

Full wwPDB X-ray Structure Validation Report (i)

Aug 21, 2023 – 04:53 PM EDT

PDB ID 2PA5

> Title Crystal structure of human protein tyrosine phosphatase PTPN9

Authors Ugochukwu, E.; Barr, A.; Pike, A.C.W.; Savitsky, P.; Papagrigoriou, E.; Turn-

> bull, A.; Uppenberg, J.; Bunkoczi, G.; Salah, E.; Das, S.; von Delft, F.; Sundstrom, M.; Arrowsmith, C.H.; Weigelt, J.; Edwards, A.; Knapp, S.; Structural

Genomics Consortium (SGC)

2007-03-27 Deposited on

Resolution 1.60 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 types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

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

MolProbity 4.02b-467

> 1.8.5 (274361), CSD as 541 be (2020)Mogul

Xtriage (Phenix) 1.13

EDS 2.35

Percentile statistics 20191225.v01 (using entries in the PDB archive December 25th 2019)

> Refmac 5.8.0158

CCP4 7.0.044 (Gargrove) Ideal geometry (proteins) Engh & Huber (2001) Ideal geometry (DNA, RNA) Parkinson et al. (1996)

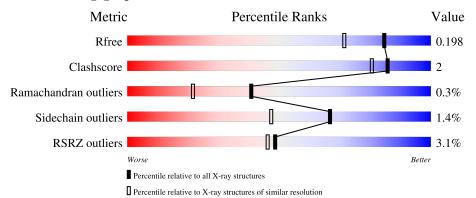
Validation Pipeline (wwPDB-VP) 2.35

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 1.60 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	Similar resolution
Metric	$(\# \mathrm{Entries})$	$(\# ext{Entries}, ext{ resolution range}(ext{Å}))$
R_{free}	130704	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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	A	314	89%	• 6%
1	В	314	5% 88%	• 7%



2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5112 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 9.

Mol	Chain	Residues	${f Atoms}$				ZeroOcc	AltConf	Trace	
1	A	296	Total 2375	C 1512	N 406	O 440	S 17	0	9	0
1	В	291	Total 2300	C 1465	N 391	O 427	S 17	0	3	0

There are 16 discrepancies between the modelled and reference sequences:

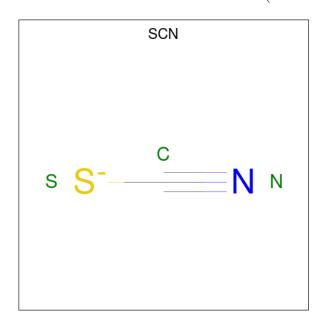
Chain	Residue	Modelled	Actual	Comment	Reference
A	276	MET	-	cloning artifact	UNP P43378
A	583	ALA	-	cloning artifact	UNP P43378
A	584	HIS	-	cloning artifact	UNP P43378
A	585	HIS	_	cloning artifact	UNP P43378
A	586	HIS	-	cloning artifact	UNP P43378
A	587	HIS	-	cloning artifact	UNP P43378
A	588	HIS	-	cloning artifact	UNP P43378
A	589	HIS	-	cloning artifact	UNP P43378
В	276	MET	-	cloning artifact	UNP P43378
В	583	ALA	-	cloning artifact	UNP P43378
В	584	HIS	-	cloning artifact	UNP P43378
В	585	HIS	-	cloning artifact	UNP P43378
В	586	HIS	-	cloning artifact	UNP P43378
В	587	HIS	-	cloning artifact	UNP P43378
В	588	HIS	-	cloning artifact	UNP P43378
В	589	HIS	-	cloning artifact	UNP P43378

• Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	3	Total Cl 3 3	0	0
2	В	1	Total Cl 1 1	0	0

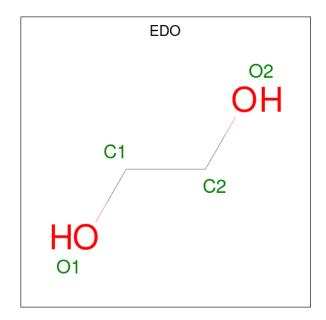


• Molecule 3 is THIOCYANATE ION (three-letter code: SCN) (formula: CNS).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total 3	C 1	N 1	S 1	0	0

 \bullet Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $\mathrm{C_2H_6O_2}).$



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 4 2 2	0	0
4	A	1	Total C O 4 2 2	0	0

Continued on next page...



 $Continued\ from\ previous\ page...$

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	В	1	Total C O 4 2 2	0	0
4	В	1	Total C O 4 2 2	0	0
4	В	1	Total C O 4 2 2	0	0

• Molecule 5 is water.

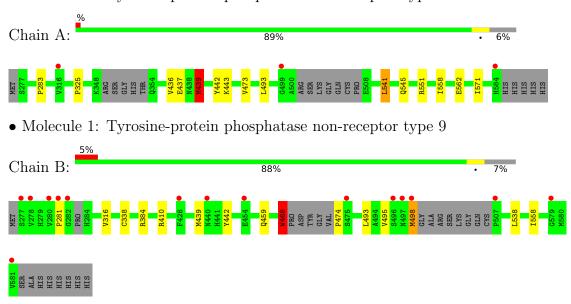
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	223	Total O 224 224	0	1
5	В	182	Total O 186 186	0	4



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 9





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants	39.97Å 57.43Å 66.45Å	Donositor
a, b, c, α , β , γ	77.44° 78.22° 80.41°	Depositor
Resolution (Å)	33.96 - 1.60	Depositor
rtesolution (A)	33.97 - 1.60	EDS
% Data completeness	96.9 (33.96-1.60)	Depositor
(in resolution range)	96.9 (33.97-1.60)	EDS
R_{merge}	0.06	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) > 1$	1.96 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
D D.	0.156 , 0.188	Depositor
R, R_{free}	0.169 , 0.198	DCC
R_{free} test set	3602 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	12.6	Xtriage
Anisotropy	0.011	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.37, 59.7	EDS
L-test for twinning ²	$ < L >=0.50, < L^2>=0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	5112	wwPDB-VP
Average B, all atoms (Å ²)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 6.27% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical values of <|L|>, $<L^2>$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



¹Intensities 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: SCN, CL, EDO

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).

Mal	Chain	Boı	nd lengths	Bond angles		
Mol Chain		RMSZ	RMSZ $ $ $\# Z > 5$		# Z > 5	
1	A	0.69	0/2474	0.76	3/3345 (0.1%)	
1	В	0.68	$2/2361 \ (0.1\%)$	0.71	0/3193	
All	All	0.68	$2/4835 \ (0.0\%)$	0.74	3/6538 (0.0%)	

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	Observed(A)	$\operatorname{Ideal}(\text{\AA})$
1	В	338	CYS	CB-SG	-6.37	1.71	1.82
1	В	468	TRP	CB-CG	6.29	1.61	1.50

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	${f Z}$	$Observed(^o)$	$Ideal(^{o})$
1	A	551	ARG	NE-CZ-NH1	9.30	124.95	120.30
1	A	439	MET	CG-SD-CE	-8.15	87.17	100.20
1	A	551	ARG	NE-CZ-NH2	-6.02	117.29	120.30

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	2375	0	2279	11	0	

Continued on next page...



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	В	2300	0	2197	9	0
2	A	3	0	0	0	0
2	В	1	0	0	1	0
3	A	3	0	0	0	0
4	A	8	0	12	0	0
4	В	12	0	18	0	0
5	A	224	0	0	2	0
5	В	186	0	0	1	0
All	All	5112	0	4506	20	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 2.

All (20) 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:541:LEU:HD13	1:A:571[B]:ILE:HD13	1.75	0.66
1:A:439:MET:HG2	1:A:442:TYR:CZ	2.32	0.65
1:B:281:PRO:HB3	1:B:538:LEU:CD1	2.27	0.64
1:A:473:VAL:HG13	1:A:562[A]:GLU:HG2	1.79	0.63
1:A:283:PRO:O	1:A:545:GLN:NE2	2.32	0.62
1:B:281:PRO:HB3	1:B:538:LEU:HD12	1.82	0.62
1:B:281:PRO:HG3	1:B:538:LEU:HD11	1.84	0.59
1:A:541:LEU:CD1	1:A:571[B]:ILE:HD13	2.38	0.54
1:A:436:VAL:HG11	1:A:443[A]:LYS:HE3	1.90	0.54
1:B:439:MET:HG3	1:B:442:TYR:CZ	2.45	0.51
1:B:495:VAL:O	1:B:498:MET:HB2	2.14	0.48
1:B:410:ARG:HD2	2:B:590:CL:CL	2.50	0.47
1:A:562[A]:GLU:HG3	5:A:800:HOH:O	2.14	0.47
1:B:281:PRO:CB	1:B:538:LEU:CD1	2.92	0.47
1:A:325:PRO:HA	5:A:927:HOH:O	2.14	0.46
1:A:473:VAL:CG1	1:A:562[A]:GLU:HG2	2.47	0.43
1:B:468:TRP:CD1	1:B:474:PRO:HD3	2.54	0.42
1:A:493:LEU:HD23	1:A:493:LEU:C	2.41	0.41
1:B:459:GLN:NE2	5:B:874:HOH:O	2.34	0.41
1:A:437:GLU:HG2	1:A:439:MET:HE3	2.03	0.41

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	ntiles
1	A	300/314~(96%)	294 (98%)	5 (2%)	1 (0%)	41	21
1	В	286/314 (91%)	279 (98%)	6 (2%)	1 (0%)	41	21
All	All	586/628 (93%)	573 (98%)	11 (2%)	2 (0%)	41	21

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	558	ILE
1	В	558	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	Percentiles		
1	A	255/278 (92%)	253 (99%)	2 (1%)	81 70	
1	В	242/278 (87%)	237 (98%)	5 (2%)	53 29	
All	All	497/556 (89%)	490 (99%)	7 (1%)	67 47	

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	439	MET
1	A	541	LEU
1	В	316	VAL
1	В	384	ARG

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type
1	В	468	TRP
1	В	493	LEU
1	В	498	MET

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

Mol	Chain	Res	Type
1	A	489	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)

Of 10 ligands modelled in this entry, 4 are monoatomic - leaving 6 for Mogul analysis.

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	Trme	e Chain	Peg	Link	Bond lengths			Bond angles		
MIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z >2
4	EDO	В	701	-	3,3,3	0.52	0	2,2,2	0.34	0
4	EDO	В	702	-	3,3,3	0.35	0	2,2,2	0.58	0
4	EDO	A	704	-	3,3,3	0.50	0	2,2,2	0.33	0
3	SCN	A	601	-	1,2,2	0.38	0	0,1,1	-	-
4	EDO	В	705	-	3,3,3	0.47	0	2,2,2	0.22	0



Mol	Type	Chain	Res	Link	\mathbf{B}_{0}	ond leng	${ m gths}$	В	ond ang	gles
MIOI	Туре	Chain		Lilik	Counts	RMSZ	# Z > 2	Counts	ints RMSZ $\# Z > 2$	
4	EDO	A	703	-	3,3,3	0.36	0	2,2,2	0.70	0

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
4	EDO	В	701	-	-	1/1/1/1	-
4	EDO	В	702	-	-	0/1/1/1	-
4	EDO	A	704	-	-	0/1/1/1	-
4	EDO	В	705	-	-	0/1/1/1	-
4	EDO	A	703	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	В	701	EDO	O1-C1-C2-O2

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	$\langle { m RSRZ} \rangle$	$\# \mathrm{RSRZ}{>}2$	$OWAB(A^2)$	Q < 0.9
1	A	296/314 (94%)	-0.27	3 (1%) 82 82	8, 12, 27, 45	0
1	В	$291/314 \ (92\%)$	-0.21	15 (5%) 27 24	5, 13, 26, 45	0
All	All	587/628 (93%)	-0.24	18 (3%) 49 46	5, 12, 27, 45	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	В	282	GLY	4.8
1	A	499	GLY	4.2
1	В	581	VAL	4.1
1	В	498	MET	3.8
1	В	278	VAL	3.3
1	В	280	VAL	3.3
1	В	475	SER	3.0
1	В	496	SER	3.0
1	В	281	PRO	2.9
1	A	316	VAL	2.8
1	В	497	ASN	2.8
1	В	579	GLY	2.5
1	В	277	SER	2.5
1	В	440	ASN	2.3
1	A	584	HIS	2.3
1	В	428	PHE	2.2
1	В	507	PRO	2.1
1	В	454	GLU	2.1

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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
2	CL	A	590	1/1	0.81	0.08	49,49,49,49	0
4	EDO	A	703	4/4	0.92	0.10	21,25,27,31	0
4	EDO	В	701	4/4	0.95	0.05	6,11,12,17	0
4	EDO	A	704	4/4	0.96	0.09	15,17,23,23	0
4	EDO	В	705	4/4	0.96	0.15	20,21,23,24	0
4	EDO	В	702	4/4	0.97	0.07	13,20,21,30	0
2	CL	A	592	1/1	0.98	0.05	24,24,24,24	0
3	SCN	A	601	3/3	0.99	0.05	11,11,13,13	0
2	CL	В	590	1/1	0.99	0.07	20,20,20,20	0
2	CL	A	591	1/1	1.00	0.04	11,11,11,11	0

6.5 Other polymers (i)

There are no such residues in this entry.

