-3')

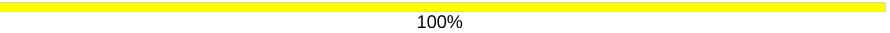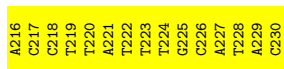Chain M:



• Molecule 1: DNA (5'-D(*TP*GP*TP*AP*TP*GP*CP*AP*AP*AP*TP*AP*AP*GP*G)-3')

Chain O:
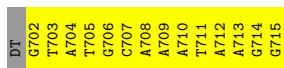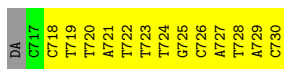


• Molecule 2: DNA (5'-D(*AP*CP*CP*TP*TP*AP*TP*TP*TP*GP*CP*AP*TP*AP*C)-3')

Chain N:



• Molecule 2: DNA (5'-D(*AP*CP*CP*TP*TP*AP*TP*TP*TP*GP*CP*AP*TP*AP*C)-3')

Chain P:



• Molecule 3: POU DOMAIN, CLASS 2, TRANSCRIPTION FACTOR 1

Chain A:

GLY SER ARG GLU **E2 P3** S4 D5 L6 L9 E10 A13 F16 F25 T26 Q27 V30 A33 M34 G35 K36 L37 N40 S43 I47 L55 N59 K64 P65 L66 L67 E68 K69 W70 L71 N72 D73 A74 E75 ASN LEU SER SER ASP SER LEU SER

SER PRO SER ALA LEU ASN SER PRO GLY ILE GLU GLY LEU SER ARG ARG R102 E117 F120 L121 Q124 E129 E130 I131 I134 A135 D136 Q137 M140 E143 R146 V147 K155 R158 I159 N160 P161
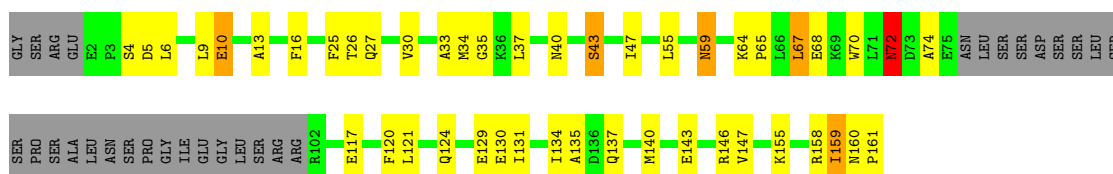
● Molecule 3: POU DOMAIN, CLASS 2, TRANSCRIPTION FACTOR 1

Chain B:        41%                    32%              6%         21%

GLY SER ARG GLU GLU PRO SER D505 L506 E507 E508 L509 E510 Q511 F512 A513 K514 T515 F516 K517 R520 I521 K522 L523 G524 F525 T526 Q527 L537 Y538 GLY N540 T545 L553 N554 L555 S556 N559 M560 L563 L566 L567 E568 K569 D573 A574 E575 ASN LEU SER SER ASP

SER SER LEU SER PRO SER ALA LEU ASN SER PRO GLY ILE GLU GLY LEU SER ARG ARG R602 K603 K604 R605 T610 R613 V614 A615 L616 E617 K618 S619 E622 K625 PRO T627 S628 I631 T632 M633 I634 A635 D636 Q637 L638 N639 M640 E641 K642 E643 V644 I645 R646 V647 W648

F649 C650 N651 R652 R653 Q654 K655 R658 I659 N660 P661

● Molecule 4: POU DOMAIN, CLASS 2, ASSOCIATING FACTOR 1

Chain I:        36%              11%    •              50%

MET LEU TRP GLN LYS PRO THR ALA PRO GLU GLN ALA PRO ALA PRO A300 Y303 Q304 R307 V312 L315 L316 G321 HIS ALA SER SER GLY ALA ALA

● Molecule 4: POU DOMAIN, CLASS 2, ASSOCIATING FACTOR 1

Chain J:        34%              18%              48%

MET LEU TRP GLN LYS PRO THR ALA PRO GLU GLN ALA PRO ALA PRO A800 Y803 Q804 G805 V806 R807 V808 P811 V812 L815 H822 ALA SER SER GLY ALA ALA
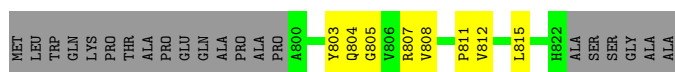
# 4 Data and refinement statistics ⓘ

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

| Property | Value | Source |
|---|---|---|
| Space group | P 41 21 2 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 93.72Å   93.72Å   152.65Å<br>90.00°   90.00°   90.00° | Depositor |
| Resolution (Å) | 30.00 – 3.20 | Depositor |
| % Data completeness<br>(in resolution range) | 99.5 (30.00-3.20) | Depositor |
| $R_{merge}$ | 0.08 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| Refinement program | X-PLOR 3.851 | Depositor |
| R, $R_{free}$ | 0.264 , 0.326 | Depositor |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| Total number of atoms | 3649 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 65.0 | wwPDB-VP |

# 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).

| Mol | Chain | Bond lengths | | Bond angles | |
|-----|-------|------|------|------|------|
| | | RMSZ | #$|Z|$ >5 | RMSZ | #$|Z|$ >5 |
| 1 | M | 0.32 | 0/349 | 0.77 | 0/538 |
| 1 | O | 0.26 | 0/330 | 0.75 | 0/508 |
| 2 | N | 0.35 | 0/333 | 0.84 | 0/511 |
| 2 | P | 0.26 | 0/312 | 0.68 | 0/478 |
| 3 | A | 0.34 | 0/1076 | 0.56 | 0/1445 |
| 3 | B | 0.35 | 0/1052 | 0.62 | 0/1405 |
| 4 | I | 0.52 | 0/181 | 0.75 | 0/240 |
| 4 | J | 0.35 | 0/186 | 0.64 | 0/247 |
| All | All | 0.34 | 0/3819 | 0.67 | 0/5372 |

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 | M | 310 | 0 | 171 | 37 | 0 |
| 1 | O | 293 | 0 | 158 | 48 | 0 |
| 2 | N | 299 | 0 | 173 | 36 | 0 |
| 2 | P | 281 | 0 | 161 | 57 | 0 |
| 3 | A | 1061 | 0 | 1046 | 35 | 0 |
| 3 | B | 1042 | 0 | 1019 | 53 | 0 |
| 4 | I | 179 | 0 | 195 | 12 | 0 |
| 4 | J | 184 | 0 | 197 | 8 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|-----|-------|-------|----------|----------|---------|--------------|
| All | All | 3649 | 0 | 3120 | 256 | 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 38.

The worst 5 of 256 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:M:209:DA:H2" | 1:M:210:DA:H5' | 1.21 | 1.12 |
| 1:M:204:DA:H2" | 1:M:205:DT:H5' | 1.22 | 1.09 |
| 1:M:204:DA:H2" | 1:M:205:DT:C5' | 1.82 | 1.08 |
| 2:N:223:DT:H2" | 2:N:224:DT:H5" | 1.08 | 1.06 |
| 2:N:219:DT:H2' | 2:N:220:DT:H72 | 1.36 | 1.06 |

There are no symmetry-related clashes.

## 5.3  Torsion angles  ⓘ

### 5.3.1  Protein backbone  ⓘ

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

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

| Mol | Chain | Analysed | Favoured | Allowed | Outliers | Percentiles | |
|-----|-------|----------|----------|---------|----------|-----|-----|
| 3 | A | 130/163 (80%) | 114 (88%) | 12 (9%) | 4 (3%) | 5 | 31 |
| 3 | B | 121/163 (74%) | 108 (89%) | 8 (7%) | 5 (4%) | 3 | 24 |
| 4 | I | 20/44 (46%) | 17 (85%) | 2 (10%) | 1 (5%) | 2 | 19 |
| 4 | J | 21/44 (48%) | 15 (71%) | 6 (29%) | 0 | 100 | 100 |
| All | All | 292/414 (70%) | 254 (87%) | 28 (10%) | 10 (3%) | 4 | 28 |

5 of 10 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | B | 574 | ALA |
| 3 | B | 631 | ILE |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | B | 526 | THR |
| 4 | I | 303 | TYR |
| 3 | A | 72 | ASN |

### 5.3.2 Protein sidechains ⓘ

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

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

| Mol | Chain | Analysed | Rotameric | Outliers | Percentiles | |
|-----|-------|----------|-----------|----------|-------------|-----|
| 3 | A | 111/148 (75%) | 104 (94%) | 7 (6%) | 21 | 59 |
| 3 | B | 110/148 (74%) | 97 (88%) | 13 (12%) | 6 | 27 |
| 4 | I | 18/34 (53%) | 17 (94%) | 1 (6%) | 25 | 64 |
| 4 | J | 18/34 (53%) | 18 (100%) | 0 | 100 | 100 |
| All | All | 257/364 (71%) | 236 (92%) | 21 (8%) | 13 | 47 |

5 of 21 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | B | 554 | ASN |
| 3 | B | 566 | LEU |
| 3 | B | 643 | GLU |
| 3 | B | 538 | TYR |
| 3 | B | 646 | ARG |

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | B | 511 | GLN |
| 4 | J | 804 | GLN |
| 3 | B | 518 | GLN |
| 3 | A | 111 | ASN |
| 3 | B | 554 | ASN |

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

# 6   Fit of model and data ⓘ

## 6.1   Protein, DNA and RNA chains ⓘ

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

## 6.3   Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

## 6.4   Ligands ⓘ

EDS was not executed - this section is therefore empty.

## 6.5   Other polymers ⓘ

EDS was not executed - this section is therefore empty.