



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 19, 2016 – 10:03 PM GMT

PDB ID : 5CAO  
Title : EGFR kinase domain mutant "TMLR" with compound 29  
Authors : Eigenbrot, C.; Yu, C.  
Deposited on : 2015-06-29  
Resolution : 2.60 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026982  
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) : rb-20026982

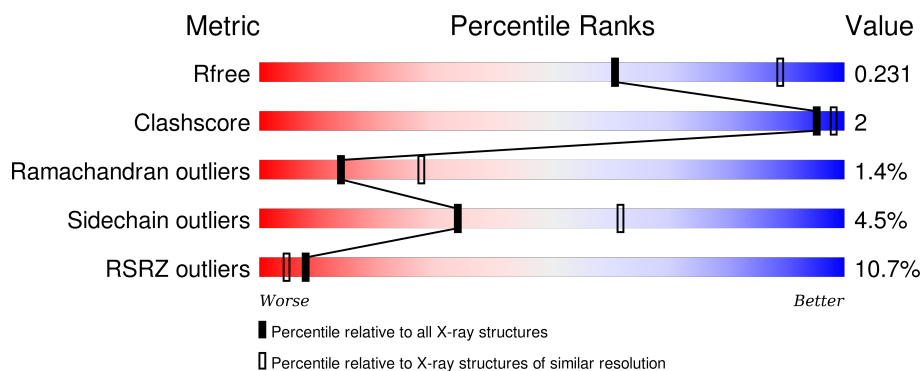
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*


The reported resolution of this entry is 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	331	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2450 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.

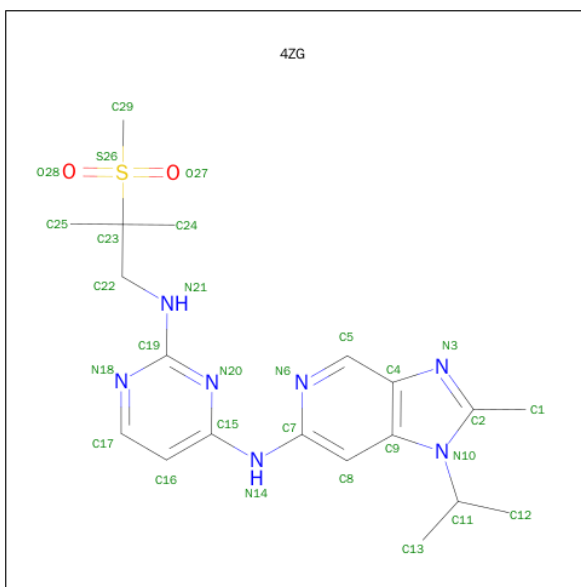
- Molecule 1 is a protein called Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	300	Total	C	N	O	S	0	0	0
			2416	1552	406	440	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 -[2-methyl-2-(methylsulfonyl)propyl]-N 4 -[2-methyl-1-(propan-2-yl)-1H-imidazo[4,5-c]pyridin-6-yl]pyrimidine-2,4-diamine (three-letter code: 4ZG) (formula: C<sub>19</sub>H<sub>27</sub>N<sub>7</sub>O<sub>2</sub>S).



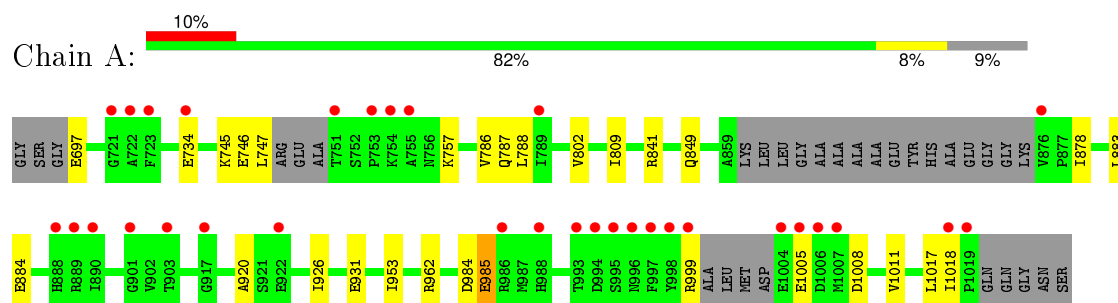
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	29	19	7	2	1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	5	Total	O	0	0
			5	5		



- Molecule 1: Epidermal growth factor receptor



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	149.16 Å   149.16 Å   149.16 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	24.80 – 2.60 24.20 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.3 (24.80-2.60) 100.0 (24.20-2.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.73 (at 2.60 Å)	Xtriage
Refinement program	BUSTER 2.11.4	Depositor
R, $R_{free}$	0.202 , 0.231 0.203 , 0.231	Depositor DCC
$R_{free}$ test set	711 reflections (4.36%)	DCC
Wilson B-factor (Å <sup>2</sup> )	69.7	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 50.9	EDS
Estimated twinning fraction	0.029 for -l,-k,-h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 17015 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2450	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.03% 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 ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 4ZG

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	0.42	0/2469	0.61	0/3342

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	2416	0	2441	7	0
2	A	29	0	27	1	0
3	A	5	0	0	0	0
All	All	2450	0	2468	8	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 2.

All (8) 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:A:883:LEU:HD22	1:A:953:ILE:HD12	1.89	0.55

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:1101:4ZG:N20	2:A:1101:4ZG:H3	2.23	0.53
1:A:984:ASP:O	1:A:985:GLU:HB2	2.14	0.48
1:A:746:GLU:HB3	1:A:787:GLN:HG2	1.96	0.47
1:A:878:ILE:HD13	1:A:920:ALA:HB1	1.99	0.43
1:A:802:VAL:HA	1:A:809:ILE:CD1	2.50	0.41
1:A:745:LYS:HD3	1:A:788:LEU:HD12	2.02	0.41
1:A:926:ILE:HG23	1:A:931:GLU:HB3	2.03	0.41

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	292/331 (88%)	275 (94%)	13 (4%)	4 (1%)	14	28

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	734	GLU
1	A	786	VAL
1	A	985	GLU
1	A	1018	ILE

### 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	269/287 (94%)	257 (96%)	12 (4%)	34 62

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	697	GLU
1	A	747	LEU
1	A	757	LYS
1	A	841	ARG
1	A	849	GLN
1	A	884	GLU
1	A	962	ARG
1	A	999	ARG
1	A	1005	GLU
1	A	1008	ASP
1	A	1011	VAL
1	A	1017	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 ⓘ

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	4ZG	A	1101	-	30,31,31	2.58	5 (16%)	30,47,47	2.92	10 (33%)

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	4ZG	A	1101	-	-	0/20/35/35	0/3/3/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1101	4ZG	C2-N3	-11.87	1.30	1.44
2	A	1101	4ZG	C2-N10	-3.34	1.39	1.46
2	A	1101	4ZG	C4-N3	-2.08	1.36	1.39
2	A	1101	4ZG	C19-N21	2.86	1.39	1.34
2	A	1101	4ZG	C29-S26	5.02	1.82	1.76

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1101	4ZG	N18-C19-N20	-5.75	120.53	126.66
2	A	1101	4ZG	C16-C17-N18	-3.42	119.88	123.88
2	A	1101	4ZG	C8-C7-N6	-2.87	119.98	123.12
2	A	1101	4ZG	C16-C15-N20	-2.45	118.80	123.22
2	A	1101	4ZG	C19-N20-C15	2.81	121.56	116.61
2	A	1101	4ZG	C5-N6-C7	4.29	122.58	117.89
2	A	1101	4ZG	C25-C23-S26	4.59	111.26	108.33
2	A	1101	4ZG	C24-C23-S26	4.99	111.52	108.33
2	A	1101	4ZG	C1-C2-N3	6.70	121.51	111.79
2	A	1101	4ZG	C17-N18-C19	7.54	121.24	115.42

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1101	4ZG	1	0

## 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

### 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	300/331 (90%)	0.29	32 (10%) <b>8</b> <b>4</b>	52, 76, 123, 150	6 (2%)

All (32) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1006	ASP	4.4
1	A	754	LYS	4.1
1	A	995	SER	3.6
1	A	888	HIS	3.6
1	A	890	ILE	3.5
1	A	722	ALA	3.4
1	A	988	HIS	3.3
1	A	999	ARG	3.3
1	A	997	PHE	3.2
1	A	1007	MET	3.1
1	A	751	THR	3.0
1	A	723	PHE	2.9
1	A	1005	GLU	2.9
1	A	994	ASP	2.8
1	A	1019	PRO	2.8
1	A	986	ARG	2.8
1	A	996	ASN	2.7
1	A	901	GLY	2.7
1	A	993	THR	2.7
1	A	903	THR	2.6
1	A	917	GLY	2.5
1	A	755	ALA	2.5
1	A	889	ARG	2.4
1	A	753	PRO	2.3
1	A	734	GLU	2.2
1	A	1004	GLU	2.2
1	A	721	GLY	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	998	TYR	2.1
1	A	1018	ILE	2.1
1	A	922	GLU	2.1
1	A	876	VAL	2.0
1	A	789	ILE	2.0

## 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](#)

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	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	4ZG	A	1101	29/29	0.95	0.17	-0.19	66,73,106,109	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.