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

Sep 13, 2020 – 09:52 AM BST

| | | |
|---|---|---|
| PDB ID | : | 1N1I |
| Title | : | The structure of MSP-1(19) from Plasmodium knowlesi |
| Authors | : | Garman, S.C.; Simcoke, W.N.; Stowers, A.W.; Garboczi, D.N. |
| Deposited on | : | 2002-10-17 |
| Resolution | : | 2.40 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 following versions of software and data (see references (i)) were used in the production of this report:

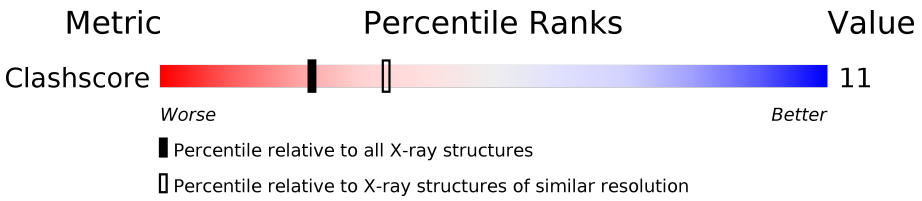
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | FAILED |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.14.4.dev1 |

# 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 2.40 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)) |
|---|---|---|
| Clashscore | 141614 | 4398 (2.40-2.40) |

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%

Note EDS failed to run properly.

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 1 | A | 105 | 72% 14% 13% |
| 1 | B | 105 | 68% 21% 11% |
| 1 | C | 105 | 72% 16% 11% |
| 1 | D | 105 | 70% 12% 17% |

# 2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3042 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 Merozoite surface protein-1.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 1 | A | 91 | 675 | 409 | 111 | 144 | 11 | 0 | 0 | 0 |
| 1 | B | 93 | 692 | 420 | 115 | 146 | 11 | 0 | 0 | 0 |
| 1 | C | 93 | 696 | 423 | 117 | 145 | 11 | 0 | 0 | 0 |
| 1 | D | 87 | 653 | 398 | 107 | 137 | 11 | 0 | 0 | 0 |

There are 52 discrepancies between the modelled and reference sequences:

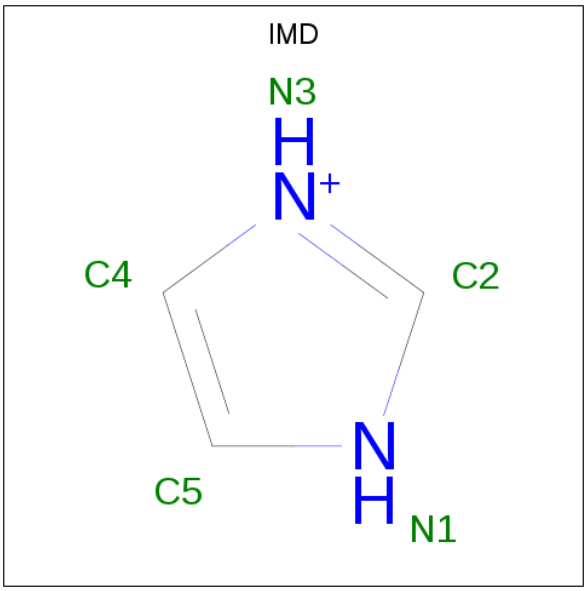
| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 1 | GLU | GLN | SEE REMARK 999 | UNP Q9GSQ9 |
| A | 2 | ALA | THR | SEE REMARK 999 | UNP Q9GSQ9 |
| A | 3 | GLU | GLN | SEE REMARK 999 | UNP Q9GSQ9 |
| A | 4 | ALA | MET | SEE REMARK 999 | UNP Q9GSQ9 |
| A | 5 | SER | LEU | SEE REMARK 999 | UNP Q9GSQ9 |
| A | 98 | GLY | - | EXPRESSION TAG | UNP Q9GSQ9 |
| A | 99 | PRO | - | EXPRESSION TAG | UNP Q9GSQ9 |
| A | 100 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| A | 101 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| A | 102 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| A | 103 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| A | 104 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| A | 105 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| B | 1 | GLU | GLN | SEE REMARK 999 | UNP Q9GSQ9 |
| B | 2 | ALA | THR | SEE REMARK 999 | UNP Q9GSQ9 |
| B | 3 | GLU | GLN | SEE REMARK 999 | UNP Q9GSQ9 |
| B | 4 | ALA | MET | SEE REMARK 999 | UNP Q9GSQ9 |
| B | 5 | SER | LEU | SEE REMARK 999 | UNP Q9GSQ9 |
| B | 98 | GLY | - | EXPRESSION TAG | UNP Q9GSQ9 |
| B | 99 | PRO | - | EXPRESSION TAG | UNP Q9GSQ9 |
| B | 100 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |

*Continued on next page...*

*Continued from previous page...*

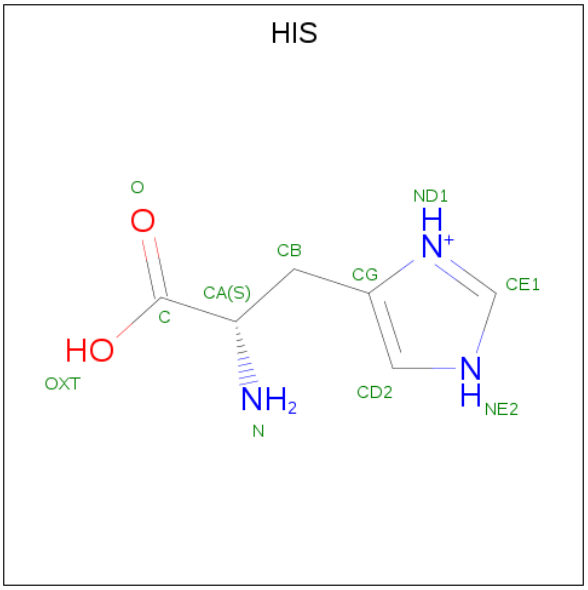
| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| B | 101 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| B | 102 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| B | 103 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| B | 104 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| B | 105 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| C | 1 | GLU | GLN | SEE REMARK 999 | UNP Q9GSQ9 |
| C | 2 | ALA | THR | SEE REMARK 999 | UNP Q9GSQ9 |
| C | 3 | GLU | GLN | SEE REMARK 999 | UNP Q9GSQ9 |
| C | 4 | ALA | MET | SEE REMARK 999 | UNP Q9GSQ9 |
| C | 5 | SER | LEU | SEE REMARK 999 | UNP Q9GSQ9 |
| C | 98 | GLY | - | EXPRESSION TAG | UNP Q9GSQ9 |
| C | 99 | PRO | - | EXPRESSION TAG | UNP Q9GSQ9 |
| C | 100 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| C | 101 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| C | 102 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| C | 103 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| C | 104 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| C | 105 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| D | 1 | GLU | GLN | SEE REMARK 999 | UNP Q9GSQ9 |
| D | 2 | ALA | THR | SEE REMARK 999 | UNP Q9GSQ9 |
| D | 3 | GLU | GLN | SEE REMARK 999 | UNP Q9GSQ9 |
| D | 4 | ALA | MET | SEE REMARK 999 | UNP Q9GSQ9 |
| D | 5 | SER | LEU | SEE REMARK 999 | UNP Q9GSQ9 |
| D | 98 | GLY | - | EXPRESSION TAG | UNP Q9GSQ9 |
| D | 99 | PRO | - | EXPRESSION TAG | UNP Q9GSQ9 |
| D | 100 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| D | 101 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| D | 102 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| D | 103 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| D | 104 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |
| D | 105 | HIS | - | EXPRESSION TAG | UNP Q9GSQ9 |

- Molecule 2 is IMIDAZOLE (three-letter code: IMD) (formula: $C_3H_5N_2$).




| Mol | Chain | Residues | Atoms | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | | | |
| 2 | B | 1 | 5 | 3 | 2 | | 0 | 0 |
| | | | Total | C | N | | | |
| 2 | D | 1 | 5 | 3 | 2 | | 0 | 0 |

- Molecule 3 is HISTIDINE (three-letter code: HIS) (formula: $C_6H_{10}N_3O_2$).




| Mol | Chain | Residues | Atoms | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | | |
| 3 | C | 1 | 10 | 6 | 3 | 1 | 0 | 0 |

- Molecule 4 is water.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | O | | |
| 4 | A | 85 | 85 | 85 | 0 | 0 |
| | | | Total | O | | |
| 4 | B | 80 | 80 | 80 | 0 | 0 |
| | | | Total | O | | |
| 4 | C | 79 | 79 | 79 | 0 | 0 |
| | | | Total | O | | |
| 4 | D | 62 | 62 | 62 | 0 | 0 |

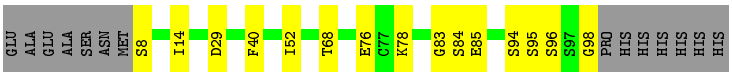
# 3 Residue-property plots (i)

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. 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.

Note EDS failed to run properly.

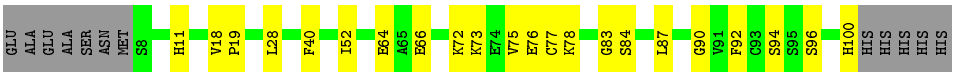
- Molecule 1: Merozoite surface protein-1

Chain A:



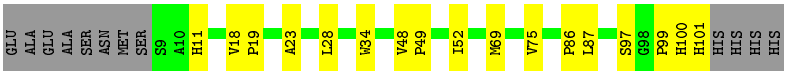

- Molecule 1: Merozoite surface protein-1

Chain B:



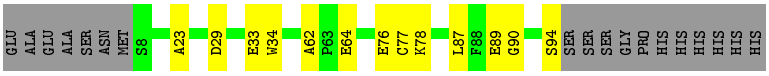

- Molecule 1: Merozoite surface protein-1

Chain C:




- Molecule 1: Merozoite surface protein-1

Chain D:



# 4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

| Property | Value | Source |
|---|---|---|
| Space group | P 1 21 1 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 33.88Å 106.46Å 62.69Å<br>90.00° 102.05° 90.00° | Depositor |
| Resolution (Å) | 20.59 – 2.40 | Depositor |
| % Data completeness<br>(in resolution range) | 97.4 (20.59-2.40) | Depositor |
| $R_{merge}$ | 0.08 | Depositor |
| $R_{sym}$ | 0.08 | Depositor |
| $< I/\sigma(I) >$ [1] | 3.95 (at 2.41Å) | Xtriage |
| Refinement program | CNS 1.0 | Depositor |
| R, $R_{free}$ | 0.234 , 0.264 | Depositor |
| Wilson B-factor (Å$^2$) | 45.5 | Xtriage |
| Anisotropy | 0.234 | Xtriage |
| L-test for twinning[2] | $< \|L\| > = 0.50$, $< L^2 > = 0.33$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| Total number of atoms | 3042 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 55.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.39% 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: IMD

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 | 0.44 | 0/686 | 0.65 | 0/923 |
| 1 | B | 0.45 | 0/705 | 0.64 | 0/950 |
| 1 | C | 0.44 | 0/710 | 0.66 | 0/957 |
| 1 | D | 0.38 | 0/664 | 0.62 | 0/894 |
| All | All | 0.43 | 0/2765 | 0.64 | 0/3724 |

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 (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 | 675 | 0 | 616 | 12 | 0 |
| 1 | B | 692 | 0 | 630 | 14 | 0 |
| 1 | C | 696 | 0 | 632 | 20 | 0 |
| 1 | D | 653 | 0 | 598 | 12 | 0 |
| 2 | B | 5 | 0 | 5 | 0 | 0 |
| 2 | D | 5 | 0 | 5 | 0 | 0 |
| 3 | C | 10 | 0 | 6 | 3 | 0 |
| 4 | A | 85 | 0 | 0 | 6 | 0 |
| 4 | B | 80 | 0 | 0 | 2 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|---|---|---|---|---|---|---|
| 4 | C | 79 | 0 | 0 | 7 | 0 |
| 4 | D | 62 | 0 | 0 | 3 | 0 |
| All | All | 3042 | 0 | 2492 | 60 | 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 11.

The worst 5 of 60 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) |
|---|---|---|---|
| 3:C:501:HIS:HA | 4:C:3027:HOH:O | 1.58 | 1.02 |
| 1:D:87:LEU:HD21 | 1:D:94:SER:HB2 | 1.55 | 0.87 |
| 1:C:101:HIS:HB3 | 4:C:3120:HOH:O | 1.75 | 0.86 |
| 1:C:69:MET:SD | 1:C:75:VAL:HG12 | 2.23 | 0.78 |
| 1:C:48:VAL:HG22 | 1:C:49:PRO:HD2 | 1.64 | 0.78 |

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)

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)

3 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 | HIS | C | 501 | - | 5,10,11 | 0.65 | 0 | 3,12,14 | 1.37 | 1 (33%) |
| 2 | IMD | B | 401 | - | 3,5,5 | 0.57 | 0 | 4,5,5 | 0.55 | 0 |
| 2 | IMD | D | 601 | - | 3,5,5 | 0.58 | 0 | 4,5,5 | 0.48 | 0 |

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 | HIS | C | 501 | - | - | 0/5/6/8 | 0/1/1/1 |
| 2 | IMD | B | 401 | - | - | - | 0/1/1/1 |
| 2 | IMD | D | 601 | - | - | - | 0/1/1/1 |

There are no bond length outliers.

All (1) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 3 | C | 501 | HIS | CD2-NE2-CE1 | 2.04 | 108.96 | 105.78 |

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 3 | C | 501 | HIS | 3 | 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)

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

## 6.3 Carbohydrates (i)

EDS failed to run properly - this section is therefore empty.

## 6.4 Ligands (i)

EDS failed to run properly - this section is therefore empty.

## 6.5 Other polymers (i)

EDS failed to run properly - this section is therefore empty.