



Full wwPDB X-ray Structure Validation Report ⓘ

Oct 19, 2017 – 05:28 PM EDT

PDB ID : 4HJP
Title : SHP-1 catalytic domain WPD loop open
Authors : Alicea-Velazquez, N.L.; Boggon, T.J.
Deposited on : unknown
Resolution : 1.40 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

<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

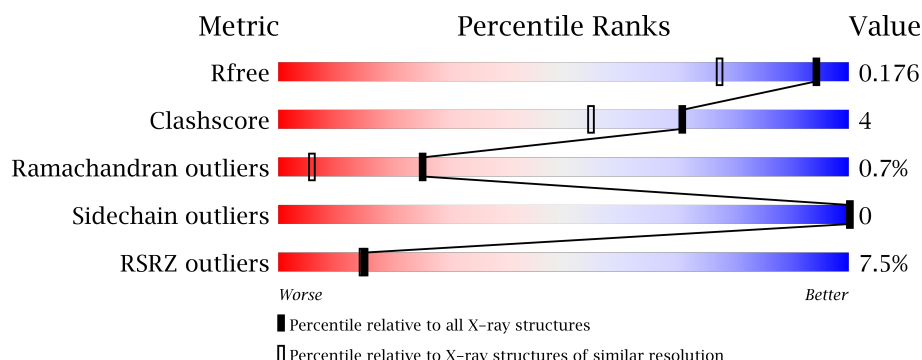
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION


The reported resolution of this entry is 1.40 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	1307 (1.40-1.40)
Clashscore	112137	1411 (1.40-1.40)
Ramachandran outliers	110173	1373 (1.40-1.40)
Sidechain outliers	110143	1372 (1.40-1.40)
RSRZ outliers	101464	1315 (1.40-1.40)

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. 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	288	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2694 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 6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	281	Total	C	N	O	S	0	27	0
			2414	1533	416	452	13			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	241	GLY	-	EXPRESSION TAG	UNP P29350
A	242	SER	-	EXPRESSION TAG	UNP P29350

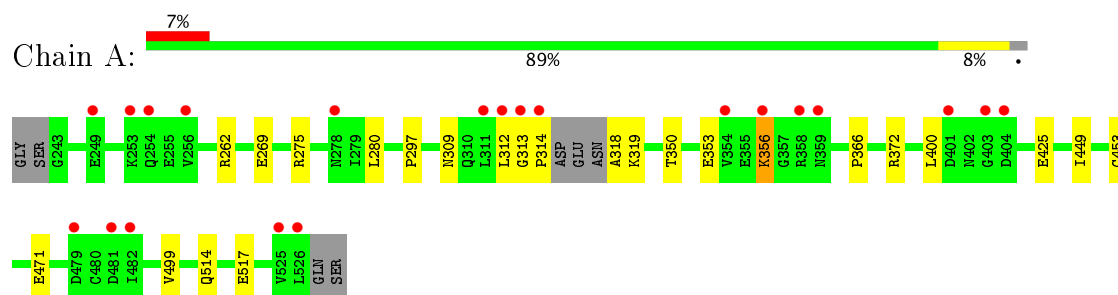
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	280	Total	O	0	0
			280	280		

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 ($\text{RSRZ} > 2$). 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 6



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 2 21	Depositor
Cell constants a, b, c, α , β , γ	45.48 Å 57.39 Å 110.81 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.08 – 1.40 42.08 – 1.40	Depositor EDS
% Data completeness (in resolution range)	99.2 (42.08-1.40) 95.6 (42.08-1.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.39 (at 1.40 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.2_869)	Depositor
R, R_{free}	0.164 , 0.186 0.152 , 0.176	Depositor DCC
R_{free} test set	2816 reflections (5.08%)	DCC
Wilson B-factor (Å ²)	18.8	Xtriage
Anisotropy	0.358	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 41.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.98	EDS
Total number of atoms	2694	wwPDB-VP
Average B, all atoms (Å ²)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.06% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.51	0/2537	0.65	0/3432

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	2414	0	2433	20	0
2	A	280	0	0	6	2
All	All	2694	0	2433	20	2

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 4.

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:309[B]:ASN:ND2	2:A:824:HOH:O	1.93	0.98
1:A:453[A]:CYS:O	2:A:879:HOH:O	1.97	0.81
1:A:262[A]:ARG:NH1	2:A:813:HOH:O	2.29	0.66
1:A:353[B]:GLU:OE1	2:A:878:HOH:O	2.13	0.65

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:319[A]:LYS:HD3	1:A:449:ILE:HD12	1.80	0.62
1:A:319[B]:LYS:NZ	1:A:471:GLU:OE1	2.34	0.61
1:A:366:PRO:HB3	1:A:372:ARG:HG3	1.83	0.60
1:A:262[A]:ARG:HD2	1:A:280:LEU:HB2	1.88	0.56
1:A:269[A]:GLU:HG2	1:A:297:PRO:HB2	1.90	0.52
1:A:312:LEU:HD11	2:A:868:HOH:O	2.13	0.48
1:A:356:LYS:HG2	1:A:356:LYS:H	1.41	0.45
1:A:350:THR:OG1	1:A:453[B]:CYS:HB3	2.16	0.45
1:A:514:GLN:HA	1:A:517[B]:GLU:HG2	1.98	0.44
1:A:400:LEU:HA	1:A:400:LEU:HD23	1.81	0.42
1:A:318:ALA:HA	1:A:471:GLU:OE2	2.19	0.42
1:A:425:GLU:OE1	2:A:816:HOH:O	2.22	0.42
1:A:262[A]:ARG:HH11	1:A:280:LEU:HB2	1.85	0.42
1:A:319[A]:LYS:HD3	1:A:449:ILE:CD1	2.48	0.42
1:A:313:GLY:HA2	1:A:314:PRO:HD3	1.94	0.40

All (2) 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)
2:A:825:HOH:O	2:A:867:HOH:O[2_555]	1.94	0.26
2:A:827:HOH:O	2:A:827:HOH:O[2_555]	2.17	0.03

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	304/288 (106%)	296 (97%)	6 (2%)	2 (1%)	25 6

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	356	LYS
1	A	499	VAL

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	275/254 (108%)	275 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

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, 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	281/288 (97%)	0.35	21 (7%) 15 15	13, 21, 41, 68	8 (2%)

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	356	LYS	6.0
1	A	359	ASN	5.6
1	A	526	LEU	4.5
1	A	479	ASP	4.2
1	A	401	ASP	4.0
1	A	313	GLY	3.9
1	A	312	LEU	3.8
1	A	256	VAL	3.7
1	A	525	VAL	3.6
1	A	253	LYS	3.6
1	A	358	ARG	3.6
1	A	404	ASP	3.4
1	A	314	PRO	3.2
1	A	403	GLY	2.9
1	A	311	LEU	2.9
1	A	481	ASP	2.7
1	A	254	GLN	2.6
1	A	482	ILE	2.6
1	A	354	VAL	2.4
1	A	249[A]	GLU	2.2
1	A	278	ASN	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 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.