



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 05:43 AM GMT

PDB ID : 2PY9  
Title : Protein-RNA Interaction involving KH1 domain from Human Poly(C)-Binding Protein-2  
Authors : James, T.L.; Du, Z.; Lee, J.K.  
Deposited on : 2007-05-15  
Resolution : 2.56 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

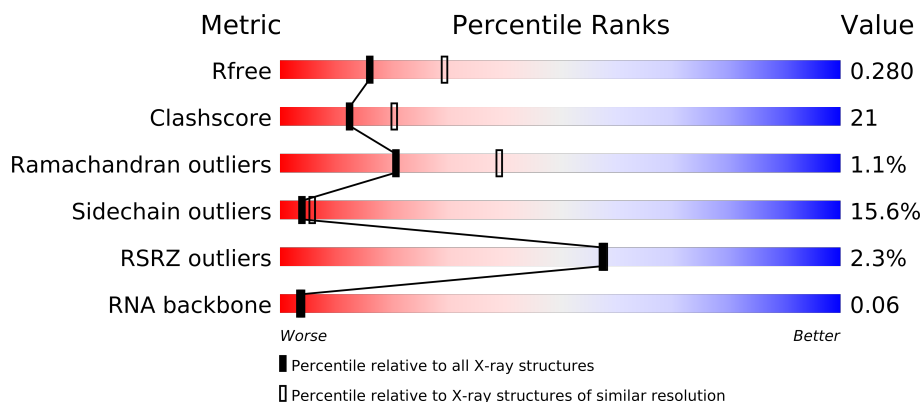
The following versions of software and data (see [references](#)) were used in the production of this report:

|                                |   |                          |
|--------------------------------|---|--------------------------|
| MolProbity                     | : | 4.02b-467                |
| Mogul                          | : | 1.15 2013                |
| Xtriage (Phenix)               | : | dev-1323                 |
| EDS                            | : | stable22639              |
| Percentile statistics          | : | 21963                    |
| Refmac                         | : | 5.8.0049                 |
| CCP4                           | : | 6.3.0 (Settle)           |
| Ideal geometry (proteins)      | : | Engh & Huber (2001)      |
| Ideal geometry (DNA, RNA)      | : | Parkinson et. al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | stable22683              |

# 1 Overall quality at a glance

The reported resolution of this entry is 2.56 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) | Similar resolution<br>(#Entries, resolution range(Å)) |
|-----------------------|-----------------------------|---|
| $R_{free}$            | 66092                       | 2347 (2.60-2.52)                                      |
| Clashscore            | 79885                       | 2876 (2.60-2.52)                                      |
| Ramachandran outliers | 78287                       | 2826 (2.60-2.52)                                      |
| Sidechain outliers    | 78261                       | 2826 (2.60-2.52)                                      |
| RSRZ outliers         | 66119                       | 2347 (2.60-2.52)                                      |
| RNA backbone          | 1838                        | 1008 (3.10-2.02)                                      |

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

| Mol | Chain | Length | Quality of chain |
|-----|-------|--------|------------------|
| 1   | E     | 12     |                  |
| 1   | F     | 12     |                  |
| 2   | A     | 73     |                  |
| 2   | B     | 73     |                  |
| 2   | C     | 73     |                  |
| 2   | D     | 73     |                  |

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2669 atoms, of which 0 are hydrogen and 0 are deuterium.

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 RNA chain called 12-mer C-rich strand of human telomeric RNA.

| Mol | Chain | Residues | Atoms |     |    |    |    | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|-----|----|----|----|---------|---------|-------|
| 1   | E     | 12       | Total | C   | N  | O  | P  | 0       | 0       | 0     |
|     |       |          | 245   | 112 | 42 | 80 | 11 |         |         |       |
| 1   | F     | 12       | Total | C   | N  | O  | P  | 0       | 0       | 0     |
|     |       |          | 245   | 112 | 42 | 80 | 11 |         |         |       |

- Molecule 2 is a protein called Poly(rC)-binding protein 2.

| Mol | Chain | Residues | Atoms |     |    |     |   |    | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|-----|----|-----|---|----|---------|---------|-------|
| 2   | A     | 72       | Total | C   | N  | O   | S | Se | 0       | 0       | 0     |
|     |       |          | 550   | 346 | 98 | 102 | 1 | 3  |         |         |       |
| 2   | B     | 69       | Total | C   | N  | O   | S | Se | 0       | 0       | 0     |
|     |       |          | 526   | 331 | 93 | 98  | 1 | 3  |         |         |       |
| 2   | C     | 69       | Total | C   | N  | O   | S | Se | 0       | 0       | 0     |
|     |       |          | 526   | 331 | 93 | 98  | 1 | 3  |         |         |       |
| 2   | D     | 69       | Total | C   | N  | O   | S | Se | 0       | 0       | 0     |
|     |       |          | 526   | 331 | 93 | 98  | 1 | 3  |         |         |       |

There are 16 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment          | Reference  |
|-------|---------|----------|--------|------------------|------------|
| A     | 10      | LYS      | -      | CLONING ARTIFACT | UNP Q15366 |
| A     | 20      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| A     | 39      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| A     | 74      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| B     | 10      | LYS      | -      | CLONING ARTIFACT | UNP Q15366 |
| B     | 20      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| B     | 39      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| B     | 74      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| C     | 10      | LYS      | -      | CLONING ARTIFACT | UNP Q15366 |
| C     | 20      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| C     | 39      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| C     | 74      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| D     | 10      | LYS      | -      | CLONING ARTIFACT | UNP Q15366 |

*Continued on next page...*

*Continued from previous page...*

| Chain | Residue | Modelled | Actual | Comment          | Reference  |
|-------|---------|----------|--------|------------------|------------|
| D     | 20      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| D     | 39      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |
| D     | 74      | MSE      | MET    | MODIFIED RESIDUE | UNP Q15366 |

- Molecule 3 is water.

| Mol | Chain | Residues | Atoms            | ZeroOcc | AltConf |
|-----|-------|----------|------------------|---------|---------|
| 3   | A     | 9        | Total O<br>9 9   | 0       | 0       |
| 3   | B     | 14       | Total O<br>14 14 | 0       | 0       |
| 3   | C     | 11       | Total O<br>11 11 | 0       | 0       |
| 3   | D     | 10       | Total O<br>10 10 | 0       | 0       |
| 3   | E     | 3        | Total O<br>3 3   | 0       | 0       |
| 3   | F     | 4        | Total O<br>4 4   | 0       | 0       |

### 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 errors 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.

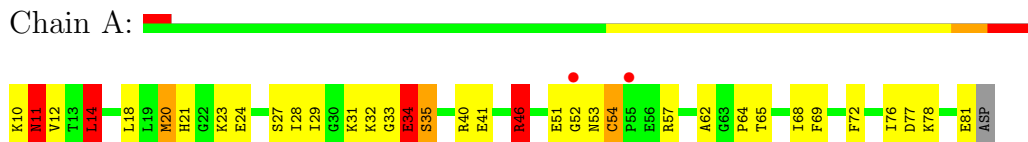
- Molecule 1: 12-mer C-rich strand of human telomeric RNA



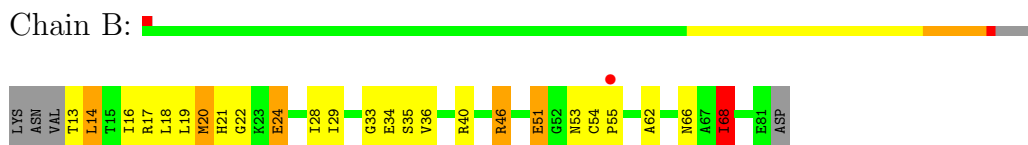
- Molecule 1: 12-mer C-rich strand of human telomeric RNA



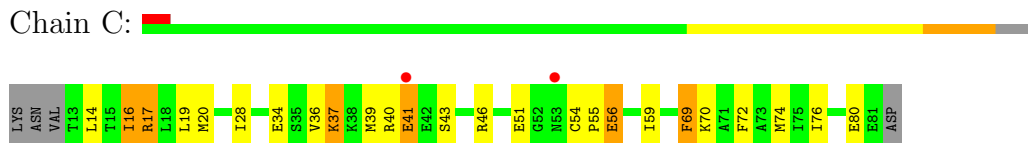
- Molecule 2: Poly(rC)-binding protein 2



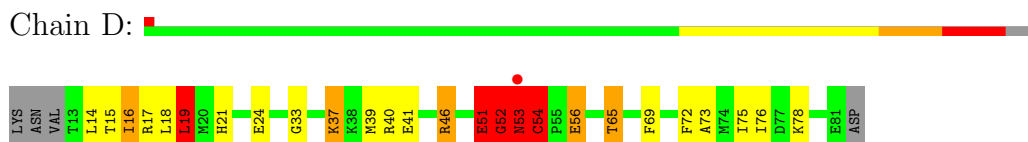
- Molecule 2: Poly(rC)-binding protein 2



- Molecule 2: Poly(rC)-binding protein 2



- Molecule 2: Poly(rC)-binding protein 2



## 4 Data and refinement statistics

| Property  | Value   | Source           |
|---|---|------------------|
| Space group   | P 21 21 2   | Depositor        |
| Cell constants<br>a, b, c, $\alpha$ , $\beta$ , $\gamma$                | 92.01Å 58.70Å 71.50Å<br>90.00° 90.00° 90.00°                | Depositor        |
| Resolution (Å)  | 40.00 – 2.56<br>71.50 – 2.56                                | Depositor<br>EDS |
| % Data completeness<br>(in resolution range)                            | 94.2 (40.00-2.56)<br>94.2 (71.50-2.56)                      | Depositor<br>EDS |
| $R_{merge}$   | 0.10  | Depositor        |
| $R_{sym}$   | 0.10  | Depositor        |
| $\langle I/\sigma(I) \rangle$ <sup>1</sup>                              | 3.11 (at 2.55Å)   | Xtriage          |
| Refinement program  | REFMAC 5.2.0019   | Depositor        |
| R, $R_{free}$   | 0.218 , 0.269<br>0.238 , 0.280                              | Depositor<br>DCC |
| $R_{free}$ test set   | 591 reflections (5.07%)                                     | DCC              |
| Wilson B-factor (Å <sup>2</sup> )                                       | 52.1  | Xtriage          |
| Anisotropy  | 0.279   | Xtriage          |
| Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> ) | 0.32 , 17.3   | EDS              |
| Estimated twinning fraction   | No twinning to report.                                      | Xtriage          |
| L-test for twinning   | $\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$ | Xtriage          |
| Outliers  | 0 of 12252 reflections                                      | Xtriage          |
| $F_o, F_c$ correlation  | 0.93  | EDS              |
| Total number of atoms   | 2669  | wwPDB-VP         |
| Average B, all atoms (Å <sup>2</sup> )                                  | 38.0  | wwPDB-VP         |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.59% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 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   | E     | 2.45         | 12/272 (4.4%)  | 4.09        | 88/420 (21.0%)  |
| 1   | F     | 2.48         | 17/272 (6.2%)  | 3.83        | 72/420 (17.1%)  |
| 2   | A     | 1.66         | 8/551 (1.5%)   | 1.46        | 10/729 (1.4%)   |
| 2   | B     | 1.48         | 3/527 (0.6%)   | 1.53        | 6/697 (0.9%)    |
| 2   | C     | 1.41         | 5/527 (0.9%)   | 1.49        | 8/697 (1.1%)    |
| 2   | D     | 1.52         | 6/527 (1.1%)   | 1.49        | 7/697 (1.0%)    |
| All | All   | 1.75         | 51/2676 (1.9%) | 2.31        | 191/3660 (5.2%) |

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 |
|-----|-------|---------------------|---------------------|
| 2   | D     | 0                   | 2                   |

The worst 5 of 51 bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z      | Observed(Å) | Ideal(Å) |
|-----|-------|-----|------|-------|--------|-------------|----------|
| 2   | A     | 54  | CYS  | CB-SG | -10.44 | 1.64        | 1.82     |
| 2   | D     | 51  | GLU  | CB-CG | -9.70  | 1.33        | 1.52     |
| 1   | F     | 8   | A    | N9-C4 | 9.40   | 1.43        | 1.37     |
| 2   | A     | 41  | GLU  | CG-CD | 8.82   | 1.65        | 1.51     |
| 1   | E     | 2   | A    | P-O5' | 8.63   | 1.68        | 1.59     |

The worst 5 of 191 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms       | Z      | Observed(°) | Ideal(°) |
|-----|-------|-----|------|-------------|--------|-------------|----------|
| 1   | E     | 10  | C    | O5'-P-OP2   | -18.63 | 88.34       | 110.70   |
| 1   | F     | 4   | C    | O4'-C1'-N1  | 15.28  | 120.42      | 108.20   |
| 1   | E     | 10  | C    | O4'-C1'-C2' | -14.44 | 91.36       | 105.80   |
| 2   | B     | 46  | ARG  | NE-CZ-NH1   | 13.51  | 127.05      | 120.30   |
| 1   | F     | 11  | C    | O4'-C1'-N1  | 13.32  | 118.86      | 108.20   |

There are no chirality outliers.

All (2) planarity outliers are listed below:

| Mol | Chain | Res | Type | Group   |
|-----|-------|-----|------|---------|
| 2   | D     | 52  | GLY  | Peptide |
| 2   | D     | 53  | ASN  | Peptide |

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|-----|-------|-------|----------|----------|---------|--------------|
| 1   | E     | 245   | 0        | 132      | 10      | 0            |
| 1   | F     | 245   | 0        | 132      | 9       | 0            |
| 2   | A     | 550   | 0        | 588      | 30      | 0            |
| 2   | B     | 526   | 0        | 560      | 21      | 0            |
| 2   | C     | 526   | 0        | 560      | 27      | 0            |
| 2   | D     | 526   | 0        | 560      | 24      | 0            |
| 3   | A     | 9     | 0        | 0        | 3       | 0            |
| 3   | B     | 14    | 0        | 0        | 1       | 0            |
| 3   | C     | 11    | 0        | 0        | 3       | 0            |
| 3   | D     | 10    | 0        | 0        | 2       | 0            |
| 3   | E     | 3     | 0        | 0        | 0       | 0            |
| 3   | F     | 4     | 0        | 0        | 0       | 0            |
| All | All   | 2669  | 0        | 2532     | 109     | 0            |

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 21.

The worst 5 of 109 close contacts within the same asymmetric unit are listed below.

| Atom-1          | Atom-2          | Distance(Å) | Clash(Å) |
|-----------------|-----------------|-------------|----------|
| 2:A:46:ARG:CG   | 2:A:46:ARG:HH11 | 1.64        | 1.10     |
| 2:A:46:ARG:HG2  | 2:A:46:ARG:HH11 | 0.87        | 1.03     |
| 2:C:56:GLU:HB3  | 3:C:86:HOH:O    | 1.57        | 1.02     |
| 2:A:46:ARG:HG2  | 2:A:46:ARG:NH1  | 1.65        | 1.00     |
| 2:A:18:LEU:HD23 | 2:B:18:LEU:HD23 | 1.45        | 0.97     |

There are no symmetry-related clashes.



## 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.

| Mol | Chain | Analysed      | Favoured  | Allowed | Outliers | Percentiles |     |
|-----|-------|---------------|-----------|---------|----------|-------------|-----|
| 2   | A     | 70/73 (96%)   | 63 (90%)  | 5 (7%)  | 2 (3%)   | 7           | 10  |
| 2   | B     | 67/73 (92%)   | 64 (96%)  | 3 (4%)  | 0        | 100         | 100 |
| 2   | C     | 67/73 (92%)   | 63 (94%)  | 4 (6%)  | 0        | 100         | 100 |
| 2   | D     | 67/73 (92%)   | 60 (90%)  | 6 (9%)  | 1 (2%)   | 15          | 28  |
| All | All   | 271/292 (93%) | 250 (92%) | 18 (7%) | 3 (1%)   | 21          | 39  |

All (3) Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 2   | D     | 52  | GLY  |
| 2   | A     | 11  | ASN  |
| 2   | A     | 65  | THR  |

### 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 |    |
|-----|-------|----------------|-----------|----------|-------------|----|
| 2   | A     | 60/58 (103%)   | 51 (85%)  | 9 (15%)  | 4           | 6  |
| 2   | B     | 57/58 (98%)    | 47 (82%)  | 10 (18%) | 3           | 4  |
| 2   | C     | 57/58 (98%)    | 51 (90%)  | 6 (10%)  | 10          | 17 |
| 2   | D     | 57/58 (98%)    | 46 (81%)  | 11 (19%) | 2           | 3  |
| All | All   | 231/232 (100%) | 195 (84%) | 36 (16%) | 4           | 6  |

5 of 36 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 2   | B     | 53  | ASN  |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 2   | C     | 17  | ARG  |
| 2   | D     | 53  | ASN  |
| 2   | B     | 68  | ILE  |
| 2   | C     | 37  | LYS  |

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 2   | A     | 11  | ASN  |
| 2   | B     | 53  | ASN  |
| 2   | B     | 66  | ASN  |
| 2   | D     | 53  | ASN  |

### 5.3.3 RNA ⓘ

| Mol | Chain | Analysed     | Backbone Outliers | Pucker Outliers |
|-----|-------|--------------|-------------------|-----------------|
| 1   | E     | 12/12 (100%) | 8 (66%)           | 3 (25%)         |
| 1   | F     | 12/12 (100%) | 5 (41%)           | 2 (16%)         |
| All | All   | 24/24 (100%) | 13 (54%)          | 5 (20%)         |

5 of 13 RNA backbone outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1   | E     | 2   | A    |
| 1   | E     | 3   | C    |
| 1   | E     | 4   | C    |
| 1   | E     | 5   | C    |
| 1   | E     | 7   | A    |

All (5) RNA pucker outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1   | E     | 1   | A    |
| 1   | E     | 2   | A    |
| 1   | E     | 3   | C    |
| 1   | F     | 1   | A    |
| 1   | F     | 9   | C    |

## 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 carbohydrates 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<sup>th</sup> 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(Å <sup>2</sup> ) | Q<0.9 |
|-----|-------|---------------|--------|--------------|-----------------------|-------|
| 1   | E     | 12/12 (100%)  | 0.53   | 1 (8%) 11 10 | 39, 52, 76, 81        | 0     |
| 1   | F     | 12/12 (100%)  | 0.29   | 0 100 100    | 44, 48, 63, 82        | 0     |
| 2   | A     | 72/73 (98%)   | 0.08   | 2 (2%) 50 51 | 18, 31, 51, 56        | 0     |
| 2   | B     | 69/73 (94%)   | 0.00   | 1 (1%) 72 74 | 19, 31, 46, 60        | 0     |
| 2   | C     | 69/73 (94%)   | 0.09   | 2 (2%) 49 49 | 21, 33, 49, 57        | 0     |
| 2   | D     | 69/73 (94%)   | 0.13   | 1 (1%) 72 74 | 23, 34, 49, 63        | 0     |
| All | All   | 303/316 (95%) | 0.10   | 7 (2%) 57 58 | 18, 33, 53, 82        | 0     |

The worst 5 of 7 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|-----|-------|-----|------|------|
| 2   | D     | 53  | ASN  | 6.4  |
| 1   | E     | 1   | A    | 3.2  |
| 2   | C     | 53  | ASN  | 2.8  |
| 2   | B     | 55  | PRO  | 2.3  |
| 2   | C     | 41  | GLU  | 2.2  |

### 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 carbohydrates 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.