wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2021 – 08:25 AM EDT

PDB ID	:	2SHP
Title	:	TYROSINE PHOSPHATASE SHP-2
Authors	:	Hof, P.; Pluskey, S.; Dhe-Paganon, S.; Eck, M.J.; Shoelson, S.E.
Deposited on	:	1997-12-01
Resolution	:	2.00 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:

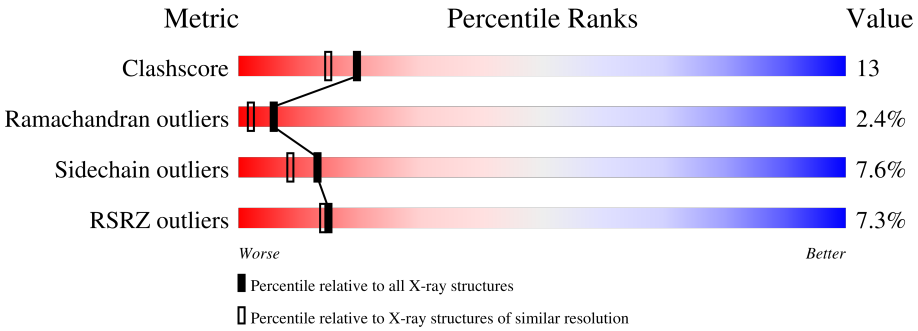
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.00 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))
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	525	6% 68% 21% • 6%
1	B	525	8% 68% 22% • 6%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 12167 atoms, of which 3440 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.

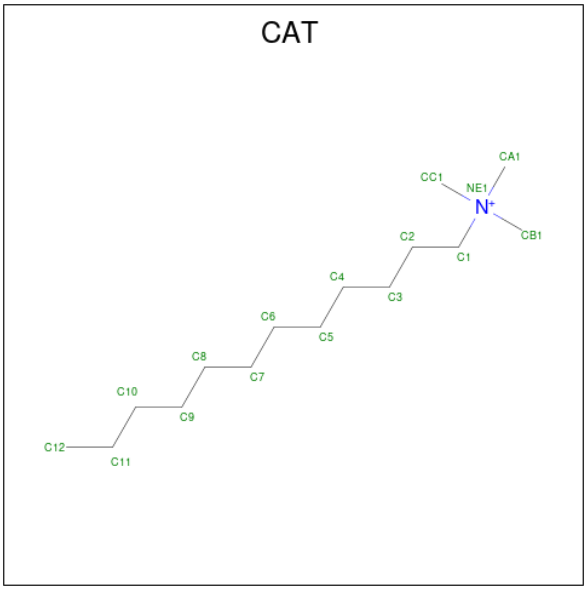
- Molecule 1 is a protein called SHP-2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	491	Total	C	H	N	O	S	2	0	0
			4902	2487	943	709	745	18			
1	B	491	Total	C	H	N	O	S	3	0	0
			4902	2487	943	709	745	18			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	LYS	THR	engineered mutation	UNP Q06124
A	41	LEU	PHE	engineered mutation	UNP Q06124
A	513	SER	PHE	engineered mutation	UNP Q06124
B	2	LYS	THR	engineered mutation	UNP Q06124
B	41	LEU	PHE	engineered mutation	UNP Q06124
B	513	SER	PHE	engineered mutation	UNP Q06124

- Molecule 2 is DODECANE-TRIMETHYLAMINE (three-letter code: CAT) (formula: $C_{15}H_{34}N$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C N 16 15 1	0	0
2	B	1	Total C N 16 15 1	0	0

- Molecule 3 is water.

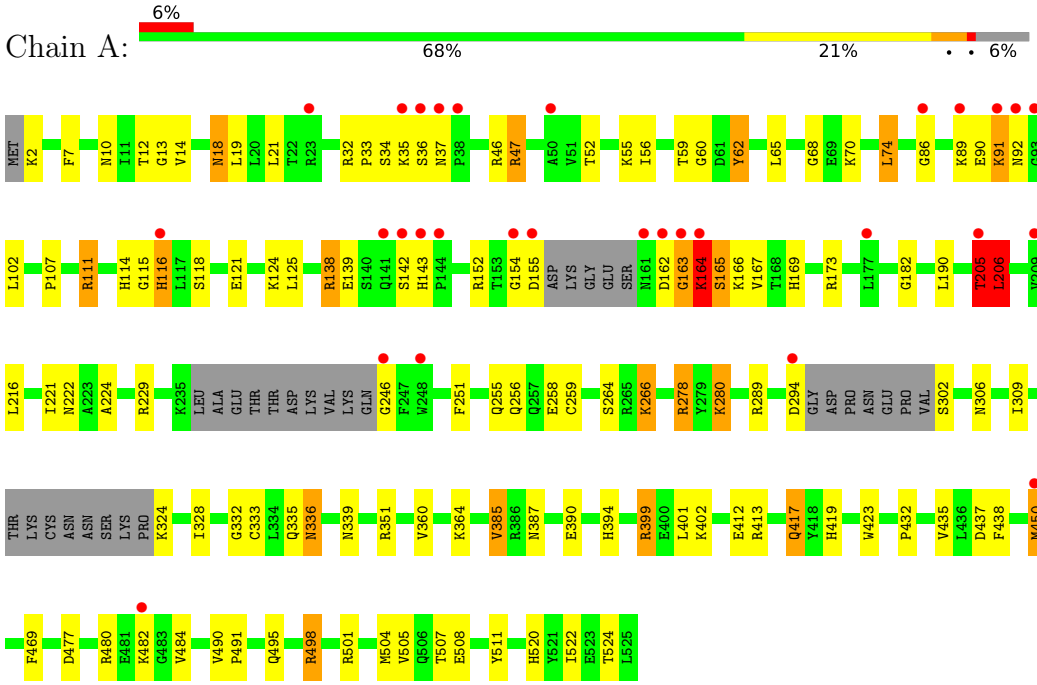
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	401	Total H O 1203 802 401	0	0
3	B	376	Total H O 1128 752 376	0	0

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.

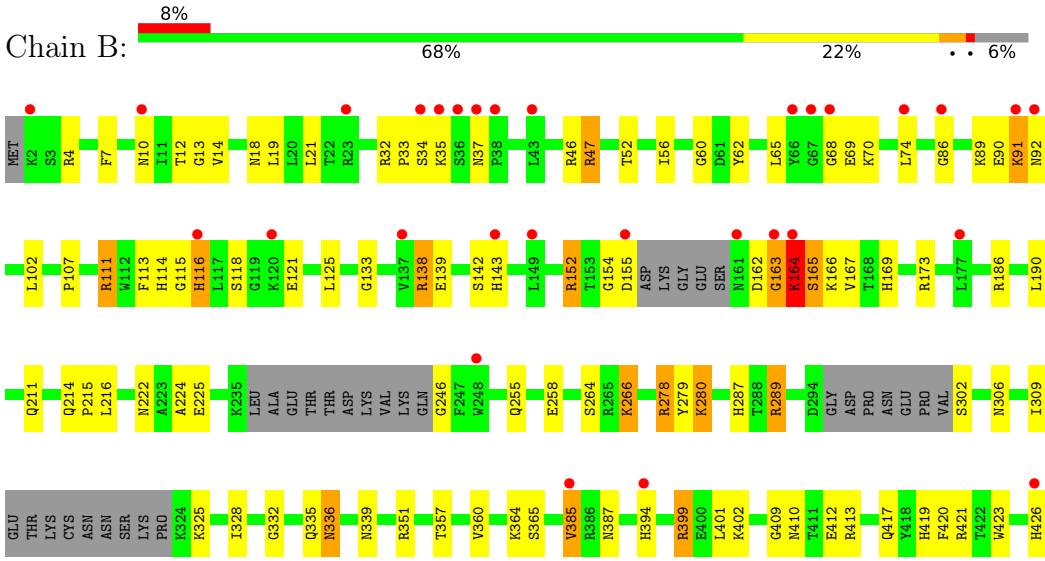
- Molecule 1: SHP-2

Chain A:



- Molecule 1: SHP-2

Chain B:




L436
D437
V442
M450
I463
G464
R465
D477
R480
V484
V490
P491
R498
R501
M504
V505
Q506
T507
E508
Y511
R512
S513
A517
V518
Q519
H520
Y521
I522
E523
T524
L525

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	45.90Å 214.50Å 55.70Å 90.00° 96.30° 90.00°	Depositor
Resolution (Å)	8.00 – 2.00 20.00 – 1.95	Depositor EDS
% Data completeness (in resolution range)	90.1 (8.00-2.00) 89.3 (20.00-1.95)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.81 (at 1.94Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.199 , 0.270 0.240 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	24.3	Xtriage
Anisotropy	0.499	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 51.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12167	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.07% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CAT

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	0.44	1/4038 (0.0%)	0.54	0/5443
1	B	0.44	0/4038	0.55	0/5443
All	All	0.44	1/8076 (0.0%)	0.55	0/10886

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	#Planarity outliers
1	A	0	6
1	B	0	5
All	All	0	11

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	333	CYS	CB-SG	-6.19	1.71	1.82

There are no bond angle outliers.

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	289	ARG	Sidechain
1	A	351	ARG	Sidechain
1	A	399	ARG	Sidechain
1	A	465	ARG	Sidechain
1	A	62	TYR	Sidechain

5.2 Too-close contacts

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	3959	943	3899	101	0
1	B	3959	943	3899	102	0
2	A	16	0	34	4	0
2	B	16	0	34	6	0
3	A	401	802	0	19	2
3	B	376	752	0	17	2
All	All	8727	3440	7866	202	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 13.

The worst 5 of 202 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:B:255:GLN:NE2	1:B:498:ARG:HH22	1.65	0.94
1:A:255:GLN:NE2	1:A:498:ARG:HH22	1.66	0.93
1:B:65:LEU:HD12	1:B:68:GLY:HA3	1.66	0.76
1:B:432:PRO:HG3	2:B:2800:CAT:H111	1.68	0.74
1:B:477:ASP:HA	1:B:480:ARG:HG2	1.70	0.73

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)
3:A:3261:HOH:O	3:B:3595:HOH:H1[2_646]	1.54	0.06
3:A:3464:HOH:O	3:B:3407:HOH:H2[1_556]	1.59	0.01

5.3 Torsion angles

5.3.1 Protein backbone

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	481/525 (92%)	452 (94%)	18 (4%)	11 (2%)	6 2
1	B	481/525 (92%)	450 (94%)	19 (4%)	12 (2%)	5 2
All	All	962/1050 (92%)	902 (94%)	37 (4%)	23 (2%)	6 2

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	142	SER
1	A	206	LEU
1	B	142	SER
1	B	206	LEU
1	A	92	ASN

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	434/467 (93%)	401 (92%)	33 (8%)	13 8
1	B	434/467 (93%)	401 (92%)	33 (8%)	13 8
All	All	868/934 (93%)	802 (92%)	66 (8%)	13 8

5 of 66 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	335	GLN
1	B	339	ASN
1	B	508	GLU
1	A	336	ASN
1	A	335	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	85	HIS
1	B	222	ASN
1	B	417	GLN
1	B	211	GLN
1	B	255	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

2 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	CAT	B	2800	-	15,15,15	0.44	0	17,17,17	0.42	0
2	CAT	A	1800	-	15,15,15	0.39	0	17,17,17	0.44	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
2	CAT	B	2800	-	-	5/13/13/13	-
2	CAT	A	1800	-	-	3/13/13/13	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1800	CAT	C2-C1-NE1-CC1
2	A	1800	CAT	C2-C1-NE1-CA1
2	A	1800	CAT	C2-C1-NE1-CB1
2	B	2800	CAT	C3-C4-C5-C6
2	B	2800	CAT	C7-C8-C9-C10

There are no ring outliers.

2 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	2800	CAT	6	0
2	A	1800	CAT	4	0

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

6.1 Protein, DNA and RNA chains

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	A	491/525 (93%)	0.67	31 (6%)	20	19	11, 27, 45, 58	0
1	B	491/525 (93%)	0.63	41 (8%)	11	10	12, 27, 45, 58	1 (0%)
All	All	982/1050 (93%)	0.65	72 (7%)	15	14	11, 27, 45, 58	1 (0%)

The worst 5 of 72 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	93	GLY	14.0
1	B	92	ASN	9.2
1	A	36	SER	8.8
1	A	163	GLY	8.1
1	B	163	GLY	8.1

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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	B-factors(Å^2)	Q<0.9
2	CAT	A	1800	16/16	0.73	0.24	29,32,38,39	0
2	CAT	B	2800	16/16	0.74	0.24	25,30,41,41	0

6.5 Other polymers (i)

There are no such residues in this entry.