



# wwPDB X-ray Structure Validation Summary Report

Nov 17, 2014 – 07:20 PM EST

PDB ID : 4RJ4  
Title : EGFR kinase (T790M/L858R) with inhibitor compound 6  
Authors : Eigenbrot, C.; Yu, C.  
Deposited on : 2014-10-08  
Resolution : 2.78 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>

---

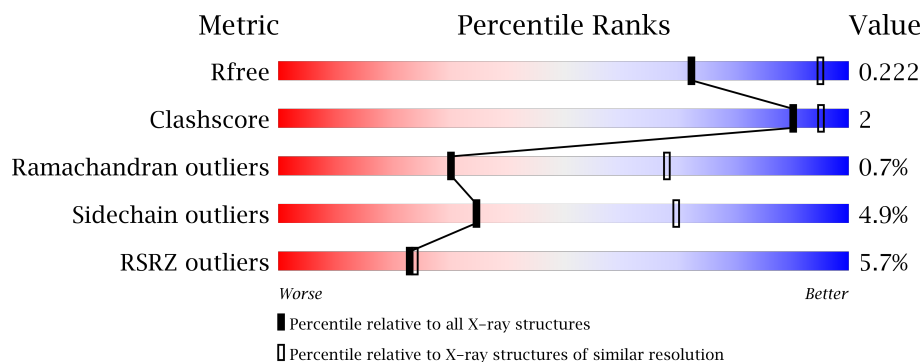
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable24195  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.1.3  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable24195

# 1 Overall quality at a glance

The reported resolution of this entry is 2.78 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))
$R_{free}$	66092	2193 (2.80-2.76)
Clashscore	79885	2751 (2.80-2.76)
Ramachandran outliers	78287	2699 (2.80-2.76)
Sidechain outliers	78261	2701 (2.80-2.76)
RSRZ outliers	66119	2196 (2.80-2.76)

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	A	331	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2438 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.

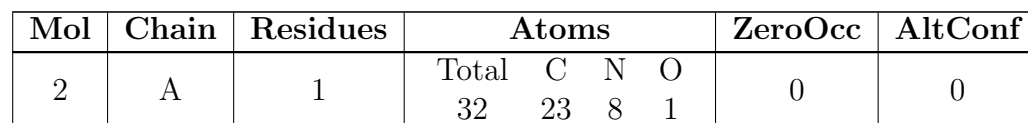
- Molecule 1 is a protein called Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	299	Total	C	N	O	S	0	0	0
			2403	1545	403	437	18			

There are 8 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
A	865	ALA	GLU	ENGINEERED MUTATION	UNP P00533
A	866	ALA	GLU	ENGINEERED MUTATION	UNP P00533
A	867	ALA	LYS	ENGINEERED MUTATION	UNP P00533
A	1023	ASN	-	EXPRESSION TAG	UNP P00533
A	1024	SER	-	EXPRESSION TAG	UNP P00533

- Molecule 2 is N-[2-(4-METHOXYPIPERIDIN-1-YL)PYRIMIDIN-4-YL]-1-(PROPAN-2-YL)-2-(1H-PYRAZOL-4-YL)-1H-PYRROLO[3,2-C]PYRIDIN-6-AMINE (three-letter code: 3QW) (formula: C<sub>23</sub>H<sub>28</sub>N<sub>8</sub>O).



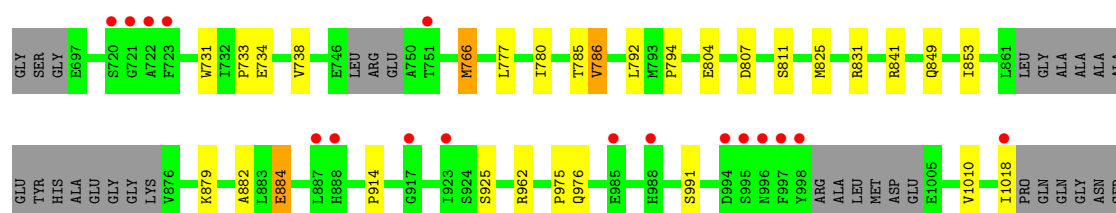
- | Mol | Chain | Residues | Atoms          | ZeroOcc | AltConf |
|-----|-------|----------|----------------|---------|---------|
| 3   | A     | 3        | Total O<br>3 3 | 0       | 0       |

### 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 ( $\text{RSRZ} > 2$ ). 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

Chain A: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	146.56Å 146.56Å 146.56Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.78 51.82 – 2.78	Depositor EDS
% Data completeness (in resolution range)	99.5 (50.00-2.78) 100.0 (51.82-2.78)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.88 (at 2.77Å)	Xtriage
Refinement program	BUSTER 2.11.4	Depositor
R, $R_{free}$	0.188 , 0.226 0.190 , 0.222	Depositor DCC
$R_{free}$ test set	570 reflections (4.48%)	DCC
Wilson B-factor (Å <sup>2</sup> )	56.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 37.6	EDS
Estimated twinning fraction	0.030 for -l,-k,-h	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 13291 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2438	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.33% 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

Bond lengths and bond angles in the following residue types are not validated in this section: 3QW

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	0.47	0/2455	0.67	0/3322

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	A	2403	0	2433	10	0
2	A	32	0	0	0	0
3	A	3	0	0	0	0
All	All	2438	0	2433	10	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 2.

The worst 5 of 10 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:882:ALA:HB1	1:A:884:GLU:OE2	1.93	0.69

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:766:MET:HB3	1:A:777:LEU:HB2	1.76	0.68
1:A:785:THR:O	1:A:786:VAL:HG12	2.14	0.47
1:A:731:TRP:CE2	1:A:733:PRO:HG3	2.52	0.45
1:A:811:SER:OG	1:A:975:PRO:HB2	2.17	0.44

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	291/331 (88%)	279 (96%)	10 (3%)	2 (1%)	30 68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	734	GLU
1	A	786	VAL

### 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	267/287 (93%)	254 (95%)	13 (5%)	35 70

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	849	GLN

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	884	GLU
1	A	991	SER
1	A	841	ARG
1	A	976	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	996	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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	3QW	A	1101	-	36,36,36	1.34	4 (11%)	51,51,51	2.91	19 (37%)

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	3QW	A	1101	-	-	0/14/28/28	0/5/5/5

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1101	3QW	C14-C13	-3.91	1.37	1.39
2	A	1101	3QW	C11-C12	-2.87	1.36	1.38
2	A	1101	3QW	C24-N26	2.47	1.40	1.35
2	A	1101	3QW	C14-N15	2.37	1.35	1.33

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1101	3QW	C13-C18-N17	-8.79	109.15	112.48
2	A	1101	3QW	C31-N26-C24	-6.50	111.69	121.84
2	A	1101	3QW	C18-N17-N15	6.50	107.79	103.95
2	A	1101	3QW	C27-N26-C24	-6.23	112.11	121.84
2	A	1101	3QW	N23-C24-N25	-5.32	120.16	126.31

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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	A	299/331 (90%)	0.11	17 (5%) 23 24	29, 52, 96, 129	6 (2%)

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	722	ALA	4.5
1	A	995	SER	4.3
1	A	1018	ILE	3.8
1	A	988	HIS	3.2
1	A	998	TYR	3.1

### 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 ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	3QW	A	1101	32/32	0.22	0.82	50,56,67,69	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.