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

Aug 7, 2023 – 06:01 pm BST

| | | |
|---|---|---|
| PDB ID | : | 7QYG |
| Title | : | Structure of the transaminase TR2 |
| Authors | : | Roda, S.; Fernandez-Lopez, L.; Benedens, M.; Bollinger, A.; Thies, S.; Schumacher, J.; Coscolin, C.; Kazemi, M.; Santiago, G.; Gertzen, C.G.; Gonzalez-Alfonso, J.; Plou, F.J.; Jaeger, K.E.; Smits, S.H.; Ferrer, M.; Guallar, V. |
| Deposited on | : | 2022-01-28 |
| Resolution | : | 3.60 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 (1)) were used in the production of this report:

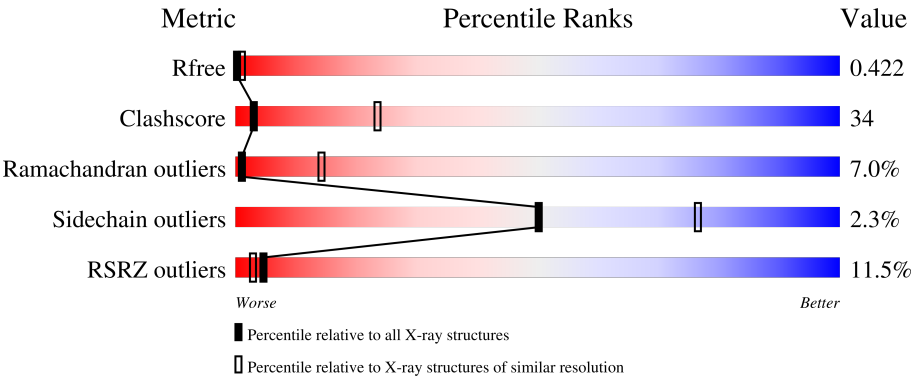
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.34 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (Gargrove) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.34 |

# 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 3.60 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 | 1257 (3.70-3.50) |
| Clashscore | 141614 | 1353 (3.70-3.50) |
| Ramachandran outliers | 138981 | 1307 (3.70-3.50) |
| Sidechain outliers | 138945 | 1307 (3.70-3.50) |
| RSRZ outliers | 127900 | 1161 (3.70-3.50) |

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 | 465 | 8% 33% 46% · 16% |
| 1 | B | 465 | 6% 38% 40% 6% 16% |
| 1 | C | 465 | 12% 34% 44% 6% 16% |
| 1 | D | 465 | 12% 38% 41% 5% 16% |

## 2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 24079 atoms, of which 12003 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 Aminotransferase TR2.

| Mol | Chain | Residues | Atoms | | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 389 | Total | C | H | N | O | S | 0 | 0 | 0 |
| | | | 6017 | 1917 | 3000 | 525 | 558 | 17 | | | |
| 1 | B | 389 | Total | C | H | N | O | S | 0 | 0 | 0 |
| | | | 6005 | 1914 | 2992 | 524 | 558 | 17 | | | |
| 1 | C | 391 | Total | C | H | N | O | S | 0 | 0 | 0 |
| | | | 6052 | 1926 | 3019 | 530 | 560 | 17 | | | |
| 1 | D | 389 | Total | C | H | N | O | S | 0 | 0 | 0 |
| | | | 6005 | 1914 | 2992 | 524 | 558 | 17 | | | |

There are 32 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 458 | LEU | - | expression tag | UNP A0A3G5BC54 |
| A | 459 | GLU | - | expression tag | UNP A0A3G5BC54 |
| A | 460 | HIS | - | expression tag | UNP A0A3G5BC54 |
| A | 461 | HIS | - | expression tag | UNP A0A3G5BC54 |
| A | 462 | HIS | - | expression tag | UNP A0A3G5BC54 |
| A | 463 | HIS | - | expression tag | UNP A0A3G5BC54 |
| A | 464 | HIS | - | expression tag | UNP A0A3G5BC54 |
| A | 465 | HIS | - | expression tag | UNP A0A3G5BC54 |
| B | 458 | LEU | - | expression tag | UNP A0A3G5BC54 |
| B | 459 | GLU | - | expression tag | UNP A0A3G5BC54 |
| B | 460 | HIS | - | expression tag | UNP A0A3G5BC54 |
| B | 461 | HIS | - | expression tag | UNP A0A3G5BC54 |
| B | 462 | HIS | - | expression tag | UNP A0A3G5BC54 |
| B | 463 | HIS | - | expression tag | UNP A0A3G5BC54 |
| B | 464 | HIS | - | expression tag | UNP A0A3G5BC54 |
| B | 465 | HIS | - | expression tag | UNP A0A3G5BC54 |
| C | 458 | LEU | - | expression tag | UNP A0A3G5BC54 |
| C | 459 | GLU | - | expression tag | UNP A0A3G5BC54 |
| C | 460 | HIS | - | expression tag | UNP A0A3G5BC54 |
| C | 461 | HIS | - | expression tag | UNP A0A3G5BC54 |
| C | 462 | HIS | - | expression tag | UNP A0A3G5BC54 |

*Continued on next page...*

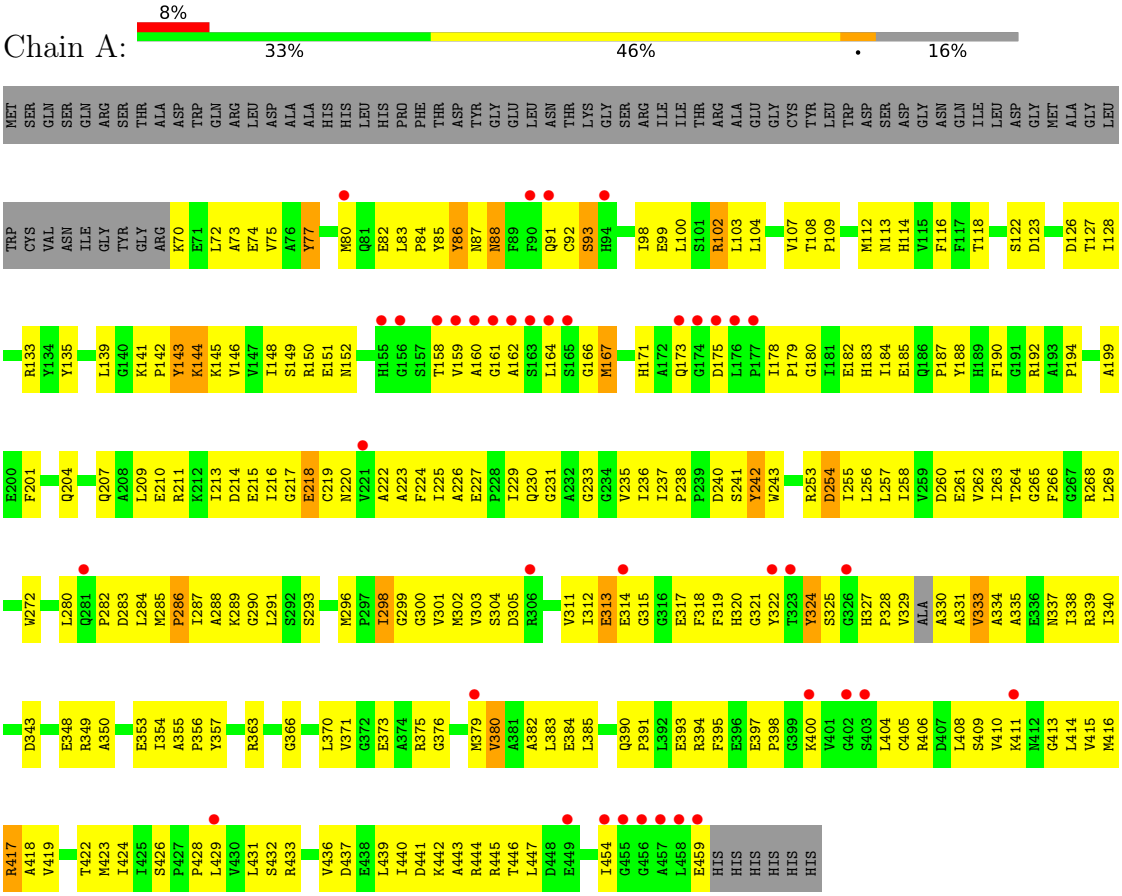
*Continued from previous page...*

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| C | 463 | HIS | - | expression tag | UNP A0A3G5BC54 |
| C | 464 | HIS | - | expression tag | UNP A0A3G5BC54 |
| C | 465 | HIS | - | expression tag | UNP A0A3G5BC54 |
| D | 458 | LEU | - | expression tag | UNP A0A3G5BC54 |
| D | 459 | GLU | - | expression tag | UNP A0A3G5BC54 |
| D | 460 | HIS | - | expression tag | UNP A0A3G5BC54 |
| D | 461 | HIS | - | expression tag | UNP A0A3G5BC54 |
| D | 462 | HIS | - | expression tag | UNP A0A3G5BC54 |
| D | 463 | HIS | - | expression tag | UNP A0A3G5BC54 |
| D | 464 | HIS | - | expression tag | UNP A0A3G5BC54 |
| D | 465 | HIS | - | expression tag | UNP A0A3G5BC54 |

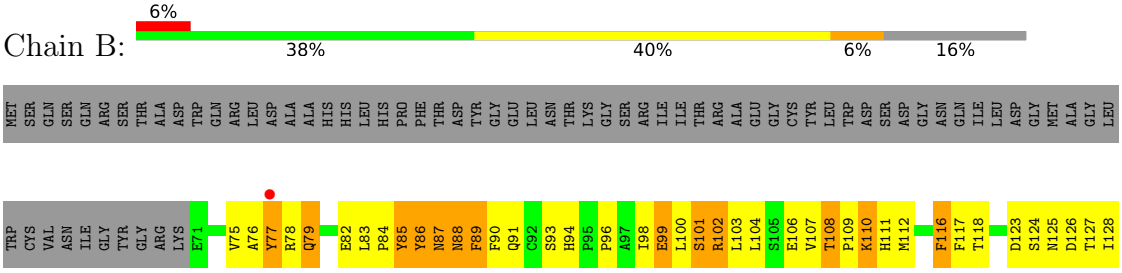
# 3 Residue-property plots (i)

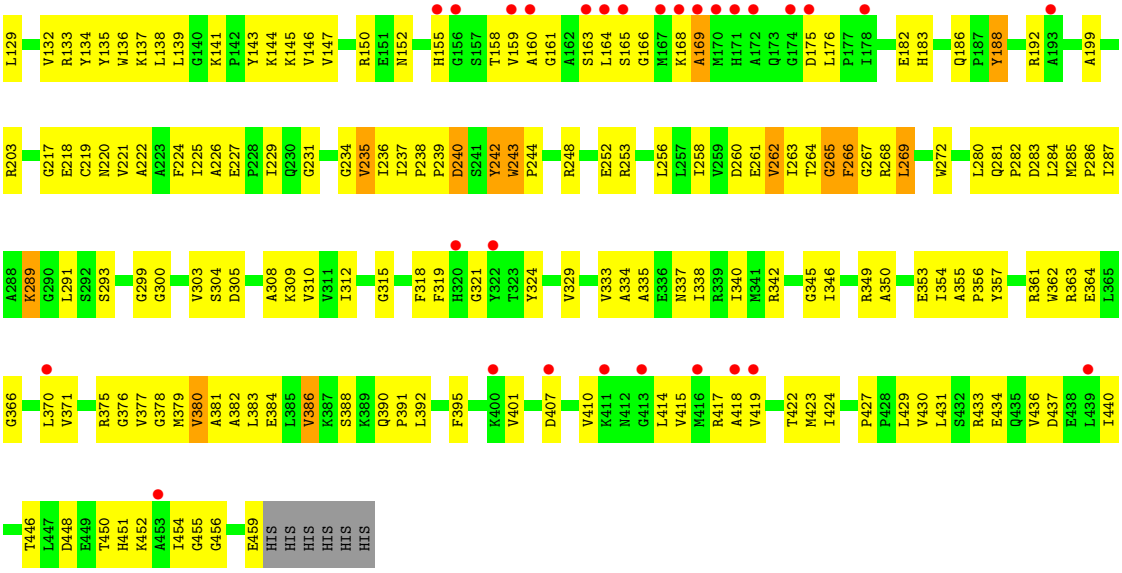
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Aminotransferase TR2



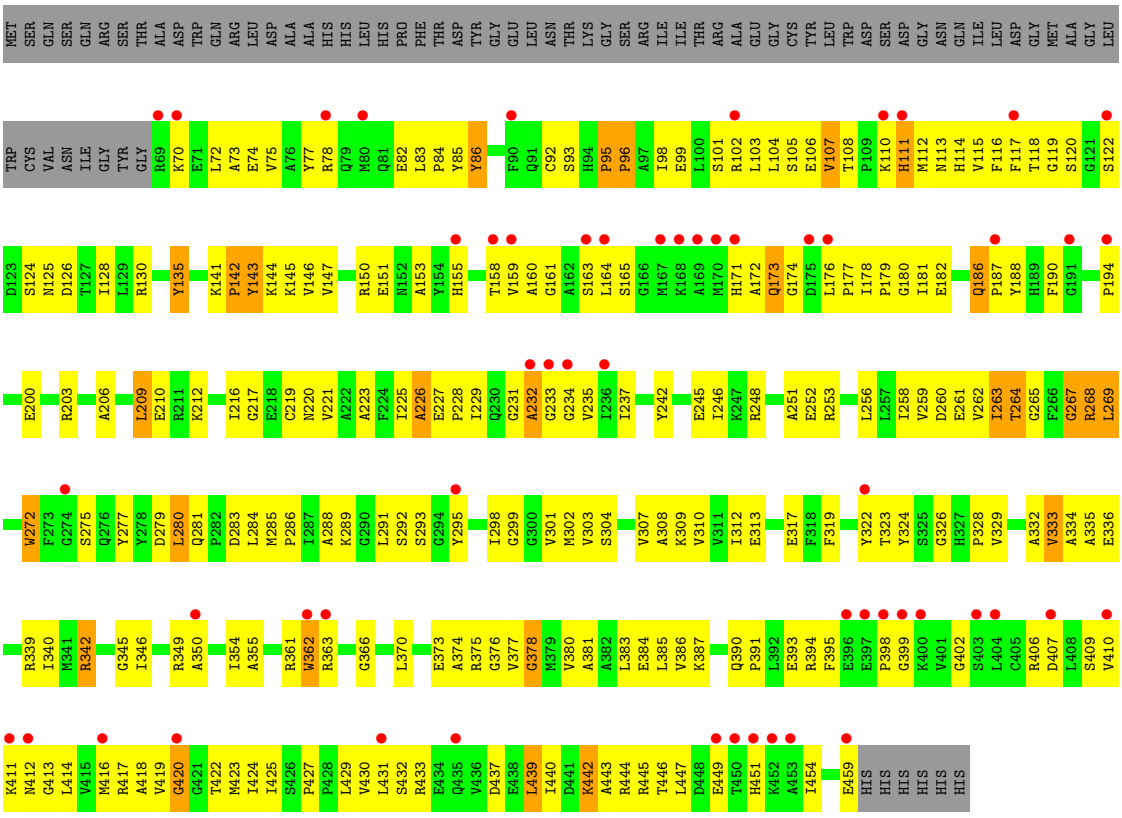

• Molecule 1: Aminotransferase TR2





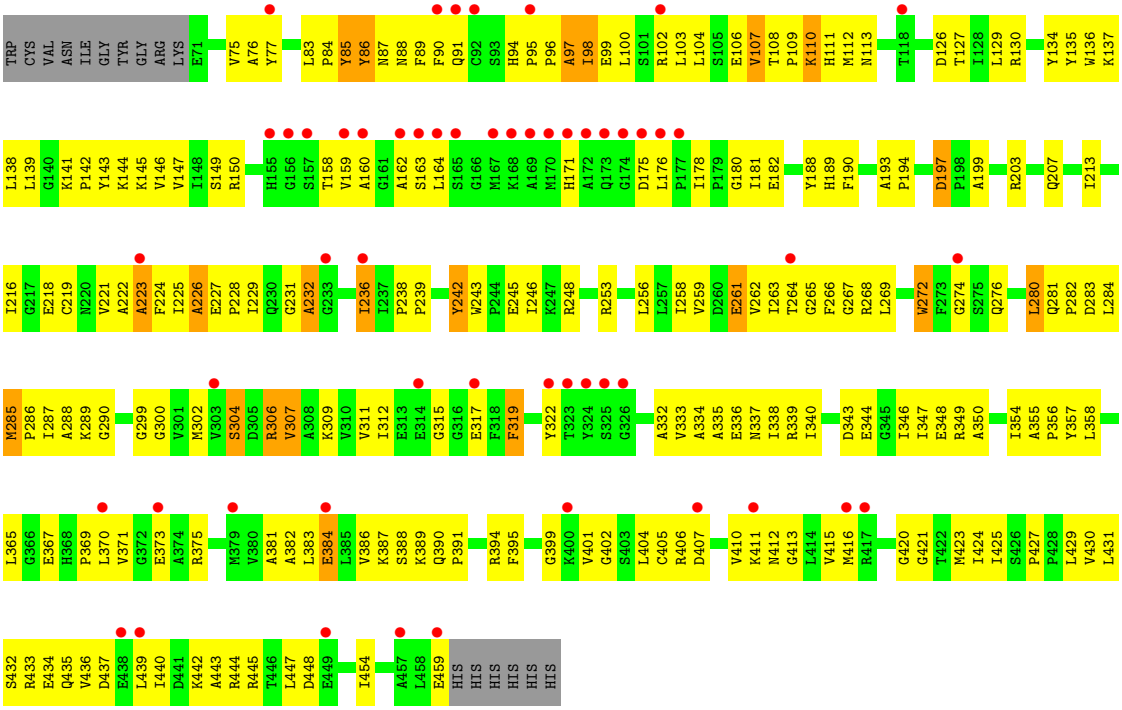
• Molecule 1: Aminotransferase TR2

Chain C:

• Molecule 1: Aminotransferase TR2

Chain D:



TRP CYS VAL ASN ILE GLY TYR GLY ARG LYS E71 V75 A76 Y77 L83 P84 Y85 Y86 N87 N88 F89 F90 Q91 C92 S93 H94 P95 P96 A97 I98 E99 L100 S101 R102 L103 L104 S105 E106 V107 T108 P109 K110 H111 M112 N113 T118 D126 T127 I128 L129 R130 Y134 Y135 W136 K137
L138 L139 G140 K141 P142 Y143 K144 K145 V146 V147 I148 S149 R150 H155 G156 S157 T158 V159 A160 G161 A162 S163 L164 S166 G166 M167 K168 A169 M170 H171 A172 Q173 G174 D175 L176 P177 I178 P179 G180 I181 E182 Y188 H189 F190 A193 P194 D197 P198 A199 R203 Q207 I213
I216 G217 E218 C219 N220 V221 A222 A223 F224 I225 A226 E227 P228 I229 Q230 G231 A232 G233 I236 I237 P238 P239 Y242 W243 P244 E245 I246 K247 R248 R253 L256 L257 I258 V259 D260 E261 V262 I263 T264 G265 F266 G267 R268 L269 W272 F273 G274 S275 Q276 L280 Q281 P282 D283 L284
M285 P286 I287 A288 K289 G290 G299 G300 V301 M302 V303 S304 D305 R306 V307 A308 K309 V310 V311 I312 E313 E314 G315 G316 E317 F318 F319 Y322 T323 Y324 S325 G326 A332 V333 A334 A335 E336 N337 I338 R339 I340 D343 E344 G345 I346 I347 E348 R349 A350 I354 A355 P356 Y357 L358
L365 G366 E367 H368 P369 L370 V371 G372 E373 A374 R375 M379 V380 A381 A382 L383 E384 L385 V386 K387 S388 K389 Q390 P391 R394 F395 G399 K400 V401 G402 S403 L404 C405 R406 D407 V410 K411 N412 G413 L414 V415 M416 R417 G420 G421 T422 M423 I424 I425 S426 P427 P428 L429 V430 L431
S432 R433 E434 Q435 V436 D437 E438 L439 I440 D441 K442 A443 R444 R445 T446 L447 D448 E449 I454 A457 L458 E459 HIS HIS HIS HIS HIS HIS

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 21 21 21 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 60.78Å 166.63Å 209.82Å<br>90.00° 90.00° 90.00° | Depositor |
| Resolution (Å) | 49.09 – 3.60<br>49.09 – 3.60 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 94.2 (49.09-3.60)<br>94.2 (49.09-3.60) | Depositor<br>EDS |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 2.15 (at 3.57Å) | Xtriage |
| Refinement program | PHENIX 1.17 | Depositor |
| R, $R_{free}$ | 0.345 , 0.422<br>0.345 , 0.422 | Depositor<br>DCC |
| $R_{free}$ test set | 1994 reflections (8.28%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 113.7 | Xtriage |
| Anisotropy | 0.935 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.29 , 67.8 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.41$, $< L^2 > = 0.24$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| $F_o$,$F_c$ correlation | 0.84 | EDS |
| Total number of atoms | 24079 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 143.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.83% 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).

| Mol | Chain | Bond lengths | | Bond angles | |
|---|---|---|---|---|---|
| | | RMSZ | #$\|Z\|>5$ | RMSZ | #$\|Z\|>5$ |
| 1 | A | 0.36 | 0/3086 | 0.57 | 0/4176 |
| 1 | B | 0.37 | 0/3082 | 0.59 | 0/4172 |
| 1 | C | 0.41 | 1/3102 (0.0%) | 0.61 | 1/4197 (0.0%) |
| 1 | D | 0.45 | 5/3082 (0.2%) | 0.60 | 2/4172 (0.0%) |
| All | All | 0.40 | 6/12352 (0.0%) | 0.59 | 3/16717 (0.0%) |

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

| Mol | Chain | #Chirality outliers | #Planarity outliers |
|---|---|---|---|
| 1 | A | 0 | 1 |

The worst 5 of 6 bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 1 | D | 384 | GLU | CD-OE1 | -7.81 | 1.17 | 1.25 |
| 1 | D | 384 | GLU | CG-CD | -6.36 | 1.42 | 1.51 |
| 1 | D | 261 | GLU | CB-CG | -6.03 | 1.40 | 1.52 |
| 1 | D | 261 | GLU | CD-OE1 | -5.65 | 1.19 | 1.25 |
| 1 | D | 261 | GLU | CG-CD | 5.24 | 1.59 | 1.51 |

All (3) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 1 | D | 285 | MET | CG-SD-CE | 7.90 | 112.84 | 100.20 |
| 1 | D | 384 | GLU | OE1-CD-OE2 | -5.87 | 116.25 | 123.30 |
| 1 | C | 439 | LEU | CA-CB-CG | 5.21 | 127.28 | 115.30 |

There are no chirality outliers.

All (1) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group |
|---|---|---|---|---|
| 1 | A | 143 | TYR | Peptide |

## 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 | 3017 | 3000 | 3000 | 209 | 0 |
| 1 | B | 3013 | 2992 | 2992 | 198 | 0 |
| 1 | C | 3033 | 3019 | 3018 | 228 | 0 |
| 1 | D | 3013 | 2992 | 2992 | 210 | 0 |
| All | All | 12076 | 12003 | 12002 | 826 | 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 34.

The worst 5 of 826 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:B:128:ILE:HD13 | 1:B:258:ILE:CD1 | 1.66 | 1.25 |
| 1:B:128:ILE:HD13 | 1:B:258:ILE:HD11 | 1.15 | 1.09 |
| 1:D:347:ILE:HD11 | 1:D:348:GLU:OE2 | 1.51 | 1.09 |
| 1:D:347:ILE:CD1 | 1:D:348:GLU:OE2 | 2.03 | 1.06 |
| 1:B:128:ILE:CD1 | 1:B:258:ILE:CD1 | 2.35 | 1.05 |

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 | 387/465 (83%) | 287 (74%) | 75 (19%) | 25 (6%) | 1 16 |
| 1 | B | 387/465 (83%) | 285 (74%) | 71 (18%) | 31 (8%) | 1 11 |
| 1 | C | 389/465 (84%) | 272 (70%) | 87 (22%) | 30 (8%) | 1 12 |
| 1 | D | 387/465 (83%) | 287 (74%) | 78 (20%) | 22 (6%) | 1 18 |
| All | All | 1550/1860 (83%) | 1131 (73%) | 311 (20%) | 108 (7%) | 1 14 |

5 of 108 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 313 | GLU |
| 1 | B | 101 | SER |
| 1 | B | 236 | ILE |
| 1 | B | 243 | TRP |
| 1 | B | 266 | PHE |

### 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 | 315/377 (84%) | 308 (98%) | 7 (2%) | 52 77 |
| 1 | B | 314/377 (83%) | 305 (97%) | 9 (3%) | 42 72 |
| 1 | C | 316/377 (84%) | 307 (97%) | 9 (3%) | 43 72 |
| 1 | D | 314/377 (83%) | 310 (99%) | 4 (1%) | 69 86 |
| All | All | 1259/1508 (84%) | 1230 (98%) | 29 (2%) | 50 76 |

5 of 29 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | B | 240 | ASP |
| 1 | D | 304 | SER |
| 1 | C | 126 | ASP |
| 1 | C | 442 | LYS |
| 1 | C | 99 | GLU |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | C | 113 | ASN |
| 1 | C | 220 | ASN |
| 1 | D | 368 | HIS |
| 1 | D | 155 | HIS |
| 1 | D | 276 | 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)

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

## 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 | 389/465 (83%) | 0.42 | 39 (10%) 7 4 | 77, 125, 171, 205 | 0 |
| 1 | B | 389/465 (83%) | 0.22 | 30 (7%) 13 8 | 79, 120, 161, 198 | 0 |
| 1 | C | 391/465 (84%) | 0.61 | 56 (14%) 2 2 | 78, 138, 172, 206 | 0 |
| 1 | D | 389/465 (83%) | 0.66 | 54 (13%) 2 2 | 82, 134, 183, 208 | 0 |
| All | All | 1558/1860 (83%) | 0.48 | 179 (11%) 4 3 | 77, 128, 174, 208 | 0 |

The worst 5 of 179 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 1 | A | 163 | SER | 14.9 |
| 1 | C | 411 | LYS | 12.4 |
| 1 | D | 175 | ASP | 9.8 |
| 1 | D | 163 | SER | 9.2 |
| 1 | C | 233 | GLY | 9.0 |

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

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

## 6.3 Carbohydrates

There are no monosaccharides in this entry.

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers (i)

There are no such residues in this entry.