



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 21, 2022 – 10:42 AM EST

PDB ID : 7MOK  
Title : Crystal Structure of Arabidopsis thaliana Plant and Fungi Atypical Dual Specificity Phosphatase 1(AtPFA-DSP1 ) in Complex with Phosphate in Conformation A (Pi(A))  
Authors : Wang, H.; Shears, S.B.  
Deposited on : 2021-05-01  
Resolution : 1.85 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 : 2.26  
buster-report : 1.1.7 (2018)  
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) : 2.26

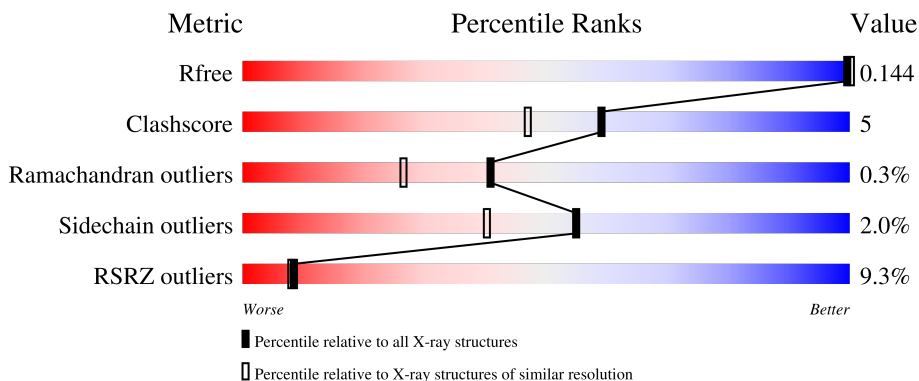
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	171	 12% 89% 8% .
1	B	171	 6% 84% 12% ..

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	BME	A	502	-	X	-	-
3	BME	B	502	-	X	-	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3210 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 DSP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	166	1411	906	245	246	14	0	8	0
1	B	166	1398	895	245	246	12	0	8	0

There are 8 discrepancies between the modelled and reference sequences:

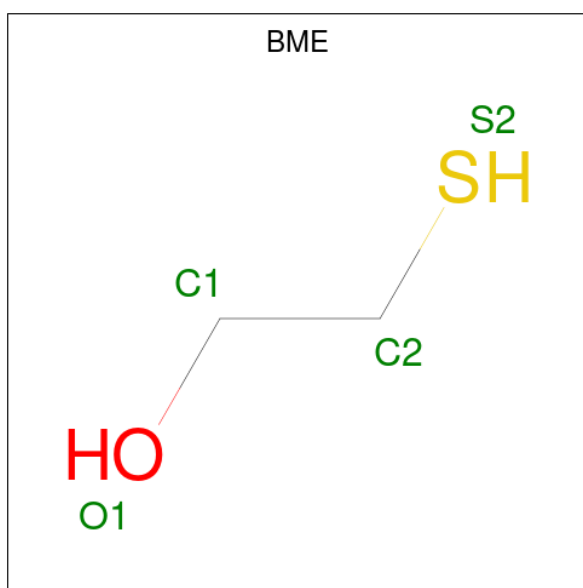
Chain	Residue	Modelled	Actual	Comment	Reference
A	45	GLY	-	expression tag	UNP Q9ZVN4
A	46	SER	-	expression tag	UNP Q9ZVN4
A	47	PHE	-	expression tag	UNP Q9ZVN4
A	48	THR	-	expression tag	UNP Q9ZVN4
B	45	GLY	-	expression tag	UNP Q9ZVN4
B	46	SER	-	expression tag	UNP Q9ZVN4
B	47	PHE	-	expression tag	UNP Q9ZVN4
B	48	THR	-	expression tag	UNP Q9ZVN4

- Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	P	0	0
			5	4	1		
2	B	1	Total	O	P	0	0
			5	4	1		

- Molecule 3 is BETA-MERCAPTOETHANOL (three-letter code: BME) (formula: C<sub>2</sub>H<sub>6</sub>OS).



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

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	B	1	Total	C	O	S	0	0
			4	2	1	1		
3	B	1	Total	C	O	S	0	0
			4	2	1	1		
3	B	1	Total	C	O	S	0	0
			4	2	1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	178	Total	O	0	0
			178	178		
4	B	193	Total	O	0	0
			193	193		



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	124.45Å 124.45Å 124.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.38 – 1.85 39.35 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.9 (39.38-1.85) 100.0 (39.35-1.85)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.74 (at 1.85Å)	Xtrriage
Refinement program	REFMAC 5.8.0258	Depositor
R, $R_{free}$	0.108 , 0.143 0.111 , 0.144	Depositor DCC
$R_{free}$ test set	2782 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.5	Xtrriage
Anisotropy	0.000	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 62.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.035 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3210	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.72% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: BME, PO4

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	0.70	0/1447	0.77	0/1949
1	B	0.65	0/1433	0.75	0/1929
All	All	0.67	0/2880	0.76	0/3878

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

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	1411	0	1402	11	0
1	B	1398	0	1387	17	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
3	A	4	0	6	0	0
3	B	16	0	24	0	0
4	A	178	0	0	3	0
4	B	193	0	0	5	0
All	All	3210	0	2819	27	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 5.

All (27) 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:159[B]:CYS:SG	1:A:194:PHE:CD1	2.69	0.85
1:A:49:GLU:HG3	1:A:52:HIS:HB3	1.72	0.71
1:B:64:ASN:OD1	1:B:212[B]:CYS:SG	2.50	0.70
1:A:104[B]:LEU:HD21	1:A:111:LEU:HB2	1.77	0.67
1:B:119:ASN:HB2	1:B:124:VAL:HG12	1.80	0.64
1:A:121:GLU:HB2	4:A:614:HOH:O	1.98	0.63
1:B:180[A]:GLN:NE2	4:B:603:HOH:O	2.33	0.61
1:A:141:GLU:HG2	4:A:693:HOH:O	2.05	0.55
1:B:49:GLU:HG3	1:B:49:GLU:O	2.07	0.54
1:B:88:ILE:HD12	1:B:104:LEU:HD21	1.90	0.52
1:B:64:ASN:CG	1:B:212[B]:CYS:SG	2.89	0.50
1:B:110:ARG:NH1	4:B:607:HOH:O	2.46	0.49
1:B:88:ILE:HD11	4:B:782:HOH:O	2.13	0.48
1:B:124:VAL:HG21	1:B:190:SER:OG	2.14	0.47
1:A:159[B]:CYS:SG	1:A:194:PHE:CE1	3.08	0.46
1:B:176:PHE:O	1:B:180[B]:GLN:HG3	2.16	0.45
1:B:152:ARG:CG	1:B:152:ARG:HH11	2.29	0.45
1:B:49:GLU:O	1:B:52:HIS:HB3	2.16	0.44
1:A:214:ILE:C	4:A:612:HOH:O	2.55	0.43
1:B:121:GLU:HG3	4:B:774:HOH:O	2.17	0.43
1:A:181:ARG:HG2	1:B:181:ARG:HG2	2.02	0.42
1:B:122:PRO:HD2	1:B:123:PHE:CD2	2.54	0.42
1:A:49:GLU:HG3	1:A:52:HIS:CB	2.45	0.42
1:A:159[B]:CYS:SG	1:A:194:PHE:HD1	2.36	0.42
1:A:181:ARG:HD3	4:B:626:HOH:O	2.20	0.42
1:B:116:ILE:HG23	1:B:127:PRO:CD	2.50	0.42
1:B:124:VAL:HG11	1:B:188[B]:ARG:NH2	2.35	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	Percentiles	
1	A	174/171 (102%)	166 (95%)	7 (4%)	1 (1%)	25	12
1	B	172/171 (101%)	167 (97%)	5 (3%)	0	100	100
All	All	346/342 (101%)	333 (96%)	12 (4%)	1 (0%)	41	26

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	50	GLU

### 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	161/155 (104%)	158 (98%)	3 (2%)	57	43
1	B	159/155 (103%)	156 (98%)	3 (2%)	57	43
All	All	320/310 (103%)	314 (98%)	6 (2%)	55	43

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	71	PHE
1	A	96	TYR
1	A	117	GLU
1	B	50	GLU
1	B	96	TYR
1	B	152	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	52	HIS
1	A	180	GLN

### 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](#)

7 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
3	BME	B	504	-	3,3,3	0.17	0	1,2,2	0.20	0
3	BME	B	502	-	3,3,3	1.40	1 (33%)	1,2,2	2.03	1 (100%)
3	BME	A	502	-	3,3,3	1.39	1 (33%)	1,2,2	2.06	1 (100%)
2	PO4	B	501	-	4,4,4	1.66	1 (25%)	6,6,6	0.49	0
3	BME	B	505	-	3,3,3	0.21	0	1,2,2	0.21	0
2	PO4	A	501	-	4,4,4	1.57	1 (25%)	6,6,6	0.36	0
3	BME	B	503	-	3,3,3	0.22	0	1,2,2	0.38	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
3	BME	B	504	-	-	1/1/1/1	-
3	BME	B	502	-	-	1/1/1/1	-

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	BME	A	502	-	-	1/1/1/1	-
3	BME	B	505	-	-	1/1/1/1	-
3	BME	B	503	-	-	1/1/1/1	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	501	PO4	P-O1	3.11	1.58	1.50
2	B	501	PO4	P-O1	2.57	1.56	1.50
3	B	502	BME	O1-C1	2.23	1.53	1.42
3	A	502	BME	O1-C1	2.10	1.52	1.42

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	502	BME	O1-C1-C2	2.06	118.96	110.83
3	B	502	BME	O1-C1-C2	2.03	118.83	110.83

There are no chirality outliers.

All (5) torsion outliers are listed below:

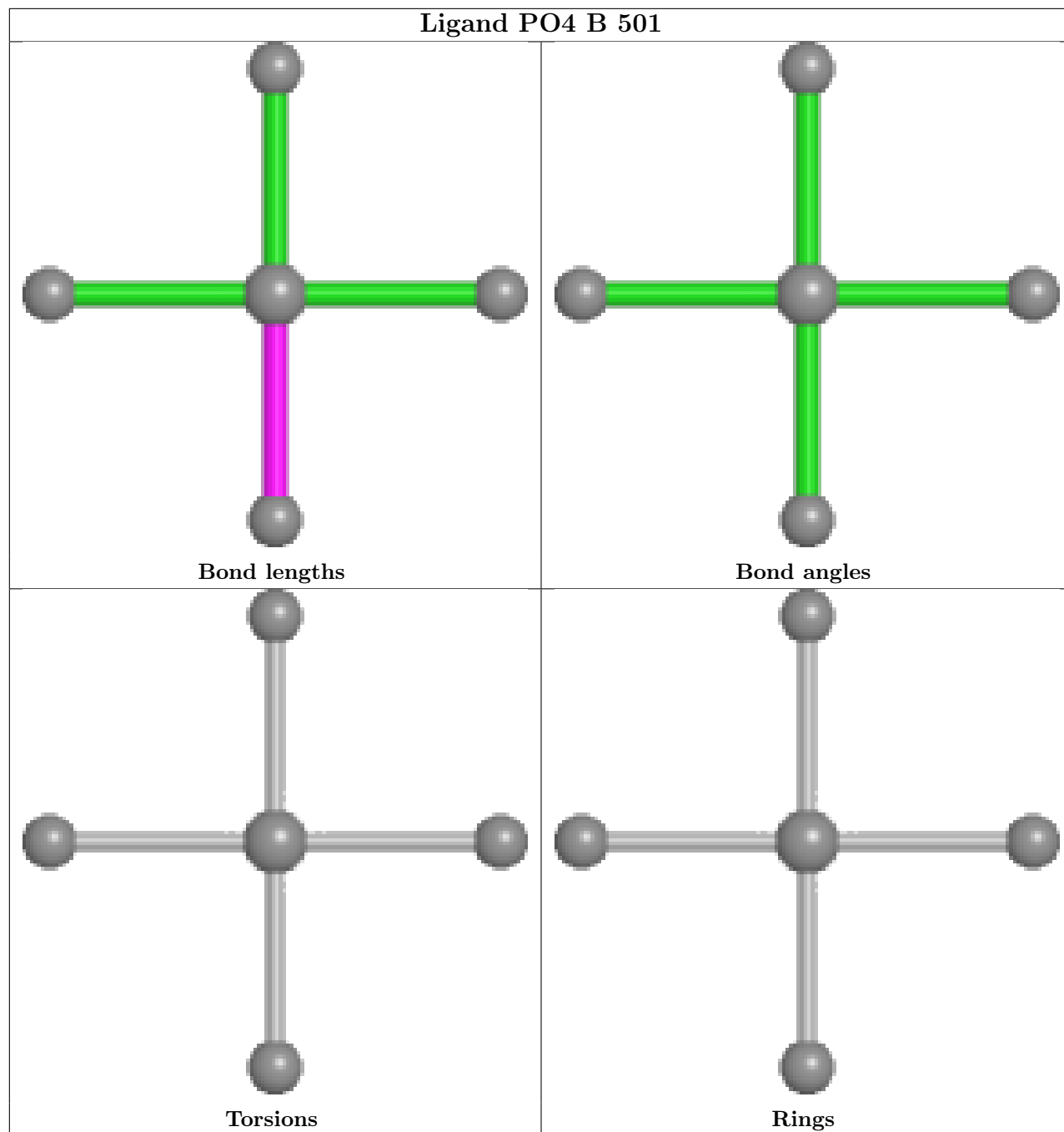
Mol	Chain	Res	Type	Atoms
3	A	502	BME	O1-C1-C2-S2
3	B	503	BME	O1-C1-C2-S2
3	B	504	BME	O1-C1-C2-S2
3	B	505	BME	O1-C1-C2-S2
3	B	502	BME	O1-C1-C2-S2

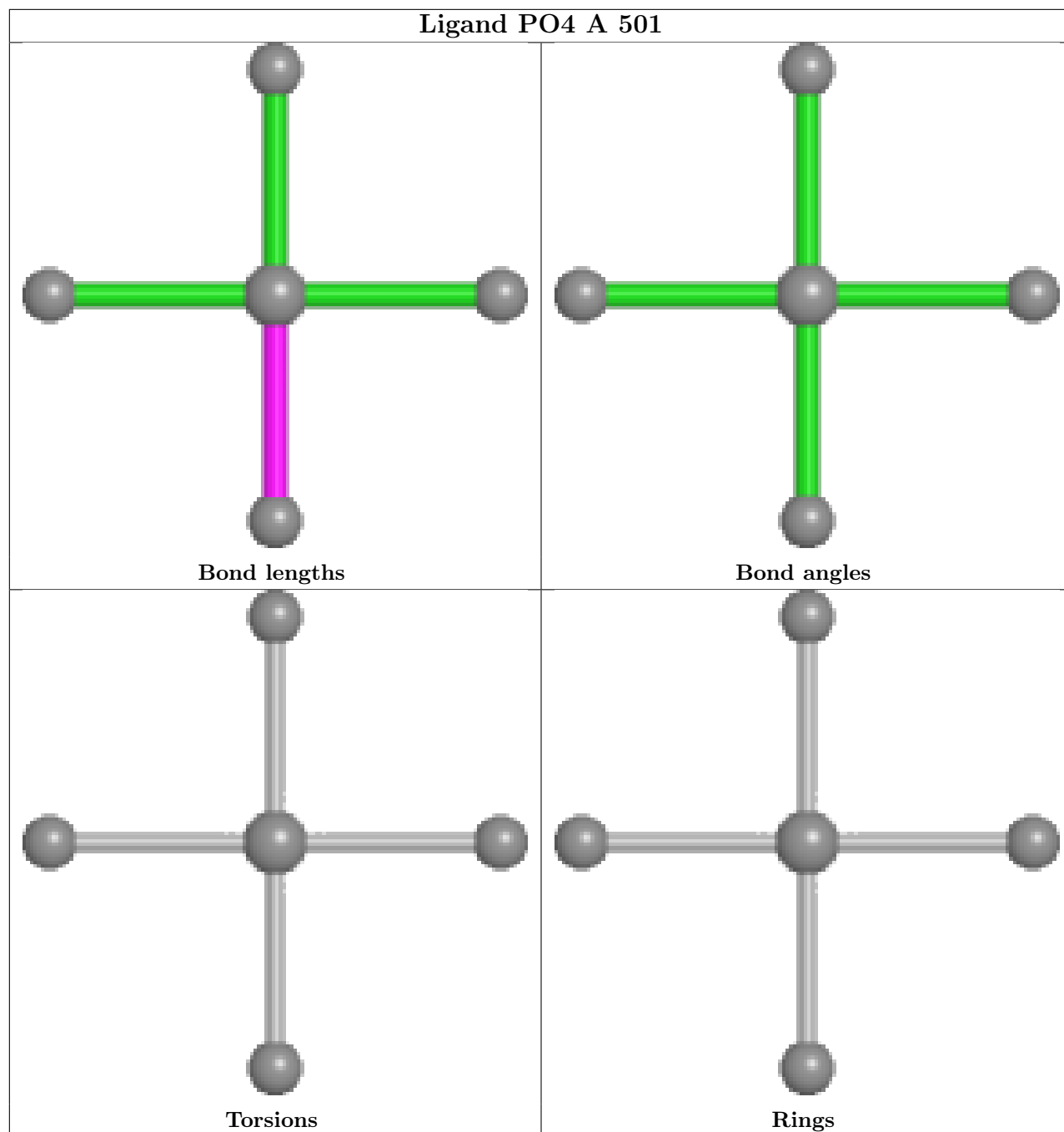
There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient

equivalents in the CSD to analyse the geometry.





### 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<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	166/171 (97%)	0.02	20 (12%) 4 4	13, 23, 75, 90	0
1	B	166/171 (97%)	-0.22	11 (6%) 18 17	11, 21, 59, 94	0
All	All	332/342 (97%)	-0.10	31 (9%) 8 8	11, 22, 71, 94	0

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	123	PHE	7.8
1	B	124	VAL	7.3
1	B	123	PHE	6.5
1	A	124	VAL	5.6
1	A	119	ASN	5.3
1	B	49	GLU	5.0
1	A	49	GLU	5.0
1	B	119	ASN	4.9
1	A	205	HIS	4.2
1	A	122	PRO	4.2
1	A	204	SER	4.1
1	B	125	ASN	3.9
1	B	118	GLY	3.9
1	B	50	GLU	3.8
1	A	207	PRO	3.8
1	A	118	GLY	3.5
1	A	125	ASN	3.2
1	B	122	PRO	3.1
1	B	126	ILE	3.1
1	A	206	ILE	3.1
1	A	209	SER	3.0
1	B	120	LYS	2.5
1	A	202	SER	2.3
1	A	117	GLU	2.3

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	121	GLU	2.2
1	B	121	GLU	2.2
1	A	200[A]	VAL	2.2
1	A	213	SER	2.1
1	A	120	LYS	2.1
1	A	214	ILE	2.0
1	A	50	GLU	2.0

## 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](#)

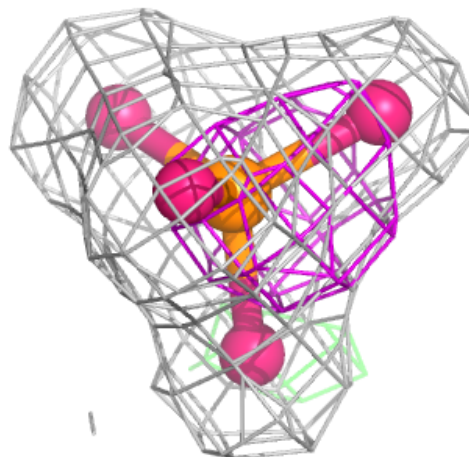
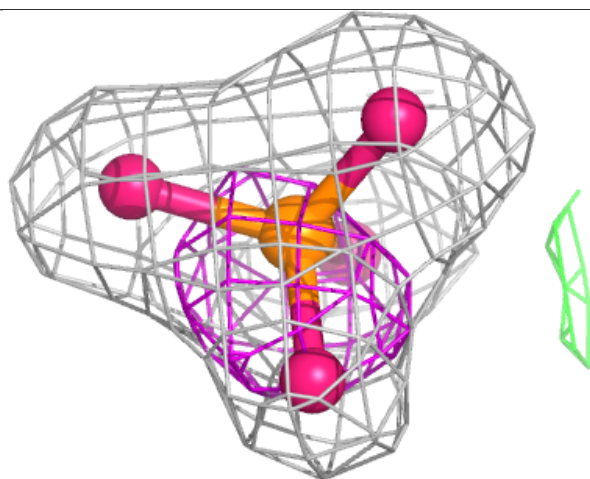
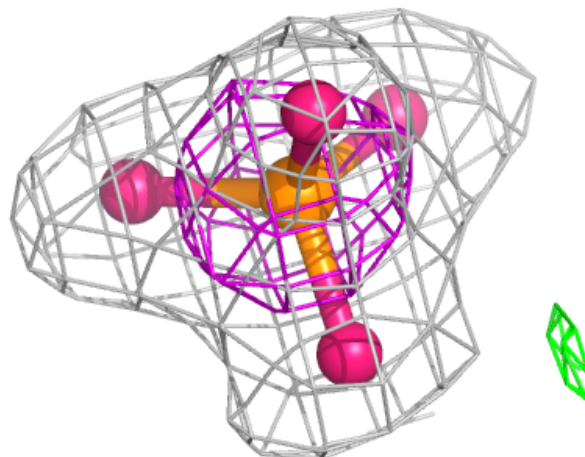
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

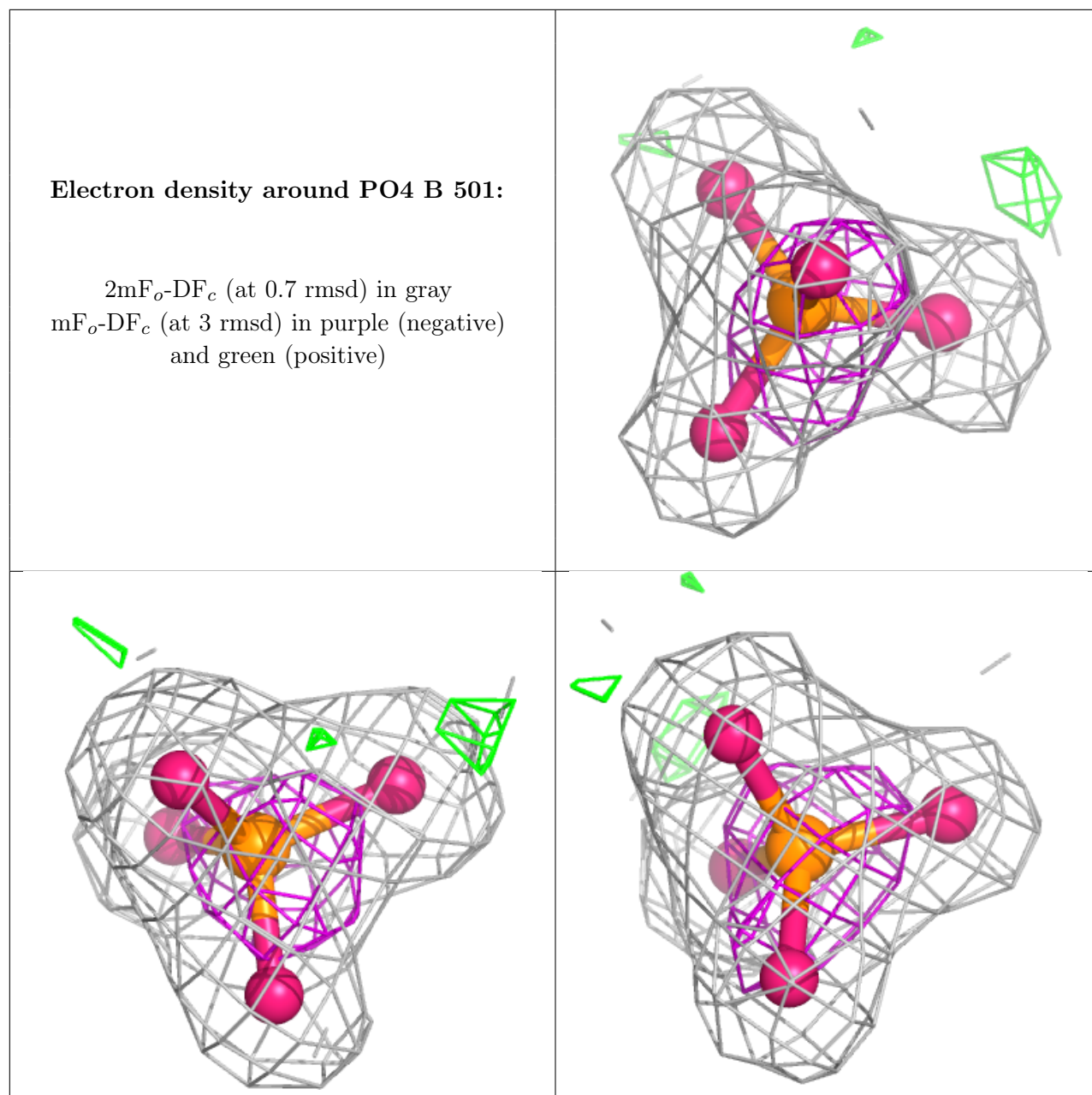
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
3	BME	B	505	4/4	0.78	0.34	72,85,92,107	0
3	BME	B	503	4/4	0.84	0.17	57,64,72,84	0
3	BME	B	502	4/4	0.92	0.13	24,39,39,52	0
3	BME	A	502	4/4	0.93	0.12	33,41,45,48	0
3	BME	B	504	4/4	0.94	0.24	79,94,95,96	0
2	PO4	A	501	5/5	0.98	0.07	23,27,34,37	0
2	PO4	B	501	5/5	0.99	0.06	19,20,27,31	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

**Electron density around PO4 A 501:**

$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)





## 6.5 Other polymers [i](#)

There are no such residues in this entry.