



# wwPDB X-ray Structure Validation Summary Report i

Feb 26, 2014 – 11:31 PM GMT

PDB ID : 4C4G  
Title : Structure-based design of orally bioavailable pyrrolopyridine inhibitors of the mitotic kinase MPS1  
Authors : Naud, S.; Westwood, I.M.; Faisal, A.; Sheldrake, P.; Bavetsias, V.; Atrash, B.; Liu, M.; Hayes, A.; Schmitt, J.; Wood, A.; Choi, V.; Boxall, K.; Mak, G.; Gurden, M.; Valenti, M.; de Haven Brandon, A.; Henley, A.; Baker, R.; McAndrew, C.; Matijssen, B.; Burke, R.; Eccles, S.A.; Raynaud, F.I.; Linardopoulos, S.; van Montfort, R.; Blagg, J.  
Deposited on : 2013-09-05  
Resolution : 2.65 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>

---

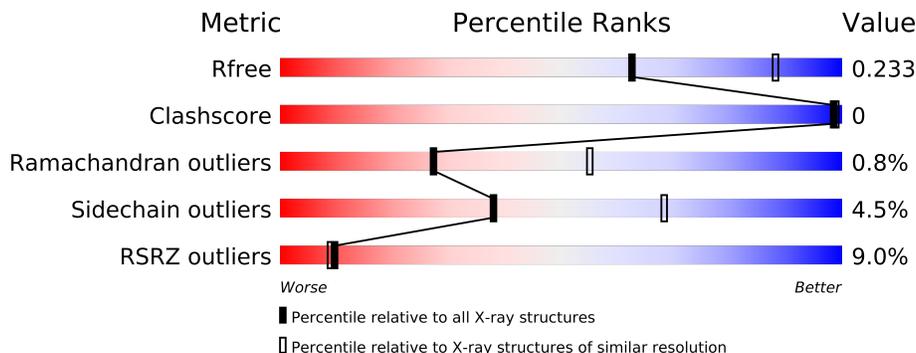
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.65 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	2232 (2.70-2.62)
Clashscore	79885	2700 (2.70-2.62)
Ramachandran outliers	78287	2657 (2.70-2.62)
Sidechain outliers	78261	2657 (2.70-2.62)
RSRZ outliers	66119	2234 (2.70-2.62)

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	313	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	7PE	A	1798	-	X
4	EDO	A	1799	-	X

## 2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 2016 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.

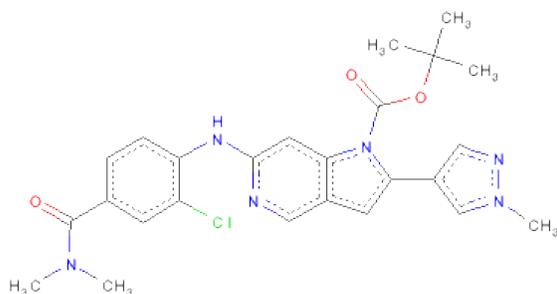
- Molecule 1 is a protein called DUAL SPECIFICITY PROTEIN KINASE TTK.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
1	A	252	1951	1261	317	361	1	11	0	0	0

There are 23 discrepancies between the modelled and reference sequences:

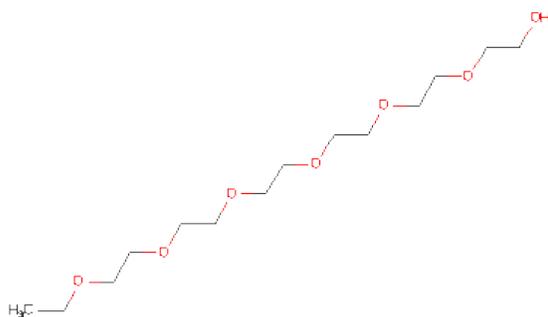
Chain	Residue	Modelled	Actual	Comment	Reference
A	496	MET	-	EXPRESSION TAG	UNP P33981
A	497	HIS	-	EXPRESSION TAG	UNP P33981
A	498	HIS	-	EXPRESSION TAG	UNP P33981
A	499	HIS	-	EXPRESSION TAG	UNP P33981
A	500	HIS	-	EXPRESSION TAG	UNP P33981
A	501	HIS	-	EXPRESSION TAG	UNP P33981
A	502	HIS	-	EXPRESSION TAG	UNP P33981
A	503	SER	-	EXPRESSION TAG	UNP P33981
A	504	SER	-	EXPRESSION TAG	UNP P33981
A	505	GLY	-	EXPRESSION TAG	UNP P33981
A	506	VAL	-	EXPRESSION TAG	UNP P33981
A	507	ASP	-	EXPRESSION TAG	UNP P33981
A	508	LEU	-	EXPRESSION TAG	UNP P33981
A	509	GLY	-	EXPRESSION TAG	UNP P33981
A	510	THR	-	EXPRESSION TAG	UNP P33981
A	511	GLU	-	EXPRESSION TAG	UNP P33981
A	512	ASN	-	EXPRESSION TAG	UNP P33981
A	513	LEU	-	EXPRESSION TAG	UNP P33981
A	514	TYR	-	EXPRESSION TAG	UNP P33981
A	515	PHE	-	EXPRESSION TAG	UNP P33981
A	516	GLN	-	EXPRESSION TAG	UNP P33981
A	517	SER	-	EXPRESSION TAG	UNP P33981
A	518	MET	-	EXPRESSION TAG	UNP P33981

- Molecule 2 is TERT-BUTYL 6-((2-CHLORO-4-(DIMETHYLCARBAMOYL)PHENYL)AMINO)-2-(1-METHYL-1H-PYRAZOL-4-YL)-1H-PYRROLO[3,2-C]PYRIDINE-1-CARBOXYLATE (three-letter code: 7RO) (formula: C<sub>25</sub>H<sub>27</sub>ClN<sub>6</sub>O<sub>3</sub>).



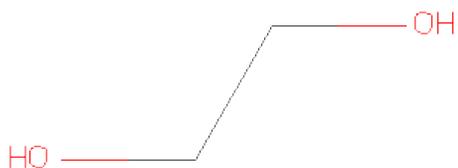
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	Cl	N	O		
2	A	1	35	25	1	6	3	0	0

- Molecule 3 is 2-(2-(2-(2-(2-(2-ETHOXYETHOXY)ETHOXY)ETHOXY)ETHOXY)ETHOXY)ETHANOL (three-letter code: 7PE) (formula: C<sub>14</sub>H<sub>30</sub>O<sub>7</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	A	1	7	4	3	0	0
3	A	1	7	4	3	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			4	2	2		
4	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	8	Total	O	0	0
			8	8		

### 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.

- Molecule 1: DUAL SPECIFICITY PROTEIN KINASE TTK

Chain A: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.93Å 111.90Å 112.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.04 – 2.65 41.04 – 2.65	Depositor EDS
% Data completeness (in resolution range)	98.8 (41.04-2.65) 98.8 (41.04-2.65)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.90 (at 2.65Å)	Xtriage
Refinement program	BUSTER-TNT 2.11.4	Depositor
R, $R_{free}$	0.193 , 0.223 0.206 , 0.233	Depositor DCC
$R_{free}$ test set	630 reflections (4.77%)	DCC
Wilson B-factor (Å <sup>2</sup> )	85.0	Xtriage
Anisotropy	0.492	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 101.4	EDS
Estimated twinning fraction	0.022 for -h,-l,-k	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 13227 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2016	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	103.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.40% of the height of the origin peak. No significant pseudotranslation is detected.*

<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: TPO, 7PE, EDO, 7RO

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	0.51	0/1984	0.68	0/2701

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 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 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	1951	0	0	1	0
2	A	35	0	27	1	0
3	A	14	0	16	0	0
4	A	8	0	12	0	0
5	A	8	0	0	0	0
All	All	2016	0	55	1	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 0.

All (1) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:603:GLU:O	2:A:1796:7RO:H7	2.14	0.48

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	245/313 (78%)	237 (97%)	6 (2%)	2 (1%)	27 55

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	687	VAL
1	A	696	LYS

### 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	198/287 (69%)	189 (96%)	9 (4%)	38 68

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	571	GLU
1	A	773	LYS
1	A	625	LYS
1	A	562	ASN
1	A	606	ASN

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 chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

1 non-standard protein/DNA/RNA residue 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
1	TPO	A	686	1	10,10,11	5.95	4 (40%)	12,14,16	1.77	3 (25%)

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
1	TPO	A	686	1	-	1/9/11/13	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	686	TPO	O-C	18.16	1.23	1.11
1	A	686	TPO	CA-C	3.38	1.54	1.48
1	A	686	TPO	CB-CA	2.39	1.57	1.53
1	A	686	TPO	OG1-CB	2.07	1.49	1.45

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	686	TPO	OG1-CB-CA	3.94	114.14	107.55

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	686	TPO	OG1-P-O1P	2.37	113.42	106.79
1	A	686	TPO	CB-CA-N	2.31	112.82	109.60

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	686	TPO	OG1-CB-CA-N

There are no ring outliers.

## 5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry (i)

5 ligands are 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	7RO	A	1796	-	38,38,38	2.07	9 (23%)	57,57,57	1.74	5 (8%)
3	7PE	A	1797	-	6,6,20	0.71	0	5,5,19	0.78	0
3	7PE	A	1798	-	6,6,20	0.63	0	5,5,19	0.78	0
4	EDO	A	1799	-	3,3,3	0.65	0	2,2,2	0.63	0
4	EDO	A	1800	-	3,3,3	0.68	0	2,2,2	0.44	0

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	7RO	A	1796	-	-	0/21/25/25	0/2/4/4
3	7PE	A	1797	-	-	0/4/4/18	0/0/0/0
3	7PE	A	1798	-	-	0/4/4/18	0/0/0/0
4	EDO	A	1799	-	-	0/1/1/1	0/0/0/0
4	EDO	A	1800	-	-	0/1/1/1	0/0/0/0

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1796	7RO	C13-N4	5.50	1.42	1.34
2	A	1796	7RO	C4-N5	4.98	1.45	1.37
2	A	1796	7RO	O2-C20	4.21	1.23	1.19
2	A	1796	7RO	C8-N2	4.03	1.42	1.34
2	A	1796	7RO	C1-C3	3.72	1.44	1.38

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1796	7RO	C8-C18-C19	-7.89	115.90	120.91
2	A	1796	7RO	O1-C20-N5	5.21	115.08	109.09
2	A	1796	7RO	C4-C5-C6	4.69	110.00	106.17
2	A	1796	7RO	C12-C11-C10	-2.63	117.62	120.11
2	A	1796	7RO	C3-C4-N5	2.33	126.92	122.88

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	252/313 (80%)	0.56	23 (9%) <b>9</b> <b>9</b>	71, 102, 139, 163	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	748	ILE	4.8
1	A	749	ASP	4.1
1	A	522	GLY	3.7
1	A	729	TYR	3.7
1	A	750	PRO	3.1

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

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
1	TPO	A	686	11/12	0.27	0.13	127,132,144,149	0

### 6.3 Carbohydrates

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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	EDO	A	1799	4/4	0.25	2.54	90,91,91,92	0
3	7PE	A	1798	7/21	0.35	2.19	104,107,111,112	0
3	7PE	A	1797	7/21	0.30	1.51	99,100,101,102	0
4	EDO	A	1800	4/4	0.33	0.96	112,113,113,113	0
2	7RO	A	1796	35/35	0.20	-0.12	72,82,92,94	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.