



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 11, 2021 – 06:47 PM EDT

PDB ID : 2HF2  
Title : Domain shifting confirms monomeric structure of Escherichia sugar phosphatase SUPH  
Authors : Patskovsky, Y.; Ramagopal, U.; Almo, S.C.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2006-06-22  
Resolution : 1.90 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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

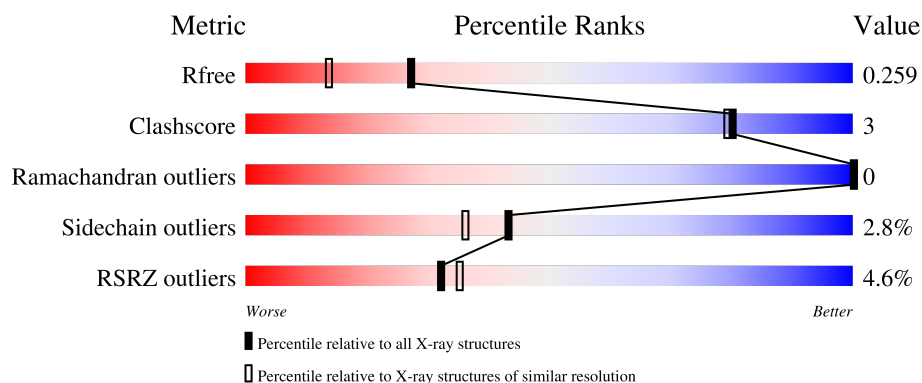
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	283	 6% 86% 7% • 6%
1	B	283	 2% 86% 9% 5%

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4760 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 Sugar phosphatase supH.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	266	Total	C	N	O	S	Se	0	7	0
			2138	1373	360	396	2	7			
1	B	270	Total	C	N	O	S	Se	0	16	0
			2224	1431	373	412	1	7			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MSE	-	SEE REMARK 999	UNP P75792
A	0	SER	-	cloning artifact	UNP P75792
A	1	LEU	MET	engineered mutation	UNP P75792
A	10	MSE	MET	modified residue	UNP P75792
A	27	MSE	MET	modified residue	UNP P75792
A	45	ASN	GLY	engineered mutation	UNP P75792
A	126	MSE	MET	modified residue	UNP P75792
A	173	MSE	MET	modified residue	UNP P75792
A	222	MSE	MET	modified residue	UNP P75792
A	225	MSE	MET	modified residue	UNP P75792
A	232	MSE	MET	modified residue	UNP P75792
A	272	GLU	-	cloning artifact	UNP P75792
A	273	GLY	-	cloning artifact	UNP P75792
A	274	GLY	-	cloning artifact	UNP P75792
A	275	SER	-	cloning artifact	UNP P75792
A	276	HIS	-	expression tag	UNP P75792
A	277	HIS	-	expression tag	UNP P75792
A	278	HIS	-	expression tag	UNP P75792
A	279	HIS	-	expression tag	UNP P75792
A	280	HIS	-	expression tag	UNP P75792
A	281	HIS	-	expression tag	UNP P75792
B	-1	MSE	-	SEE REMARK 999	UNP P75792
B	0	SER	-	cloning artifact	UNP P75792
B	1	LEU	MET	engineered mutation	UNP P75792
B	10	MSE	MET	modified residue	UNP P75792

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	27	MSE	MET	modified residue	UNP P75792
B	45	ASN	GLY	engineered mutation	UNP P75792
B	126	MSE	MET	modified residue	UNP P75792
B	173	MSE	MET	modified residue	UNP P75792
B	222	MSE	MET	modified residue	UNP P75792
B	225	MSE	MET	modified residue	UNP P75792
B	232	MSE	MET	modified residue	UNP P75792
B	272	GLU	-	cloning artifact	UNP P75792
B	273	GLY	-	cloning artifact	UNP P75792
B	274	GLY	-	cloning artifact	UNP P75792
B	275	SER	-	cloning artifact	UNP P75792
B	276	HIS	-	expression tag	UNP P75792
B	277	HIS	-	expression tag	UNP P75792
B	278	HIS	-	expression tag	UNP P75792
B	279	HIS	-	expression tag	UNP P75792
B	280	HIS	-	expression tag	UNP P75792
B	281	HIS	-	expression tag	UNP P75792

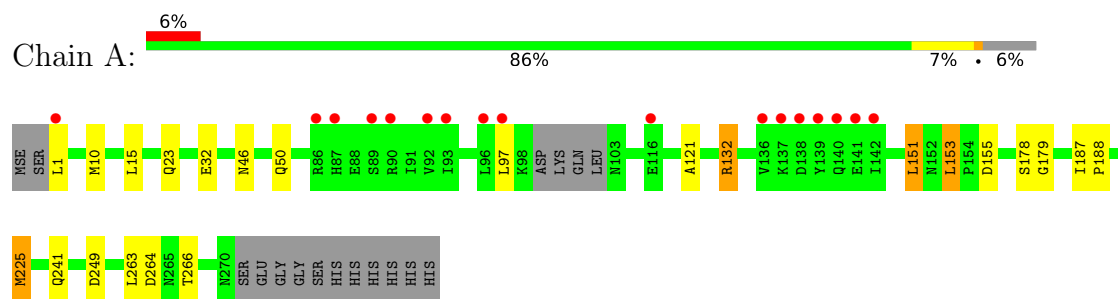
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	189	Total O 189 189	0	0
2	B	209	Total O 209 209	0	0

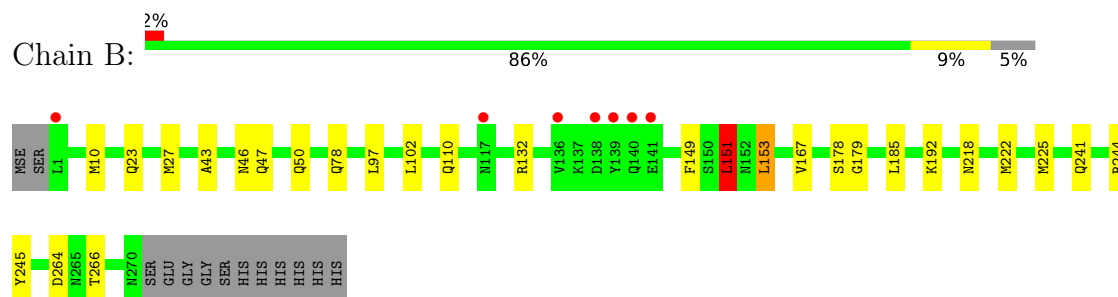
### 3 Residue-property plots [i](#)

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: Sugar phosphatase supH



#### • Molecule 1: Sugar phosphatase supH



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	38.77Å 111.30Å 68.80Å 90.00° 106.23° 90.00°	Depositor
Resolution (Å)	20.00 – 1.90 29.07 – 1.90	Depositor EDS
% Data completeness (in resolution range)	95.7 (20.00-1.90) 95.1 (29.07-1.90)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.09 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.204 , 0.252 0.210 , 0.259	Depositor DCC
$R_{free}$ test set	1330 reflections (3.14%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.6	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 32.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.200 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4760	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.84% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<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.

## 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	0.53	0/2193	0.71	5/2954 (0.2%)
1	B	0.60	1/2309 (0.0%)	0.69	5/3112 (0.2%)
All	All	0.57	1/4502 (0.0%)	0.70	10/6066 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	10	MSE	SE-CE	-15.91	1.01	1.95

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	132	ARG	NE-CZ-NH2	-8.72	115.94	120.30
1	A	132	ARG	NE-CZ-NH1	-8.47	116.06	120.30
1	A	132	ARG	NE-CZ-NH2	8.22	124.41	120.30
1	B	132	ARG	NE-CZ-NH1	7.89	124.25	120.30
1	A	153	LEU	CA-CB-CG	6.76	130.84	115.30
1	B	153[A]	LEU	CA-CB-CG	5.93	128.95	115.30
1	B	153[B]	LEU	CA-CB-CG	5.93	128.95	115.30
1	A	151	LEU	CA-CB-CG	5.86	128.77	115.30
1	B	151	LEU	CA-CB-CG	5.86	128.77	115.30
1	A	249	ASP	CB-CG-OD2	5.34	123.10	118.30

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	2138	0	2147	11	0
1	B	2224	0	2262	13	0
2	A	189	0	0	2	0
2	B	209	0	0	3	0
All	All	4760	0	4409	24	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 3.

All (24) 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:23:GLN:HB3	2:A:325:HOH:O	1.85	0.76
1:A:10:MSE:HE3	1:A:15:LEU:HG	1.71	0.72
1:A:264:ASP:HB3	1:A:266[B]:THR:HG23	1.79	0.65
1:B:222:MSE:O	1:B:225:MSE:HG3	1.99	0.61
1:A:46:ASN:HB3	1:A:50:GLN:HG3	1.87	0.57
1:A:264:ASP:CB	1:A:266[B]:THR:HG23	2.38	0.53
1:B:102:LEU:HD21	1:B:151:LEU:HG	1.91	0.52
1:B:264:ASP:HB3	1:B:266[B]:THR:HG23	1.92	0.51
1:B:23[B]:GLN:HE22	1:B:27:MSE:HE2	1.75	0.50
1:A:32:GLU:HG3	1:A:263:LEU:HD22	1.94	0.49
1:A:187:ILE:HA	1:A:188[B]:PRO:HD3	1.69	0.46
1:B:149:PHE:HB2	1:B:185:LEU:HB2	1.98	0.46
1:B:47:GLN:H	1:B:50:GLN:NE2	2.14	0.45
1:A:225:MSE:HB3	1:A:225:MSE:HE2	1.59	0.44
1:B:47:GLN:H	1:B:50:GLN:HE21	1.65	0.44
1:B:43:ALA:HB1	1:B:192:LYS:HG2	2.02	0.42
1:B:78:GLN:NE2	2:B:459:HOH:O	2.52	0.41
1:B:178[B]:SER:HB2	1:B:179:GLY:H	1.61	0.41
1:B:46:ASN:HB3	1:B:50:GLN:HG3	2.02	0.41
1:A:1:LEU:HD22	2:A:387:HOH:O	2.21	0.41
1:A:121:ALA:HA	2:B:489:HOH:O	2.19	0.41
1:A:178[B]:SER:HB2	1:A:179:GLY:H	1.62	0.41
1:B:110:GLN:NE2	2:B:434:HOH:O	2.52	0.41
1:B:244[B]:ARG:HD3	1:B:245:TYR:CZ	2.56	0.40

There are no symmetry-related clashes.



## 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	269/283 (95%)	255 (95%)	14 (5%)	0	100	100
1	B	284/283 (100%)	275 (97%)	9 (3%)	0	100	100
All	All	553/566 (98%)	530 (96%)	23 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains

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	234/234 (100%)	227 (97%)	7 (3%)	41	33
1	B	247/234 (106%)	239 (97%)	8 (3%)	39	30
All	All	481/468 (103%)	466 (97%)	15 (3%)	43	32

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	97	LEU
1	A	132	ARG
1	A	151	LEU
1	A	153	LEU
1	A	155	ASP
1	A	225	MSE
1	A	241	GLN
1	B	97	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	151	LEU
1	B	153[A]	LEU
1	B	153[B]	LEU
1	B	167	VAL
1	B	218	ASN
1	B	241[A]	GLN
1	B	241[B]	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	140	GLN
1	A	209	ASN
1	B	50	GLN
1	B	87	HIS
1	B	110	GLN
1	B	140	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 ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	259/283 (91%)	0.20	17 (6%) 18 20	19, 32, 62, 103	0
1	B	263/283 (92%)	0.07	7 (2%) 54 57	18, 32, 58, 104	0
All	All	522/566 (92%)	0.13	24 (4%) 32 35	18, 32, 62, 104	0

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1	LEU	10.4
1	A	139	TYR	6.9
1	A	87	HIS	6.6
1	B	139	TYR	6.4
1	A	140	GLN	6.4
1	A	86	ARG	4.5
1	A	138	ASP	4.5
1	A	141	GLU	4.3
1	A	93	ILE	4.0
1	A	90	ARG	3.6
1	A	137	LYS	3.6
1	A	97	LEU	3.6
1	A	1	LEU	3.1
1	B	140	GLN	3.0
1	B	138	ASP	2.9
1	A	96	LEU	2.6
1	A	136	VAL	2.5
1	B	117	ASN	2.4
1	A	116	GLU	2.4
1	B	141	GLU	2.4
1	A	142	ILE	2.3
1	B	136	VAL	2.2
1	A	89	SER	2.2
1	A	92	VAL	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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.