# wwPDB EM Map/Model Validation Report (i)

Apr 26, 2016 – 04:25 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 5IRE |
| EMDB ID: | : | EMD-8116 |
| Title | : | The cryo-EM structure of Zika Virus |
| Authors | : | Sirohi, D.; Chen, Z.; Sun, L.; Klose, T.; Pierson, T.; Rossmann, M.; Kuhn, R. |
| Deposited on | : | 2016-03-13 |
| Resolution | : | 3.80 Å(reported) |

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

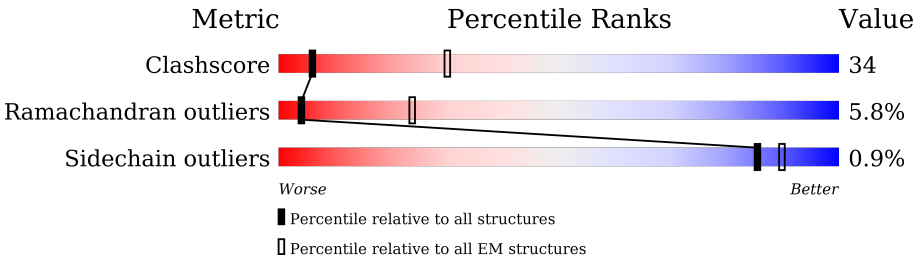
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.7.1 (RC1), CSD as537be (2016) |
| Percentile statistics | : | 20151230.v01 (using entries in the PDB archive December 30th 2015) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et. al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | rb-20027457 |

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.80 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) | EM structures (#Entries) |
|---|---|---|
| Clashscore | 114402 | 924 |
| Ramachandran outliers | 111179 | 726 |
| Sidechain outliers | 111093 | 686 |

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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%

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 1 | A | 504 | 66% 28% |
| 1 | C | 504 | 62% 32% 6% |
| 1 | E | 504 | 67% 27% 5% |
| 2 | B | 75 | 80% 19% |
| 2 | D | 75 | 73% 24% |
| 2 | F | 75 | 73% 23% |

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 | NAG | C | 600 | - | - | X | - |
| 3 | NAG | E | 600 | X | - | - | - |

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11024 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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.

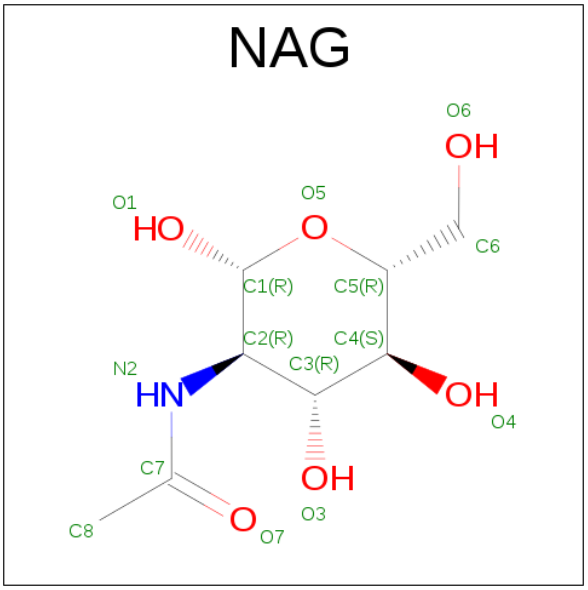
- Molecule 1 is a protein called E protein.

| Mol | Chain | Residues | Atoms | | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 501 | Total | C | N | O | S | 0 | 0 |
| | | | 3167 | 1993 | 576 | 582 | 16 | | |
| 1 | C | 501 | Total | C | N | O | S | 0 | 0 |
| | | | 3225 | 2035 | 589 | 586 | 15 | | |
| 1 | E | 501 | Total | C | N | O | S | 0 | 0 |
| | | | 3186 | 2003 | 581 | 585 | 17 | | |

- Molecule 2 is a protein called M protein.

| Mol | Chain | Residues | Atoms | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|
| 2 | B | 75 | Total | C | N | O | 0 | 0 |
| | | | 456 | 285 | 85 | 86 | | |
| 2 | D | 75 | Total | C | N | O | 0 | 0 |
| | | | 456 | 285 | 85 | 86 | | |
| 2 | F | 75 | Total | C | N | O | 0 | 0 |
| | | | 450 | 282 | 82 | 86 | | |

- Molecule 3 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



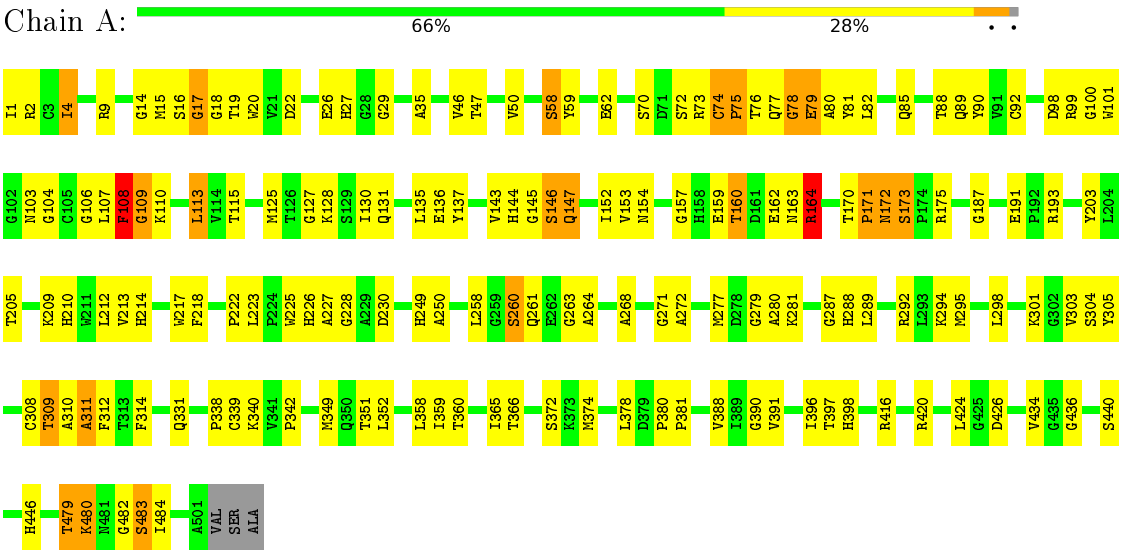
| Mol | Chain | Residues | Atoms | | | | AltConf |
|---|---|---|---|---|---|---|---|
| 3 | A | 1 | Total | C | N | O | 0 |
| | | | 28 | 16 | 2 | 10 | |
| 3 | A | 1 | Total | C | N | O | 0 |
| | | | 28 | 16 | 2 | 10 | |
| 3 | C | 1 | Total | C | N | O | 0 |
| | | | 28 | 16 | 2 | 10 | |
| 3 | C | 1 | Total | C | N | O | 0 |
| | | | 28 | 16 | 2 | 10 | |
| 3 | E | 1 | Total | C | N | O | 0 |
| | | | 28 | 16 | 2 | 10 | |
| 3 | E | 1 | Total | C | N | O | 0 |
| | | | 28 | 16 | 2 | 10 | |

# 3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

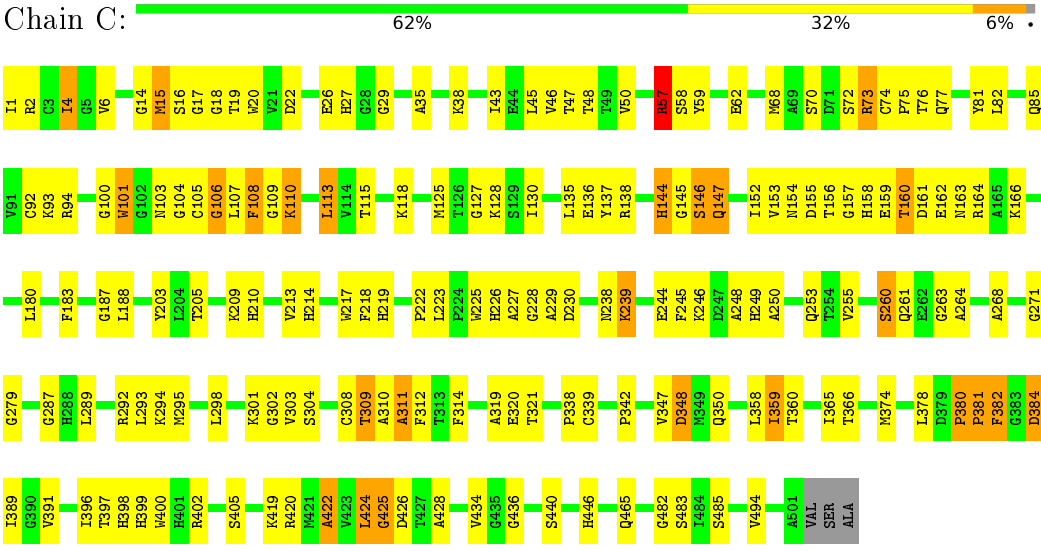
- Molecule 1: E protein

Chain A:



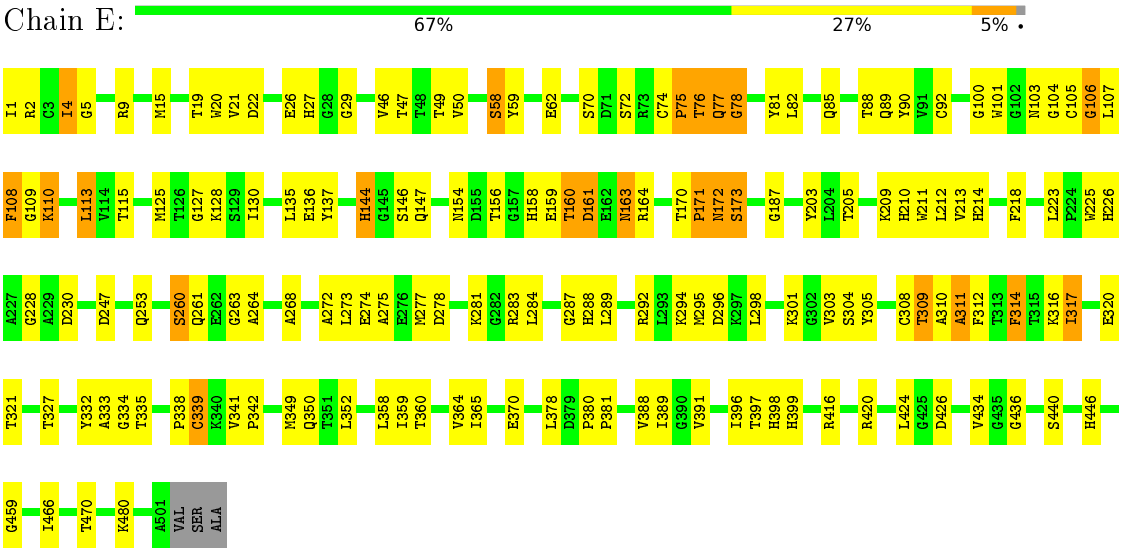

- Molecule 1: E protein

Chain C:



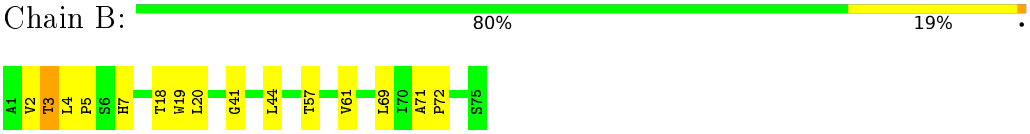

- Molecule 1: E protein

Chain E:



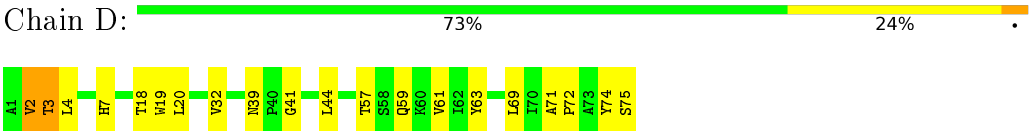

• Molecule 2: M protein

Chain B:



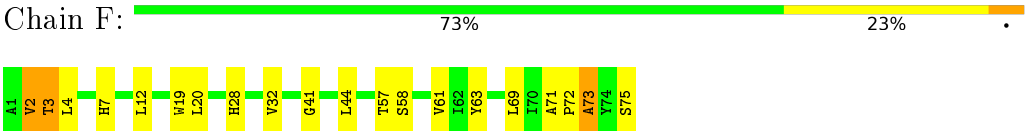

• Molecule 2: M protein

Chain D:




• Molecule 2: M protein

Chain F:



# 4 Experimental information (i)

| Property | Value | Source |
|---|---|---|
| Reconstruction method | SINGLE PARTICLE | Depositor |
| Imposed symmetry | POINT, I | Depositor |
| Number of images | 11842 | Depositor |
| Resolution determination method | FSC 0.143 CUT-OFF | Depositor |
| CTF correction method | Not provided | Depositor |
| Microscope | FEI TITAN KRIOS | Depositor |
| Voltage (kV) | 300 | Depositor |
| Electron dose ($e^-/Å^2$) | Not provided | Depositor |
| Minimum defocus (nm) | 1000 | Depositor |
| Maximum defocus (nm) | 2500 | Depositor |
| Magnification | 14000 | Depositor |
| Image detector | Not provided | Depositor |

# 5 Model quality

## 5.1 Standard geometry

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

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 | #$\|Z\|>2$ | RMSZ | #$\|Z\|>2$ |
| 1 | A | 0.56 | 2/3228 (0.1%) | 0.64 | 3/4439 (0.1%) |
| 1 | C | 0.60 | 2/3289 (0.1%) | 0.62 | 3/4514 (0.1%) |
| 1 | E | 0.53 | 0/3247 | 0.60 | 1/4460 (0.0%) |
| 2 | B | 0.41 | 0/466 | 0.48 | 0/650 |
| 2 | D | 0.47 | 0/466 | 0.54 | 1/650 (0.2%) |
| 2 | F | 0.42 | 0/460 | 0.48 | 0/643 |
| All | All | 0.55 | 4/11156 (0.0%) | 0.60 | 8/15356 (0.1%) |

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 |
| 1 | C | 0 | 1 |
| 1 | E | 0 | 1 |
| All | All | 0 | 3 |

All (4) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 1 | C | 73 | ARG | NE-CZ | -7.87 | 1.22 | 1.33 |
| 1 | C | 73 | ARG | CZ-NH1 | -7.82 | 1.22 | 1.33 |
| 1 | A | 164 | ARG | CZ-NH1 | -7.61 | 1.23 | 1.33 |
| 1 | A | 164 | ARG | NE-CZ | -7.48 | 1.23 | 1.33 |

The worst 5 of 8 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 1 | A | 74 | CYS | C-N-CD | -13.49 | 90.92 | 120.60 |
| 1 | C | 73 | ARG | NE-CZ-NH1 | -5.56 | 117.52 | 120.30 |
| 1 | A | 113 | LEU | CA-CB-CG | 5.36 | 127.64 | 115.30 |
| 1 | C | 113 | LEU | CA-CB-CG | 5.35 | 127.61 | 115.30 |
| 1 | E | 113 | LEU | CA-CB-CG | 5.25 | 127.37 | 115.30 |

There are no chirality outliers.

All (3) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group |
|---|---|---|---|---|
| 1 | A | 359 | ILE | Peptide |
| 1 | C | 359 | ILE | Peptide |
| 1 | E | 359 | ILE | 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 | 3167 | 0 | 2593 | 176 | 0 |
| 1 | C | 3225 | 0 | 2708 | 307 | 0 |
| 1 | E | 3186 | 0 | 2621 | 223 | 0 |
| 2 | B | 456 | 0 | 329 | 16 | 0 |
| 2 | D | 456 | 0 | 329 | 24 | 0 |
| 2 | F | 450 | 0 | 318 | 24 | 0 |
| 3 | A | 28 | 0 | 25 | 1 | 0 |
| 3 | C | 28 | 0 | 25 | 9 | 0 |
| 3 | E | 28 | 0 | 24 | 1 | 0 |
| All | All | 11024 | 0 | 8972 | 677 | 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 677 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:C:138:ARG:HD3 | 1:C:166:LYS:CE | 1.23 | 1.61 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|---|---|---|---|
| 1:C:382:PHE:CE2 | 1:C:405:SER:HA | 1.41 | 1.56 |
| 1:A:222:PRO:HG3 | 2:B:3:THR:CB | 1.38 | 1.53 |
| 1:C:27:HIS:CE1 | 1:C:45:LEU:HD21 | 1.44 | 1.52 |
| 1:C:465:GLN:HE21 | 1:C:494:VAL:CG1 | 1.26 | 1.49 |

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 PDB entries followed by that with respect to all EM entries.

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 | 499/504 (99%) | 407 (82%) | 61 (12%) | 31 (6%) | 2 26 |
| 1 | C | 499/504 (99%) | 394 (79%) | 71 (14%) | 34 (7%) | 1 24 |
| 1 | E | 499/504 (99%) | 407 (82%) | 65 (13%) | 27 (5%) | 2 30 |
| 2 | B | 73/75 (97%) | 63 (86%) | 7 (10%) | 3 (4%) | 3 37 |
| 2 | D | 73/75 (97%) | 63 (86%) | 8 (11%) | 2 (3%) | 6 48 |
| 2 | F | 73/75 (97%) | 61 (84%) | 9 (12%) | 3 (4%) | 3 37 |
| All | All | 1716/1737 (99%) | 1395 (81%) | 221 (13%) | 100 (6%) | 4 28 |

5 of 100 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 15 | MET |
| 1 | A | 17 | GLY |
| 1 | A | 75 | PRO |
| 1 | A | 78 | GLY |
| 1 | A | 79 | GLU |

### 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 PDB entries followed by that with respect to all EM entries.

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 | 221/410 (54%) | 220 (100%) | 1 (0%) | 92 97 |
| 1 | C | 235/410 (57%) | 232 (99%) | 3 (1%) | 76 91 |
| 1 | E | 225/410 (55%) | 222 (99%) | 3 (1%) | 76 91 |
| 2 | B | 25/64 (39%) | 25 (100%) | 0 | 100 100 |
| 2 | D | 25/64 (39%) | 25 (100%) | 0 | 100 100 |
| 2 | F | 24/64 (38%) | 24 (100%) | 0 | 100 100 |
| All | All | 755/1422 (53%) | 748 (99%) | 7 (1%) | 85 93 |

5 of 7 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | C | 382 | PHE |
| 1 | E | 339 | CYS |
| 1 | E | 108 | PHE |
| 1 | C | 57 | ARG |
| 1 | E | 154 | ASN |

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

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | C | 399 | HIS |
| 1 | C | 465 | GLN |
| 1 | E | 147 | GLN |
| 1 | C | 154 | ASN |
| 1 | E | 398 | HIS |

### 5.3.3 RNA

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 carbohydrates in this entry.

## 5.6 Ligand geometry (i)

6 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 | NAG | A | 600 | 1,3 | 14,14,15 | 0.53 | 0 | 15,19,21 | 0.99 | 0 |
| 3 | NAG | A | 601 | 3 | 14,14,15 | 0.54 | 0 | 15,19,21 | 0.85 | 1 (6%) |
| 3 | NAG | C | 600 | 1,3 | 14,14,15 | 0.58 | 0 | 15,19,21 | 0.66 | 0 |
| 3 | NAG | C | 601 | 3 | 14,14,15 | 0.54 | 0 | 15,19,21 | 0.68 | 1 (6%) |
| 3 | NAG | E | 600 | 1,3 | 14,14,15 | 0.76 | 0 | 15,19,21 | 0.84 | 1 (6%) |
| 3 | NAG | E | 601 | 3 | 14,14,15 | 0.83 | 0 | 15,19,21 | 0.95 | 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 | NAG | A | 600 | 1,3 | - | 0/6/23/26 | 0/1/1/1 |
| 3 | NAG | A | 601 | 3 | - | 0/6/23/26 | 0/1/1/1 |
| 3 | NAG | C | 600 | 1,3 | - | 0/6/23/26 | 0/1/1/1 |
| 3 | NAG | C | 601 | 3 | - | 0/6/23/26 | 0/1/1/1 |
| 3 | NAG | E | 600 | 1,3 | 1/1/5/7 | 0/6/23/26 | 0/1/1/1 |
| 3 | NAG | E | 601 | 3 | - | 1/6/23/26 | 0/1/1/1 |

There are no bond length outliers.

All (4) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 3 | A | 601 | NAG | C2-N2-C7 | -2.59 | 119.73 | 123.11 |
| 3 | E | 601 | NAG | C2-N2-C7 | -2.38 | 120.01 | 123.11 |
| 3 | E | 600 | NAG | C2-N2-C7 | -2.37 | 120.02 | 123.11 |
| 3 | C | 601 | NAG | C2-N2-C7 | -2.17 | 120.28 | 123.11 |

All (1) chirality outliers are listed below:

| Mol | Chain | Res | Type | Atom |
|---|---|---|---|---|
| 3 | E | 600 | NAG | C1 |

All (1) torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|---|---|---|---|---|
| 3 | E | 601 | NAG | O7-C7-N2-C2 |

There are no ring outliers.

4 monomers are involved in 11 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 3 | A | 600 | NAG | 1 | 0 |
| 3 | C | 600 | NAG | 9 | 0 |
| 3 | C | 601 | NAG | 2 | 0 |
| 3 | E | 600 | NAG | 1 | 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.