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

Jun 17, 2024 – 06:44 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 5Y0E |
| Title | : | Crystal structure of TssK |
| Authors | : | Tong, H.; Xu, D.; Ran, T.; Wang, W.W. |
| Deposited on | : | 2017-07-17 |
| Resolution | : | 2.50 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:

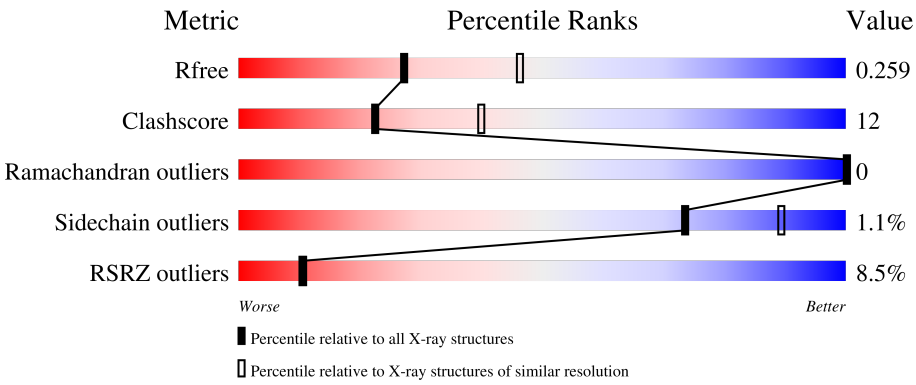
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.37.1 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (Gargrove) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.37.1 |

# 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 2.50 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 | 4661 (2.50-2.50) |
| Clashscore | 141614 | 5346 (2.50-2.50) |
| Ramachandran outliers | 138981 | 5231 (2.50-2.50) |
| Sidechain outliers | 138945 | 5233 (2.50-2.50) |
| RSRZ outliers | 127900 | 4559 (2.50-2.50) |

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 | 456 | 6%, 81%, 16%, • |
| 1 | B | 456 | 3%, 78%, 19%, • |
| 1 | C | 456 | 16%, 65%, 28%, 7% |

# 2 Entry composition (i)

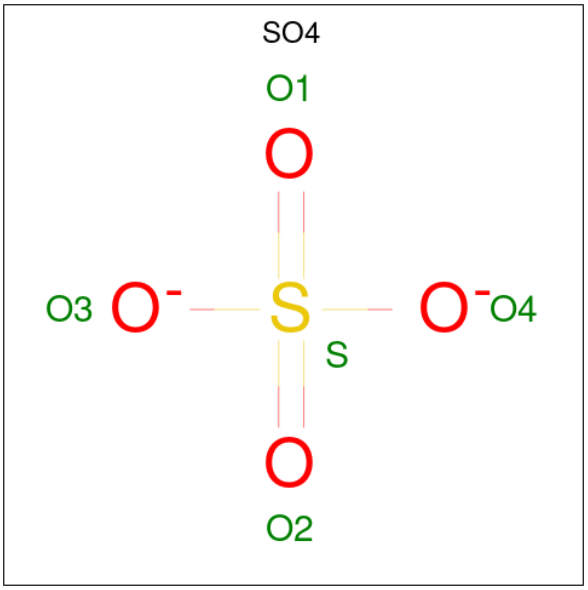
There are 3 unique types of molecules in this entry. The entry contains 10406 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 TssK.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 1 | A | 442 | 3476 | 2208 | 617 | 633 | 18 | 0 | 0 | 0 |
| 1 | B | 440 | 3477 | 2212 | 616 | 631 | 18 | 0 | 0 | 0 |
| 1 | C | 425 | 3288 | 2095 | 578 | 599 | 16 | 0 | 0 | 0 |

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: $O_4S$).




| Mol | Chain | Residues | Atoms | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|
| | | | Total | O | S | | |
| 2 | A | 1 | 5 | 4 | 1 | 0 | 0 |
| 2 | B | 1 | 5 | 4 | 1 | 0 | 0 |
| 2 | B | 1 | 5 | 4 | 1 | 0 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Residues | Atoms | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|
| | | | Total | O | S | | |
| 2 | C | 1 | 5 | 4 | 1 | 0 | 0 |
| | | | Total | O | S | | |
| 2 | C | 1 | 5 | 4 | 1 | 0 | 0 |

- Molecule 3 is water.

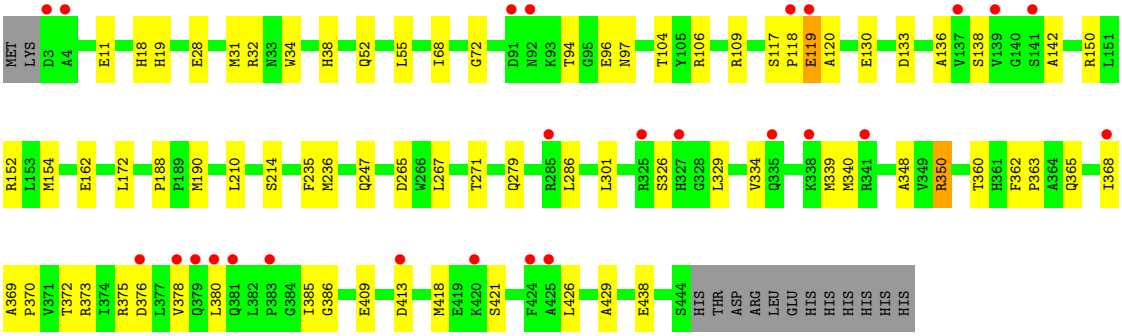
| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | O | | |
| 3 | A | 55 | 55 | 55 | 0 | 0 |
| | | | Total | O | | |
| 3 | B | 49 | 49 | 49 | 0 | 0 |
| | | | Total | O | | |
| 3 | C | 36 | 36 | 36 | 0 | 0 |

# 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.

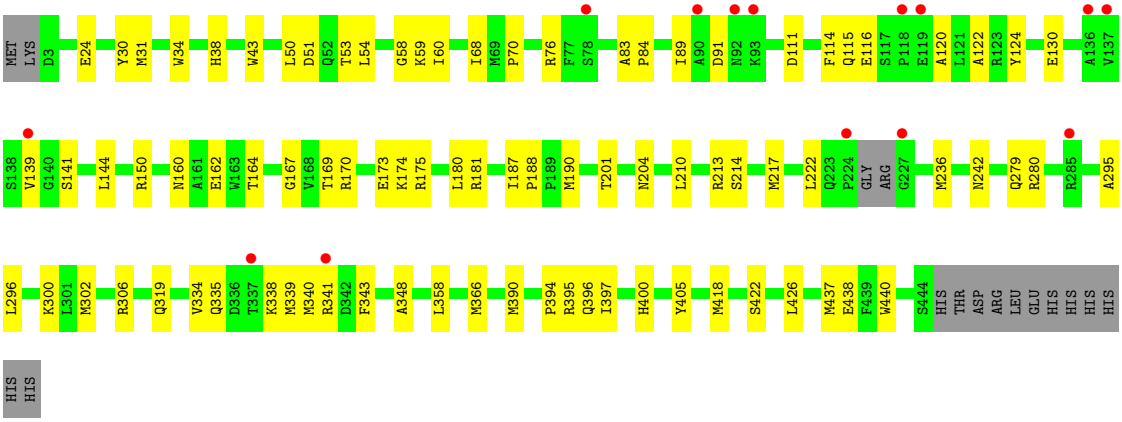
• Molecule 1: TssK

Chain A:



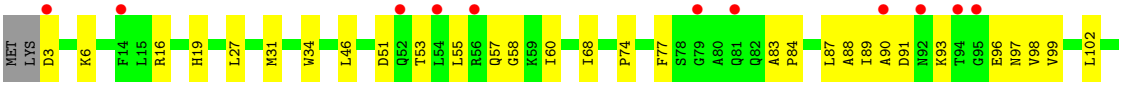

• Molecule 1: TssK

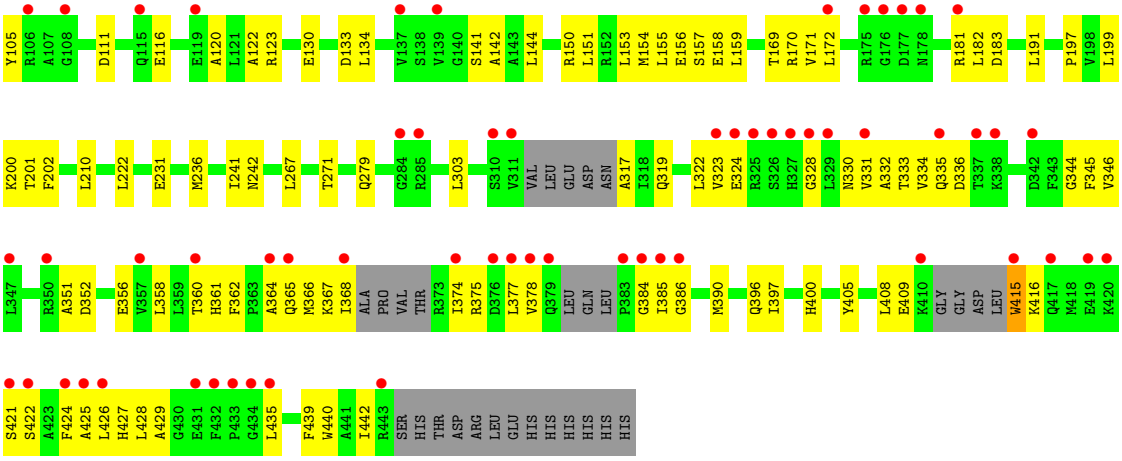
Chain B:




• Molecule 1: TssK

Chain C:




Y105 R106 A107 G108 D111 Q115 E116 E119 A120 L121 A122 R123 E130 D133 L134 V137 S138 V139 G140 S141 A142 A143 L144 R150 L151 R152 L153 M154 L155 E156 S157 E158 L159 T169 R170 V171 L172 R175 G176 D177 N178 R181 L182 D183 L191 P197 V198 L199
K200 T201 F202 L210 L222 E231 M236 I241 N242 L267 T271 Q279 G284 R285 L303 S310 V311 VAL LEU GLU ASP ASN A317 I318 Q319 L322 V323 E324 R325 S326 H327 G328 L329 N330 V331 A332 T333 V334 Q335 D336 T337 K338 D342 F343 G344 F345 V346
L347 R350 A351 D352 E356 V357 L358 L359 T360 H361 F362 P363 A364 Q365 M366 K367 I368 ALA PRO VAL THR R373 I374 R375 D376 L377 V378 Q379 LEU GLN LEU P383 G384 I385 G386 M390 Q396 I397 H400 Y405 L408 E409 K410 GLY GLY ASP LEU W415 K416 Q417 M418 E419 K420
S421 S422 A423 F424 A425 L426 H427 L428 A429 G430 E431 F432 P433 G434 L435 F439 W440 A441 I442 R443 SER HIS THR ASP ARG LEU GLU HIS HIS HIS HIS HIS HIS

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 65 2 2 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 151.83Å 151.83Å 313.19Å<br>90.00° 90.00° 120.00° | Depositor |
| Resolution (Å) | 19.97 – 2.50<br>19.97 – 2.50 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 100.0 (19.97-2.50)<br>100.0 (19.97-2.50) | Depositor<br>EDS |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 2.43 (at 2.50Å) | Xtriage |
| Refinement program | PHENIX (DEV_2247: ???) | Depositor |
| R, $R_{free}$ | 0.219 , 0.262<br>0.225 , 0.259 | Depositor<br>DCC |
| $R_{free}$ test set | 3594 reflections (4.84%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 57.7 | Xtriage |
| Anisotropy | 0.204 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.34 , 43.8 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.48$, $< L^2 > = 0.31$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| $F_o$,$F_c$ correlation | 0.94 | EDS |
| Total number of atoms | 10406 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 63.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.97% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

# 5 Model quality (i)

## 5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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 | 0.74 | 0/3555 | 0.65 | 0/4823 |
| 1 | B | 0.71 | 0/3555 | 0.64 | 0/4820 |
| 1 | C | 0.59 | 1/3361 (0.0%) | 0.66 | 0/4562 |
| All | All | 0.69 | 1/10471 (0.0%) | 0.65 | 0/14205 |

All (1) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 1 | C | 222 | LEU | C-N | -5.15 | 1.22 | 1.34 |

There are no bond angle outliers.

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 | 3476 | 0 | 3424 | 76 | 0 |
| 1 | B | 3477 | 0 | 3447 | 70 | 0 |
| 1 | C | 3288 | 0 | 3180 | 119 | 0 |
| 2 | A | 5 | 0 | 0 | 0 | 0 |
| 2 | B | 10 | 0 | 0 | 0 | 0 |
| 2 | C | 10 | 0 | 0 | 0 | 0 |
| 3 | A | 55 | 0 | 0 | 0 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|---|---|---|---|---|---|---|
| 3 | B | 49 | 0 | 0 | 0 | 0 |
| 3 | C | 36 | 0 | 0 | 2 | 0 |
| All | All | 10406 | 0 | 10051 | 236 | 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 12.

The worst 5 of 236 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:118:PRO:O | 1:A:119:GLU:HG3 | 1.42 | 1.16 |
| 1:A:118:PRO:C | 1:A:119:GLU:HG3 | 1.57 | 1.09 |
| 1:A:18:HIS:HD2 | 1:C:142:ALA:HB3 | 1.14 | 1.06 |
| 1:A:18:HIS:HD2 | 1:C:142:ALA:CB | 1.69 | 1.05 |
| 1:C:328:GLY:O | 1:C:427:HIS:HA | 1.60 | 1.00 |

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 | 440/456 (96%) | 425 (97%) | 15 (3%) | 0 | 100 | 100 |
| 1 | B | 436/456 (96%) | 418 (96%) | 18 (4%) | 0 | 100 | 100 |
| 1 | C | 415/456 (91%) | 394 (95%) | 21 (5%) | 0 | 100 | 100 |
| All | All | 1291/1368 (94%) | 1237 (96%) | 54 (4%) | 0 | 100 | 100 |

There are no Ramachandran outliers to report.

### 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 | 363/381 (95%) | 359 (99%) | 4 (1%) | 73 | 89 |
| 1 | B | 366/381 (96%) | 364 (100%) | 2 (0%) | 88 | 96 |
| 1 | C | 334/381 (88%) | 328 (98%) | 6 (2%) | 59 | 81 |
| All | All | 1063/1143 (93%) | 1051 (99%) | 12 (1%) | 73 | 89 |

5 of 12 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | C | 91 | ASP |
| 1 | C | 181 | ARG |
| 1 | C | 415 | TRP |
| 1 | C | 352 | ASP |
| 1 | A | 413 | ASP |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 18 | HIS |
| 1 | C | 82 | GLN |
| 1 | C | 239 | GLN |

### 5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry (i)

5 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

| Mol | Type | Chain | Res | Link | Bond lengths | | | Bond angles | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Counts | RMSZ | $\#\|Z\| > 2$ | Counts | RMSZ | $\#\|Z\| > 2$ |
| 2 | SO4 | C | 502 | - | 4,4,4 | 0.31 | 0 | 6,6,6 | 0.81 | 0 |
| 2 | SO4 | A | 501 | - | 4,4,4 | 0.33 | 0 | 6,6,6 | 0.94 | 0 |
| 2 | SO4 | B | 502 | - | 4,4,4 | 0.31 | 0 | 6,6,6 | 0.91 | 0 |
| 2 | SO4 | B | 501 | - | 4,4,4 | 0.14 | 0 | 6,6,6 | 0.59 | 0 |
| 2 | SO4 | C | 501 | - | 4,4,4 | 0.54 | 0 | 6,6,6 | 0.64 | 0 |

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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)

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 | 442/456 (96%) | 0.10 | 26 (5%) | 22 | 23 | 31, 55, 105, 134 | 0 |
| 1 | B | 440/456 (96%) | 0.06 | 14 (3%) | 47 | 51 | 32, 50, 83, 118 | 0 |
| 1 | C | 425/456 (93%) | 0.80 | 71 (16%) | 1 | 1 | 35, 71, 138, 157 | 0 |
| All | All | 1307/1368 (95%) | 0.31 | 111 (8%) | 10 | 10 | 31, 55, 126, 157 | 0 |

The worst 5 of 111 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 1 | C | 377 | LEU | 8.7 |
| 1 | C | 420 | LYS | 8.3 |
| 1 | C | 421 | SER | 8.2 |
| 1 | C | 384 | GLY | 7.4 |
| 1 | C | 108 | GLY | 5.4 |

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

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

## 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 |
|---|---|---|---|---|---|---|---|---|
| 2 | SO4 | B | 501 | 5/5 | 0.83 | 0.22 | 81,82,89,116 | 0 |
| 2 | SO4 | A | 501 | 5/5 | 0.88 | 0.35 | 30,30,30,30 | 0 |
| 2 | SO4 | C | 502 | 5/5 | 0.88 | 0.39 | 30,30,30,30 | 0 |
| 2 | SO4 | B | 502 | 5/5 | 0.91 | 0.47 | 30,30,30,30 | 0 |
| 2 | SO4 | C | 501 | 5/5 | 0.99 | 0.09 | 44,45,49,54 | 0 |

## 6.5 Other polymers (i)

There are no such residues in this entry.