



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:21 PM GMT

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)

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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

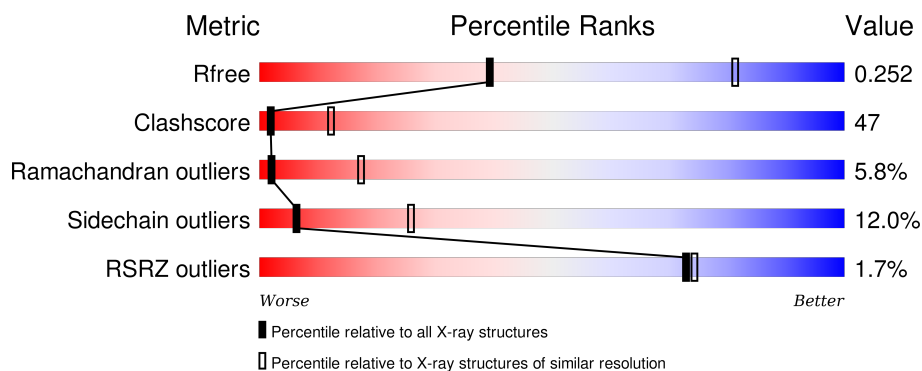
# 1 Overall quality at a glance ⓘ

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1084 (3.46-3.30)
Clashscore	102246	1158 (3.46-3.30)
Ramachandran outliers	100387	1139 (3.46-3.30)
Sidechain outliers	100360	1138 (3.46-3.30)
RSRZ outliers	91569	1089 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	<div> <div>29%</div> <div>48%</div> <div>11%</div> <div>12%</div> </div>
1	B	329	<div> <div>2%</div> <div>31%</div> <div>49%</div> <div>12%</div> <div>8%</div> </div>
2	C	62	<div> <div>21%</div> <div>19%</div> <div>8%</div> <div>52%</div> </div>
2	D	62	<div> <div>21%</div> <div>15%</div> <div>5%</div> <div>58%</div> </div>

## 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			

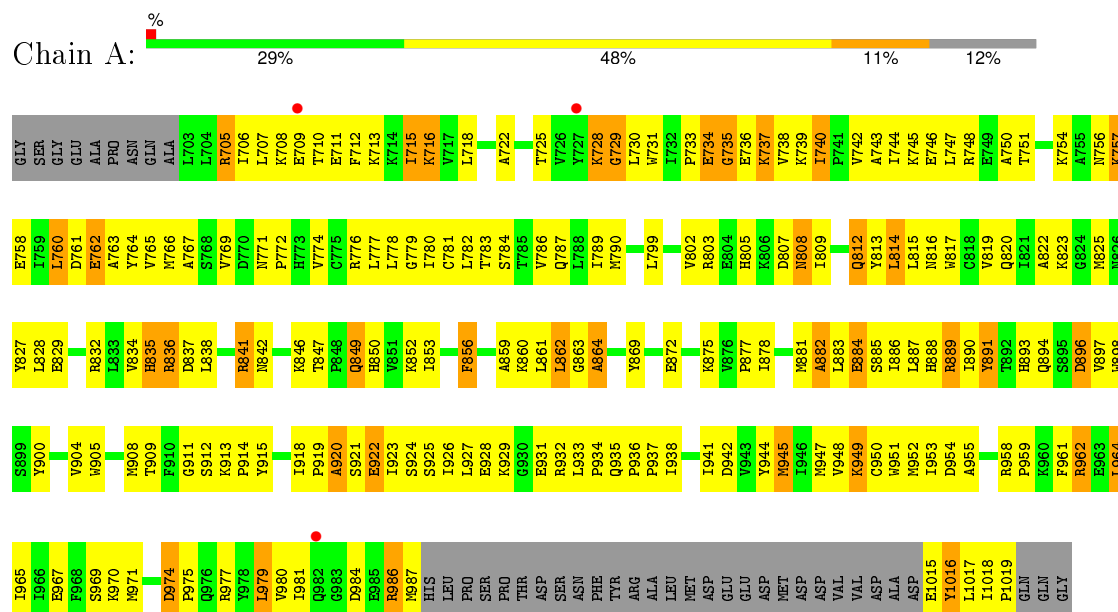
There are 4 discrepancies between the modelled and reference sequences:

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

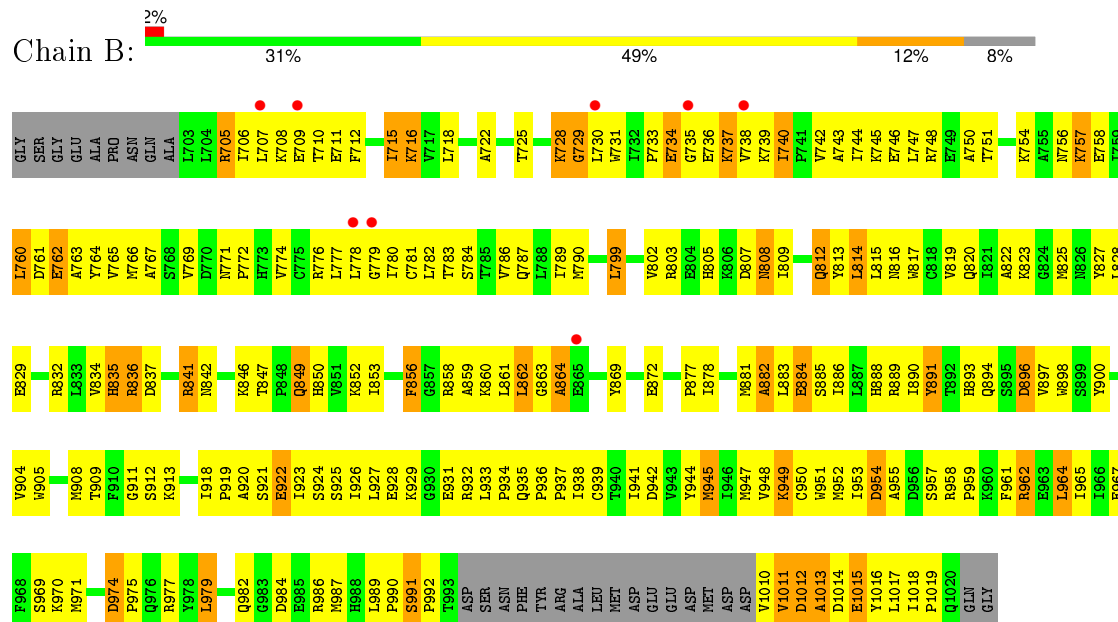
### 3 Residue-property plots

These plots are drawn for all protein, RNA 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.

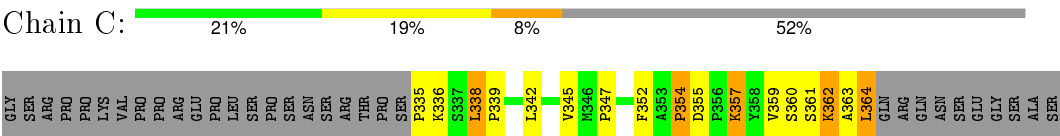
#### • Molecule 1: Epidermal growth factor receptor



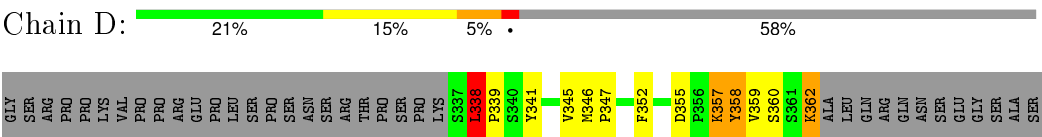
#### • Molecule 1: Epidermal growth factor receptor



● Molecule 2: ERBB receptor feedback inhibitor 1



● Molecule 2: ERBB receptor feedback inhibitor 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	38.67Å 66.05Å 94.52Å 71.85° 87.38° 74.26°	Depositor
Resolution (Å)	37.18 – 3.37 37.18 – 3.37	Depositor EDS
% Data completeness (in resolution range)	97.1 (37.18-3.37) 86.8 (37.18-3.37)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.73 (at 3.40Å)	Xtriage
Refinement program	CNX 2005	Depositor
R, $R_{free}$	0.230 , 0.279 0.210 , 0.252	Depositor DCC
$R_{free}$ test set	605 reflections (5.43%)	DCC
Wilson B-factor (Å <sup>2</sup> )	85.1	Xtriage
Anisotropy	0.679	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 64.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 11738 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	5182	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	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.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.41	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 [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	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 [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	286/329 (87%)	231 (81%)	41 (14%)	14 (5%)	3	22
1	B	298/329 (91%)	232 (78%)	47 (16%)	19 (6%)	2	15
2	C	28/62 (45%)	21 (75%)	4 (14%)	3 (11%)	0	4
2	D	24/62 (39%)	16 (67%)	7 (29%)	1 (4%)	3	27
All	All	636/782 (81%)	500 (79%)	99 (16%)	37 (6%)	2	17

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 [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	256/288 (89%)	226 (88%)	30 (12%)	7	27
1	B	267/288 (93%)	236 (88%)	31 (12%)	7	28
2	C	27/56 (48%)	23 (85%)	4 (15%)	4	17
2	D	24/56 (43%)	20 (83%)	4 (17%)	3	13
All	All	574/688 (83%)	505 (88%)	69 (12%)	6	26

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

Mol	Chain	Res	Type
1	B	716	LYS
1	B	762	GLU
2	C	362	LYS
1	B	725	THR
1	B	740	ILE

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

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

### 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 carbohydrates 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 [i](#)

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

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	290/329 (88%)	0.03	3 (1%) 84 86	32, 66, 105, 129	0
1	B	302/329 (91%)	0.06	8 (2%) 59 61	34, 68, 106, 129	0
2	C	30/62 (48%)	0.24	0 100 100	36, 70, 93, 97	0
2	D	26/62 (41%)	0.09	0 100 100	44, 63, 78, 83	0
All	All	648/782 (82%)	0.06	11 (1%) 73 74	32, 67, 106, 129	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	709	GLU	3.9
1	B	865	GLU	3.3
1	A	709	GLU	2.7
1	B	778	LEU	2.6
1	B	779	GLY	2.5

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.