

# Full wwPDB X-ray Structure Validation Report (i)

#### Aug 9, 2021 – 12:03 pm BST

PDB ID	:	6ZZ4
$\operatorname{Title}$	:	Crystal structure of the PTPN2 C216G mutant
Authors	:	Mechaly, A.E.; Berthelet, J.; Nian, Q.; Parlato, M.; Cerf-Bensussan, N.;
		Haouz, A.; Rodrigues-Lima, F.
Deposited on	:	2020-08-03
$\operatorname{Resolution}$	:	2.43  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* A user guide is available at https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

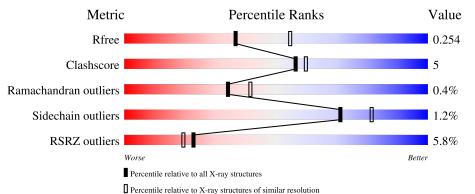
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.23.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	$7.0.044 (\mathrm{Gargrove})$
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.1

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: X-RAY DIFFRACTION

The reported resolution of this entry is 2.43 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries},{ m resolution\ range}({ m \AA}))$
R <sub>free</sub>	130704	$1564 \ (2.46-2.42)$
Clashscore	141614	1631(2.46-2.42)
Ramachandran outliers	138981	1617(2.46-2.42)
Sidechain outliers	138945	1617(2.46-2.42)
RSRZ outliers	127900	1547(2.46-2.42)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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 <=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	А	314	79%	9%	11%
1	В	314	5% 74%	13% •	12%



## 2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4758 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.

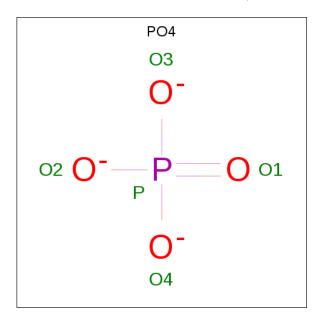
• Molecule 1 is a protein called Tyrosine-protein phosphatase non-receptor type 2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
1	А	279	Total 2286	C 1449	N 399	O 426	S 12	0	0	0
1	В	277	Total 2269	C 1437	N 396	0 424	S 12	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	216	GLY	CYS	engineered mutation	UNP P17706
В	216	GLY	CYS	engineered mutation	UNP P17706

• Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P).



Mo	Chain	Residues	Atoms			ZeroOcc	AltConf
2	А	1	Total 5	0 4	Р 1	0	0

Continued on next page...



Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	А	1	Total O P 5 4 1	0	0
			Total O P	0	0
2	A	1	5 4 1	0	0
2	В	1	Total O P 5 4 1	0	0
	D	- 1	Total O P	0	0
2	В	1	5 4 1	U	0
2	В	1	Total O P 5 4 1	0	0

• Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	67	Total         O           67         67	0	0
3	В	106	Total O 106 106	0	0



# 3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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 2



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants	142.89Å $55.98$ Å $101.42$ Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $120.74^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	48.71 - 2.43	Depositor
	48.71 - 2.43	EDS
% Data completeness	98.2(48.71-2.43)	Depositor
(in resolution range)	98.3(48.71-2.43)	EDS
R <sub>merge</sub>	0.12	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.10 (at 2.42 \text{\AA})$	Xtriage
Refinement program	PHENIX 1.17.1_3660, BUSTER	Depositor
$R, R_{free}$	0.198 , $0.253$	Depositor
II, II, <i>free</i>	0.199 , $0.254$	DCC
$R_{free}$ test set	1290 reflections $(5.00%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	44.9	Xtriage
Anisotropy	0.446	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.38 , $56.3$	EDS
L-test for twinning <sup>2</sup>	$ L  > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4758	wwPDB-VP
Average B, all atoms $(Å^2)$	53.0	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 22.55 % 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 5.5717e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.

<sup>&</sup>lt;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.



<sup>&</sup>lt;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:  $\mathrm{PO4}$ 

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		
	Cham	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.45	0/2342	0.61	1/3172~(0.0%)	
1	В	0.46	0/2325	0.62	1/3150~(0.0%)	
All	All	0.45	0/4667	0.61	2/6322~(0.0%)	

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	<b>#Planarity outliers</b>
1	А	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
1	В	237	GLU	CA-CB-CG	6.57	127.86	113.40
1	A	41	GLU	CA-CB-CG	6.09	126.79	113.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	А	41	GLU	Sidechain

### 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



Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	А	2286	0	2238	20	0
1	В	2269	0	2214	26	0
2	А	15	0	0	0	0
2	В	15	0	0	0	0
3	А	67	0	0	2	0
3	В	106	0	0	3	0
All	All	4758	0	4452	45	0

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.

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 (45) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom 1	Atom 9	Interatomic	Clash
Atom-1	Atom-2	${ m distance}~({ m \AA})$	overlap (Å)
1:B:7:ARG:NH2	3:B:501:HOH:O	2.03	0.91
1:B:50:ASP:OD1	3:B:502:HOH:O	2.06	0.72
1:A:43:ARG:HH22	1:A:49:ARG:HE	1.46	0.63
1:A:122:LYS:HB3	1:A:217:SER:HB3	1.81	0.62
1:B:138:GLU:H	1:B:138:GLU:CD	2.05	0.58
1:A:137:LYS:HG3	3:A:517:HOH:O	2.03	0.58
1:B:4:THR:O	1:B:8:GLU:HG3	2.05	0.57
1:B:180:TRP:CE2	1:B:222:ARG:HG2	2.43	0.53
1:B:25:ILE:HG12	1:B:245:LYS:HG3	1.90	0.53
1:B:101:VAL:HG13	1:B:106:THR:HB	1.91	0.53
1:A:38:LYS:HA	1:A:43:ARG:HD3	1.92	0.52
1:B:106:THR:HG21	1:B:213:VAL:HG23	1.91	0.52
1:A:51:VAL:HG13	1:A:220:ILE:HD12	1.92	0.50
1:A:111:MET:HG3	1:A:215:HIS:CE1	2.46	0.50
1:B:122:LYS:HB3	1:B:217:SER:HB3	1.94	0.50
1:B:234:VAL:HA	1:B:237:GLU:OE2	2.11	0.50
1:B:117:GLU:OE2	1:B:122:LYS:HD2	2.12	0.50
1:B:232:CYS:O	1:B:236:MET:HG3	2.12	0.50
1:A:121:VAL:HG21	1:B:131:ASP:HB3	1.97	0.47
1:A:180:TRP:CE2	1:A:222:ARG:HG2	2.50	0.47
1:B:180:TRP:NE1	1:B:222:ARG:HG2	2.30	0.47
1:B:188:SER:HB3	1:B:191:SER:H	1.79	0.47
1:A:122:LYS:NZ	1:A:182:ASP:OD2	2.49	0.46
1:A:153:TYR:CZ	1:A:187:GLU:OE2	2.68	0.46
1:A:142:SER:HB3	1:A:164:ILE:HD11	1.98	0.45

Continued on next page...



Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:41:GLU:OE1	1:A:41:GLU:N	2.39	0.45
1:A:122:LYS:HB3	1:A:217:SER:CB	2.46	0.45
1:A:38:LYS:NZ	3:A:505:HOH:O	2.49	0.45
1:A:266:ARG:NH1	1:A:270:MET:HE1	2.33	0.43
1:A:43:ARG:HH12	1:A:49:ARG:NE	2.16	0.43
1:A:51:VAL:HG13	1:A:220:ILE:CD1	2.48	0.43
1:B:220:ILE:HD12	1:B:256:MET:O	2.19	0.43
1:B:225:THR:HG23	1:B:268:SER:OG	2.19	0.43
1:B:157:HIS:HB2	1:B:174:HIS:HB3	2.01	0.43
1:B:107:LYS:HE2	1:B:202:SER:O	2.18	0.42
1:B:72:SER:HB2	1:B:255:ARG:HD2	2.01	0.42
1:B:263:ASP:HB3	3:B:540:HOH:O	2.19	0.42
1:B:147:SER:HB3	1:B:158:LEU:HB2	2.02	0.41
1:A:28:GLU:OE1	1:A:245:LYS:NZ	2.41	0.41
1:A:72:SER:HB2	1:A:255:ARG:HD2	2.03	0.41
1:B:9:PHE:CE2	1:B:270:MET:HG3	2.56	0.41
1:B:228:LEU:HD22	1:B:248:LEU:HD13	2.02	0.41
1:B:95:CYS:HB3	1:B:138:GLU:OE2	2.20	0.40
1:B:180:TRP:CE2	1:B:186:PRO:HD3	2.56	0.40
1:A:5:ILE:HD12	1:A:244:ILE:HD12	2.04	0.40

Continued from previous page...

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	Perce	$\mathbf{ntiles}$
1	А	277/314 (88%)	266~(96%)	10 (4%)	1 (0%)	34	41
1	В	275/314~(88%)	268~(98%)	6(2%)	1 (0%)	34	41
All	All	552/628~(88%)	534 (97%)	16 (3%)	2(0%)	34	41

All (2) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	А	259	ILE
1	В	259	ILE

#### 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	А	255/288~(88%)	252~(99%)	3 (1%)	71 81		
1	В	253/288~(88%)	250~(99%)	3 (1%)	71 81		
All	All	508/576~(88%)	502~(99%)	6 (1%)	71 81		

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	120	SER
1	А	137	LYS
1	А	240	ASP
1	В	83	TYR
1	В	191	SER
1	В	253	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	243	ASN

#### 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 monosaccharides in this entry.

### 5.6 Ligand geometry (i)

6 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	Tune	Chain	Res	Bog Link Bond lengths			gths	Bond angles		
	Type	Cham	nes	Link	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	PO4	В	403	-	4,4,4	0.82	0	$^{6,6,6}$	0.44	0
2	PO4	А	403	-	4,4,4	0.87	0	$^{6,6,6}$	0.38	0
2	PO4	В	401	-	4, 4, 4	0.96	0	$^{6,6,6}$	0.28	0
2	PO4	А	402	-	4,4,4	0.94	0	$^{6,6,6}$	0.68	0
2	PO4	А	401	-	4,4,4	0.96	0	$^{6,6,6}$	0.40	0
2	PO4	В	402	-	4,4,4	0.79	0	$^{6,6,6}$	0.52	0

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,  $95^{th}$  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		$OWAB(Å^2)$	Q<0.9
1	А	279/314~(88%)	0.21	15 (5%) 25	22	30, 50, 80, 105	0
1	В	277/314 (88%)	0.31	17 (6%) 21	17	32, 49, 85, 108	0
All	All	556/628~(88%)	0.26	32 (5%) 23	19	30, 49, 83, 108	0

All (32) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ	
1	В	208	ASP	4.2	
1	В	207	PRO	3.8	
1	В	241	ASP	3.7	
1	А	3	THR	3.6	
1	В	16	ARG	3.4	
1	В	246 GLN		3.2	
1	А	241	ASP	3.2	
1	А	278	CYS	3.1	
1	В	78	GLU	3.1	
1	А	163	ASN	3.0	
1	В	75	ASP	3.0	
1	В	7	ARG	2.9	
1	В	14	THR	2.9	
1	В	11	GLU	2.8	
1	А	275	GLY	2.8	
1	А	102	TRP	2.6	
1	В	209	HIS	2.5	
1	А	167	GLY	2.5	
1	А	279	ILE	2.5	
1	В	19	GLN	2.4	
1	А	201	GLU	2.4	
1	А	4	THR	2.4	
1	В	111	MET	2.3	
1	А	30	HIS	2.2	

Continued on next page...



Mol	Chain	Res	Type	RSRZ
1	В	79	ALA	2.2
1	В	206	ASN	2.2
1	В	239	GLY	2.2
1	А	237	GLU	2.1
1	А	187	GLU	2.1
1	А	280	LYS	2.1
1	В	33	PRO	2.0
1	А	2	PRO	2.0

Continued from previous page...

#### 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 monosaccharides 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. The B-factors column lists the minimum, median,  $95^{th}$  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	${f B} ext{-factors}({ m \AA}^2)$	Q<0.9
2	PO4	А	403	5/5	0.85	0.13	$94,\!101,\!101,\!103$	0
2	PO4	В	403	5/5	0.94	0.17	$78,\!81,\!86,\!90$	0
2	PO4	А	402	5/5	0.97	0.17	$54,\!54,\!58,\!59$	0
2	PO4	В	402	5/5	0.98	0.16	$29,\!47,\!60,\!61$	0
2	PO4	А	401	5/5	0.99	0.15	$37,\!38,\!43,\!44$	0
2	PO4	В	401	5/5	0.99	0.20	33,35,41,48	0

### 6.5 Other polymers (i)

There are no such residues in this entry.

