



## wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Oct 10, 2017 – 06:06 PM EDT

PDB ID : 6B45  
EMDB ID: : EMD-7049  
Title : Cryo-EM structure of Type I-F CRISPR crRNA-guided Csy surveillance complex  
Authors : Guo, T.W.; Bartesaghi, A.; Yang, H.; Falconieri, V.; Rao, P.; Merk, A.; Fox, T.; Earl, L.; Patel, D.J.; Subramaniam, S.  
Deposited on : unknown  
Resolution : 3.50 Å(reported)

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/EMValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030345

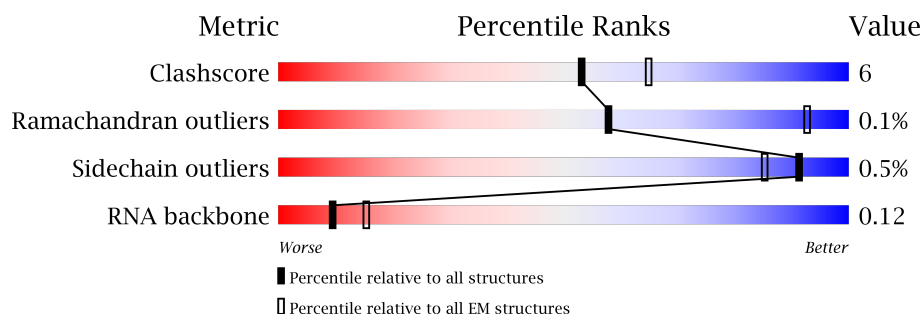
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.50 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<br>(#Entries) | EM structures<br>(#Entries) |
|-----------------------|-----------------------------|-----------------------------|
| Clashscore            | 125131                      | 1336                        |
| Ramachandran outliers | 121729                      | 1120                        |
| Sidechain outliers    | 121581                      | 1026                        |
| RNA backbone          | 3398                        | 335                         |

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  $\geq 3$ , 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  $\leq 5\%$

| Mol | Chain | Length | Quality of chain |
|-----|-------|--------|------------------|
| 1   | A     | 436    | 91% 6% .         |
| 2   | B     | 329    | 71% 21% 7%       |
| 3   | C     | 344    | 68% 17% 15%      |
| 3   | D     | 344    | 77% 19% .        |
| 3   | E     | 344    | 82% 15% .        |
| 3   | F     | 344    | 83% 13% . .      |
| 3   | G     | 344    | 82% 15% .        |
| 3   | H     | 344    | 81% 16% .        |

Continued on next page...

Continued from previous page...

| Mol | Chain | Length | Quality of chain   |
|-----|-------|--------|--|
| 4   | L     | 189    | <div><div></div><div>97%</div><div></div></div>                  |
| 5   | M     | 60     | <div><div></div><div>20%</div><div>60%</div><div>20%</div></div> |

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 21547 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 CRISPR-associated protein Csy1.

| Mol | Chain | Residues | Atoms |      |     |     | AltConf | Trace |
|-----|-------|----------|-------|------|-----|-----|---------|-------|
| 1   | A     | 424      | Total | C    | N   | O   | 0       | 0     |
|     |       |          | 2079  | 1145 | 473 | 461 |         |       |

There are 2 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment        | Reference  |
|-------|---------|----------|--------|----------------|------------|
| A     | -1      | GLY      | -      | expression tag | UNP Q02ML9 |
| A     | 0       | SER      | -      | expression tag | UNP Q02ML9 |

- Molecule 2 is a protein called CRISPR-associated protein Csy2.

| Mol | Chain | Residues | Atoms |      |     |     |   | AltConf | Trace |
|-----|-------|----------|-------|------|-----|-----|---|---------|-------|
| 2   | B     | 305      | Total | C    | N   | O   | S | 0       | 0     |
|     |       |          | 2374  | 1504 | 440 | 425 | 5 |         |       |

There are 2 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment               | Reference  |
|-------|---------|----------|--------|-----------------------|------------|
| B     | -1      | MET      | -      | initiating methionine | UNP Q02MM0 |
| B     | 0       | ALA      | -      | expression tag        | UNP Q02MM0 |

- Molecule 3 is a protein called CRISPR-associated protein Csy3.

| Mol | Chain | Residues | Atoms |      |     |     |   | AltConf | Trace |
|-----|-------|----------|-------|------|-----|-----|---|---------|-------|
| 3   | C     | 293      | Total | C    | N   | O   | S | 0       | 0     |
|     |       |          | 2272  | 1430 | 409 | 431 | 2 |         |       |
| 3   | D     | 333      | Total | C    | N   | O   | S | 0       | 0     |
|     |       |          | 2554  | 1603 | 466 | 483 | 2 |         |       |
| 3   | E     | 334      | Total | C    | N   | O   | S | 0       | 0     |
|     |       |          | 2561  | 1611 | 466 | 482 | 2 |         |       |
| 3   | F     | 335      | Total | C    | N   | O   | S | 0       | 0     |
|     |       |          | 2566  | 1614 | 467 | 483 | 2 |         |       |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Residues | Atoms |      |     |     |   | AltConf | Trace |
|-----|-------|----------|-------|------|-----|-----|---|---------|-------|
| 3   | G     | 333      | Total | C    | N   | O   | S | 0       | 0     |
|     |       |          | 2557  | 1608 | 466 | 481 | 2 |         |       |
| 3   | H     | 333      | Total | C    | N   | O   | S | 0       | 0     |
|     |       |          | 2554  | 1603 | 466 | 483 | 2 |         |       |

There are 12 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment               | Reference  |
|-------|---------|----------|--------|-----------------------|------------|
| C     | -1      | MET      | -      | initiating methionine | UNP Q02MM1 |
| C     | 0       | ALA      | -      | expression tag        | UNP Q02MM1 |
| D     | -1      | MET      | -      | initiating methionine | UNP Q02MM1 |
| D     | 0       | ALA      | -      | expression tag        | UNP Q02MM1 |
| E     | -1      | MET      | -      | initiating methionine | UNP Q02MM1 |
| E     | 0       | ALA      | -      | expression tag        | UNP Q02MM1 |
| F     | -1      | MET      | -      | initiating methionine | UNP Q02MM1 |
| F     | 0       | ALA      | -      | expression tag        | UNP Q02MM1 |
| G     | -1      | MET      | -      | initiating methionine | UNP Q02MM1 |
| G     | 0       | ALA      | -      | expression tag        | UNP Q02MM1 |
| H     | -1      | MET      | -      | initiating methionine | UNP Q02MM1 |
| H     | 0       | ALA      | -      | expression tag        | UNP Q02MM1 |

- Molecule 4 is a protein called CRISPR-associated endonuclease Cas6/Csy4.

| Mol | Chain | Residues | Atoms |     |     |     | AltConf | Trace |
|-----|-------|----------|-------|-----|-----|-----|---------|-------|
| 4   | L     | 189      | Total | C   | N   | O   | 0       | 0     |
|     |       |          | 758   | 380 | 189 | 189 |         |       |

There are 2 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment               | Reference  |
|-------|---------|----------|--------|-----------------------|------------|
| L     | -1      | MET      | -      | initiating methionine | UNP Q02MM2 |
| L     | 0       | ALA      | -      | expression tag        | UNP Q02MM2 |

- Molecule 5 is a RNA chain called Pseudomonas aeruginosa strain SMC4485 CRISPR repeat sequence.

| Mol | Chain | Residues | Atoms |     |     |     |    | AltConf | Trace |
|-----|-------|----------|-------|-----|-----|-----|----|---------|-------|
| 5   | M     | 60       | Total | C   | N   | O   | P  | 0       | 0     |
|     |       |          | 1272  | 569 | 223 | 421 | 59 |         |       |

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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.

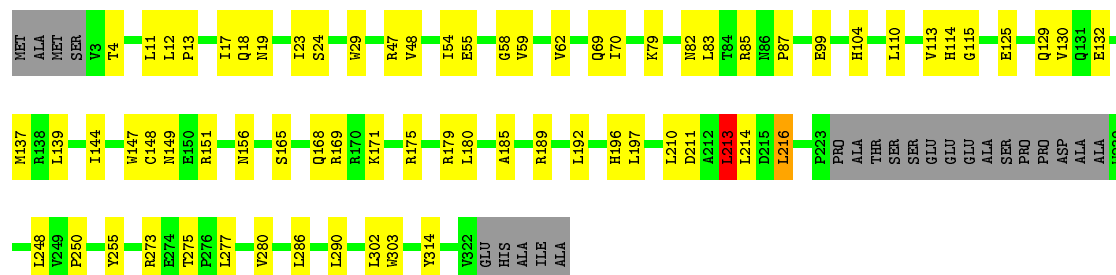
#### • Molecule 1: CRISPR-associated protein Csy1

Chain A: 



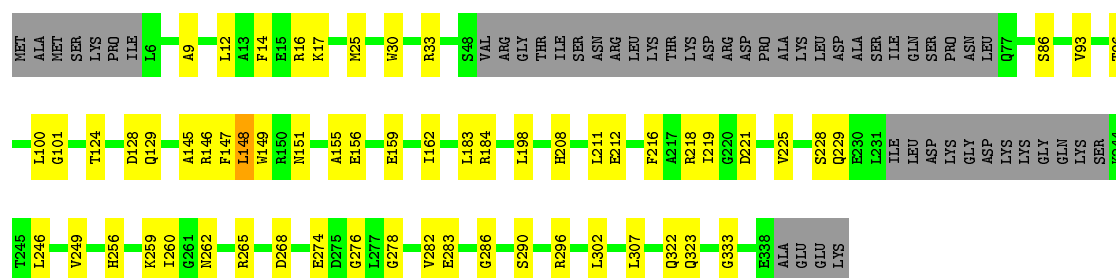
#### • Molecule 2: CRISPR-associated protein Csy2

Chain B: 




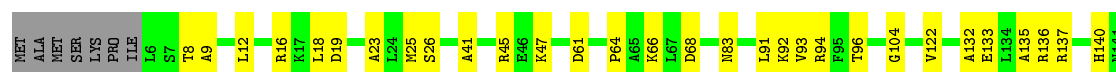
#### • Molecule 3: CRISPR-associated protein Csy3

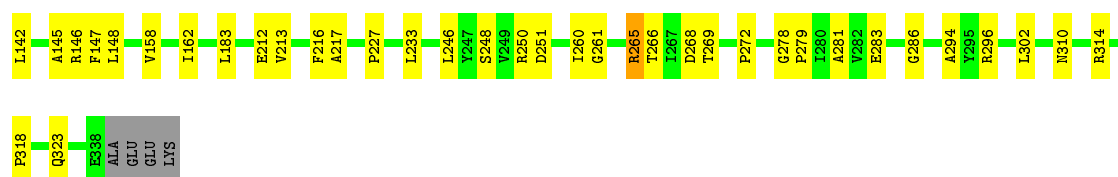
Chain C: 



#### • Molecule 3: CRISPR-associated protein Csy3

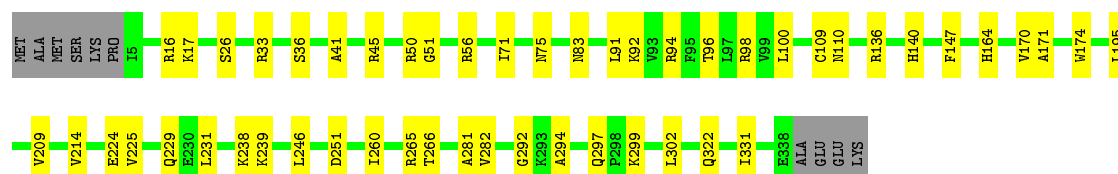
Chain D: 





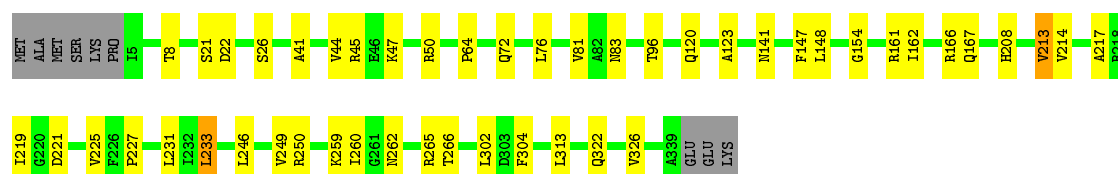
• Molecule 3: CRISPR-associated protein Csy3

Chain E: 82% 15% .



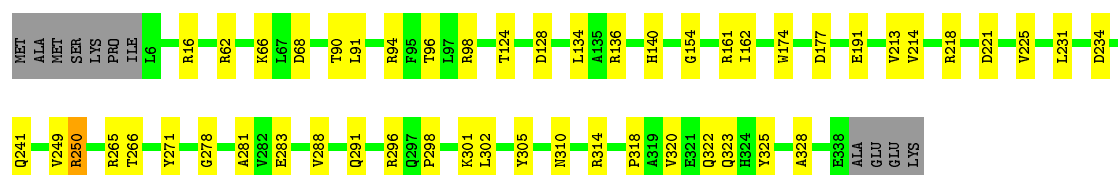
• Molecule 3: CRISPR-associated protein Csy3

Chain F: 83% 13% . .



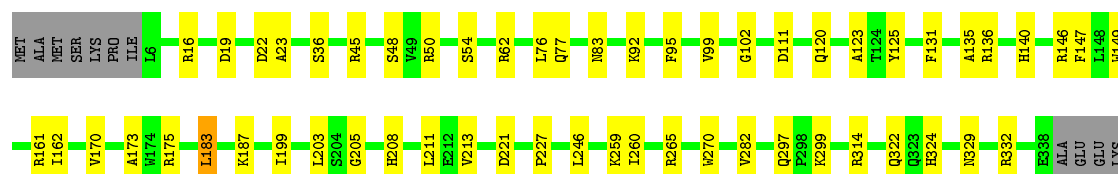
• Molecule 3: CRISPR-associated protein Csy3

Chain G: 82% 15% .



• Molecule 3: CRISPR-associated protein Csy3

Chain H: 81% 16% .

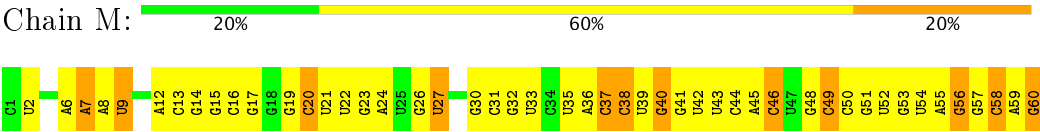


• Molecule 4: CRISPR-associated endonuclease Cas6/Csy4

Chain L: 97% .



- Molecule 5: Pseudomonas aeruginosa strain SMC4485 CRISPR repeat sequence





## 4 Experimental information

| Property                             | Value                                   | Source    |
|--------------------------------------|---|-----------|
| Reconstruction method                | SINGLE PARTICLE                         | Depositor |
| Imposed symmetry                     | POINT, C1                               | Depositor |
| Number of particles used             | 31488                                   | Depositor |
| Resolution determination method      | FSC 0.143 CUT-OFF                       | Depositor |
| CTF correction method                | PHASE FLIPPING AND AMPLITUDE CORRECTION | Depositor |
| Microscope                           | FEI TITAN KRIOS                         | Depositor |
| Voltage (kV)                         | 300                                     | Depositor |
| Electron dose ( $e^-/\text{\AA}^2$ ) | 40                                      | Depositor |
| Minimum defocus (nm)                 | Not provided                            | Depositor |
| Maximum defocus (nm)                 | Not provided                            | Depositor |
| Magnification                        | Not provided                            | Depositor |
| Image detector                       | GATAN K2 SUMMIT (4k 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         | $\# Z  > 2$ | RMSZ        | $\# Z  > 2$     |
| 1   | A     | 0.26         | 0/2104      | 0.56        | 3/2715 (0.1%)   |
| 2   | B     | 0.31         | 0/2431      | 0.63        | 2/3310 (0.1%)   |
| 3   | C     | 0.30         | 0/2315      | 0.65        | 3/3143 (0.1%)   |
| 3   | D     | 0.33         | 0/2601      | 0.63        | 1/3532 (0.0%)   |
| 3   | E     | 0.34         | 0/2608      | 0.62        | 1/3540 (0.0%)   |
| 3   | F     | 0.35         | 0/2613      | 0.63        | 1/3547 (0.0%)   |
| 3   | G     | 0.34         | 0/2604      | 0.60        | 0/3533          |
| 3   | H     | 0.32         | 0/2601      | 0.61        | 1/3532 (0.0%)   |
| 4   | L     | 0.24         | 0/757       | 0.51        | 0/946           |
| 5   | M     | 0.49         | 0/1420      | 1.14        | 14/2212 (0.6%)  |
| All | All   | 0.33         | 0/22054     | 0.67        | 26/30010 (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                   | 2                   |
| 2   | B     | 0                   | 1                   |
| 3   | C     | 0                   | 1                   |
| 3   | G     | 0                   | 1                   |
| All | All   | 0                   | 5                   |

There are no bond length outliers.

The worst 5 of 26 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms     | Z     | Observed(°) | Ideal(°) |
|-----|-------|-----|------|-----------|-------|-------------|----------|
| 5   | M     | 58  | C    | N1-C2-O2  | 9.05  | 124.33      | 118.90   |
| 5   | M     | 58  | C    | C2-N1-C1' | 8.86  | 128.54      | 118.80   |
| 5   | M     | 58  | C    | C6-N1-C2  | -7.23 | 117.41      | 120.30   |
| 5   | M     | 58  | C    | N3-C2-O2  | -7.20 | 116.86      | 121.90   |
| 5   | M     | 46  | C    | C6-N1-C2  | -7.05 | 117.48      | 120.30   |

There are no chirality outliers.

All (5) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group   |
|-----|-------|-----|------|---------|
| 1   | A     | 227 | SER  | Peptide |
| 1   | A     | 263 | TRP  | Peptide |
| 2   | B     | 12  | LEU  | Peptide |
| 3   | C     | 249 | VAL  | Peptide |
| 3   | G     | 301 | LYS  | Peptide |

## 5.2 Too-close contacts

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     | 2079  | 0        | 1094     | 17      | 0            |
| 2   | B     | 2374  | 0        | 2345     | 53      | 0            |
| 3   | C     | 2272  | 0        | 2232     | 39      | 0            |
| 3   | D     | 2554  | 0        | 2522     | 40      | 0            |
| 3   | E     | 2561  | 0        | 2542     | 35      | 0            |
| 3   | F     | 2566  | 0        | 2547     | 31      | 0            |
| 3   | G     | 2557  | 0        | 2542     | 31      | 0            |
| 3   | H     | 2554  | 0        | 2522     | 38      | 0            |
| 4   | L     | 758   | 0        | 209      | 5       | 0            |
| 5   | M     | 1272  | 0        | 644      | 24      | 0            |
| All | All   | 21547 | 0        | 19199    | 261     | 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 6.

The worst 5 of 261 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:H:199:ILE:O   | 3:H:203:LEU:HB2 | 1.85                     | 0.77              |
| 3:C:148:LEU:HB2 | 3:C:151:ASN:HB3 | 1.68                     | 0.75              |
| 3:C:124:THR:O   | 3:C:128:ASP:HB2 | 1.88                     | 0.73              |
| 1:A:221:ALA:HB1 | 1:A:227:SER:HB2 | 1.71                     | 0.72              |
| 4:L:-1:MET:CB   | 4:L:183:PHE:C   | 2.58                     | 0.71              |

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     | 422/436 (97%)   | 355 (84%)  | 66 (16%)  | 1 (0%)   | 51          | 85  |
| 2   | B     | 301/329 (92%)   | 266 (88%)  | 35 (12%)  | 0        | 100         | 100 |
| 3   | C     | 287/344 (83%)   | 255 (89%)  | 32 (11%)  | 0        | 100         | 100 |
| 3   | D     | 331/344 (96%)   | 297 (90%)  | 34 (10%)  | 0        | 100         | 100 |
| 3   | E     | 332/344 (96%)   | 292 (88%)  | 40 (12%)  | 0        | 100         | 100 |
| 3   | F     | 333/344 (97%)   | 302 (91%)  | 31 (9%)   | 0        | 100         | 100 |
| 3   | G     | 331/344 (96%)   | 295 (89%)  | 34 (10%)  | 2 (1%)   | 28          | 70  |
| 3   | H     | 331/344 (96%)   | 297 (90%)  | 34 (10%)  | 0        | 100         | 100 |
| 4   | L     | 187/189 (99%)   | 143 (76%)  | 44 (24%)  | 0        | 100         | 100 |
| All | All   | 2855/3018 (95%) | 2502 (88%) | 350 (12%) | 3 (0%)   | 58          | 88  |

All (3) Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3   | G     | 320 | VAL  |
| 3   | G     | 250 | ARG  |
| 1   | A     | 34  | PRO  |

### 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     | 78/366 (21%)    | 78 (100%)   | 0        | 100         | 100 |
| 2   | B     | 245/271 (90%)   | 242 (99%)   | 3 (1%)   | 75          | 90  |
| 3   | C     | 230/274 (84%)   | 230 (100%)  | 0        | 100         | 100 |
| 3   | D     | 258/274 (94%)   | 256 (99%)   | 2 (1%)   | 85          | 94  |
| 3   | E     | 259/274 (94%)   | 259 (100%)  | 0        | 100         | 100 |
| 3   | F     | 259/274 (94%)   | 258 (100%)  | 1 (0%)   | 93          | 97  |
| 3   | G     | 259/274 (94%)   | 257 (99%)   | 2 (1%)   | 85          | 94  |
| 3   | H     | 258/274 (94%)   | 257 (100%)  | 1 (0%)   | 93          | 97  |
| All | All   | 1846/2281 (81%) | 1837 (100%) | 9 (0%)   | 91          | 96  |

5 of 9 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3   | D     | 265 | ARG  |
| 3   | H     | 314 | ARG  |
| 3   | G     | 94  | ARG  |
| 2   | B     | 286 | LEU  |
| 3   | F     | 213 | VAL  |

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3   | E     | 164 | HIS  |
| 3   | E     | 300 | GLN  |
| 3   | F     | 322 | GLN  |
| 3   | C     | 322 | GLN  |
| 3   | F     | 223 | GLN  |

### 5.3.3 RNA ⓘ

| Mol | Chain | Analysed    | Backbone Outliers | Pucker Outliers |
|-----|-------|-------------|-------------------|-----------------|
| 5   | M     | 59/60 (98%) | 38 (64%)          | 0               |

5 of 38 RNA backbone outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 5   | M     | 6   | A    |
| 5   | M     | 7   | A    |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 5   | M     | 9   | U    |
| 5   | M     | 12  | A    |
| 5   | M     | 15  | G    |

There are no RNA pucker outliers to report.

#### 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](#)

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.