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

Jan 31, 2016 – 07:49 PM GMT

| | | |
|---|---|---|
| PDB ID | : | 1HEG |
| Title | : | The crystal structures at 2.2 angstroms resolution of hydroxyethylene-based inhibitors bound to human immunodeficiency virus type 1 protease show that the inhibitors are present in two distinct orientations |
| Authors | : | Murthy, K.; Winborne, E.L.; Minnich, M.D.; Culp, J.S.; Debouck, C. |
| Deposited on | : | 1992-09-21 |
| Resolution | : | 2.20 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

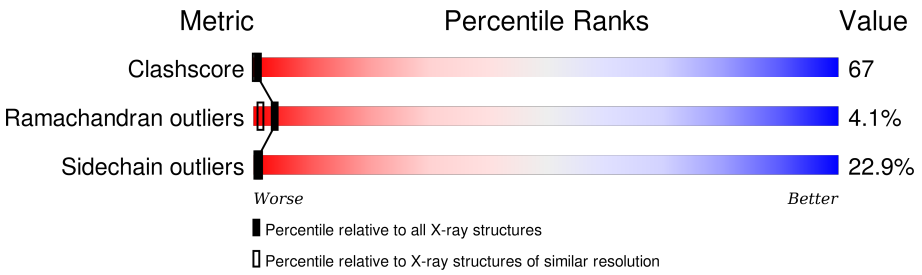
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.7 (RC4), CSD as536be (2015) |
| Xtriage (Phenix) | : | NOT EXECUTED |
| EDS | : | NOT EXECUTED |
| Percentile statistics | : | 20151230.v01 (using entries in the PDB archive December 30th 2015) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | trunk26865 |

# 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 2.20 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 | 102246 | 4477 (2.20-2.20) |
| Ramachandran outliers | 100387 | 4404 (2.20-2.20) |
| Sidechain outliers | 100360 | 4405 (2.20-2.20) |

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 >=3, 2, 1 and 0 types of geometric quality criteria. 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.

Note EDS was not executed.

| Mol | Chain | Length | Quality of chain |
|---|---|---|---|
| 1 | E | 99 | 29% 41% 24% 5% |

# 2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 832 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.

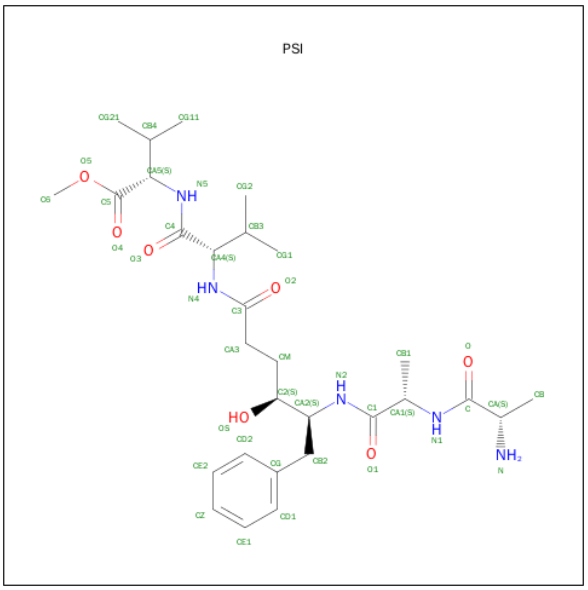
- Molecule 1 is a protein called HIV-1 PROTEASE.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | S | | | |
| 1 | E | 99 | 758 | 488 | 131 | 136 | 3 | 0 | 0 | 0 |

There is a discrepancy between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|---|---|---|---|---|---|
| E | 36 | ASN | MET | CONFLICT | UNP P03366 |

- Molecule 2 is METHYL N-{(4S,5S)-5-[(L-ALANYL-L-ALANYL)AMINO]-4-HYDROXY-6-PHENYLHEXANOYL}-L-VALYL-L-VALINATE (three-letter code: PSI) (formula: $C_{29}H_{47}N_5O_7$).




| Mol | Chain | Residues | Atoms | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|
| | | | Total | C | N | O | | |
| 2 | E | 1 | 41 | 29 | 5 | 7 | 0 | 0 |

- Molecule 3 is water.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| | | | Total | O | | |
| 3 | E | 33 | 33 | 33 | 0 | 0 |

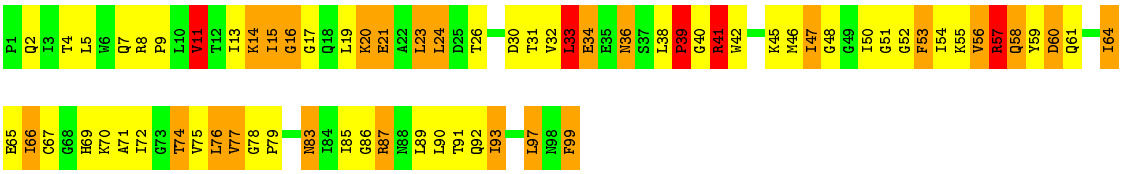
# 3 Residue-property plots (i)

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.

Note EDS was not executed.

- Molecule 1: HIV-1 PROTEASE

Chain E: 29% 41% 24% 5%



# 4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section will therefore be incomplete.

| Property | Value | Source |
|---|---|---|
| Space group | P 61 2 2 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 64.14Å 64.14Å 84.24Å<br>90.00° 90.00° 120.00° | Depositor |
| Resolution (Å) | 5.00 – 2.20 | Depositor |
| % Data completeness<br>(in resolution range) | (Not available) (5.00-2.20) | Depositor |
| $R_{merge}$ | (Not available) | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| Refinement program | PROLSQ | Depositor |
| R, $R_{free}$ | 0.186 , (Not available) | Depositor |
| Estimated twinning fraction | No twinning to report. | Xtriage |
| Total number of atoms | 832 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 20.0 | wwPDB-VP |

# 5 Model quality (i)

## 5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: PSI

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 | 1.13 | 0/771 | 2.17 | 36/1044 (3.4%) |

There are no bond length outliers.

The worst 5 of 36 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 1 | E | 8 | ARG | CD-NE-CZ | 12.53 | 141.15 | 123.60 |
| 1 | E | 8 | ARG | NE-CZ-NH1 | 10.84 | 125.72 | 120.30 |
| 1 | E | 60 | ASP | CA-CB-CG | 8.56 | 132.24 | 113.40 |
| 1 | E | 71 | ALA | CB-CA-C | 8.38 | 122.67 | 110.10 |
| 1 | E | 58 | GLN | C-N-CA | 7.34 | 140.04 | 121.70 |

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 | E | 758 | 0 | 803 | 108 | 16 |
| 2 | E | 41 | 0 | 43 | 7 | 1 |
| 3 | E | 33 | 0 | 0 | 39 | 17 |
| All | All | 832 | 0 | 846 | 111 | 20 |

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 67.

The worst 5 of 111 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:E:21:GLU:HG2 | 3:E:124:HOH:O | 1.24 | 1.31 |
| 1:E:42:TRP:CD2 | 1:E:57:ARG:HG2 | 1.74 | 1.22 |
| 1:E:42:TRP:CE3 | 1:E:57:ARG:HG2 | 1.77 | 1.18 |
| 1:E:34:GLU:O | 1:E:77:VAL:HG12 | 1.59 | 1.02 |
| 1:E:50:ILE:HG12 | 2:E:201:PSI:HG12 | 1.42 | 1.00 |

The worst 5 of 20 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:E:39:PRO:CG | 3:E:103:HOH:O[2_665] | 0.76 | 1.44 |
| 1:E:39:PRO:CB | 3:E:118:HOH:O[2_665] | 1.37 | 0.83 |
| 3:E:131:HOH:O | 3:E:131:HOH:O[11_555] | 1.43 | 0.77 |
| 1:E:53:PHE:CD2 | 3:E:132:HOH:O[9_555] | 1.60 | 0.60 |
| 1:E:53:PHE:CD2 | 1:E:53:PHE:CD2[10_555] | 1.64 | 0.56 |

## 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 | E | 97/99 (98%) | 90 (93%) | 3 (3%) | 4 (4%) | 3 1 |

All (4) Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | E | 34 | GLU |
| 1 | E | 39 | PRO |
| 1 | E | 16 | GLY |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | E | 17 | GLY |

### 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 | E | 83/83 (100%) | 64 (77%) | 19 (23%) | 1 1 |

5 of 19 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | E | 41 | ARG |
| 1 | E | 53 | PHE |
| 1 | E | 72 | ILE |
| 1 | E | 39 | PRO |
| 1 | E | 77 | VAL |

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | E | 58 | GLN |
| 1 | E | 92 | GLN |
| 1 | E | 98 | 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 carbohydrates in this entry.

## 5.6 Ligand geometry (i)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

| Mol | Type | Chain | Res | Link | Bond lengths | | | Bond angles | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Counts | RMSZ | $\#\|Z\| > 2$ | Counts | RMSZ | $\#\|Z\| > 2$ |
| 2 | PSI | E | 201 | - | 40,41,41 | 1.06 | 2 (5%) | 54,55,55 | 1.70 | 15 (27%) |

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

| Mol | Type | Chain | Res | Link | Chirals | Torsions | Rings |
|---|---|---|---|---|---|---|---|
| 2 | PSI | E | 201 | - | - | 0/55/55/55 | 0/1/1/1 |

All (2) bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 2 | E | 201 | PSI | CM-C2 | 2.74 | 1.57 | 1.52 |
| 2 | E | 201 | PSI | O5-C5 | 3.45 | 1.42 | 1.33 |

The worst 5 of 15 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 2 | E | 201 | PSI | C6-O5-C5 | -3.21 | 108.46 | 115.99 |
| 2 | E | 201 | PSI | CA3-CM-C2 | -3.14 | 109.19 | 113.23 |
| 2 | E | 201 | PSI | CA4-C4-N5 | -2.54 | 110.85 | 116.44 |
| 2 | E | 201 | PSI | CA-C-N1 | -2.14 | 111.82 | 116.83 |
| 2 | E | 201 | PSI | CB2-CG-CD2 | -2.13 | 116.45 | 120.90 |

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 8 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 2 | E | 201 | PSI | 7 | 1 |

## 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)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

## 6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

## 6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

## 6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.