# wwPDB EM Validation Summary Report

Oct 20, 2024 – 12:08 AM JST

| | | |
|---|---|---|
| PDB ID | : | 8Z0K |
| EMDB ID | : | EMD-39706 |
| Title | : | Cryo-EM structure of Cas8-HNH system at full R-loop state |
| Authors | : | Zhang, H.; Zhu, H.; Li, X.; Liu, Y. |
| Deposited on | : | 2024-04-09 |
| Resolution | : | 2.51 Å(reported) |

This is a wwPDB EM 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/EMValidationReportHelp
with specific help available everywhere you see the ⓘ 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 ⓘ) were used in the production of this report:

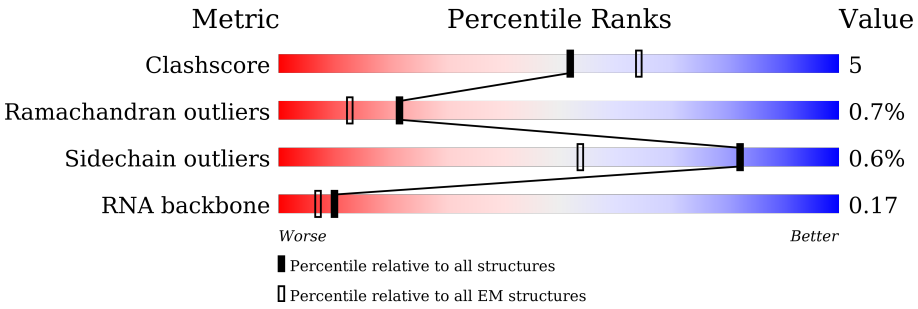
| | | |
|---|---|---|
| EMDB validation analysis | : | FAILED |
| MolProbity | : | 4.02b-467 |
| Percentile statistics | : | 20231227.v01 (using entries in the PDB archive December 27th 2023) |
| MapQ | : | FAILED |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.39 |

# 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 2.51 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 | 210492 | 15764 |
| Ramachandran outliers | 207382 | 16835 |
| Sidechain outliers | 206894 | 16415 |
| RNA backbone | 6643 | 2191 |

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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%

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 1 | A | 325 | 94% 6% |
| 1 | B | 325 | 95% • • |
| 1 | C | 325 | 95% 5% |
| 1 | D | 325 | 94% • • • |
| 1 | G | 325 | 90% 9% • |
| 1 | H | 325 | 77% 12% • 10% |
| 2 | E | 255 | 91% 9% |
| 3 | F | 344 | 78% 17% • • |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 4 | I | 37 | 57% 43% |
| 5 | J | 7 | 86% 14% |
| 6 | L | 69 | 25% 48% 12% 16% |
| 7 | M | 181 | 56% 25% 14% |

# 2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 23037 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 type I-F CRISPR-associated protein Csy3.

| Mol | Chain | Residues | Atoms | | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 325 | Total | C | N | O | S | 0 | 0 |
| | | | 2620 | 1673 | 431 | 501 | 15 | | |
| 1 | B | 321 | Total | C | N | O | S | 0 | 0 |
| | | | 2588 | 1655 | 426 | 492 | 15 | | |
| 1 | C | 324 | Total | C | N | O | S | 0 | 0 |
| | | | 2613 | 1669 | 430 | 499 | 15 | | |
| 1 | D | 319 | Total | C | N | O | S | 0 | 0 |
| | | | 2580 | 1653 | 426 | 486 | 15 | | |
| 1 | G | 321 | Total | C | N | O | S | 0 | 0 |
| | | | 2597 | 1662 | 428 | 492 | 15 | | |
| 1 | H | 292 | Total | C | N | O | S | 0 | 0 |
| | | | 2346 | 1503 | 384 | 446 | 13 | | |

- Molecule 2 is a protein called hypothetical protein J6N51_11000.

| Mol | Chain | Residues | Atoms | | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 2 | E | 255 | Total | C | N | O | S | 0 | 0 |
| | | | 1979 | 1259 | 331 | 377 | 12 | | |

- Molecule 3 is a protein called HNH endonuclease.

| Mol | Chain | Residues | Atoms | | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 3 | F | 332 | Total | C | N | O | S | 0 | 0 |
| | | | 2458 | 1545 | 427 | 478 | 8 | | |

- Molecule 4 is a DNA chain called DNA (37-MER).

| Mol | Chain | Residues | Atoms | | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 4 | I | 37 | Total | C | N | O | P | 0 | 0 |
| | | | 750 | 357 | 135 | 221 | 37 | | |

- Molecule 5 is a DNA chain called DNA (5'-D(P*GP*TP*GP*CP*GP*GP*A)-3').

| Mol | Chain | Residues | Atoms | | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 5 | J | 7 | Total | C | N | O | P | 0 | 0 |
| | | | 148 | 69 | 30 | 42 | 7 | | |

- Molecule 6 is a RNA chain called RNA (69-MER).

| Mol | Chain | Residues | Atoms | | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 6 | L | 58 | Total | C | N | O | P | 0 | 0 |
| | | | 1249 | 556 | 229 | 406 | 58 | | |

- Molecule 7 is a protein called type I-F CRISPR-associated endoribonuclease Cas6/Csy4.

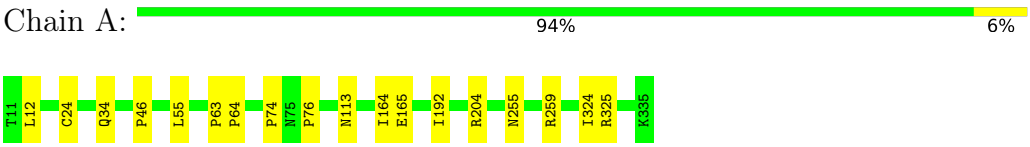
| Mol | Chain | Residues | Atoms | | | | | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|
| 7 | M | 155 | Total | C | N | O | S | 0 | 0 |
| | | | 1109 | 705 | 175 | 226 | 3 | | |

# 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.

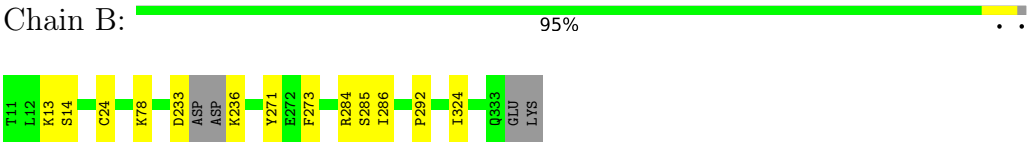
• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain A:



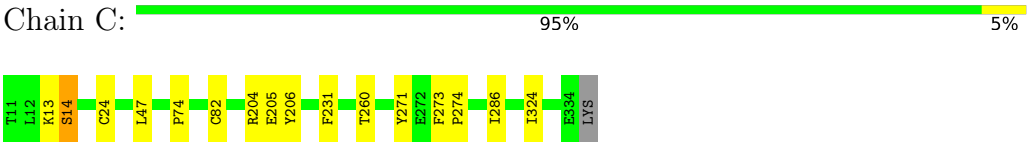

• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain B:



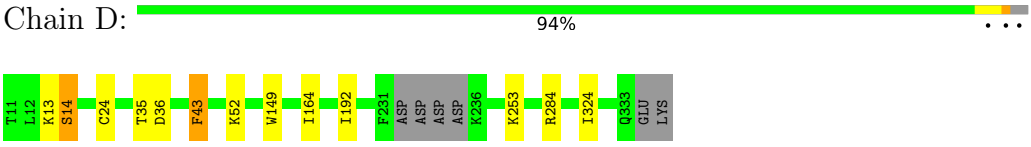

• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain C:



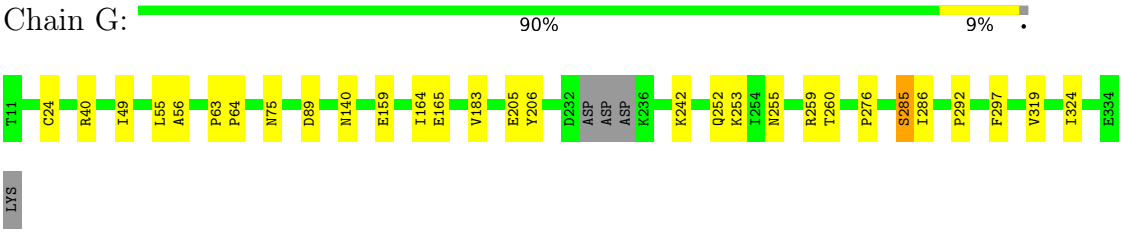

• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain D:



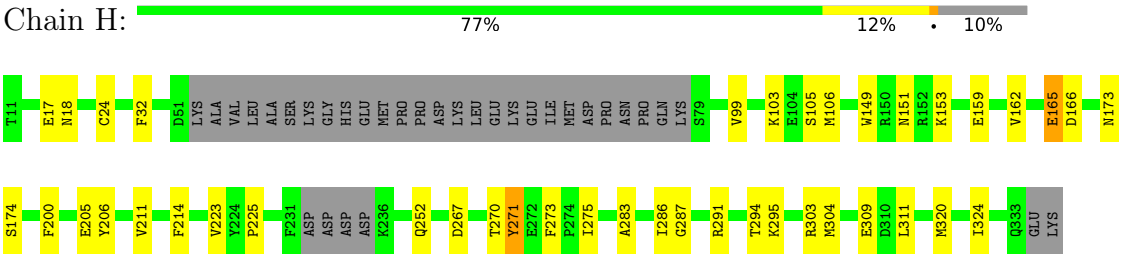

• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain G:



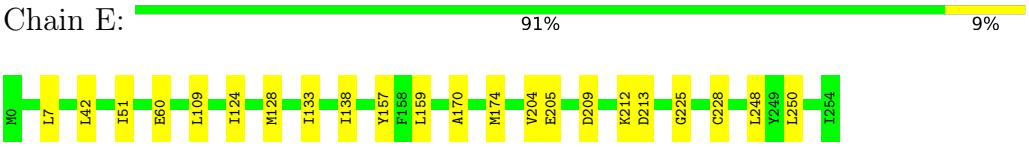

• Molecule 1: type I-F CRISPR-associated protein Csy3

Chain H:



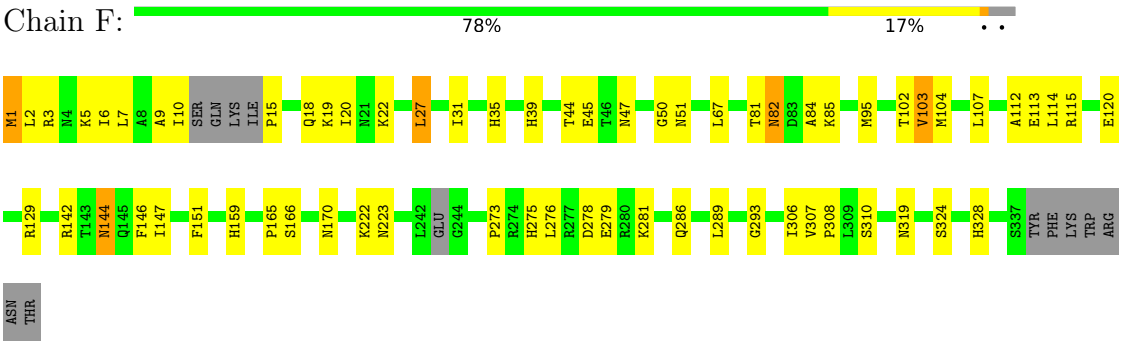

• Molecule 2: hypothetical protein J6N51_11000

Chain E:




• Molecule 3: HNH endonuclease

Chain F:

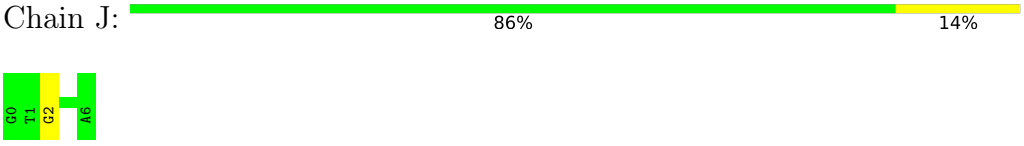



• Molecule 4: DNA (37-MER)

Chain I:

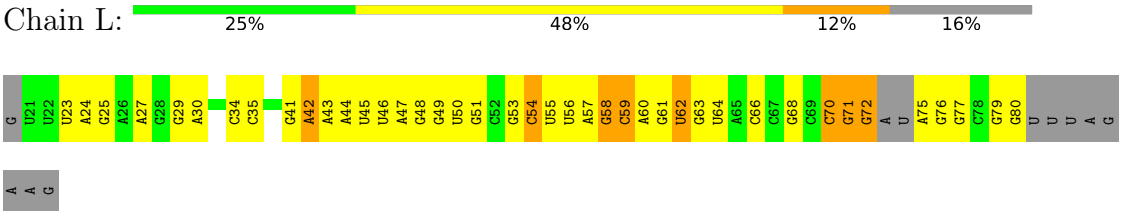

• Molecule 5: DNA (5'-D(P*GP*TP*GP*CP*GP*GP*A)-3')

Chain J:



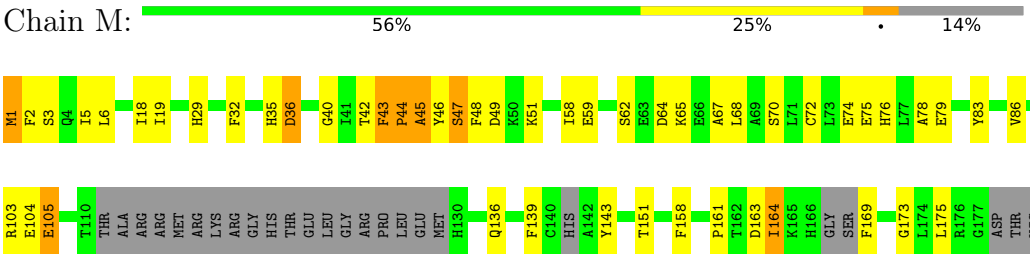

• Molecule 6: RNA (69-MER)

Chain L:



• Molecule 7: type I-F CRISPR-associated endoribonuclease Cas6/Csy4

Chain M:



# 4 Experimental information (i)

| Property | Value | Source |
|---|---|---|
| EM reconstruction method | SINGLE PARTICLE | Depositor |
| Imposed symmetry | POINT, Not provided | |
| Number of particles used | 173047 | Depositor |
| Resolution determination method | FSC 0.143 CUT-OFF | Depositor |
| CTF correction method | PHASE FLIPPING ONLY | Depositor |
| Microscope | JEOL CRYO ARM 300 | Depositor |
| Voltage (kV) | 300 | Depositor |
| Electron dose ($e^-/Å^2$) | 60 | Depositor |
| Minimum defocus (nm) | 1200 | Depositor |
| Maximum defocus (nm) | 1600 | Depositor |
| Magnification | Not provided | |
| Image detector | GATAN K3 (6k x 4k) | Depositor |

# 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 | 0/2679 | 0.59 | 0/3624 |
| 1 | B | 0.34 | 0/2646 | 0.60 | 0/3579 |
| 1 | C | 0.37 | 0/2672 | 0.60 | 0/3615 |
| 1 | D | 0.36 | 0/2638 | 0.61 | 0/3565 |
| 1 | G | 0.37 | 0/2655 | 0.66 | 0/3588 |
| 1 | H | 0.37 | 0/2398 | 0.69 | 0/3246 |
| 2 | E | 0.36 | 0/2014 | 0.62 | 0/2723 |
| 3 | F | 0.42 | 0/2505 | 0.66 | 0/3409 |
| 4 | I | 0.53 | 0/839 | 0.84 | 0/1290 |
| 5 | J | 0.66 | 0/166 | 0.92 | 0/255 |
| 6 | L | 0.38 | 0/1397 | 0.88 | 0/2176 |
| 7 | M | 0.51 | 0/1125 | 0.74 | 0/1535 |
| All | All | 0.39 | 0/23734 | 0.67 | 0/32605 |

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 | H | 0 | 1 |

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group |
|---|---|---|---|---|
| 1 | H | 286 | 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 | 2620 | 0 | 2551 | 13 | 0 |
| 1 | B | 2588 | 0 | 2528 | 9 | 0 |
| 1 | C | 2613 | 0 | 2550 | 11 | 0 |
| 1 | D | 2580 | 0 | 2542 | 9 | 0 |
| 1 | G | 2597 | 0 | 2552 | 18 | 0 |
| 1 | H | 2346 | 0 | 2261 | 25 | 0 |
| 2 | E | 1979 | 0 | 1952 | 16 | 0 |
| 3 | F | 2458 | 0 | 2277 | 57 | 0 |
| 4 | I | 750 | 0 | 416 | 18 | 0 |
| 5 | J | 148 | 0 | 79 | 1 | 0 |
| 6 | L | 1249 | 0 | 627 | 18 | 0 |
| 7 | M | 1109 | 0 | 976 | 61 | 0 |
| All | All | 23037 | 0 | 21311 | 232 | 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 5.

The worst 5 of 232 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) |
|---|---|---|---|
| 7:M:64:ASP:O | 7:M:67:ALA:CB | 1.82 | 1.26 |
| 7:M:64:ASP:C | 7:M:67:ALA:HB3 | 1.70 | 1.12 |
| 7:M:64:ASP:O | 7:M:67:ALA:HB3 | 0.94 | 1.11 |
| 7:M:64:ASP:HB3 | 7:M:67:ALA:HB2 | 1.30 | 1.10 |
| 7:M:64:ASP:CB | 7:M:67:ALA:HB2 | 1.81 | 1.09 |

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 | 323/325 (99%) | 311 (96%) | 12 (4%) | 0 | 100 | 100 |
| 1 | B | 317/325 (98%) | 304 (96%) | 11 (4%) | 2 (1%) | 22 | 39 |
| 1 | C | 322/325 (99%) | 311 (97%) | 10 (3%) | 1 (0%) | 37 | 56 |
| 1 | D | 315/325 (97%) | 302 (96%) | 12 (4%) | 1 (0%) | 37 | 56 |
| 1 | G | 317/325 (98%) | 303 (96%) | 12 (4%) | 2 (1%) | 22 | 39 |
| 1 | H | 286/325 (88%) | 260 (91%) | 22 (8%) | 4 (1%) | 9 | 17 |
| 2 | E | 253/255 (99%) | 239 (94%) | 13 (5%) | 1 (0%) | 30 | 49 |
| 3 | F | 326/344 (95%) | 292 (90%) | 33 (10%) | 1 (0%) | 37 | 56 |
| 7 | M | 147/181 (81%) | 122 (83%) | 19 (13%) | 6 (4%) | 2 | 3 |
| All | All | 2606/2730 (96%) | 2444 (94%) | 144 (6%) | 18 (1%) | 21 | 35 |

5 of 18 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 2 | E | 170 | ALA |
| 7 | M | 47 | SER |
| 1 | B | 285 | SER |
| 1 | C | 14 | SER |
| 3 | F | 103 | 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 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 | 285/292 (98%) | 285 (100%) | 0 | 100 | 100 |
| 1 | B | 282/292 (97%) | 282 (100%) | 0 | 100 | 100 |
| 1 | C | 285/292 (98%) | 284 (100%) | 1 (0%) | 89 | 96 |
| 1 | D | 282/292 (97%) | 281 (100%) | 1 (0%) | 89 | 96 |
| 1 | G | 284/292 (97%) | 284 (100%) | 0 | 100 | 100 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Analysed | Rotameric | Outliers | Percentiles |
|---|---|---|---|---|---|
| 1 | H | 252/292 (86%) | 251 (100%) | 1 (0%) | 89 96 |
| 2 | E | 211/219 (96%) | 211 (100%) | 0 | 100 100 |
| 3 | F | 244/304 (80%) | 238 (98%) | 6 (2%) | 42 69 |
| 7 | M | 106/160 (66%) | 102 (96%) | 4 (4%) | 28 53 |
| All | All | 2231/2435 (92%) | 2218 (99%) | 13 (1%) | 82 94 |

5 of 13 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 3 | F | 306 | ILE |
| 1 | H | 271 | TYR |
| 7 | M | 43 | PHE |
| 7 | M | 29 | HIS |
| 7 | M | 36 | ASP |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 3 | F | 170 | ASN |
| 3 | F | 193 | ASN |
| 1 | H | 208 | ASN |
| 1 | G | 140 | ASN |
| 3 | F | 82 | ASN |

### 5.3.3 RNA (i)

| Mol | Chain | Analysed | Backbone Outliers | Pucker Outliers |
|---|---|---|---|---|
| 6 | L | 56/69 (81%) | 34 (60%) | 1 (1%) |

5 of 34 RNA backbone outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 6 | L | 23 | U |
| 6 | L | 24 | A |
| 6 | L | 25 | G |
| 6 | L | 29 | G |
| 6 | L | 30 | A |

All (1) RNA pucker outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 6 | L | 24 | A |

## 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 oligosaccharides 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.