



# Full wwPDB X-ray Structure Validation Report i

May 28, 2014 – 12:40 AM EDT

PDB ID : 4PPB  
Title : ITK kinase domain with compound 28 (N-{1-[(1S)-3-(DIMETHYLAMINO)-1-PHENYLPROPYL]-1H-PYRAZOL-4-YL}-6-(1H-PYRAZOL-4-YL)-1H-INDAZOLE-3-CARBOXAMIDE)  
Authors : Eigenbrot, C.; Shia, S.  
Deposited on : 2014-02-26  
Resolution : 2.82 Å(reported)

This is a full wwPDB 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/ValidationPDFNotes.html>

---

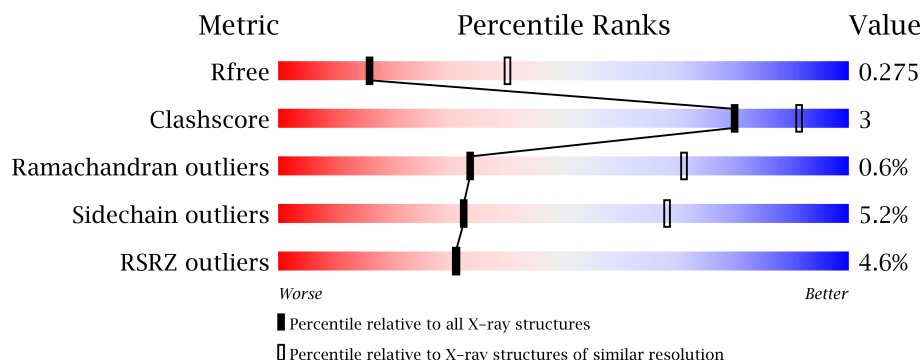
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	<b>FAILED</b>
Xtriage (Phenix)	:	dev-1439
EDS	:	stable23161
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)	:	stable23161

# 1 Overall quality at a glance

The reported resolution of this entry is 2.82 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	1963 (2.84-2.80)
Clashscore	79885	2478 (2.84-2.80)
Ramachandran outliers	78287	2429 (2.84-2.80)
Sidechain outliers	78261	2431 (2.84-2.80)
RSRZ outliers	66119	1966 (2.84-2.80)

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	266	
1	B	266	

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 3888 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.

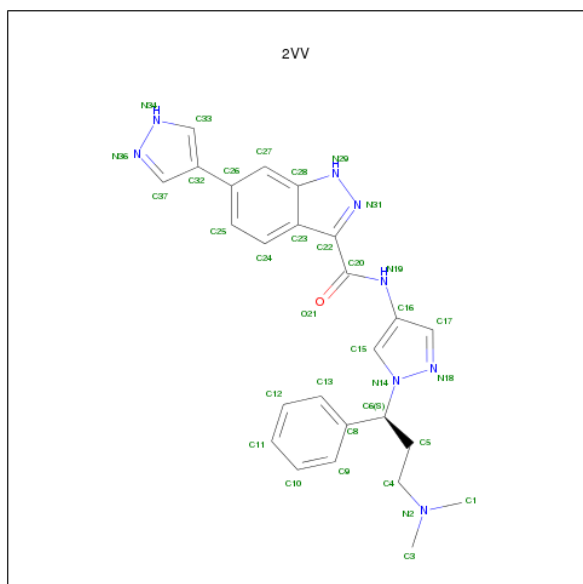
- Molecule 1 is a protein called Tyrosine-protein kinase ITK/TSK.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	239	Total	C	N	O	S	0	0	0
			1910	1223	317	355	15			
1	B	239	Total	C	N	O	S	0	0	0
			1910	1223	317	355	15			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	356	SER	-	EXPRESSION TAG	UNP Q08881
B	356	SER	-	EXPRESSION TAG	UNP Q08881

- Molecule 2 is N-{1-[(1S)-3-(DIMETHYLAMINO)-1-PHENYLPROPYL]-1H-PYRAZOL-4-YL}-6-(1H-PYRAZOL-4-YL)-1H-INDAZOLE-3-CARBOXAMIDE (three-letter code: 2VV) (formula: C<sub>25</sub>H<sub>26</sub>N<sub>8</sub>O).



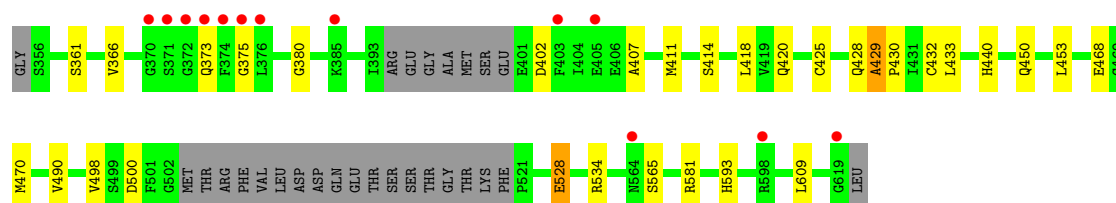
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			34	25	8	1		
2	B	1	Total	C	N	O	0	0
			34	25	8	1		

### 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 ( $RSRZ > 2$ ). 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.

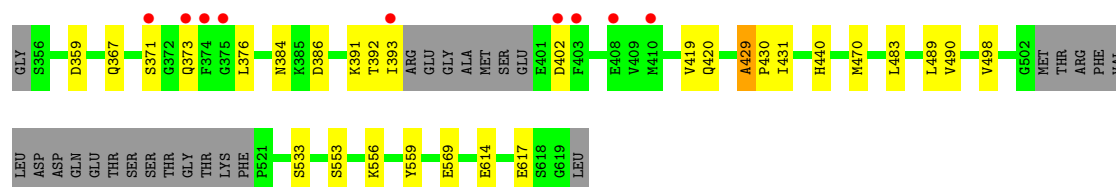
- Molecule 1: Tyrosine-protein kinase ITK/TSK

Chain A: 



- Molecule 1: Tyrosine-protein kinase ITK/TSK

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	39.94Å 94.41Å 156.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.40 – 2.82 40.40 – 2.82	Depositor EDS
% Data completeness (in resolution range)	98.2 (40.40-2.82) 98.2 (40.40-2.82)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.24 (at 2.81Å)	Xtriage
Refinement program	BUSTER 2.11.4	Depositor
R, $R_{free}$	0.211 , 0.266 0.222 , 0.275	Depositor DCC
$R_{free}$ test set	745 reflections (5.35%)	DCC
Wilson B-factor (Å <sup>2</sup> )	57.7	Xtriage
Anisotropy	0.424	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 41.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	1 of 14680 reflections (0.007%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3888	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 67.78 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 4.8690e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: 2VV

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	0.43	0/1953	0.62	0/2637
1	B	0.44	0/1953	0.64	0/2637
All	All	0.43	0/3906	0.63	0/5274

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	1910	0	1870	14	0
1	B	1910	0	1870	12	0
2	A	34	0	0	1	0
2	B	34	0	0	0	0
All	All	3888	0	3740	26	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 3.

All (26) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:429:ALA:HB1	1:B:430:PRO:HD3	1.65	0.79
1:A:470:MET:HE3	1:A:498:VAL:HG11	1.70	0.72
1:A:468:GLU:HG3	1:A:609:LEU:HD11	1.84	0.60
1:B:429:ALA:CB	1:B:430:PRO:HD3	2.32	0.59
1:B:419:VAL:HG21	1:B:489:LEU:HD12	1.85	0.57
1:A:450:GLN:O	1:A:453:LEU:HB2	2.07	0.54
1:A:429:ALA:HB1	1:A:430:PRO:CD	2.38	0.53
1:A:440:HIS:HB2	1:A:490:VAL:HB	1.92	0.52
1:A:375:GLY:HA3	1:B:391:LYS:HE2	1.92	0.51
1:B:429:ALA:CB	1:B:430:PRO:CD	2.90	0.49
1:A:418:LEU:HD23	1:A:498:VAL:HB	1.95	0.49
1:A:429:ALA:HB1	1:A:430:PRO:HD3	1.94	0.49
1:A:425:CYS:HB3	1:A:432:CYS:HB2	1.96	0.47
1:A:414:SER:HA	1:A:420:GLN:HE22	1.80	0.47
1:B:371:SER:HB3	1:B:376:LEU:HD23	1.95	0.47
1:A:407:ALA:O	1:A:411:MET:HG2	2.18	0.44
1:A:450:GLN:HG3	1:A:453:LEU:HD22	2.00	0.44
1:B:553:SER:HB3	1:B:556:LYS:HB2	1.99	0.43
1:B:470:MET:CE	1:B:498:VAL:HG11	2.49	0.43
1:B:470:MET:HE3	1:B:498:VAL:HG11	2.00	0.42
1:B:470:MET:HE3	1:B:483:LEU:HD13	2.01	0.42
1:B:440:HIS:HB2	1:B:490:VAL:HB	2.01	0.42
1:A:366:VAL:HG23	1:A:380:GLY:HA2	2.01	0.42
1:A:528:GLU:HB2	1:A:534:ARG:HB2	2.01	0.41
1:B:392:THR:HG22	1:B:393:ILE:HG22	2.03	0.41
2:A:701:2VV:C17	2:A:701:2VV:O21	2.68	0.40

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	233/266 (88%)	219 (94%)	12 (5%)	2 (1%)	25	61
1	B	233/266 (88%)	220 (94%)	12 (5%)	1 (0%)	43	79
All	All	466/532 (88%)	439 (94%)	24 (5%)	3 (1%)	33	71



All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	429	ALA
1	A	429	ALA
1	A	428	GLN

### 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	210/233 (90%)	201 (96%)	9 (4%)	40	76
1	B	210/233 (90%)	197 (94%)	13 (6%)	26	59
All	All	420/466 (90%)	398 (95%)	22 (5%)	32	68

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	361	SER
1	A	373	GLN
1	A	402	ASP
1	A	433	LEU
1	A	500	ASP
1	A	528	GLU
1	A	565	SER
1	A	581	ARG
1	A	593	HIS
1	B	359	ASP
1	B	367	GLN
1	B	373	GLN
1	B	384	ASN
1	B	386	ASP
1	B	402	ASP
1	B	420	GLN
1	B	431	ILE
1	B	533	SER
1	B	559	TYR
1	B	569	GLU
1	B	614	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	617	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	420	GLN
1	B	384	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

Mogul failed to run properly - this section will therefore be empty.

## 5.5 Carbohydrates ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.6 Ligand geometry ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.7 Other polymers ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 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	239/266 (89%)	0.08	13 (5%)	25 25	36, 59, 117, 149	0
1	B	239/266 (89%)	0.09	9 (3%)	38 39	37, 60, 115, 135	0
All	All	478/532 (89%)	0.08	22 (4%)	31 31	36, 60, 116, 149	0

All (22) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	374	PHE	8.4
1	A	375	GLY	5.9
1	B	375	GLY	4.5
1	B	374	PHE	4.2
1	A	373	GLN	4.2
1	B	373	GLN	4.1
1	A	405	GLU	3.9
1	B	403	PHE	3.7
1	A	403	PHE	3.3
1	B	393	ILE	3.3
1	A	619	GLY	2.9
1	A	564	ASN	2.8
1	A	598	ARG	2.7
1	B	371	SER	2.6
1	B	408	GLU	2.6
1	A	371	SER	2.5
1	A	376	LEU	2.5
1	B	410	MET	2.5
1	A	372	GLY	2.5
1	A	370	GLY	2.4
1	A	385	LYS	2.0
1	B	402	ASP	2.0

## 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	2VV	A	701	34/34	0.23	0.49	38,44,57,57	0
2	2VV	B	701	34/34	0.23	0.21	43,52,66,67	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.