



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 07:01 AM GMT

PDB ID : 2Z72  
Title : New Structure Of Cold-Active Protein Tyrosine Phosphatase At 1.1 Angstrom  
Authors : Tsuruta, H.; Mikami, B.; Yamamoto, C.; Yamagata, H.  
Deposited on : 2007-08-10  
Resolution : 1.10 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 (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

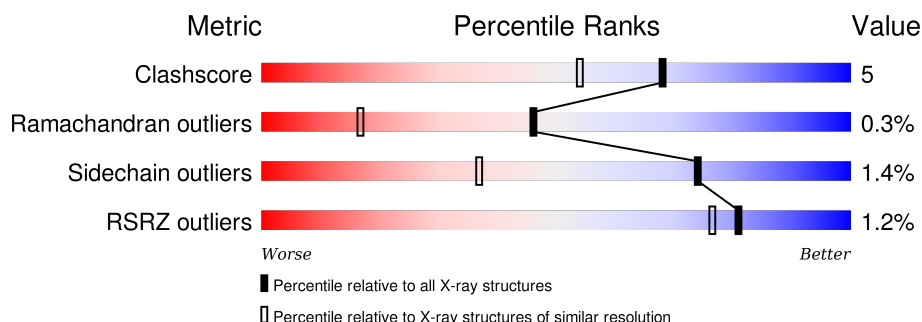
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.10 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))
Clashscore	102246	1055 (1.14-1.06)
Ramachandran outliers	100387	1016 (1.14-1.06)
Sidechain outliers	100360	1014 (1.14-1.06)
RSRZ outliers	91569	1009 (1.14-1.06)

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	342	<div> <div></div> <div>83%</div> <div>14%</div> <div>..</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3657 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 Protein-tyrosine-phosphatase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	338	Total	C	N	O	S	0	39	0
			2883	1838	486	544	15			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP Q9S427
A	2	GLY	-	EXPRESSION TAG	UNP Q9S427
A	182	SER	GLY	SEE REMARK 999	UNP Q9S427

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Zn	0	0
			2	2		

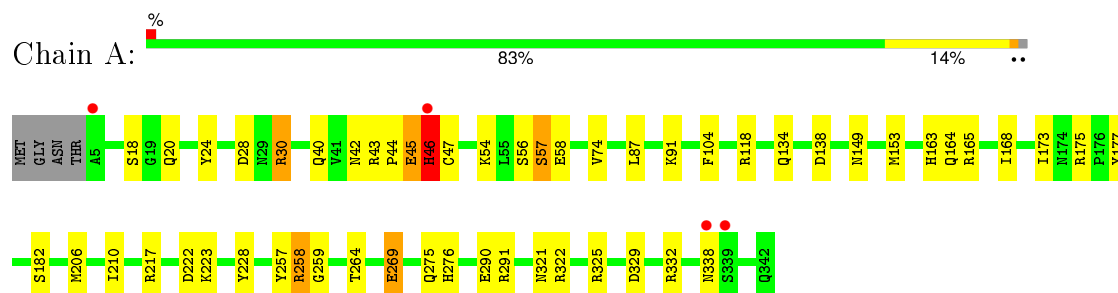
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	772	Total	O	0	0
			772	772		

### 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: Protein-tyrosine-phosphatase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	55.56Å 77.31Å 81.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 1.10 8.00 – 1.10	Depositor EDS
% Data completeness (in resolution range)	84.7 (8.00-1.10) 81.1 (8.00-1.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.31 (at 1.10Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, $R_{free}$	0.116 , 0.161 0.124 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	9.4	Xtriage
Anisotropy	0.709	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 69.9	EDS
Estimated twinning fraction	0.013 for -h,l,k	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 119401 reflections	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	3657	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.60% 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: ZN

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	0.83	4/3083 (0.1%)	1.46	41/4161 (1.0%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	332	ARG	CD-NE	12.05	1.67	1.46
1	A	332	ARG	CG-CD	5.89	1.66	1.51
1	A	24	TYR	CE2-CZ	-5.09	1.31	1.38
1	A	163	HIS	ND1-CE1	-5.06	1.22	1.34

All (41) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	258	ARG	NE-CZ-NH1	24.21	132.41	120.30
1	A	43	ARG	CD-NE-CZ	13.09	141.93	123.60
1	A	332	ARG	NE-CZ-NH2	10.54	125.57	120.30
1	A	43	ARG	NE-CZ-NH1	10.38	125.49	120.30
1	A	332	ARG	CG-CD-NE	-10.19	90.41	111.80
1	A	258	ARG	NE-CZ-NH2	-9.71	115.45	120.30
1	A	258	ARG	CD-NE-CZ	9.56	136.99	123.60
1	A	269	GLU	CB-CG-CD	8.73	137.77	114.20
1	A	291	ARG	NE-CZ-NH2	-8.07	116.26	120.30
1	A	58[A]	GLU	OE1-CD-OE2	-8.07	113.62	123.30
1	A	58[B]	GLU	OE1-CD-OE2	-8.07	113.62	123.30
1	A	30	ARG	NE-CZ-NH2	-7.99	116.30	120.30
1	A	164	GLN	O-C-N	-7.37	110.90	122.70
1	A	28	ASP	CB-CG-OD2	-7.09	111.92	118.30
1	A	30	ARG	CD-NE-CZ	7.01	133.41	123.60
1	A	325	ARG	NE-CZ-NH2	-6.98	116.81	120.30
1	A	138	ASP	CB-CG-OD2	6.76	124.39	118.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	258	ARG	NH1-CZ-NH2	-6.59	112.15	119.40
1	A	163	HIS	CG-ND1-CE1	6.47	117.26	108.20
1	A	332	ARG	CD-NE-CZ	-6.45	114.57	123.60
1	A	46[A]	HIS	CA-CB-CG	6.20	124.14	113.60
1	A	46[B]	HIS	CA-CB-CG	6.20	124.14	113.60
1	A	28	ASP	CB-CG-OD1	5.93	123.64	118.30
1	A	332	ARG	NH1-CZ-NH2	-5.84	112.98	119.40
1	A	322	ARG	NE-CZ-NH2	-5.82	117.39	120.30
1	A	165	ARG	NE-CZ-NH1	5.77	123.19	120.30
1	A	177	TYR	CB-CG-CD2	5.76	124.46	121.00
1	A	217	ARG	NE-CZ-NH1	5.61	123.10	120.30
1	A	222	ASP	CB-CG-OD2	5.58	123.32	118.30
1	A	228	TYR	CB-CG-CD1	5.54	124.32	121.00
1	A	18	SER	C-N-CA	-5.48	110.80	122.30
1	A	332	ARG	CB-CG-CD	-5.44	97.46	111.60
1	A	329	ASP	CB-CG-OD1	-5.38	113.46	118.30
1	A	45[A]	GLU	O-C-N	-5.33	114.18	122.70
1	A	45[B]	GLU	O-C-N	-5.33	114.18	122.70
1	A	325	ARG	CD-NE-CZ	5.29	131.01	123.60
1	A	165	ARG	N-CA-CB	5.25	120.06	110.60
1	A	264	THR	CA-CB-CG2	-5.21	105.10	112.40
1	A	321	ASN	CA-CB-CG	-5.21	101.94	113.40
1	A	57[A]	SER	O-C-N	-5.10	114.54	122.70
1	A	57[B]	SER	O-C-N	-5.10	114.54	122.70

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	2883	0	2851	27	0
2	A	2	0	0	0	0
3	A	772	0	0	17	0
All	All	3657	0	2851	27	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 5.

All (27) 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:87:LEU:O	1:A:91[A]:LYS:HG3	1.95	0.66
1:A:276:HIS:HD2	3:A:1088:HOH:O	1.87	0.57
1:A:223[A]:LYS:HD3	3:A:1570:HOH:O	2.05	0.56
1:A:91[A]:LYS:HE2	3:A:1188:HOH:O	2.05	0.55
1:A:44:PRO:HG2	1:A:47[A]:CYS:SG	2.46	0.55
1:A:40[B]:GLN:HG3	1:A:42:ASN:OD1	2.08	0.54
1:A:168:ILE:HD11	3:A:1500:HOH:O	2.10	0.51
1:A:257:TYR:CZ	1:A:259:GLY:HA3	2.46	0.51
1:A:275:GLN:HG3	3:A:1683:HOH:O	2.12	0.50
1:A:91[A]:LYS:HE2	3:A:1396:HOH:O	2.13	0.48
1:A:45[B]:GLU:HG3	1:A:46[B]:HIS:CD2	2.47	0.48
1:A:91[A]:LYS:NZ	3:A:1754:HOH:O	2.44	0.47
1:A:149:ASN:O	1:A:153[B]:MET:HG3	2.15	0.47
1:A:91[A]:LYS:NZ	3:A:1125:HOH:O	2.48	0.46
1:A:57[B]:SER:HB2	3:A:1764:HOH:O	2.15	0.46
1:A:206[A]:MET:SD	1:A:210[A]:ILE:HG12	2.56	0.45
1:A:269:GLU:HG2	3:A:1637:HOH:O	2.14	0.45
1:A:258:ARG:NH1	3:A:1359:HOH:O	2.49	0.45
1:A:175:ARG:NH1	3:A:1744:HOH:O	2.47	0.44
1:A:118[B]:ARG:NH2	3:A:1588:HOH:O	2.50	0.44
1:A:30:ARG:NE	3:A:1622:HOH:O	2.44	0.43
1:A:57[B]:SER:HB3	3:A:1581:HOH:O	2.18	0.43
1:A:258:ARG:NE	3:A:1358:HOH:O	2.49	0.43
1:A:173:ILE:HG22	1:A:175:ARG:NE	2.34	0.43
1:A:20[B]:GLN:NE2	3:A:1631:HOH:O	2.51	0.43
1:A:56[A]:SER:OG	1:A:134:GLN:OE1	2.28	0.42
1:A:74[B]:VAL:HG13	1:A:104:PHE:CZ	2.56	0.41

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	375/342 (110%)	364 (97%)	10 (3%)	1 (0%)	46 17

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	338	ASN

### 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	335/299 (112%)	327 (98%)	8 (2%)	57 15

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	46[A]	HIS
1	A	46[B]	HIS
1	A	54[A]	LYS
1	A	54[B]	LYS
1	A	182[A]	SER
1	A	182[B]	SER
1	A	290[A]	GLU
1	A	290[B]	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	231	ASN
1	A	275	GLN
1	A	276	HIS
1	A	289	GLN

### 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](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 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	338/342 (98%)	-0.44	4 (1%) 81 76	8, 13, 30, 65	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	338	ASN	3.6
1	A	5	ALA	2.8
1	A	339	SER	2.6
1	A	46[A]	HIS	2.5

### 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	ZN	A	401	1/1	1.00	0.02	-1.63	9,9,9,9	0
2	ZN	A	402	1/1	1.00	0.02	-4.53	9,9,9,9	0

## 6.5 Other polymers

There are no such residues in this entry.