



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 27, 2017 – 05:47 PM EST

PDB ID : 5MJY  
Title : Crystal structure of the His Domain Protein Tyrosine Phosphatase (HD-PTP/PTPN23) Bro1 domain (SARA complex structure)  
Authors : Levy, C.; Gahloth, D.  
Deposited on : unknown  
Resolution : 2.25 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030345

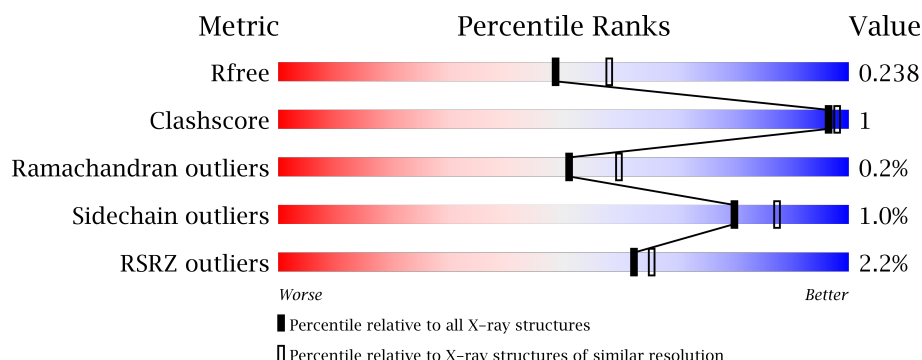
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.25 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))
$R_{free}$	100719	1062 (2.26-2.26)
Clashscore	112137	1178 (2.26-2.26)
Ramachandran outliers	110173	1145 (2.26-2.26)
Sidechain outliers	110143	1146 (2.26-2.26)
RSRZ outliers	101464	1066 (2.26-2.26)

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	361	<div> <div>2%</div> <div>94%</div> <div>5%</div> </div>
1	B	361	<div> <div>%</div> <div>98%</div> <div>.</div> </div>
1	C	361	<div> <div>%</div> <div>97%</div> <div>.</div> </div>
1	D	361	<div> <div>4%</div> <div>95%</div> <div>..</div> </div>
2	E	22	<div> <div>9%</div> <div>86%</div> <div>5%</div> <div>9%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	22	 A horizontal bar chart showing the quality of chain F. The bar is divided into three segments: a green segment representing 82%, a yellow segment representing 9%, and a grey segment representing 9%. The percentages are labeled below the bar.

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 12216 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 phosphatase non-receptor type 23.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	358	Total	C	N	O	S	0	0	0
			2835	1813	487	517	18			
1	B	361	Total	C	N	O	S	0	0	0
			2857	1826	490	522	19			
1	C	360	Total	C	N	O	S	0	0	0
			2850	1821	489	521	19			
1	D	358	Total	C	N	O	S	0	0	0
			2835	1813	487	517	18			

- Molecule 2 is a protein called Zinc finger FYVE domain-containing protein 9.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	20	Total	C	N	O	0	0	0
			174	111	25	38			
2	F	20	Total	C	N	O	0	0	0
			174	111	25	38			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	108	Total	O	0	0
			108	108		
3	B	120	Total	O	0	0
			120	120		
3	C	134	Total	O	0	0
			134	134		
3	D	115	Total	O	0	0
			115	115		
3	E	5	Total	O	0	0
			5	5		
3	F	9	Total	O	0	0
			9	9		

### 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 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 phosphatase non-receptor type 23

Chain A: 



- Molecule 1: Tyrosine-protein phosphatase non-receptor type 23

Chain B: 



- Molecule 1: Tyrosine-protein phosphatase non-receptor type 23

Chain C: 




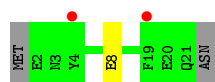
- Molecule 1: Tyrosine-protein phosphatase non-receptor type 23

Chain D: 

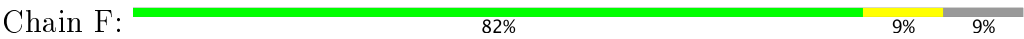


- Molecule 2: Zinc finger FYVE domain-containing protein 9

Chain E: 



- Molecule 2: Zinc finger FYVE domain-containing protein 9



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.33 Å 70.36 Å 83.48 Å 111.48° 90.06° 93.00°	Depositor
Resolution (Å)	65.37 – 2.25 62.58 – 2.25	Depositor EDS
% Data completeness (in resolution range)	95.8 (65.37-2.25) 84.4 (62.58-2.25)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.75 (at 2.25 Å)	Xtriage
Refinement program	PHENIX (1.11_2563: ???)	Depositor
R, $R_{free}$	0.187 , 0.235 0.191 , 0.238	Depositor DCC
$R_{free}$ test set	3005 reflections (4.71%)	DCC
Wilson B-factor (Å <sup>2</sup> )	27.0	Xtriage
Anisotropy	0.553	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 44.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	12216	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.36% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

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	0.25	0/2897	0.40	0/3914
1	B	0.25	0/2919	0.39	0/3943
1	C	0.25	0/2912	0.40	0/3933
1	D	0.25	0/2897	0.39	0/3914
2	E	0.24	0/177	0.31	0/238
2	F	0.25	0/177	0.32	0/238
All	All	0.25	0/11979	0.39	0/16180

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	2835	0	2850	7	0
1	B	2857	0	2873	3	0
1	C	2850	0	2864	6	0
1	D	2835	0	2850	10	0
2	E	174	0	149	0	0
2	F	174	0	149	1	0
3	A	108	0	0	1	0
3	B	120	0	0	1	1
3	C	134	0	0	1	1
3	D	115	0	0	6	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	E	5	0	0	0	0
3	F	9	0	0	1	0
All	All	12216	0	11735	27	1

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 1.

The worst 5 of 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:D:292:ASP:OD2	3:D:401:HOH:O	2.08	0.71
1:A:137:GLU:N	3:D:402:HOH:O	2.23	0.71
1:D:43:GLU:OE2	3:D:402:HOH:O	2.10	0.69
1:D:37:ASN:OD1	1:D:39:GLU:N	2.30	0.65
2:F:20:GLU:OE1	3:F:101:HOH:O	2.15	0.63

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:B:483:HOH:O	3:C:532:HOH:O[1_454]	2.10	0.10

## 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	356/361 (99%)	347 (98%)	6 (2%)	3 (1%)	22	20
1	B	359/361 (99%)	355 (99%)	4 (1%)	0	100	100
1	C	358/361 (99%)	353 (99%)	5 (1%)	0	100	100
1	D	356/361 (99%)	350 (98%)	6 (2%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	E	18/22 (82%)	18 (100%)	0	0	100	100
2	F	18/22 (82%)	18 (100%)	0	0	100	100
All	All	1465/1488 (98%)	1441 (98%)	21 (1%)	3 (0%)	51	60

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	36	GLU
1	A	326	ALA
1	A	327	LEU

### 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	302/304 (99%)	297 (98%)	5 (2%)	66	76
1	B	304/304 (100%)	300 (99%)	4 (1%)	73	83
1	C	303/304 (100%)	302 (100%)	1 (0%)	94	97
1	D	302/304 (99%)	301 (100%)	1 (0%)	94	97
2	E	18/20 (90%)	17 (94%)	1 (6%)	25	25
2	F	18/20 (90%)	17 (94%)	1 (6%)	25	25
All	All	1247/1256 (99%)	1234 (99%)	13 (1%)	80	87

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	23	GLN
1	B	172	GLN
1	D	57	ARG
1	A	328	ASP
1	C	197	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no

such sidechains identified.

### 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](#)

There are no ligands in this entry.

### 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	358/361 (99%)	-0.02	9 (2%) 58 61	18, 33, 71, 104	0
1	B	361/361 (100%)	-0.14	4 (1%) 80 83	18, 32, 56, 77	0
1	C	360/361 (99%)	-0.14	5 (1%) 75 78	17, 33, 57, 71	0
1	D	358/361 (99%)	0.05	13 (3%) 43 46	19, 33, 71, 130	0
2	E	20/22 (90%)	0.25	2 (10%) 8 8	26, 43, 55, 58	0
2	F	20/22 (90%)	0.12	0 100 100	23, 38, 51, 58	0
All	All	1477/1488 (99%)	-0.06	33 (2%) 62 65	17, 33, 64, 130	0

The worst 5 of 33 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	35	GLY	9.2
1	D	331	GLN	5.0
1	A	326	ALA	5.0
1	D	33	ASN	4.8
1	D	41	TYR	3.7

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.