wwPDB X-ray Structure Validation Summary Report (i)

May 29, 2020 – 08:48 am BST

PDB ID	:	5QG2
Title	:	PanDDA analysis group deposition – Crystal structure of PTP1B in complex with compound_FMOPL000275a
Authors	:	Keedy, D.A.; Hill, Z.B.; Biel, J.T.; Kang, E.; Rettenmaier, T.J.; Brandao-Neto, J.; von Delft, F.; Wells, J.A.; Fraser, J.S.
Deposited on	:	2018-08-30
Resolution	:	2.12 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*
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:

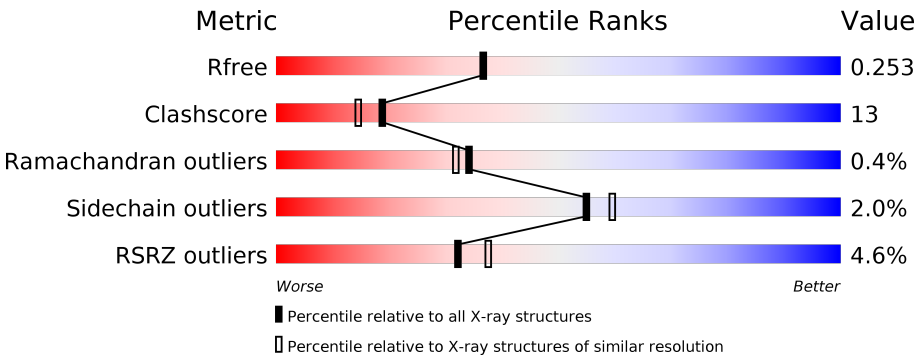
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.12 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}	130704	6241 (2.14-2.10)
Clashscore	141614	6778 (2.14-2.10)
Ramachandran outliers	138981	6705 (2.14-2.10)
Sidechain outliers	138945	6706 (2.14-2.10)
RSRZ outliers	127900	6112 (2.14-2.10)

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 >=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	321	4% 61% 25% • 12%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 16805 atoms, of which 8238 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.

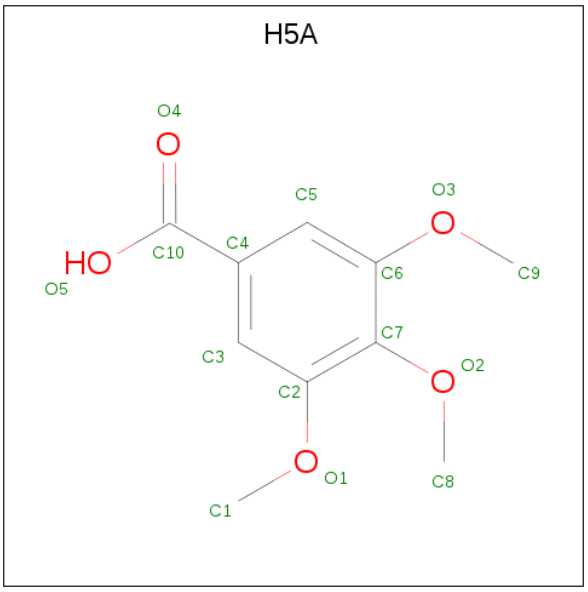
- Molecule 1 is a protein called Tyrosine-protein phosphatase non-receptor type 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	283	16493	5252	8204	1419	1561	57	0	240	0

There are 2 discrepancies between the modelled and reference sequences:

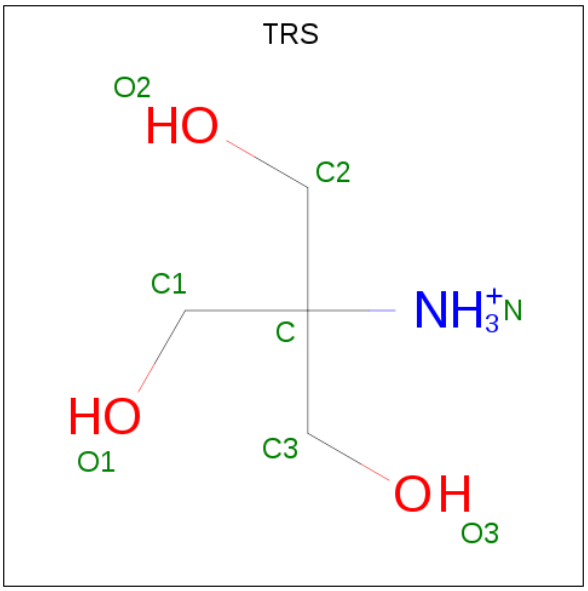
Chain	Residue	Modelled	Actual	Comment	Reference
A	32	SER	CYS	engineered mutation	UNP P18031
A	92	VAL	CYS	engineered mutation	UNP P18031

- Molecule 2 is 3,4,5-trimethoxybenzoic acid (three-letter code: H5A) (formula: $C_{10}H_{12}O_5$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
2	A	1	52	20	22	10	0	1

- Molecule 3 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (three-letter code: TRS) (formula: $C_4H_{12}NO_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	N	O		
3	A	1	20	4	12	1	3	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	240	240	240	0	0

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 1

Chain A: 4% 61% 25% • 12%

M1 E2 M3 E4 K5 E6 Q9 I10 D11 K12 S13 G14 S15 W16 R24 S28 D29 F30 P31 S32 R33 R45 Y46 R47 D48 V49 Q61 E62 D63 I72 K73 N74 L83 L88 V92 W100 M109 R112 E115 K120 C121 Y124 K131

F135 N139 I145 I149 Y152 Y153 T154 V155 R156 Q157 N162 L163 T164 T165 Q166 R169 H173 W179 F182 E186 F191 L195 F196 K197 E200 S205 P206 E207 V211 V212 V213 H214 C215 S216 C226 D229 M235 D236 K237 R238 V244

K247 K248 M253 R254 L260 I261 D265 S270 Y271 V274 I275 A278 K279 F280 I281 M282 G283 ASP SER SER VAL GLN ASP GLN TRP LYS GLU LEU SER HIS GLU ASP LEU GLU PRO PRO PRO GLU HIS ILE PRO PRO PRO PRO ARG PRO PRO LYS ARG ILE LEU GLU PRO

HIS ASN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	90.14Å 90.14Å 106.84Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	78.07 – 2.12 78.07 – 2.12	Depositor EDS
% Data completeness (in resolution range)	99.7 (78.07-2.12) 99.7 (78.07-2.12)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.74 (at 2.12Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.215 , 0.251 0.216 , 0.253	Depositor DCC
R_{free} test set	1136 reflections (3.94%)	wwPDB-VP
Wilson B-factor (Å^2)	30.9	Xtriage
Anisotropy	0.374	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 60.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.022 for -h,-k,l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	16805	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.29% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical 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.

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: TRS, H5A

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	0.53	0/8474	0.66	4/11400 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	121[A]	CYS	CA-CB-SG	5.70	124.26	114.00
1	A	121[B]	CYS	CA-CB-SG	5.70	124.26	114.00
1	A	121[C]	CYS	CA-CB-SG	5.70	124.26	114.00
1	A	121[D]	CYS	CA-CB-SG	5.70	124.26	114.00

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	8289	8204	8136	217	0
2	A	30	22	0	2	0
3	A	8	12	12	0	0
4	A	240	0	0	30	1
All	All	8567	8238	8148	217	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 13.

The worst 5 of 217 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:74[C]:MET:CE	1:A:253[C]:MET:HG2	1.50	1.39
1:A:74[D]:MET:CE	1:A:253[D]:MET:HG2	1.50	1.39
1:A:9[C]:GLN:O	1:A:13[C]:SER:OG	1.54	1.22
1:A:9[D]:GLN:O	1:A:13[D]:SER:OG	1.54	1.22
1:A:162[D]:ASN:OD1	1:A:164[D]:THR:HB	1.35	1.21

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)
4:A:556:HOH:O	4:A:662:HOH:O[5_664]	2.19	0.01

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	995/321 (310%)	930 (94%)	61 (6%)	4 (0%)	34 32

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	261[A]	ILE
1	A	261[B]	ILE
1	A	261[C]	ILE
1	A	261[D]	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	A	931/294 (317%)	914 (98%)	17 (2%)	59 63

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	72[D]	ILE
1	A	213	VAL
1	A	281[A]	ILE
1	A	72[C]	ILE
1	A	281[B]	ILE

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

Mol	Chain	Res	Type
1	A	68	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

3 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	H5A	A	401[C]	-	13,15,15	3.51	7 (53%)	17,20,20	1.34	2 (11%)
3	TRS	A	402	-	7,7,7	0.51	0	9,9,9	0.79	0
2	H5A	A	401[D]	-	13,15,15	3.51	7 (53%)	17,20,20	1.34	2 (11%)

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	H5A	A	401[C]	-	-	3/6/10/10	0/1/1/1
3	TRS	A	402	-	-	3/9/9/9	-
2	H5A	A	401[D]	-	-	3/6/10/10	0/1/1/1

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401[C]	H5A	C4-C10	8.14	1.55	1.47
2	A	401[D]	H5A	C4-C10	8.14	1.55	1.47
2	A	401[C]	H5A	O3-C6	-5.54	1.28	1.37
2	A	401[D]	H5A	O3-C6	-5.54	1.28	1.37
2	A	401[C]	H5A	C6-C7	-3.66	1.33	1.41

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	401[C]	H5A	C8-O2-C7	2.52	121.69	114.78
2	A	401[D]	H5A	C8-O2-C7	2.52	121.69	114.78
2	A	401[C]	H5A	O1-C2-C7	2.12	118.89	115.16
2	A	401[D]	H5A	O1-C2-C7	2.12	118.89	115.16

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	402	TRS	C2-C-C1-O1
3	A	402	TRS	C3-C-C1-O1
3	A	402	TRS	N-C-C1-O1
2	A	401[C]	H5A	C7-C2-O1-C1
2	A	401[D]	H5A	C7-C2-O1-C1

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	401[C]	H5A	1	0
2	A	401[D]	H5A	1	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	10

The worst 5 of 10 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	82[C]:ILE	C	83:LEU	N	1.20
1	A	82[D]:ILE	C	83:LEU	N	1.20
1	A	263[C]:THR	C	264:ALA	N	1.19
1	A	263[D]:THR	C	264:ALA	N	1.19
1	A	216[C]:SER	C	217:ALA	N	1.18

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, 95th 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	A	283/321 (88%)	0.51	13 (4%) 32 37	18, 30, 58, 76	1 (0%)

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1[A]	MET	7.4
1	A	280[A]	PHE	7.3
1	A	281[A]	ILE	4.9
1	A	282[A]	MET	4.9
1	A	152[A]	TYR	4.7

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)

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, 95th 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	B-factors(Å^2)	Q<0.9
3	TRS	A	402	8/8	0.64	0.28	46,78,98,111	0
2	H5A	A	401[C]	15/15	0.84	0.17	35,59,74,74	26
2	H5A	A	401[D]	15/15	0.84	0.17	35,59,74,74	26

6.5 Other polymers (i)

There are no such residues in this entry.