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

Aug 27, 2023 – 05:04 AM EDT

| | | |
|---|---|---|
| PDB ID | : | 3HFZ |
| Title | : | Crystal structure of Thermus thermophilus Phenylalanyl-tRNA synthetase complexed with m-tyrosine |
| Authors | : | Klipcan, L.; Moor, N.; Kessler, N.; Safro, M.G. |
| Deposited on | : | 2009-05-13 |
| Resolution | : | 2.90 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:

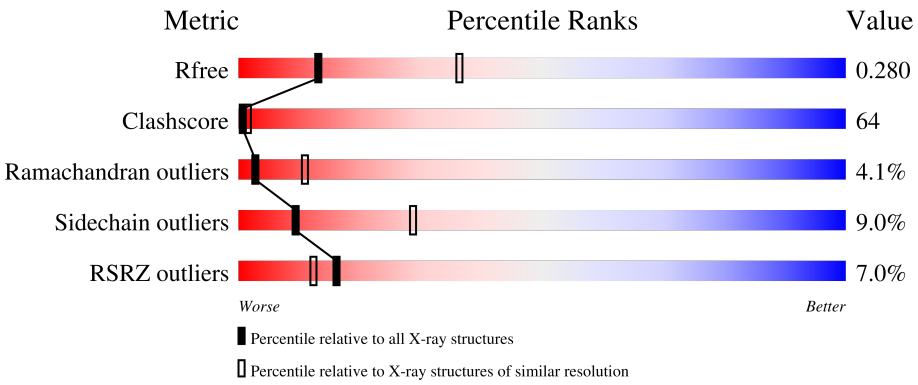
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.35 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (Gargrove) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.35 |

# 1 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 2.90 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 | 1957 (2.90-2.90) |
| Clashscore | 141614 | 2172 (2.90-2.90) |
| Ramachandran outliers | 138981 | 2115 (2.90-2.90) |
| Sidechain outliers | 138945 | 2117 (2.90-2.90) |
| RSRZ outliers | 127900 | 1906 (2.90-2.90) |

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 | 350 | 3%; 23%; 45%; 7%; 24% |
| 2 | B | 785 | 8%; 37%; 55%; 8% |

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

| Mol | Type | Chain | Res | Chirality | Geometry | Clashes | Electron density |
|---|---|---|---|---|---|---|---|
| 3 | MTY | A | 351 | - | - | X | - |
| 3 | MTY | B | 786 | - | - | X | - |

# 2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8517 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.

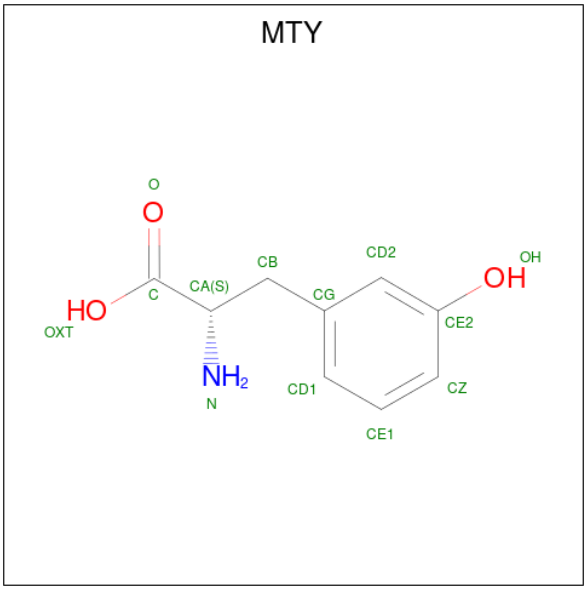
- Molecule 1 is a protein called Phenylalanyl-tRNA synthetase alpha chain.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 1 | A | 266 | 2123 | 1388 | 363 | 365 | 7 | 11 | 0 | 0 |

- Molecule 2 is a protein called Phenylalanyl-tRNA synthetase beta chain.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 2 | B | 785 | 6127 | 3925 | 1091 | 1101 | 10 | 46 | 0 | 0 |

- Molecule 3 is META-TYROSINE (three-letter code: MTY) (formula: $C_9H_{11}NO_3$).




| Mol | Chain | Residues | Atoms | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | | |
| 3 | A | 1 | 13 | 9 | 1 | 3 | 0 | 0 |
| 3 | B | 1 | 13 | 9 | 1 | 3 | 0 | 0 |

- Molecule 4 is water.

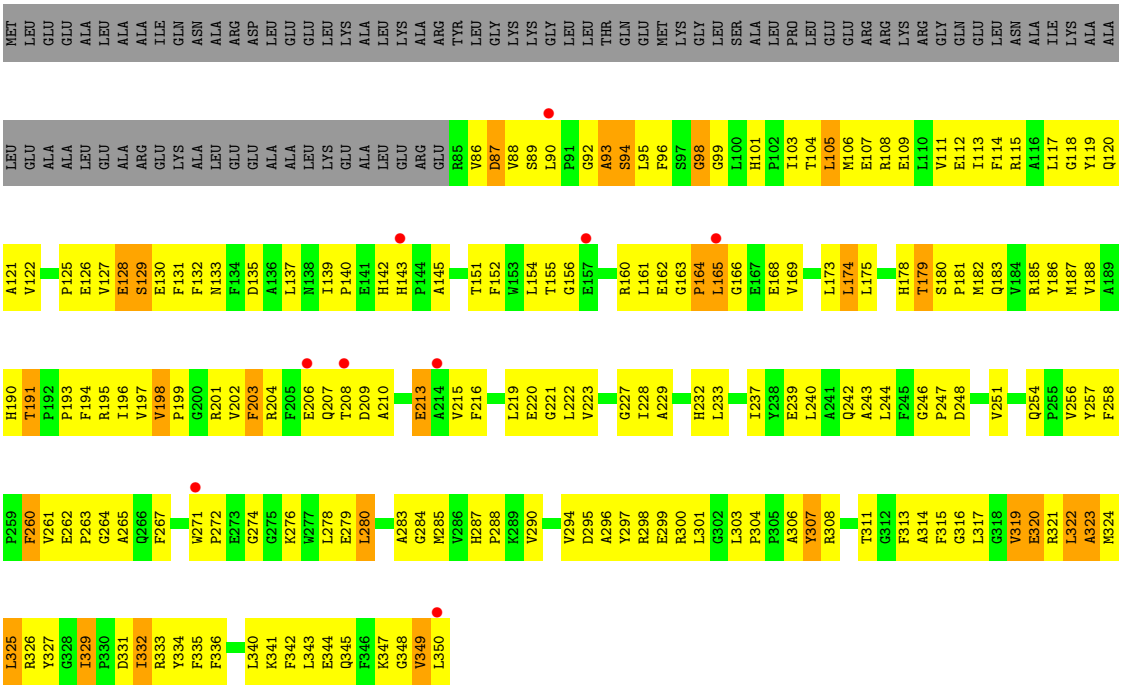
| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | O | | |
| 4 | A | 64 | 64 | 64 | 0 | 0 |
| | | | Total | O | | |
| 4 | B | 177 | 177 | 177 | 0 | 0 |

# 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.

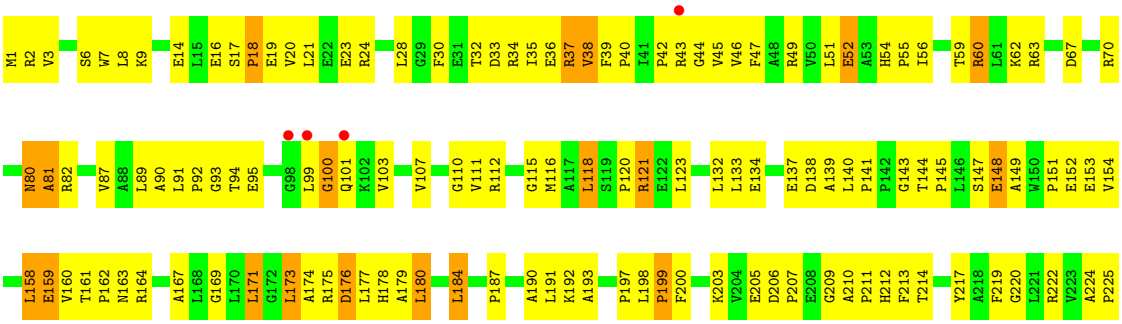
• Molecule 1: Phenylalanyl-tRNA synthetase alpha chain

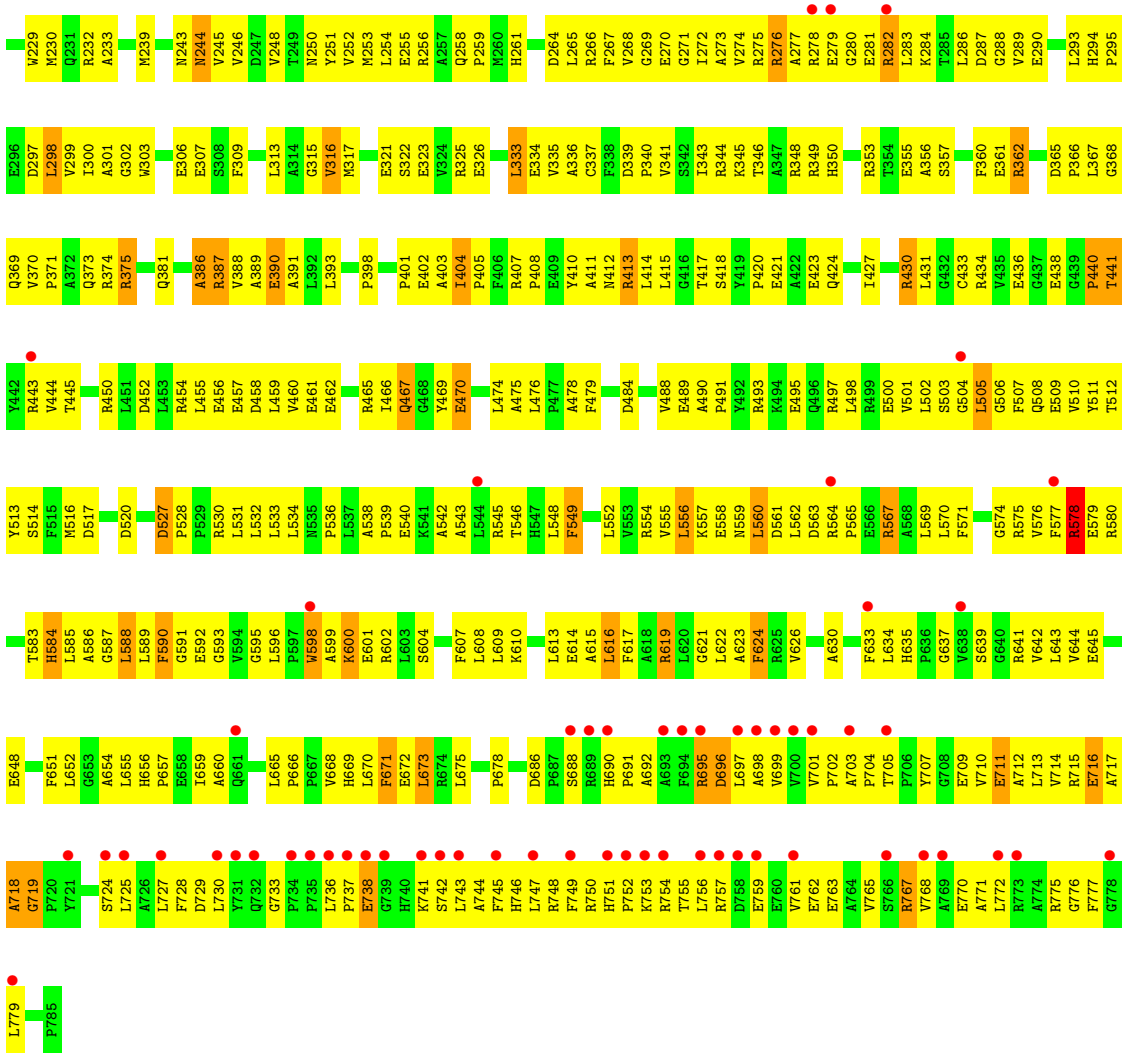
Chain A: 3% 23% 45% 7% 24%




• Molecule 2: Phenylalanyl-tRNA synthetase beta chain

Chain B: 8% 37% 55% 8%




W229 M230 Q231 R232 A233 M239 N243 N244 V245 V246 D247 V248 T249 N250 Y251 V252 M253 L254 E255 R256 A257 Q258 P259 M260 H261 D264 L265 R266 F267 V268 G269 E270 G271 I272 A273 V274 R275 R276 A277 R278 E279 G280 E281 R282 L283 K284 T285 L286 D287 G288 V289 E290 L293 H294 P295
E296 D297 L298 V299 I300 A301 G302 W303 E306 E307 S308 F309 L313 A314 G315 V316 M317 E321 S322 E323 V324 R325 E326 L333 E334 V335 A336 C337 F338 D339 P340 V341 S342 I343 R344 K345 T346 A347 R348 R349 H350 R353 T354 E355 A356 S357 F360 E361 R362 D365 P366 L367 G368
Q369 V370 P371 A372 Q373 R374 R375 Q381 A386 R387 V388 A389 E390 A391 L392 L393 P398 P401 E402 A403 I404 P405 F406 R407 P408 E409 Y410 A411 N412 R413 L414 L415 G416 T417 S418 Y419 P420 E421 A422 E423 Q424 I427 R430 L431 G432 C433 R434 V435 E436 G437 E438 G439 P440 T441
Y442 R443 V444 T445 R450 L451 D452 L453 R454 L455 E456 E457 D458 L459 V460 E461 E462 R465 I466 Q467 G468 Y469 E470 L474 A475 L476 P477 A478 F479 D484 V488 E489 A490 P491 Y492 R493 K494 E495 Q496 R497 L498 R499 E500 V501 L502 S503 G504 L505 G506 F507 Q508 E509 V510 Y511 T512
Y513 S514 F515 M516 D517 D520 D527 P528 P529 R530 L531 L532 L533 L534 N535 P536 L537 A538 P539 E540 K541 A542 A543 L544 R545 T546 H547 L548 F549 L552 V553 R554 V555 L556 K557 E558 N559 L560 D561 L562 D563 R564 P565 E566 R567 A568 L569 L570 F571 G574 R575 V576 F577 R578 E579 R580
T583 H584 L585 A586 G587 L588 L589 F590 G591 E592 G593 V594 G595 L596 P597 W598 A599 K600 E601 R602 L603 S604 F607 L608 L609 K610 L613 E614 A615 L616 F617 A618 R619 L620 G621 L622 A623 F624 R625 V626 A630 F633 L634 H635 P636 G637 V638 S639 G640 R641 V642 L643 V644 E645
E648 F651 L652 G653 A654 L655 H656 P657 E658 I659 A660 Q661 L665 P666 P667 V668 H669 L670 P671 E672 L673 R674 L675 P678 D686 P687 S688 R689 H690 P691 A692 A693 F694 R695 D696 L697 A698 V699 V700 V701 P702 A703 P704 T705 P706 Y707 G708 E709 V710 E711 A712 L713 V714 R715 E716 A717
A718 G719 P720 Y721 S724 L725 A726 L727 F728 D729 L730 Y731 G732 G733 P734 P735 L736 P737 E738 G739 H740 K741 S742 L743 A744 F745 H746 L747 R748 F749 R750 H751 P752 K753 R754 T755 L756 R757 D758 E759 E760 V761 E762 E763 A764 V765 S766 R767 V768 A769 E770 A771 L772 R773 A774 R775 G776 F777 G778
L779 P785

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 32 2 1 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 173.18Å 173.18Å 138.85Å<br>90.00° 90.00° 120.00° | Depositor |
| Resolution (Å) | 47.04 – 2.90<br>47.04 – 2.90 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 99.2 (47.04-2.90)<br>99.2 (47.04-2.90) | Depositor<br>EDS |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 2.36 (at 2.91Å) | Xtriage |
| Refinement program | REFMAC 5.4.0067 | Depositor |
| R, $R_{free}$ | 0.269 , 0.282<br>0.267 , 0.280 | Depositor<br>DCC |
| $R_{free}$ test set | 2688 reflections (5.06%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 69.4 | Xtriage |
| Anisotropy | 0.295 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.34 , 43.1 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.52$, $< L^2 > = 0.36$ | Xtriage |
| Estimated twinning fraction | 0.005 for -h,-k,l | Xtriage |
| $F_o$,$F_c$ correlation | 0.91 | EDS |
| Total number of atoms | 8517 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 69.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.68% 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: MTY

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 | 0.51 | 0/2191 | 0.70 | 0/2971 |
| 2 | B | 0.44 | 0/6280 | 0.71 | 3/8536 (0.0%) |
| All | All | 0.46 | 0/8471 | 0.71 | 3/11507 (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 |

There are no bond length outliers.

All (3) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 2 | B | 387 | ARG | N-CA-C | -5.86 | 95.18 | 111.00 |
| 2 | B | 38 | VAL | N-CA-C | 5.48 | 125.80 | 111.00 |
| 2 | B | 37 | ARG | N-CA-C | -5.35 | 96.55 | 111.00 |

There are no chirality outliers.

All (1) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group |
|---|---|---|---|---|
| 1 | A | 186 | TYR | Sidechain |

## 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 | 2123 | 0 | 2075 | 299 | 0 |
| 2 | B | 6127 | 0 | 6180 | 766 | 1 |
| 3 | A | 13 | 0 | 9 | 8 | 0 |
| 3 | B | 13 | 0 | 9 | 11 | 0 |
| 4 | A | 64 | 0 | 0 | 120 | 0 |
| 4 | B | 177 | 0 | 0 | 254 | 1 |
| All | All | 8517 | 0 | 8273 | 1051 | 1 |

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 64.

The worst 5 of 1051 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:284:GLY:HA3 | 4:A:391:HOH:O | 1.16 | 1.31 |
| 4:A:408:HOH:O | 2:B:579:GLU:HB3 | 1.18 | 1.31 |
| 2:B:9:LYS:HE2 | 4:B:951:HOH:O | 1.22 | 1.29 |
| 2:B:33:ASP:HB3 | 4:B:843:HOH:O | 1.18 | 1.27 |
| 1:A:112:GLU:HG2 | 4:A:377:HOH:O | 1.15 | 1.26 |

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|---|---|---|---|
| 2:B:110:GLY:O | 4:B:808:HOH:O[5_565] | 2.13 | 0.07 |

## 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 | 264/350 (75%) | 223 (84%) | 25 (10%) | 16 (6%) | 1 4 |
| 2 | B | 783/785 (100%) | 654 (84%) | 102 (13%) | 27 (3%) | 3 15 |
| All | All | 1047/1135 (92%) | 877 (84%) | 127 (12%) | 43 (4%) | 3 11 |

5 of 43 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 93 | ALA |
| 1 | A | 94 | SER |
| 1 | A | 319 | VAL |
| 1 | A | 320 | GLU |
| 1 | A | 349 | VAL |

### 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 | 214/277 (77%) | 190 (89%) | 24 (11%) | 6 18 |
| 2 | B | 630/630 (100%) | 578 (92%) | 52 (8%) | 11 32 |
| All | All | 844/907 (93%) | 768 (91%) | 76 (9%) | 9 29 |

5 of 76 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 2 | B | 549 | PHE |
| 2 | B | 673 | LEU |
| 2 | B | 562 | LEU |
| 2 | B | 588 | LEU |
| 2 | B | 767 | ARG |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 2 | B | 467 | GLN |
| 2 | B | 732 | GLN |
| 2 | B | 746 | HIS |
| 2 | B | 669 | HIS |
| 2 | B | 178 | HIS |

### 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)

2 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 | $\#\|Z\| > 2$ |
| 3 | MTY | B | 786 | - | 12,13,13 | 0.60 | 0 | 16,17,17 | 0.89 | 1 (6%) |
| 3 | MTY | A | 351 | - | 12,13,13 | 0.44 | 0 | 16,17,17 | 0.97 | 1 (6%) |

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 |
|---|---|---|---|---|---|---|---|
| 3 | MTY | B | 786 | - | - | 6/8/8/8 | 0/1/1/1 |
| 3 | MTY | A | 351 | - | - | 0/8/8/8 | 0/1/1/1 |

There are no bond length outliers.

All (2) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 3 | B | 786 | MTY | CG-CB-CA | -2.52 | 108.89 | 114.13 |
| 3 | A | 351 | MTY | CG-CB-CA | -2.22 | 109.51 | 114.13 |

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|---|---|---|---|---|
| 3 | B | 786 | MTY | CA-CB-CG-CD2 |
| 3 | B | 786 | MTY | CA-CB-CG-CD1 |
| 3 | B | 786 | MTY | N-CA-CB-CG |
| 3 | B | 786 | MTY | OXT-C-CA-CB |
| 3 | B | 786 | MTY | C-CA-CB-CG |

There are no ring outliers.

2 monomers are involved in 19 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 3 | B | 786 | MTY | 11 | 0 |
| 3 | A | 351 | MTY | 8 | 0 |

## 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 | 265/350 (75%) | 0.32 | 9 (3%) 45 40 | 29, 61, 94, 104 | 0 |
| 2 | B | 779/785 (99%) | 0.48 | 64 (8%) 11 9 | 35, 66, 119, 120 | 0 |
| All | All | 1044/1135 (91%) | 0.44 | 73 (6%) 16 12 | 29, 65, 115, 120 | 0 |

The worst 5 of 73 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 2 | B | 98 | GLY | 6.0 |
| 2 | B | 761 | VAL | 5.8 |
| 2 | B | 756 | LEU | 5.8 |
| 2 | B | 99 | LEU | 5.3 |
| 2 | B | 731 | TYR | 4.9 |

## 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 |
|---|---|---|---|---|---|---|---|---|
| 3 | MTY | B | 786 | 13/13 | 0.90 | 0.42 | 20,20,20,20 | 0 |
| 3 | MTY | A | 351 | 13/13 | 0.96 | 0.20 | 20,20,20,20 | 0 |

## 6.5 Other polymers (i)

There are no such residues in this entry.