



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:45 PM GMT

PDB ID : 4PUZ  
Title : Crystal structure of spleen tyrosine kinase (Syk) in complex with GS-9973  
Authors : Lansdon, E.B.; Mitchell, S.A.  
Deposited on : 2014-03-14  
Resolution : 2.08 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

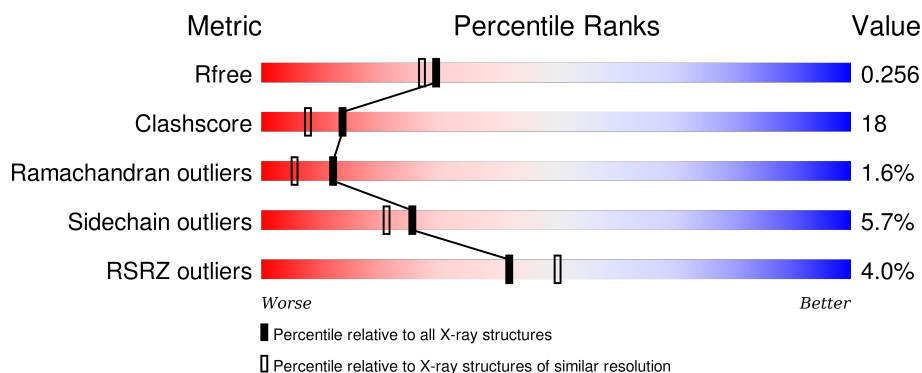
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.08 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	4546 (2.10-2.06)
Clashscore	102246	5101 (2.10-2.06)
Ramachandran outliers	100387	5048 (2.10-2.06)
Sidechain outliers	100360	5049 (2.10-2.06)
RSRZ outliers	91569	4556 (2.10-2.06)

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	291	<div> <div>3%</div> <div>57%</div> <div>30%</div> <div>•</div> <div>10%</div> </div>
1	B	291	<div> <div>4%</div> <div>60%</div> <div>29%</div> <div>•</div> <div>9%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4479 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.

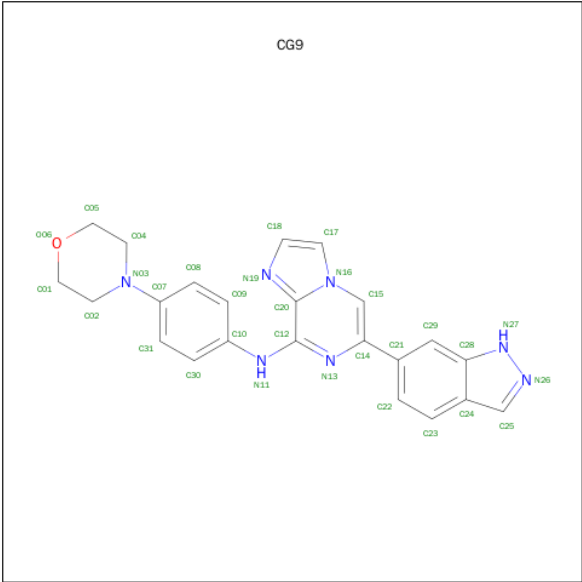
- Molecule 1 is a protein called Tyrosine-protein kinase SYK.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	263	Total	C	N	O	S	0	0	0
			2145	1381	359	386	19			
1	B	265	Total	C	N	O	S	0	0	0
			2160	1390	362	389	19			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	353	MET	ALA	CONFLICT	UNP P43405
A	354	ALA	ASP	CONFLICT	UNP P43405
A	355	LEU	PRO	CONFLICT	UNP P43405
A	636	GLU	-	EXPRESSION TAG	UNP P43405
A	637	GLY	-	EXPRESSION TAG	UNP P43405
A	638	HIS	-	EXPRESSION TAG	UNP P43405
A	639	HIS	-	EXPRESSION TAG	UNP P43405
A	640	HIS	-	EXPRESSION TAG	UNP P43405
A	641	HIS	-	EXPRESSION TAG	UNP P43405
A	642	HIS	-	EXPRESSION TAG	UNP P43405
A	643	HIS	-	EXPRESSION TAG	UNP P43405
B	353	MET	ALA	CONFLICT	UNP P43405
B	354	ALA	ASP	CONFLICT	UNP P43405
B	355	LEU	PRO	CONFLICT	UNP P43405
B	636	GLU	-	EXPRESSION TAG	UNP P43405
B	637	GLY	-	EXPRESSION TAG	UNP P43405
B	638	HIS	-	EXPRESSION TAG	UNP P43405
B	639	HIS	-	EXPRESSION TAG	UNP P43405
B	640	HIS	-	EXPRESSION TAG	UNP P43405
B	641	HIS	-	EXPRESSION TAG	UNP P43405
B	642	HIS	-	EXPRESSION TAG	UNP P43405
B	643	HIS	-	EXPRESSION TAG	UNP P43405

- Molecule 2 is 6-(1H-INDAZOL-6-YL)-N-[4-(MORPHOLIN-4-YL)PHENYL]IMIDAZO[1,2-A]PYRAZIN-8-AMINE (three-letter code: CG9) (formula: C<sub>23</sub>H<sub>21</sub>N<sub>7</sub>O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			31	23	7	1		
2	B	1	Total	C	N	O	0	0
			31	23	7	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	48	Total	O	0	0
			48	48		
3	B	64	Total	O	0	0
			64	64		

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.

Chain A:

57% 30% 3% 10%

Sequence logo for Chain A (48 residues):

Residue	Color Group	Red Dot
V433	Yellow	No
A441	Yellow	No
E442	Yellow	No
S443	Yellow	No
W444	Yellow	No
M450	Yellow	No
A451	Yellow	No
E452	Yellow	No
Y459	Yellow	No
Q462	Green	Yes
N463	Green	No
H465	Orange	Yes
A466	Orange	No
K467	Orange	No
D468	Orange	No
I471	Orange	No
I472	Orange	No
E473	Orange	No
Q477	Orange	No
N482	Orange	No
K483	Orange	No
E486	Orange	No
E487	Orange	No
F490	Orange	No
V491	Orange	No
H492	Orange	No
K493	Orange	No
Q505	Orange	Yes
H506	Orange	No
Y507	Orange	No
A508	Orange	No
K509	Orange	No
Y510	Orange	No
S516	Orange	No
Y517	Orange	No
A518	Orange	No
L519	Orange	No
R520	Orange	No
A521	Orange	No
D522	Orange	No
Q529	Orange	Yes
THR	Grey	No
HIS	Grey	No
GLY	Grey	No
K533	Grey	Yes
H534	Grey	No
D535	Grey	No

Chain B:

Amino Acid	Count
MET	1
ALA	1
LEU	1
GLU	1
GLU	1
ILE	1
ARG	1
PRO	1
LYS	1
GLU	1
V363	1
D366	1
R367	1
K368	1
L369	1
L370	1
T371	1
D374	1
L377	1
G378	1
N381	1
F382	1
G383	1
K386	1
K387	1
G388	1
Q391	1
M392	1
K393	1
K394	1
V395	1
V396	1
L404	1
K405	1
ASN	1
GLU	1
ALA	1
ASN	1
ASP	1
P411	1
A412	1
L413	1
K414	1
D415	1
E416	1
L417	1
L418	1
A419	1
E420	1
A421	1
N422	1
Y423	1
M424	1
Q425	1
Q426	1
N429	1
P430	1
M435	1
I436	1
G437	1
I438	1
A441	1
E442	1
S443	1
N444	1
N445	1
L446	1
E449	1
M450	1
A451	1
E452	1
N463	1
R464	1
H465	1
V466	1
K467	1
D468	1
K469	1
I472	1
H476	1
K483	1
Y484	1
L485	1
E486	1
E487	1
H492	1
R493	1
V500	1
L501	1
L502	1
V503	1
T504	1
Y507	1
L515	1
S516	1
K517	1
N524	1
Y525	1
Y526	1

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	39.88Å 41.88Å 87.41Å 98.67° 90.43° 101.38°	Depositor
Resolution (Å)	43.17 – 2.08 43.17 – 2.09	Depositor EDS
% Data completeness (in resolution range)	81.3 (43.17-2.08) 78.7 (43.17-2.09)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.80 (at 2.08Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, $R_{free}$	0.209 , 0.261 0.208 , 0.256	Depositor DCC
$R_{free}$ test set	2000 reflections (8.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	33.1	Xtriage
Anisotropy	0.675	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 48.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 28645 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4479	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.87% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CG9

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	0.44	0/2193	0.59	0/2952
1	B	0.46	0/2208	0.59	0/2973
All	All	0.45	0/4401	0.59	0/5925

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	2145	0	2143	79	0
1	B	2160	0	2158	81	0
2	A	31	0	21	3	0
2	B	31	0	21	4	0
3	A	48	0	0	3	0
3	B	64	0	0	3	0
All	All	4479	0	4343	161	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 18.

The worst 5 of 161 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:414:LYS:HE3	1:A:444:TRP:NE1	1.75	0.99
1:A:382:PHE:HA	1:A:413:LEU:HD23	1.51	0.92
1:A:414:LYS:HE3	1:A:444:TRP:HE1	1.33	0.88
1:B:594:PRO:HG2	1:B:597:CYS:HB2	1.61	0.83
1:A:450:MET:HG3	1:A:452:GLU:HG2	1.62	0.80

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	257/291 (88%)	233 (91%)	18 (7%)	6 (2%)	8	3
1	B	259/291 (89%)	243 (94%)	14 (5%)	2 (1%)	24	17
All	All	516/582 (89%)	476 (92%)	32 (6%)	8 (2%)	12	5

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	412	ALA
1	A	467	LYS
1	B	394	LYS
1	A	595	ALA
1	A	628	ASN

### 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	229/253 (90%)	215 (94%)	14 (6%)	23	18
1	B	231/253 (91%)	219 (95%)	12 (5%)	29	25
All	All	460/506 (91%)	434 (94%)	26 (6%)	25	21

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	507	TYR
1	B	368	LYS
1	B	524	ASN
1	A	577	LYS
1	A	599	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	505	GLN
1	B	465	HIS
1	A	506	HIS
1	A	470	ASN
1	B	463	ASN

### 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

2 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	CG9	A	701	-	29,36,36	1.20	3 (10%)	35,51,51	1.25	3 (8%)
2	CG9	B	701	-	29,36,36	1.24	2 (6%)	35,51,51	1.21	3 (8%)

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	CG9	A	701	-	-	0/12/20/20	0/6/6/6
2	CG9	B	701	-	-	0/12/20/20	0/6/6/6

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	701	CG9	C15-C14	-2.74	1.35	1.39
2	B	701	CG9	C15-C14	-2.61	1.35	1.39
2	A	701	CG9	C23-C22	2.25	1.41	1.36
2	A	701	CG9	C14-N13	3.86	1.40	1.34
2	B	701	CG9	C14-N13	4.45	1.41	1.34

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	701	CG9	C21-C29-C28	-3.87	116.79	120.80
2	A	701	CG9	C18-C17-N16	-3.52	103.97	107.02
2	B	701	CG9	C21-C29-C28	-2.92	117.77	120.80
2	B	701	CG9	C18-C17-N16	-2.41	104.93	107.02
2	A	701	CG9	C22-C21-C29	2.03	121.20	118.11

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	701	CG9	3	0
2	B	701	CG9	4	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 [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	263/291 (90%)	0.28	10 (3%)	44 52	34, 50, 75, 88	0
1	B	265/291 (91%)	0.12	11 (4%)	40 48	33, 47, 71, 87	0
All	All	528/582 (90%)	0.20	21 (3%)	42 50	33, 49, 74, 88	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	630	TYR	6.7
1	B	631	TYR	5.2
1	B	415	ASP	4.1
1	A	624	LEU	3.9
1	A	631	TYR	3.9

### 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( $\text{\AA}^2$ )	Q<0.9
2	CG9	A	701	31/31	0.95	0.12	-0.17	40,47,64,70	0
2	CG9	B	701	31/31	0.94	0.12	-0.24	36,45,56,59	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.