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

Sep 14, 2023 – 03:36 PM EDT

| | | |
|---|---|---|
| PDB ID | : | 2FS3 |
| Title | : | Bacteriophage HK97 K169Y Head I |
| Authors | : | Gan, L.; Speir, J.A.; Conway, J.F.; Lander, G.; Cheng, N.; Firek, B.A.; Hendrix, R.W.; Duda, R.L.; Liljas, L.; Johnson, J.E. |
| Deposited on | : | 2006-01-20 |
| Resolution | : | 4.20 Å(reported) |

This is a wwPDB X-ray Structure 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/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

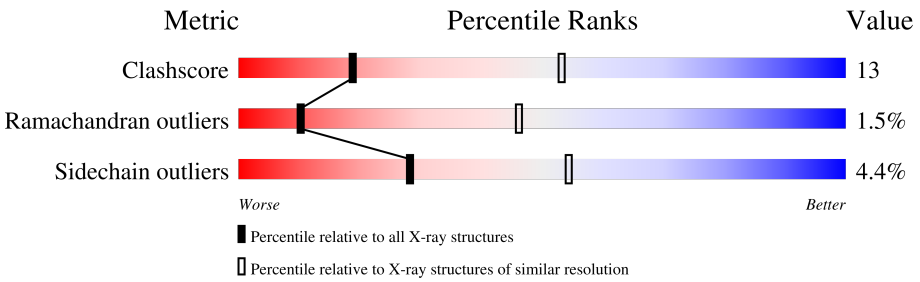
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.35.1 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (Gargrove) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.35.1 |

# 1 Overall quality at a glance (i)

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

The reported resolution of this entry is 4.20 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)) |
|---|---|---|
| Clashscore | 141614 | 1044 (4.60-3.80) |
| Ramachandran outliers | 138981 | 1000 (4.60-3.80) |
| Sidechain outliers | 138945 | 1007 (4.62-3.78) |

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%

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 1 | A | 282 | 71% 26% • • |
| 1 | B | 282 | 73% 24% • • |
| 1 | C | 282 | 68% 27% • • |
| 1 | D | 282 | 70% 28% • • |
| 1 | E | 282 | 74% 23% • • |
| 1 | F | 282 | 64% 25% • 9% |
| 1 | G | 282 | 59% 24% • 14% |

# 2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 14664 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 Major capsid protein.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 1 | A | 280 | 2154 | 1347 | 374 | 423 | 10 | 0 | 0 | 0 |
| 1 | B | 280 | 2154 | 1347 | 374 | 423 | 10 | 0 | 0 | 0 |
| 1 | C | 280 | 2154 | 1347 | 374 | 423 | 10 | 0 | 0 | 0 |
| 1 | D | 280 | 2154 | 1347 | 374 | 423 | 10 | 0 | 0 | 0 |
| 1 | E | 280 | 2154 | 1347 | 374 | 423 | 10 | 0 | 0 | 0 |
| 1 | F | 258 | 2004 | 1254 | 350 | 391 | 9 | 0 | 0 | 0 |
| 1 | G | 243 | 1890 | 1181 | 333 | 368 | 8 | 0 | 0 | 0 |

There are 7 discrepancies between the modelled and reference sequences:

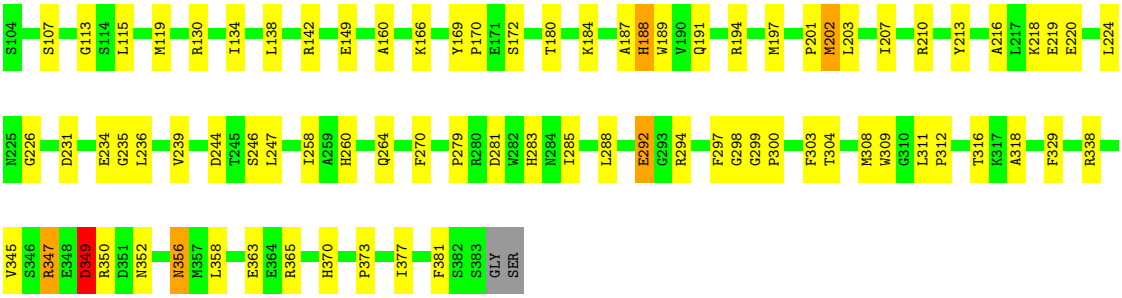
| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| A | 169 | TYR | LYS | engineered mutation | UNP P49861 |
| B | 169 | TYR | LYS | engineered mutation | UNP P49861 |
| C | 169 | TYR | LYS | engineered mutation | UNP P49861 |
| D | 169 | TYR | LYS | engineered mutation | UNP P49861 |
| E | 169 | TYR | LYS | engineered mutation | UNP P49861 |
| F | 169 | TYR | LYS | engineered mutation | UNP P49861 |
| G | 169 | TYR | LYS | engineered mutation | UNP P49861 |

# 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.

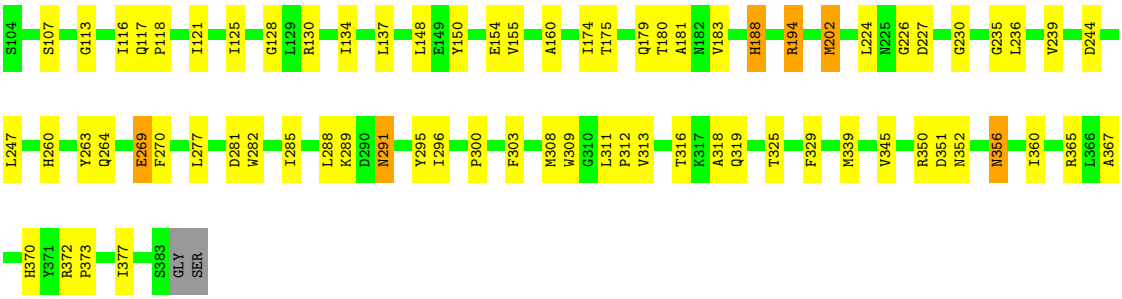
• Molecule 1: Major capsid protein

Chain A:




• Molecule 1: Major capsid protein

Chain B:

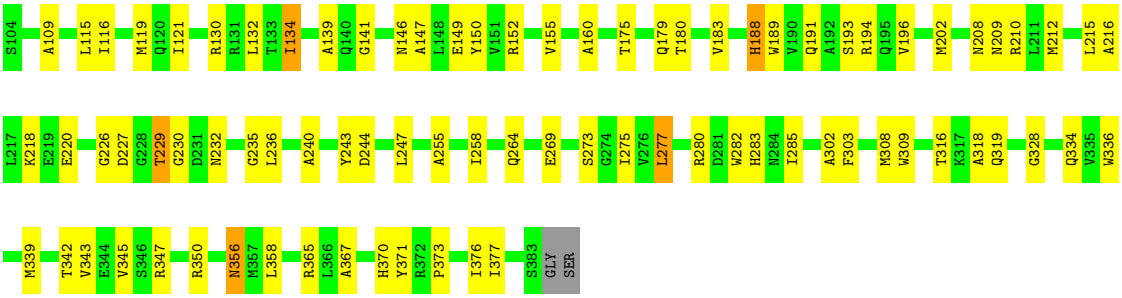



• Molecule 1: Major capsid protein

Chain C:

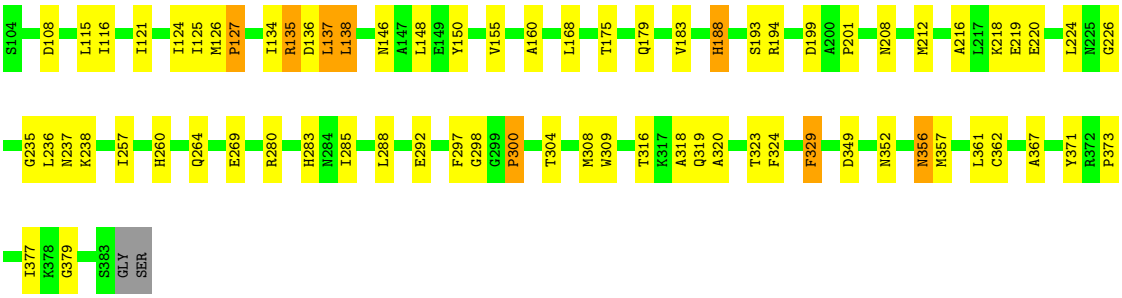
- Molecule 1: Major capsid protein

Chain D:



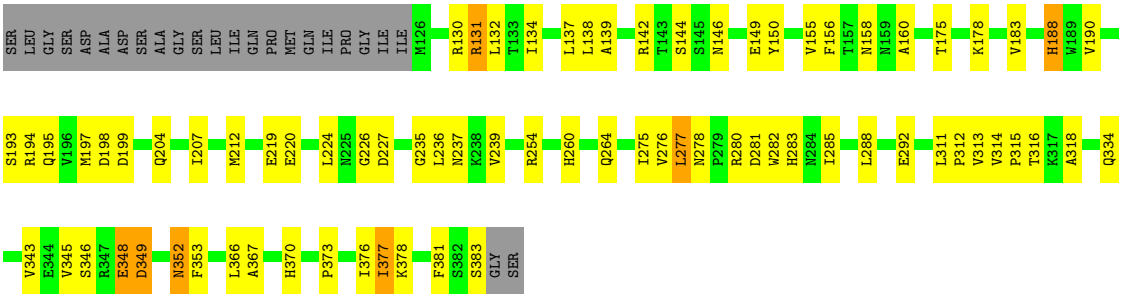

- Molecule 1: Major capsid protein

Chain E:



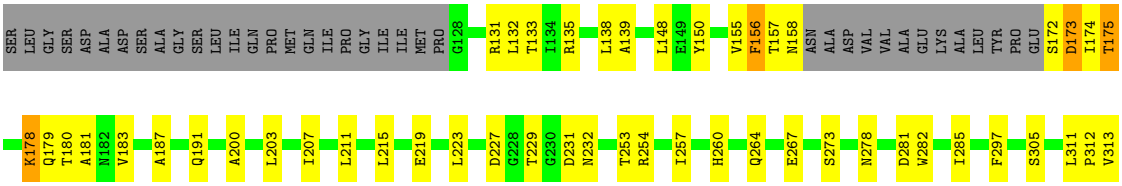

- Molecule 1: Major capsid protein

Chain F:




- Molecule 1: Major capsid protein

Chain G:



T316
K317
A318
V326
G327
G328
F329
A332
S333
Q334
V335
W336
M339
V345
D349
R350
D351
N352
F353
V354
K355
N356
M357
L358
T359
C362
E363
E364
R365
L366
A367
L368
A369
H370
Y371
I376
I377
F381
S382
S383
GLY
SER

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 43 21 2 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 1009.64Å 1009.64Å 729.81Å<br>90.00° 90.00° 90.00° | Depositor |
| Resolution (Å) | 30.00 – 4.20<br>29.98 – 4.20 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 63.1 (30.00-4.20)<br>63.1 (29.98-4.20) | Depositor<br>EDS |
| $R_{merge}$ | 0.16 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 1.95 (at 4.26Å) | Xtriage |
| Refinement program | CNS | Depositor |
| R, $R_{free}$ | 0.384 , (Not available)<br>0.313 , (Not available) | Depositor<br>DCC |
| $R_{free}$ test set | No test flags present. | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 20.0 | Xtriage |
| Anisotropy | 2.024 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.20 , 6.3 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.42$, $< L^2 > = 0.24$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| $F_o$,$F_c$ correlation | 0.69 | EDS |
| Total number of atoms | 14664 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 80.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.10% 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 | 0.25 | 0/2192 | 0.48 | 0/2976 |
| 1 | B | 0.25 | 0/2192 | 0.49 | 0/2976 |
| 1 | C | 0.25 | 0/2192 | 0.50 | 1/2976 (0.0%) |
| 1 | D | 0.25 | 0/2192 | 0.48 | 0/2976 |
| 1 | E | 0.25 | 0/2192 | 0.50 | 0/2976 |
| 1 | F | 0.26 | 0/2040 | 0.49 | 0/2769 |
| 1 | G | 0.28 | 0/1922 | 0.51 | 0/2605 |
| All | All | 0.25 | 0/14922 | 0.49 | 1/20254 (0.0%) |

There are no bond length outliers.

All (1) bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 1 | C | 356 | ASN | CB-CA-C | 6.05 | 122.50 | 110.40 |

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 | 2154 | 0 | 2115 | 62 | 0 |
| 1 | B | 2154 | 0 | 2115 | 54 | 0 |
| 1 | C | 2154 | 0 | 2115 | 65 | 0 |
| 1 | D | 2154 | 0 | 2115 | 69 | 0 |
| 1 | E | 2154 | 0 | 2115 | 53 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|---|---|---|---|---|---|---|
| 1 | F | 2004 | 0 | 1963 | 58 | 0 |
| 1 | G | 1890 | 0 | 1851 | 66 | 0 |
| All | All | 14664 | 0 | 14389 | 380 | 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 13.

The worst 5 of 380 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:F:130:ARG:HD3 | 1:F:131:ARG:H | 1.33 | 0.92 |
| 1:D:264:GLN:HE21 | 1:D:377:ILE:HD12 | 1.40 | 0.85 |
| 1:E:297:PHE:HB3 | 1:E:304:THR:HG21 | 1.59 | 0.84 |
| 1:G:157:THR:HB | 1:G:173:ASP:O | 1.76 | 0.83 |
| 1:G:191:GLN:HG3 | 1:G:359:THR:HG22 | 1.61 | 0.83 |

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 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 | 278/282 (99%) | 241 (87%) | 32 (12%) | 5 (2%) | 8 42 |
| 1 | B | 278/282 (99%) | 239 (86%) | 36 (13%) | 3 (1%) | 14 52 |
| 1 | C | 278/282 (99%) | 248 (89%) | 27 (10%) | 3 (1%) | 14 52 |
| 1 | D | 278/282 (99%) | 247 (89%) | 27 (10%) | 4 (1%) | 11 47 |
| 1 | E | 278/282 (99%) | 244 (88%) | 29 (10%) | 5 (2%) | 8 42 |
| 1 | F | 256/282 (91%) | 228 (89%) | 24 (9%) | 4 (2%) | 9 45 |
| 1 | G | 239/282 (85%) | 204 (85%) | 30 (13%) | 5 (2%) | 7 39 |
| All | All | 1885/1974 (96%) | 1651 (88%) | 205 (11%) | 29 (2%) | 10 46 |

5 of 29 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 201 | PRO |
| 1 | E | 300 | PRO |
| 1 | F | 352 | ASN |
| 1 | A | 202 | MET |
| 1 | B | 352 | ASN |

### 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 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 | 230/231 (100%) | 220 (96%) | 10 (4%) | 29 | 55 |
| 1 | B | 230/231 (100%) | 221 (96%) | 9 (4%) | 32 | 57 |
| 1 | C | 230/231 (100%) | 217 (94%) | 13 (6%) | 20 | 48 |
| 1 | D | 230/231 (100%) | 224 (97%) | 6 (3%) | 46 | 67 |
| 1 | E | 230/231 (100%) | 221 (96%) | 9 (4%) | 32 | 57 |
| 1 | F | 213/231 (92%) | 202 (95%) | 11 (5%) | 23 | 50 |
| 1 | G | 201/231 (87%) | 190 (94%) | 11 (6%) | 21 | 49 |
| All | All | 1564/1617 (97%) | 1495 (96%) | 69 (4%) | 28 | 54 |

5 of 69 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | F | 377 | ILE |
| 1 | G | 156 | PHE |
| 1 | G | 336 | TRP |
| 1 | C | 224 | LEU |
| 1 | C | 188 | HIS |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 52 such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | D | 264 | GLN |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | E | 204 | GLN |
| 1 | G | 232 | ASN |
| 1 | D | 291 | ASN |
| 1 | D | 356 | ASN |

#### 5.3.3 RNA (i)

There are no RNA molecules in this entry.

### 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 monosaccharides 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.

# 6 Fit of model and data (i)

## 6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

## 6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

## 6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

## 6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.