# W O R L D W I D E
# wwPDB
## P R O T E I N   D A T A   B A N K

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

Feb 18, 2024 – 06:48 PM EST

| | | |
|---|---|---|
| PDB ID | : | 4I21 |
| Title | : | Crystal structure of L858R + T790M EGFR kinase domain in complex with MIG6 peptide |
| Authors | : | Gajiwala, K.S.; Feng, J.; Ferre, R.; Ryan, K.; Brodsky, O.; Stewart, A. |
| Deposited on | : | 2012-11-21 |
| Resolution | : | 3.37 Å(reported) |

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

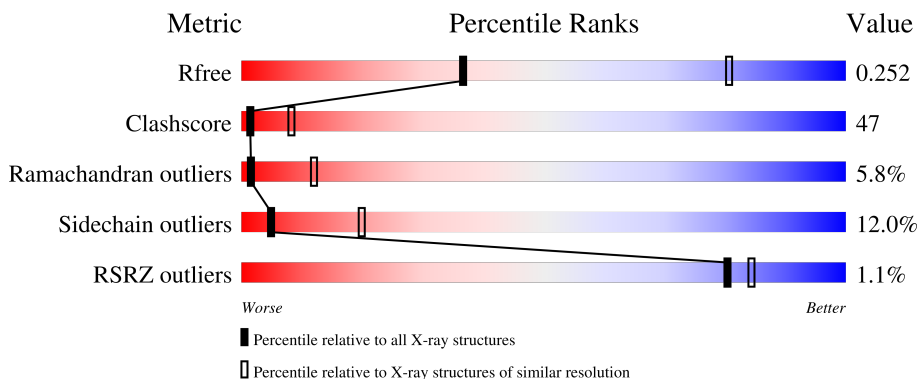| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.36 |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (Gargrove) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.36 |

# 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 3.37 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 |
|---|---|---|
| Rfree | | 0.252 |
| Clashscore | | 47 |
| Ramachandran outliers | | 5.8% |
| Sidechain outliers | | 12.0% |
| RSRZ outliers | | 1.1% |

*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)) |
|---|---|---|
| $R_{free}$ | 130704 | 1691 (3.46-3.30) |
| Clashscore | 141614 | 1762 (3.46-3.30) |
| Ramachandran outliers | 138981 | 1732 (3.46-3.30) |
| Sidechain outliers | 138945 | 1731 (3.46-3.30) |
| RSRZ outliers | 127900 | 1635 (3.46-3.30) |

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 | 329 | %  29%  48%  11%  12% |
| 1 | B | 329 | 2%  31%  49%  12%  8% |
| 2 | C | 62 | 21%  19%  8%  52% |
| 2 | D | 62 | 21%  15%  5% •  58% |

## 2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 5182 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 Epidermal growth factor receptor.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---|---|---|---|---------|---------|-------|
| 1 | A | 290 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 2335 | 1506 | 396 | 416 | 17 | | | |
| 1 | B | 302 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 2424 | 1561 | 411 | 435 | 17 | | | |

There are 6 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|---------|-----------|
| A | 694 | GLY | - | expression tag | UNP P00533 |
| A | 790 | MET | THR | engineered mutation | UNP P00533 |
| A | 858 | ARG | LEU | engineered mutation | UNP P00533 |
| B | 694 | GLY | - | expression tag | UNP P00533 |
| B | 790 | MET | THR | engineered mutation | UNP P00533 |
| B | 858 | ARG | LEU | engineered mutation | UNP P00533 |

- Molecule 2 is a protein called ERBB receptor feedback inhibitor 1.

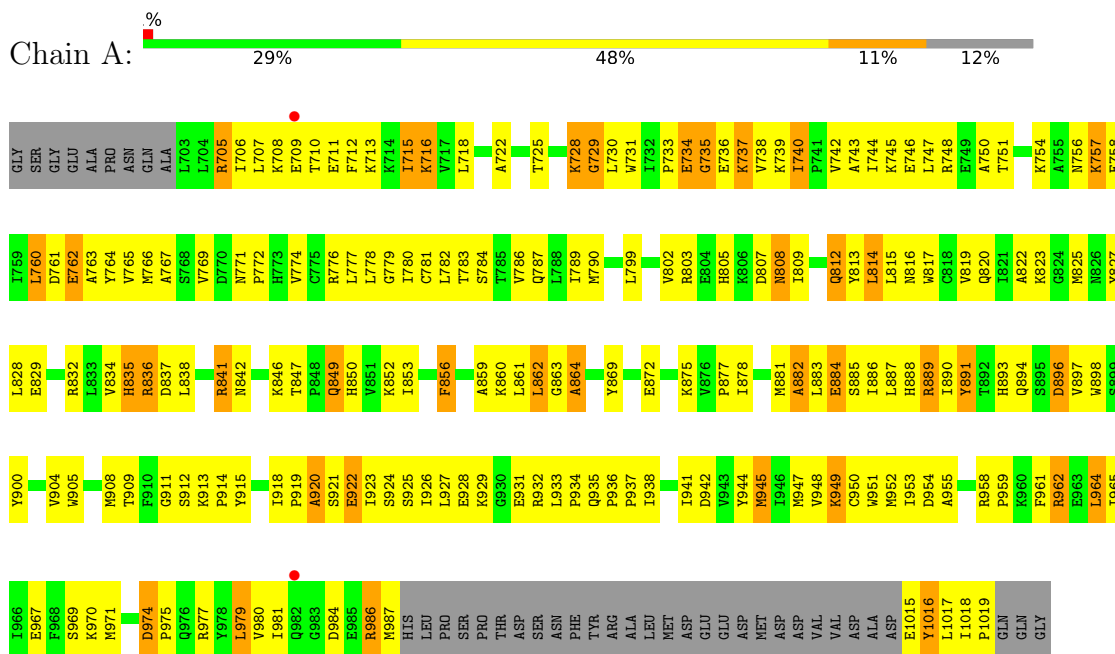| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|---|---|---|---|---------|---------|-------|
| 2 | C | 30 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 226 | 148 | 35 | 42 | 1 | | | |
| 2 | D | 26 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 197 | 128 | 30 | 38 | 1 | | | |

There are 4 discrepancies between the modelled and reference sequences:

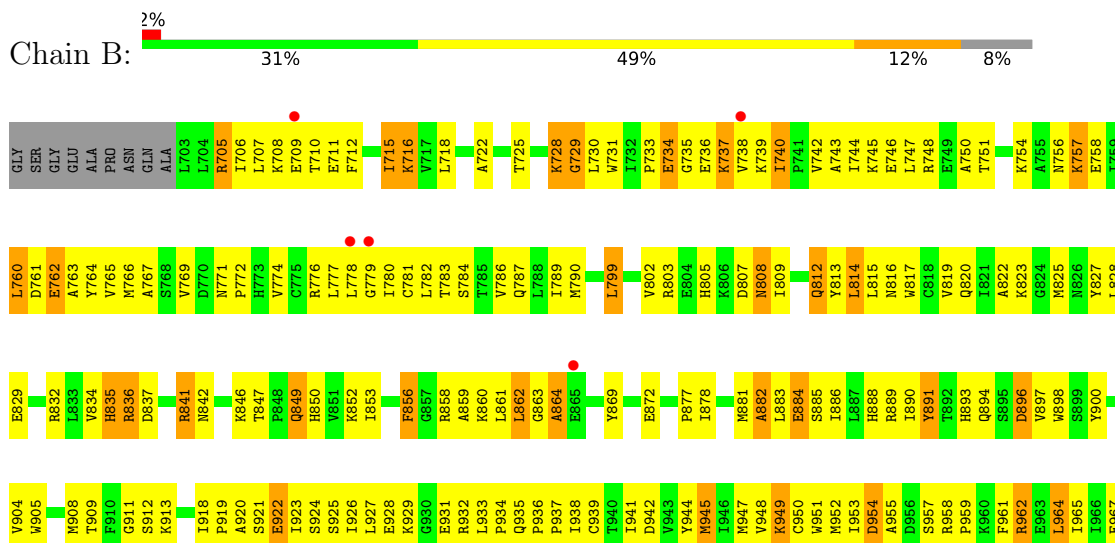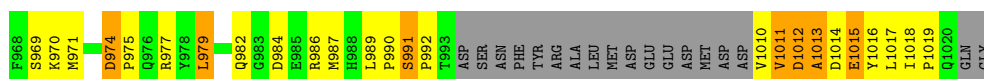| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|---------|-----------|
| C | 313 | GLY | - | expression tag | UNP Q9UJM3 |
| C | 314 | SER | - | expression tag | UNP Q9UJM3 |
| D | 313 | GLY | - | expression tag | UNP Q9UJM3 |
| D | 314 | SER | - | expression tag | UNP Q9UJM3 |

# 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.
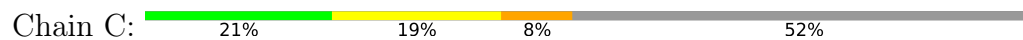
- Molecule 1: Epidermal growth factor receptor



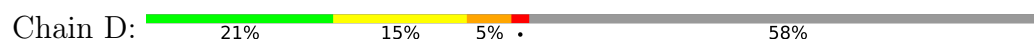- Molecule 1: Epidermal growth factor receptor

F968 S969 K970 M971 D974 P975 Q976 R977 Y978 L979 Q982 G983 D984 E985 R986 M987 H988 L989 P990 S991 P992 T993 ASP SER ASN PHE TYR ARG ALA LEU MET ASP GLU GLU ASP MET ASP V1010 V1011 D1012 A1013 D1014 E1015 Y1016 L1017 I1018 P1019 Q1020 GLN GLY

• Molecule 2: ERBB receptor feedback inhibitor 1

Chain C: 21%  19%  8%  52%

GLY SER ARG PRO PRO LYS VAL PRO PRO ARG GLU PRO LEU SER PRO SER ASN SER ARG THR PRO SER P335 K336 S337 L338 P339 L342 V345 M346 P347 F352 A353 P354 D355 P356 K357 Y358 V359 S360 S361 K362 A363 L364 GLN ARG GLN ASN SER GLU GLY SER ALA SER

• Molecule 2: ERBB receptor feedback inhibitor 1

Chain D: 21%  15%  5% •  58%

GLY SER ARG PRO PRO LYS VAL PRO PRO ARG GLU PRO LEU SER PRO SER ASN SER ARG THR PRO SER PRO LYS S337 L338 P339 S340 Y341 V345 M346 P347 F352 D355 P356 K357 Y358 V359 S360 S361 K362 ALA LEU GLN ARG GLN ASN SER GLU GLY SER ALA SER

# 4 Data and refinement statistics ⓘ

| Property | Value | Source |
|---|---|---|
| Space group | P 1 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 38.67Å   66.05Å   94.52Å<br>71.85°   87.38°   74.26° | Depositor |
| Resolution (Å) | 37.18 – 3.37<br>37.18 – 3.37 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 97.1 (37.18-3.37)<br>97.4 (37.18-3.37) | Depositor<br>EDS |
| $R_{merge}$ | 0.09 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 2.73 (at 3.40Å) | Xtriage |
| Refinement program | CNX | Depositor |
| R, $R_{free}$ | 0.230 , 0.279<br>0.210 , 0.252 | Depositor<br>DCC |
| $R_{free}$ test set | 605 reflections (5.15%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 85.1 | Xtriage |
| Anisotropy | 0.679 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.34 , 66.1 | EDS |
| L-test for twinning[2] | $< |L| > = 0.48, < L^2 > = 0.31$ | Xtriage |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| $F_o$,$F_c$ correlation | 0.93 | EDS |
| Total number of atoms | 5182 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 69.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.57% 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 | 0.41 | 0/2386 | 0.56 | 0/3223 |
| 1 | B | 0.42 | 0/2478 | 0.57 | 0/3352 |
| 2 | C | 0.52 | 0/234 | 0.70 | 0/319 |
| 2 | D | 0.49 | 0/204 | 0.65 | 0/279 |
| All | All | 0.42 | 0/5302 | 0.58 | 0/7173 |

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 ⓘ

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 | 2335 | 0 | 2396 | 230 | 0 |
| 1 | B | 2424 | 0 | 2479 | 238 | 0 |
| 2 | C | 226 | 0 | 231 | 28 | 0 |
| 2 | D | 197 | 0 | 194 | 26 | 0 |
| All | All | 5182 | 0 | 5300 | 490 | 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 47.

The worst 5 of 490 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:B:1010:VAL:HG23 | 1:B:1011:VAL:H | 1.09 | 1.17 |
| 1:B:730:LEU:HD11 | 1:B:739:LYS:HB2 | 1.42 | 1.02 |
| 1:A:730:LEU:HD11 | 1:A:739:LYS:HB2 | 1.41 | 1.01 |
| 1:B:1011:VAL:O | 1:B:1012:ASP:HB2 | 1.63 | 0.95 |
| 2:C:362:LYS:HG2 | 2:C:363:ALA:HA | 1.46 | 0.94 |

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 | |
|---|---|---|---|---|---|---|---|
| 1 | A | 286/329 (87%) | 231 (81%) | 41 (14%) | 14 (5%) | 2 | 15 |
| 1 | B | 298/329 (91%) | 232 (78%) | 47 (16%) | 19 (6%) | 1 | 10 |
| 2 | C | 28/62 (45%) | 21 (75%) | 4 (14%) | 3 (11%) | 0 | 3 |
| 2 | D | 24/62 (39%) | 16 (67%) | 7 (29%) | 1 (4%) | 3 | 19 |
| All | All | 636/782 (81%) | 500 (79%) | 99 (16%) | 37 (6%) | 1 | 12 |

5 of 37 Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 734 | GLU |
| 1 | A | 836 | ARG |
| 1 | A | 882 | ALA |
| 1 | A | 1016 | TYR |
| 1 | B | 734 | GLU |

### 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 | 256/288 (89%) | 226 (88%) | 30 (12%) | 5 | 21 | |
| 1 | B | 267/288 (93%) | 236 (88%) | 31 (12%) | 5 | 21 | |
| 2 | C | 27/56 (48%) | 23 (85%) | 4 (15%) | 3 | 13 | |
| 2 | D | 24/56 (43%) | 20 (83%) | 4 (17%) | 2 | 9 | |
| All | All | 574/688 (83%) | 505 (88%) | 69 (12%) | 5 | 20 | |

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

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | B | 962 | ARG |
| 1 | B | 979 | LEU |
| 2 | D | 338 | LEU |
| 1 | A | 949 | LYS |
| 1 | A | 945 | MET |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | B | 870 | HIS |
| 1 | B | 849 | GLN |
| 1 | B | 756 | ASN |
| 1 | B | 816 | ASN |
| 1 | A | 893 | HIS |

### 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(Å$^2$) | Q<0.9 |
|-----|-------|----------|--------|---------|-----|-----|-------------|-------|
| 1 | A | 290/329 (88%) | -0.01 | 2 (0%) | 87 | 91 | 32, 66, 105, 129 | 0 |
| 1 | B | 302/329 (91%) | 0.03 | 5 (1%) | 70 | 74 | 34, 68, 106, 129 | 0 |
| 2 | C | 30/62 (48%) | 0.20 | 0 | 100 | 100 | 36, 70, 93, 97 | 0 |
| 2 | D | 26/62 (41%) | 0.05 | 0 | 100 | 100 | 44, 63, 78, 83 | 0 |
| All | All | 648/782 (82%) | 0.02 | 7 (1%) | 80 | 84 | 32, 67, 106, 129 | 0 |

The worst 5 of 7 RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|-----|-------|-----|------|------|
| 1 | B | 709 | GLU | 3.8 |
| 1 | B | 865 | GLU | 3.2 |
| 1 | A | 709 | GLU | 2.6 |
| 1 | B | 778 | LEU | 2.5 |
| 1 | A | 982 | GLN | 2.4 |

## 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.