# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 20, 2024 – 08:27 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 9CQI |
| Title | : | CRYSTAL STRUCTURE OF GAGA-DOG HSP47(36-418) IN COMPLEX WITH ADNECTIN-44 |
| Authors | : | Sheriff, S. |
| Deposited on | : | 2024-07-19 |
| Resolution | : | 1.94 Å(reported) |

---

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

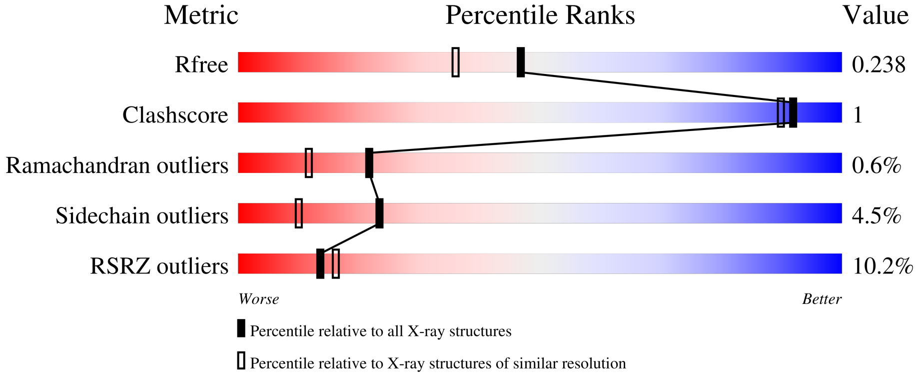
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Xtriage (Phenix) | : | 1.20.1 |
| EDS | : | 3.0 |
| Percentile statistics | : | 20231227.v01 (using entries in the PDB archive December 27th 2023) |
| CCP4 | : | 9.0.003 (Gargrove) |
| Density-Fitness | : | 1.0.11 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.39 |

# 1 Overall quality at a glance

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

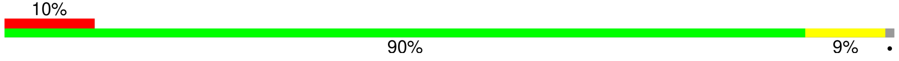
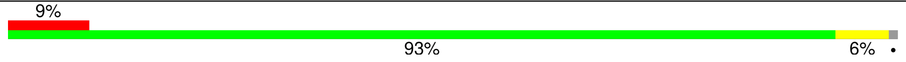
The reported resolution of this entry is 1.94 Å.

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) | Similar resolution (#Entries, resolution range(Å)) |
|---|---|---|
| $R_{free}$ | 164625 | 1306 (1.94-1.94) |
| Clashscore | 180529 | 1400 (1.94-1.94) |
| Ramachandran outliers | 177936 | 1387 (1.94-1.94) |
| Sidechain outliers | 177891 | 1387 (1.94-1.94) |
| RSRZ outliers | 164620 | 1306 (1.94-1.94) |

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 | 387 |  |
| 2 | B | 107 |  |

# 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7686 atoms, of which 3715 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 Serpin H1.

| Mol | Chain | Residues | Atoms | | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | S | | | |
| 1 | A | 384 | 5918 | 1893 | 2948 | 511 | 553 | 13 | 2948 | 1 | 0 |

There are 4 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 32 | GLY | - | expression tag | UNP C7C419 |
| A | 33 | ALA | - | expression tag | UNP C7C419 |
| A | 34 | GLY | - | expression tag | UNP C7C419 |
| A | 35 | ALA | - | expression tag | UNP C7C419 |

- Molecule 2 is a protein called anti-HSP47 Adnectin-44.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | H | N | O | | | |
| 2 | B | 106 | 1598 | 536 | 767 | 133 | 162 | 767 | 0 | 0 |

- Molecule 3 is water.

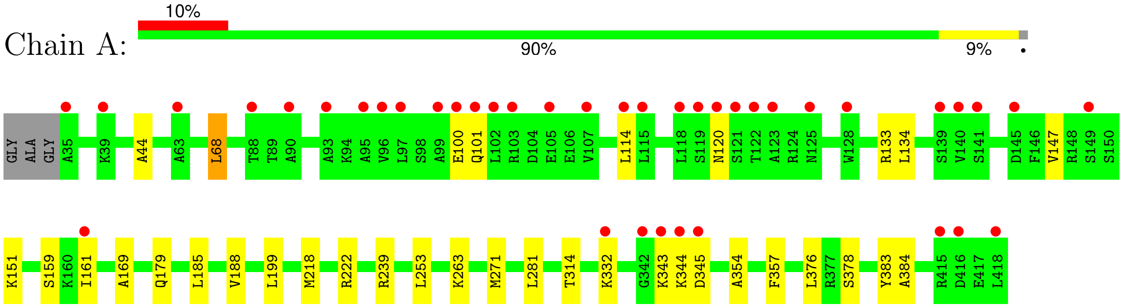
| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | O | | |
| 3 | A | 120 | 120 | 120 | 0 | 0 |
| | | | Total | O | | |
| 3 | B | 50 | 50 | 50 | 0 | 0 |

# 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.

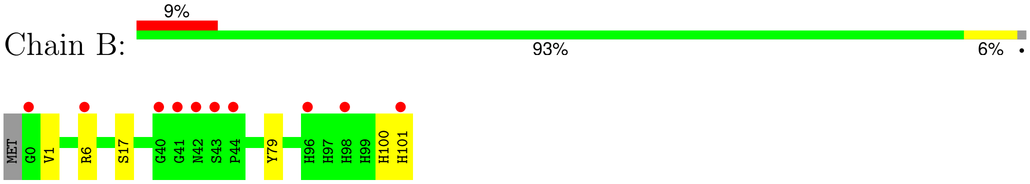
- Molecule 1: Serpin H1

Chain A:



- Molecule 2: anti-HSP47 Adnectin-44

Chain B:



# 4 Data and refinement statistics ⓘ

| Property | Value | Source |
|---|---|---|
| Space group | C 1 2 1 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 94.33Å 82.52Å 98.32Å<br>90.00° 116.14° 90.00° | Depositor |
| Resolution (Å) | 26.71 – 1.94<br>26.71 – 1.94 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 72.9 (26.71-1.94)<br>72.9 (26.71-1.94) | Depositor<br>EDS |
| $R_{merge}$ | 0.09 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 1.30 (at 1.95Å) | Xtriage |
| Refinement program | BUSTER 2.11.8 | Depositor |
| R, $R_{free}$ | 0.214 , 0.245<br>0.206 , 0.238 | Depositor<br>DCC |
| $R_{free}$ test set | 1833 reflections (5.02%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 29.0 | Xtriage |
| Anisotropy | 0.050 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.40 , 38.6 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.50$, $< L^2 > = 0.34$ | Xtriage |
| Estimated twinning fraction | 0.014 for h,-k,-h-l | Xtriage |
| $F_o$,$F_c$ correlation | 0.95 | EDS |
| Total number of atoms | 7686 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 34.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.05% 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 (i)

## 5.1 Standard geometry (i)

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 | 0.43 | 0/3031 | 0.61 | 0/4096 |
| 2 | B | 0.43 | 0/863 | 0.65 | 0/1187 |
| All | All | 0.43 | 0/3894 | 0.62 | 0/5283 |

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

## 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 | 2970 | 2948 | 2949 | 11 | 1 |
| 2 | B | 831 | 767 | 770 | 1 | 0 |
| 3 | A | 120 | 0 | 0 | 0 | 0 |
| 3 | B | 50 | 0 | 0 | 0 | 0 |
| All | All | 3971 | 3715 | 3719 | 11 | 1 |

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 1.

All (11) 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:A:44:ALA:HB2 | 1:A:114:LEU:HD21 | 1.73 | 0.71 |

*Continued on next page...*

*Continued from previous page...*

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|---|---|---|---|
| 1:A:188:VAL:O | 1:A:314:THR:HG21 | 2.07 | 0.55 |
| 1:A:222:ARG:HD2 | 1:A:383:TYR:OH | 2.08 | 0.53 |
| 1:A:185:LEU:HD21 | 1:A:354:ALA:HB1 | 1.91 | 0.53 |
| 1:A:253:LEU:HD22 | 1:A:281:LEU:HD11 | 1.98 | 0.46 |
| 1:A:271:MET:SD | 1:A:384:ALA:HA | 2.55 | 0.46 |
| 1:A:68:LEU:HD11 | 1:A:357:PHE:HB2 | 1.99 | 0.45 |
| 1:A:218:MET:HG3 | 2:B:79:TYR:O | 2.17 | 0.45 |
| 1:A:185:LEU:HD21 | 1:A:354:ALA:CB | 2.47 | 0.44 |
| 1:A:147:VAL:O | 1:A:151:LYS:HB2 | 2.19 | 0.42 |
| 1:A:161:ILE:HD11 | 1:A:169:ALA:HA | 2.03 | 0.41 |

All (1) 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) |
|---|---|---|---|
| 1:A:263:LYS:NZ | 1:A:263:LYS:HZ2[2_556] | 1.50 | 0.10 |

## 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 | |
|---|---|---|---|---|---|---|---|
| 1 | A | 383/387 (99%) | 370 (97%) | 11 (3%) | 2 (0%) | 25 | 15 |
| 2 | B | 104/107 (97%) | 100 (96%) | 3 (3%) | 1 (1%) | 13 | 5 |
| All | All | 487/494 (99%) | 470 (96%) | 14 (3%) | 3 (1%) | 22 | 12 |

All (3) Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 2 | B | 1 | VAL |
| 1 | A | 344 | LYS |
| 1 | A | 120 | ASN |

### 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 | 312/328 (95%) | 298 (96%) | 14 (4%) | 23 | 10 |
| 2 | B | 90/95 (95%) | 86 (96%) | 4 (4%) | 24 | 11 |
| All | All | 402/423 (95%) | 384 (96%) | 18 (4%) | 23 | 10 |

All (18) residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 68 | LEU |
| 1 | A | 100 | GLU |
| 1 | A | 101 | GLN |
| 1 | A | 133 | ARG |
| 1 | A | 134 | LEU |
| 1 | A | 159 | SER |
| 1 | A | 179 | GLN |
| 1 | A | 199 | LEU |
| 1 | A | 239 | ARG |
| 1 | A | 332 | LYS |
| 1 | A | 343 | LYS |
| 1 | A | 345 | ASP |
| 1 | A | 376 | LEU |
| 1 | A | 378 | SER |
| 2 | B | 6 | ARG |
| 2 | B | 17 | SER |
| 2 | B | 100 | HIS |
| 2 | B | 101 | HIS |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 oligosaccharides 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(Å$^2$) | Q<0.9 |
|---|---|---|---|---|---|---|---|---|
| 1 | A | 384/387 (99%) | 0.54 | 40 (10%) | 13 | 15 | 7, 17, 35, 50 | 1 (0%) |
| 2 | B | 106/107 (99%) | 0.24 | 10 (9%) | 15 | 18 | 8, 14, 29, 40 | 0 |
| All | All | 490/494 (99%) | 0.47 | 50 (10%) | 13 | 16 | 7, 17, 33, 50 | 1 (0%) |

All (50) RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 1 | A | 121 | SER | 6.1 |
| 1 | A | 118 | LEU | 5.6 |
| 1 | A | 102 | LEU | 4.7 |
| 1 | A | 122 | THR | 4.6 |
| 1 | A | 345 | ASP | 4.6 |
| 1 | A | 416 | ASP | 4.5 |
| 1 | A | 128 | TRP | 4.5 |
| 1 | A | 120 | ASN | 4.5 |
| 2 | B | 42 | ASN | 4.2 |
| 1 | A | 343 | LYS | 4.2 |
| 2 | B | 44 | PRO | 4.0 |
| 1 | A | 97 | LEU | 3.9 |
| 1 | A | 96 | VAL | 3.8 |
| 1 | A | 140 | VAL | 3.8 |
| 1 | A | 344 | LYS | 3.5 |
| 1 | A | 123 | ALA | 3.5 |
| 1 | A | 119 | SER | 3.4 |
| 1 | A | 145 | ASP | 3.4 |
| 1 | A | 101 | GLN | 3.4 |
| 1 | A | 342 | GLY | 3.3 |
| 1 | A | 63 | ALA | 3.3 |
| 1 | A | 35 | ALA | 3.2 |
| 1 | A | 139 | SER | 3.0 |
| 1 | A | 418 | LEU | 3.0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type | RSRZ |
|---|---|---|---|---|
| 1 | A | 415 | ARG | 3.0 |
| 2 | B | 0 | GLY | 2.9 |
| 1 | A | 114 | LEU | 2.9 |
| 2 | B | 43 | SER | 2.8 |
| 1 | A | 99 | ALA | 2.7 |
| 2 | B | 41 | GLY | 2.7 |
| 2 | B | 98 | HIS | 2.7 |
| 1 | A | 103 | ARG | 2.6 |
| 1 | A | 100 | GLU | 2.6 |
| 1 | A | 105 | GLU | 2.5 |
| 1 | A | 149 | SER | 2.5 |
| 1 | A | 88 | THR | 2.5 |
| 1 | A | 161 | ILE | 2.5 |
| 1 | A | 95 | ALA | 2.4 |
| 2 | B | 40 | GLY | 2.3 |
| 1 | A | 141 | SER | 2.3 |
| 1 | A | 332 | LYS | 2.2 |
| 1 | A | 90 | ALA | 2.2 |
| 2 | B | 96 | HIS | 2.2 |
| 1 | A | 93 | ALA | 2.1 |
| 2 | B | 101 | HIS | 2.1 |
| 1 | A | 39 | LYS | 2.1 |
| 2 | B | 6 | ARG | 2.1 |
| 1 | A | 107 | VAL | 2.0 |
| 1 | A | 115 | LEU | 2.0 |
| 1 | A | 125 | ASN | 2.0 |

## 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.