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

Jun 11, 2024 – 10:08 PM EDT

| | | |
|---:|:---:|:---|
| PDB ID | : | 1JJO |
| Title | : | Crystal Structure of Mouse Neuroserpin (Cleaved form) |
| Authors | : | Briand, C.; Kozlov, S.V.; Sonderegger, P.; Gruetter, M.G. |
| Deposited on | : | 2001-07-09 |
| Resolution | : | 3.06 Å(reported) |

The following versions of software and data (see references (i)) were used in the production of this report:

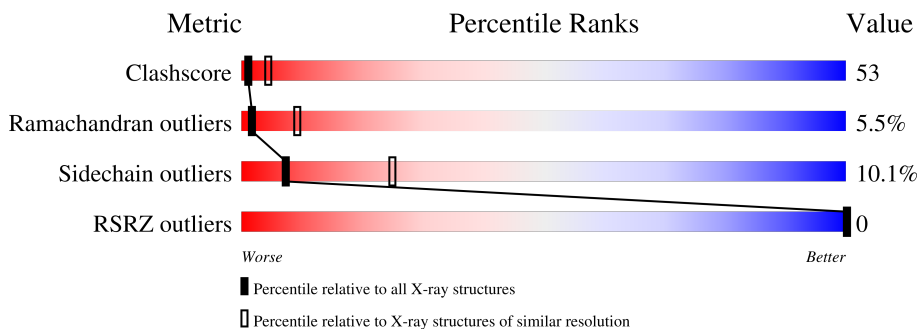| | | |
|---:|:---:|:---|
| MolProbity | : | 4.02b-467 |
| Xtriage (Phenix) | : | 1.20.1 |
| EDS | : | 2.36.2 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (Gargrove) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.36.2 |

# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.06 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 | Percentile Ranks | Value |
|--------|------------------|-------|
| Clashscore | | 53 |
| Ramachandran outliers | | 5.5% |
| Sidechain outliers | | 10.1% |
| RSRZ outliers | | 0 |

*Worse*     *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

| Metric | Whole archive (#Entries) | Similar resolution (#Entries, resolution range(Å)) |
|--------|--------------------------|------------------------------------------------------|
| Clashscore | 141614 | 1864 (3.10-3.02) |
| Ramachandran outliers | 138981 | 1794 (3.10-3.02) |
| Sidechain outliers | 138945 | 1793 (3.10-3.02) |
| RSRZ outliers | 127900 | 1713 (3.10-3.02) |

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 | 40 | 42%   48%   10% |
| 1 | B | 40 | 42%   48%   10% |
| 2 | C | 261 | 31%   54%   8% · 7% |
| 2 | D | 261 | 32%   53%   7% · 7% |
| 3 | E | 33 | 27%   61%   12% |
| 3 | F | 33 | 30%   58%   12% |

## 2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 5074 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 NEUROSERPIN.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|-----|-----|-----|-----|---------|---------|-------|
| 1 | A | 40 | Total 307 | C 193 | N 49 | O 61 | S 4 | 0 | 0 | 0 |
| 1 | B | 40 | Total 307 | C 193 | N 49 | O 61 | S 4 | 0 | 0 | 0 |

- Molecule 2 is a protein called NEUROSERPIN.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|-----|-----|-----|-----|---------|---------|-------|
| 2 | C | 244 | Total 1950 | C 1244 | N 311 | O 386 | S 9 | 0 | 0 | 0 |
| 2 | D | 244 | Total 1950 | C 1244 | N 311 | O 386 | S 9 | 0 | 0 | 0 |

- Molecule 3 is a protein called NEUROSERPIN.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|-----|-----|-----|-----|---------|---------|-------|
| 3 | E | 33 | Total 280 | C 184 | N 52 | O 42 | S 2 | 0 | 0 | 0 |
| 3 | F | 33 | Total 280 | C 184 | N 52 | O 42 | S 2 | 0 | 0 | 0 |

There are 4 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|---------|-----------|
| E | 392 | HIS | - | SEE REMARK 999 | UNP O35684 |
| E | 393 | HIS | - | SEE REMARK 999 | UNP O35684 |
| F | 392 | HIS | - | SEE REMARK 999 | UNP O35684 |
| F | 393 | HIS | - | SEE REMARK 999 | UNP O35684 |

# 3 Residue-property plots ⓘ

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.
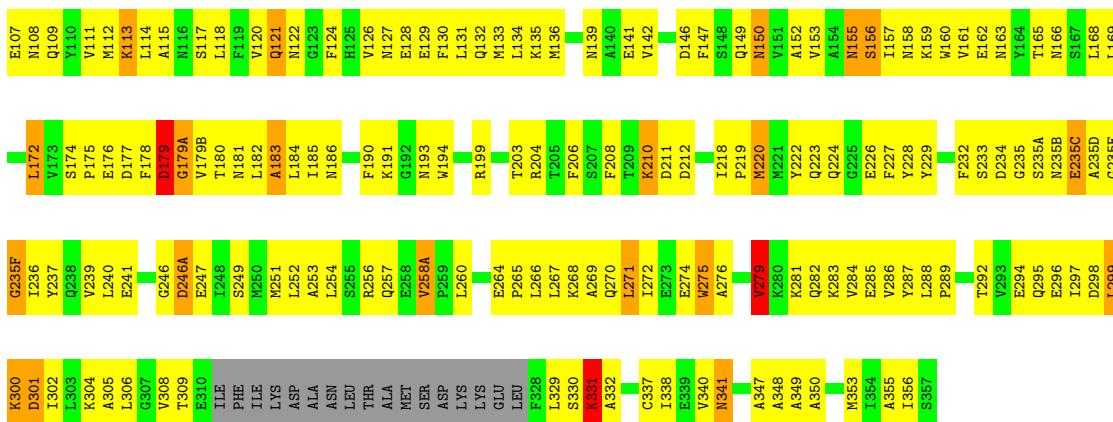
• Molecule 1: NEUROSERPIN
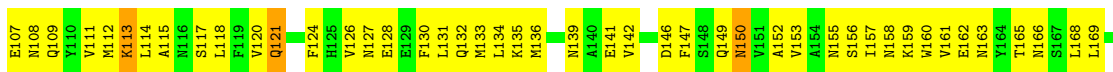
Chain A:

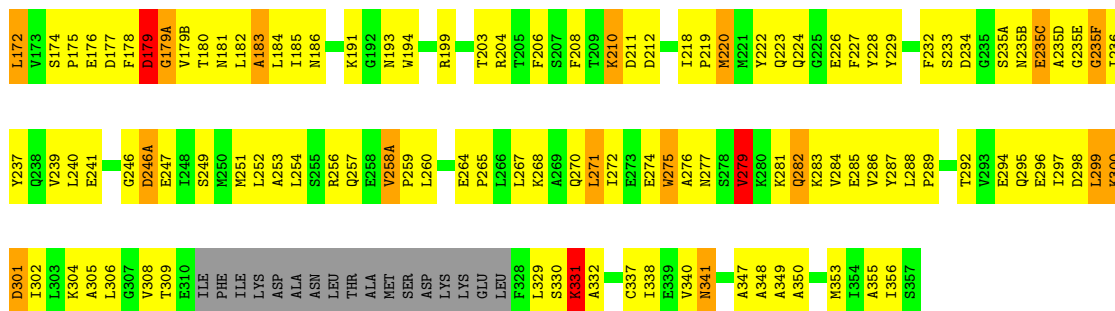• Molecule 1: NEUROSERPIN

Chain B:

• Molecule 2: NEUROSERPIN

Chain C:

• Molecule 2: NEUROSERPIN

Chain D:

L172 V173 S174 P175 E176 D177 F178 D179 G179A V179B T180 N181 L182 A183 L184 I185 N186 K191 G192 N193 W194 R199 T203 R204 T205 S207 F206 F208 T209 K210 D211 D212 I218 P219 M220 M221 Y222 Q223 Q224 G225 E226 F227 Y228 Y229 F232 S233 D234 G235 S235A N235B E235C A235D G235E G235F I236

Y237 Q238 V239 L240 E241 G246 D246A E247 I248 S249 M250 M251 L252 A253 L254 S255 R256 Q257 E258 V258A P259 L260 E264 P265 L266 L267 K268 A269 Q270 L271 I272 E273 E274 W275 A276 N277 S278 V279 K280 K281 Q282 K283 V284 E285 V286 Y287 L288 P289 T292 V293 E294 Q295 E296 I297 D298 L299 K300

D301 I302 L303 K304 A305 L306 G307 V308 T309 E310 ILE PHE ILE LYS ASP ALA ASN LEU THR ALA MET SER ASP LYS GLU LEU F328 L329 S330 K331 A332 C337 I338 E339 V340 N341 A347 A348 A349 A350 M353 I354 A355 I356 S357

• Molecule 3: NEUROSERPIN

Chain E:    27%    61%    12%

Y361 P362 Q363 V364 I365 V366 D367 H368 P369 F370 L371 Y372 L373 I374 R377 K378 S379 G380 I381 I382 G386 R387 V388 M389 N390 P391 H392 H393

• Molecule 3: NEUROSERPIN

Chain F:    30%    58%    12%

Y361 P362 Q363 V364 I365 V366 D367 H368 P369 F370 L371 Y372 L373 I374 R377 K378 S379 G380 I381 I382 G386 R387 V388 M389 N390 P391 H392 H393

# 4 Data and refinement statistics ⓘ

| Property | Value | Source |
|---|---|---|
| Space group | P 1 21 1 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 64.29Å    108.69Å    45.99Å<br>90.00°    101.28°    90.00° | Depositor |
| Resolution (Å) | 16.89   –   3.06<br>16.89   –   3.06 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 79.7 (16.89-3.06)<br>80.8 (16.89-3.06) | Depositor<br>EDS |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | 0.15 | Depositor |
| $< I/\sigma(I) >$ [1] | 2.33 (at 3.08Å) | Xtriage |
| Refinement program | CNS 0.9 | Depositor |
| R, $R_{free}$ | 0.232   ,        0.308<br>0.251   ,   (Not available) | Depositor<br>DCC |
| $R_{free}$ test set | No test flags present. | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 49.5 | Xtriage |
| Anisotropy | 0.595 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.28 , 40.4 | EDS |
| L-test for twinning[2] | $< |L| > = 0.47, < L^2 > = 0.30$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| $F_o,F_c$ correlation | 0.90 | EDS |
| Total number of atoms | 5074 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 28.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.25% 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 ⓘ

### 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 | 0.42 | 0/312 | 0.64 | 0/420 |
| 1 | B | 0.42 | 0/312 | 0.64 | 0/420 |
| 2 | C | 0.46 | 1/1988 (0.1%) | 0.70 | 1/2687 (0.0%) |
| 2 | D | 0.46 | 1/1988 (0.1%) | 0.70 | 1/2687 (0.0%) |
| 3 | E | 0.45 | 0/289 | 0.76 | 1/389 (0.3%) |
| 3 | F | 0.45 | 0/289 | 0.76 | 1/389 (0.3%) |
| All | All | 0.45 | 2/5178 (0.0%) | 0.70 | 4/6992 (0.1%) |

All (2) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|-----|-------|--------|------|-------|------|-------------|----------|
| 2 | D | 179(A) | GLY | CA-C | 5.35 | 1.60 | 1.51 |
| 2 | C | 179(A) | GLY | CA-C | 5.29 | 1.60 | 1.51 |

All (4) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|-----|-------|-----|------|-------|-------|----------------|-------------|
| 2 | D | 179 | ASP | O-C-N | -6.56 | 112.05 | 123.20 |
| 2 | C | 179 | ASP | O-C-N | -6.53 | 112.09 | 123.20 |
| 3 | E | 374 | ILE | N-CA-C | -5.07 | 97.30 | 111.00 |
| 3 | F | 374 | ILE | N-CA-C | -5.07 | 97.32 | 111.00 |

There are no chirality outliers.

There are no planarity outliers.

### 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 | 307 | 0 | 296 | 60 | 0 |
| 1 | B | 307 | 0 | 296 | 67 | 0 |
| 2 | C | 1950 | 0 | 1887 | 239 | 5 |
| 2 | D | 1950 | 0 | 1887 | 223 | 3 |
| 3 | E | 280 | 0 | 286 | 37 | 2 |
| 3 | F | 280 | 0 | 286 | 37 | 0 |
| All | All | 5074 | 0 | 4938 | 529 | 6 |

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 53.

The worst 5 of 529 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) |
|--------|--------|--------------------------|-------------------|
| 2:C:136:MET:CE | 1:B:55:LEU:HD12 | 1.10 | 1.55 |
| 2:C:136:MET:CE | 1:B:55:LEU:CD1 | 1.90 | 1.42 |
| 2:C:136:MET:HE2 | 1:B:55:LEU:CD1 | 1.46 | 1.41 |
| 1:A:55:LEU:HD12 | 2:D:136:MET:CE | 1.50 | 1.39 |
| 1:A:30:ILE:CB | 2:D:132:GLN:HG2 | 1.79 | 1.12 |

The worst 5 of 6 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:C:235(B):ASN:ND2 | 2:D:309:THR:OG1[2_655] | 1.82 | 0.38 |
| 2:C:109:GLN:NE2 | 2:C:309:THR:OG1[1_556] | 1.95 | 0.25 |
| 3:E:392:HIS:CD2 | 2:D:277:ASN:O[2_555] | 1.98 | 0.22 |
| 2:C:235:GLY:O | 2:D:304:LYS:CE[2_655] | 2.09 | 0.11 |
| 2:C:109:GLN:CG | 2:C:304:LYS:NZ[1_556] | 2.13 | 0.07 |

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

| Mol | Chain | Analysed | Favoured | Allowed | Outliers | Percentiles | |
|-----|-------|----------|----------|---------|----------|-----|-----|
| 1 | A | 38/40 (95%) | 32 (84%) | 4 (10%) | 2 (5%) | 2 | 10 |
| 1 | B | 38/40 (95%) | 32 (84%) | 4 (10%) | 2 (5%) | 2 | 10 |
| 2 | C | 240/261 (92%) | 189 (79%) | 38 (16%) | 13 (5%) | 2 | 9 |
| 2 | D | 240/261 (92%) | 189 (79%) | 38 (16%) | 13 (5%) | 2 | 9 |
| 3 | E | 31/33 (94%) | 26 (84%) | 3 (10%) | 2 (6%) | 1 | 6 |
| 3 | F | 31/33 (94%) | 26 (84%) | 3 (10%) | 2 (6%) | 1 | 6 |
| All | All | 618/668 (92%) | 494 (80%) | 90 (15%) | 34 (6%) | 2 | 9 |

5 of 34 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | A | 30 | ILE |
| 2 | C | 235(C) | GLU |
| 2 | C | 299 | LEU |
| 2 | C | 300 | LYS |
| 1 | B | 30 | ILE |

### 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 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 | 33/33 (100%) | 29 (88%) | 4 (12%) | 5 | 17 |
| 1 | B | 33/33 (100%) | 29 (88%) | 4 (12%) | 5 | 17 |
| 2 | C | 214/229 (93%) | 192 (90%) | 22 (10%) | 7 | 24 |
| 2 | D | 214/229 (93%) | 192 (90%) | 22 (10%) | 7 | 24 |
| 3 | E | 31/31 (100%) | 29 (94%) | 2 (6%) | 17 | 44 |
| 3 | F | 31/31 (100%) | 29 (94%) | 2 (6%) | 17 | 44 |
| All | All | 556/586 (95%) | 500 (90%) | 56 (10%) | 7 | 25 |

5 of 56 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | B | 39 | ASN |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | F | 377 | ARG |
| 2 | D | 166 | ASN |
| 3 | F | 364 | VAL |
| 2 | D | 284 | VAL |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 3 | E | 390 | ASN |
| 3 | F | 390 | ASN |
| 2 | D | 139 | ASN |
| 2 | D | 270 | GLN |
| 2 | D | 121 | GLN |

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

# 6 Fit of model and data ⓘ

## 6.1 Protein, DNA and RNA chains ⓘ

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(Å²) | Q<0.9 |
|-----|-------|----------|--------|---|-----|-----|----------|-------|
| 1 | A | 40/40 (100%) | -0.47 | 0 | 100 | 100 | 16, 29, 50, 51 | 0 |
| 1 | B | 40/40 (100%) | -0.50 | 0 | 100 | 100 | 16, 29, 50, 51 | 0 |
| 2 | C | 244/261 (93%) | -0.49 | 0 | 100 | 100 | 9, 27, 51, 60 | 0 |
| 2 | D | 244/261 (93%) | -0.47 | 0 | 100 | 100 | 9, 27, 51, 60 | 0 |
| 3 | E | 33/33 (100%) | -0.70 | 0 | 100 | 100 | 11, 20, 33, 47 | 0 |
| 3 | F | 33/33 (100%) | -0.77 | 0 | 100 | 100 | 11, 20, 33, 47 | 0 |
| All | All | 634/668 (94%) | -0.51 | 0 | 100 | 100 | 9, 26, 51, 60 | 0 |

There are no RSRZ outliers to report.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 6.4 Ligands ⓘ

There are no ligands in this entry.

## 6.5 Other polymers ⓘ

There are no such residues in this entry.