

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

#### May 15, 2020 – 09:48 pm BST

PDB ID : 5AAC

Title: Structure of C1156Y Mutant Human Anaplastic Lymphoma Kinase in Com-

plex with Crizotinib

Authors : McTigue, M.; Deng, Y.; Liu, W.; Brooun, A.; Stewart, A.

Deposited on : 2015-07-23

Resolution : 1.70 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
A user guide is available at

https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity : 4.02b-467

Mogul: 1.8.5 (274361), CSD as541be (2020)

Xtriage (Phenix) : 1.13 EDS : 2.11

115/2

buster-report : 1.1.7 (2018)
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)

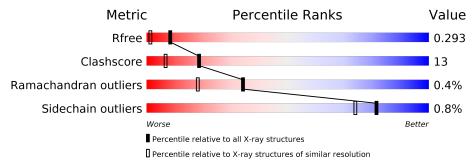
Validation Pipeline (wwPDB-VP) : 2.11

# 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 1.70 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	Similar resolution
Metric	$(\# \mathrm{Entries})$	$\mid \; (\#  ext{Entries},   ext{resolution range}( ext{Å})) \; \mid \;$
$R_{free}$	130704	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)

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 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%

Mol	Chain	Length	Quality of chain			
1	A	327	72%	17%	•	10%



# 2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2590 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 ALK TYROSINE KINASE RECEPTOR.

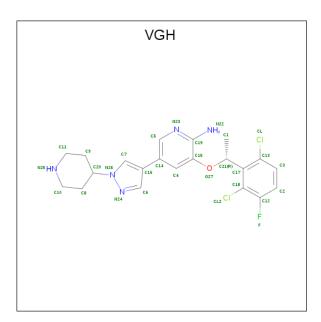
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	Α	293	Total	С	Ν	О	S	0	3	1
1	11	250	2330	1485	397	426	22			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1085	MET	-	expression tag	UNP Q9UM73
A	1086	ALA	-	expression tag	UNP Q9UM73
A	1087	HIS	-	expression tag	UNP Q9UM73
A	1088	HIS	-	expression tag	UNP Q9UM73
A	1089	HIS	_	expression tag	UNP Q9UM73
A	1090	HIS	_	expression tag	UNP Q9UM73
A	1091	HIS	=	expression tag	UNP Q9UM73
A	1092	HIS	=	expression tag	UNP Q9UM73
A	1156	TYR	CYS	engineered mutation	UNP Q9UM73

• Molecule 2 is 3-[(1R)-1-(2,6-dichloro-3-fluorophenyl)ethoxy]-5-(1-piperidin-4-yl-1H-pyrazol-4-yl)pyridin-2-amine (three-letter code: VGH) (formula: C<sub>21</sub>H<sub>22</sub>Cl<sub>2</sub>FN<sub>5</sub>O).





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf		
2	A	1	Total 30		Cl 2			O 1	0	0

### • Molecule 3 is water.

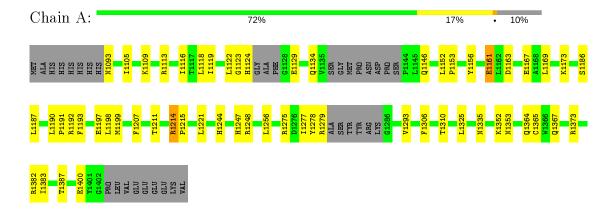
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	230	Total O 230 230	0	0



# 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 the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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: ALK TYROSINE KINASE RECEPTOR





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	51.54Å 57.45Å 104.79Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.66 - 1.70	Depositor
Resolution (A)	38.71 - 1.63	EDS
% Data completeness	99.9 (38.66-1.70)	Depositor
(in resolution range)	98.9 (38.71-1.63)	EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.15 (at 1.63Å)	Xtriage
Refinement program	CNX 2005	Depositor
P. P.	0.207 , $0.232$	Depositor
$R, R_{free}$	0.276 , $0.293$	DCC
$R_{free}$ test set	1205 reflections $(3.05\%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.6	Xtriage
Anisotropy	0.293	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.37, 50.2	EDS
L-test for twinning <sup>2</sup>	$ < L >=0.48, < L^2>=0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	2590	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 6.52% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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: VGH

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).

Mal	Mol Chain		lengths	Bond angles		
MIOI	Moi Chain	RMSZ	# Z >5	RMSZ	# Z  > 5	
1	A	0.26	0/2386	0.44	0/3233	

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	2330	0	2302	61	0
2	A	30	0	22	4	0
3	A	230	0	0	18	0
All	All	2590	0	2324	62	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 13.

All (62) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1 Atom-2		$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	$egin{array}{c}  ext{Clash} \  ext{overlap } ( ext{Å}) \end{array}$
1:A:1116:ILE:HD12	3:A:2046:HOH:O	1.58	1.02

Continued on next page...



Continued from previous page...

Atom 1		Interatomic	Clash
Atom-1	Atom-2	${ m distance}({ m \AA})$	$ \text{overlap } (\text{\AA})$
1:A:1093:ASN:HD22	1:A:1105:ILE:HB	1.42	0.82
1:A:1367:GLN:HB2	1:A:1373:ARG:HG2	1.71	0.71
1:A:1093:ASN:HB2	1:A:1105:ILE:HD13	1.73	0.70
1:A:1192:ARG:HA	3:A:2047:HOH:O	1.93	0.69
1:A:1365:CYS:O	1:A:1373:ARG:HD2	1.94	0.68
1:A:1248:ARG:CZ	1:A:1277:ILE:HG21	2.25	0.65
1:A:1146:GLN:HG3	3:A:2051:HOH:O	1.95	0.65
1:A:1113:ARG:HA	3:A:2046:HOH:O	1.96	0.65
1:A:1190:LEU:HA	3:A:2086:HOH:O	1.98	0.64
1:A:1191:PRO:HD3	3:A:2061:HOH:O	1.99	0.62
1:A:1190:LEU:HD23	3:A:2086:HOH:O	2.00	0.61
1:A:1119:ILE:HD11	1:A:1134:GLN:HG3	1.81	0.61
1:A:1146:GLN:HG2	1:A:1198:LEU:HD22	1.82	0.61
1:A:1211:THR:HG23	1:A:1221:LEU:HD22	1.81	0.61
1:A:1244:HIS:HD2	3:A:2007:HOH:O	1.82	0.61
1:A:1207:PHE:O	1:A:1211:THR:HG22	2.01	0.60
1:A:1187:LEU:HA	3:A:2047:HOH:O	2.03	0.57
1:A:1197:GLU:HG2	1:A:1199:MET:HE3	1.89	0.55
1:A:1105:ILE:N	1:A:1105:ILE:HD12	2.22	0.55
1:A:1279:ARG:HE	1:A:1279:ARG:HA	1.72	0.55
1:A:1382:ARG:HD3	3:A:2207:HOH:O	2.07	0.54
1:A:1146:GLN:HG2	1:A:1198:LEU:CD2	2.38	0.54
1:A:1105:ILE:HD12	1:A:1169:LEU:HD11	1.91	0.53
1:A:1199:MET:O	2:A:9000:VGH:H5	2.09	0.52
1:A:1152:LEU:HD21	1:A:1161:GLU:HG2	1.92	0.52
1:A:1198:LEU:O	1:A:1199:MET:HE3	2.10	0.51
1:A:1199:MET:HA	1:A:1199:MET:HE2	1.92	0.51
1:A:1122:LEU:HD22	2:A:9000:VGH:H92C	1.92	0.50
1:A:1214:ARG:HG3	1:A:1215:PRO:HD2	1.94	0.50
1:A:1247:HIS:O	1:A:1248:ARG:HB2	2.12	0.50
1:A:1214:ARG:HE	1:A:1400:GLU:HG3	1.77	0.49
1:A:1197:GLU:HG2	1:A:1199:MET:CE	2.43	0.48
1:A:1163:ASP:HA	1:A:1275:ARG:HD3	1.96	0.48
1:A:1325:LEU:N	1:A:1325:LEU:HD22	2.29	0.48
1:A:1118:LEU:HD12	1:A:1118:LEU:N	2.28	0.48
1:A:1256:LEU:HD11	2:A:9000:VGH:CL2	2.51	0.48
1:A:1383:ILE:O	1:A:1387:THR:HG23	2.14	0.48
1:A:1192:ARG:HD3	3:A:2047:HOH:O	2.13	0.48
1:A:1367:GLN:O	1:A:1373:ARG:HD3	2.13	0.48
1:A:1364:GLN:HG2	3:A:2022:HOH:O	2.15	0.47
1:A:1167:GLU:CD	3:A:2066:HOH:O	2.53	0.46

Continued on next page...



 $Continued\ from\ previous\ page...$ 

Atom-1	Atom 2	Interatomic	Clash
Atom-1	Atom-2	${ m distance}({ m \AA})$	overlap (Å)
1:A:1382:ARG:HG2	1:A:1382:ARG:HH11	1.80	0.46
1:A:1186:SER:HB3	1:A:1193:PHE:HB2	1.97	0.46
1:A:1278:TYR:O	1:A:1279:ARG:HB2	2.16	0.46
1:A:1199:MET:HA	1:A:1199:MET:CE	2.45	0.46
1:A:1248:ARG:NH2	1:A:1277:ILE:HG21	2.33	0.44
1:A:1153:PRO:HG2	1:A:1156:TYR:CZ	2.52	0.44
1:A:1105:ILE:HD12	1:A:1105:ILE:H	1.82	0.44
1:A:1293:VAL:HG21	1:A:1335:ASN:HB3	2.00	0.43
1:A:1109:LYS:HE3	3:A:2041:HOH:O	2.18	0.43
1:A:1163:ASP:CA	1:A:1275:ARG:HD3	2.48	0.43
1:A:1173:LYS:HE2	3:A:2034:HOH:O	2.17	0.43
1:A:1279:ARG:HA	1:A:1279:ARG:NE	2.32	0.42
1:A:1113:ARG:NE	3:A:2046:HOH:O	2.52	0.42
1:A:1192:ARG:HG2	3:A:2086:HOH:O	2.18	0.42
1:A:1352:LYS:O	1:A:1353:ASN:HB2	2.18	0.42
1:A:1093:ASN:ND2	1:A:1105:ILE:HB	2.22	0.41
2:A:9000:VGH:CL2	2:A:9000:VGH:H12C	2.57	0.41
1:A:1306:PHE:HD2	1:A:1310:THR:HG21	1.86	0.41
1:A:1124:HIS:HB3	1:A:1129:GLU:HA	2.02	0.41
1:A:1199:MET:HE1	3:A:2089:HOH:O	2.22	0.40

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	288/327 (88%)	280 (97%)	7 (2%)	1 (0%)	41 24

#### All (1) Ramachandran outliers are listed below:

$\mathbf{Mol}$	Chain	$\operatorname{Res}$	Type
1	A	1123	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	A	259/285 (91%)	257 (99%)	2 (1%)	81 74	

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1161	GLU
1	A	1214	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1093	ASN
1	A	1095	ASN
1	A	1115	ASN
1	A	1134	GLN
1	A	1159	GLN
1	A	1217	GLN
1	A	1244	HIS

#### 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	В	ond leng	$\operatorname{gths}$	В	ond ang	les
IVIOI	туре	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	VGH	A	9000	-	31,33,33	2.43	12 (38%)	36,47,47	1.62	6 (16%)

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	VGH	A	9000	-	-	0/12/24/24	0/4/4/4

All (12) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\operatorname{\AA})$	$Ideal(\AA)$
2	A	9000	VGH	C7-N26	5.49	1.41	1.35
2	A	9000	VGH	C18-C12	5.15	1.44	1.38
2	A	9000	VGH	C13-C17	5.06	1.46	1.39
2	A	9000	VGH	C18-C17	5.03	1.47	1.39
2	A	9000	VGH	C4-C15	3.34	1.44	1.38
2	A	9000	VGH	C19-N22	2.98	1.41	1.34
2	A	9000	VGH	C17-C21	2.53	1.57	1.53
2	A	9000	VGH	C5-C14	2.22	1.43	1.39
2	A	9000	VGH	O27-C15	2.20	1.41	1.37
2	A	9000	VGH	C2-C12	2.19	1.42	1.37
2	A	9000	VGH	C4-C14	2.14	1.43	1.39
2	A	9000	VGH	O27-C21	2.04	1.48	1.45

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$Observed(^o)$	$\operatorname{Ideal}({}^o)$
2	A	9000	VGH	O27-C21-C17	4.67	114.54	107.64
2	A	9000	VGH	C15-O27-C21	4.18	125.53	118.17

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^{o})$
2	A	9000	VGH	C17-C13-CL	3.35	123.73	120.49
2	A	9000	VGH	C13-C17-C18	2.68	118.00	114.90
2	A	9000	VGH	C17-C18-CL2	2.47	123.16	120.32
2	A	9000	VGH	C4-C15-C19	-2.12	115.84	119.25

There are no chirality outliers.

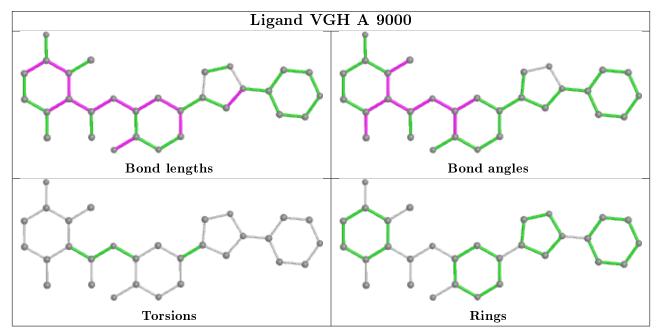
There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	9000	VGH	4	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.





# 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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

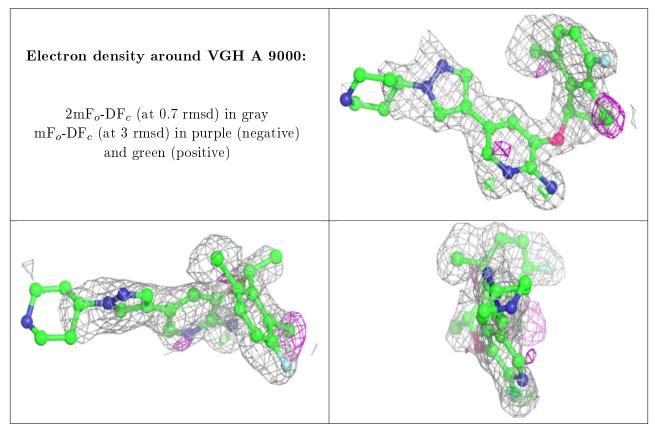
## 6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

## 6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.





# 6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

