



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 4, 2026 – 10:43 PM UTC

PDB ID : 1G6C / pdb_00001g6c
Title : THIAMIN PHOSPHATE SYNTHASE
Authors : Peapus, D.H.; Chiu, H.-J.; Campobasso, N.; Reddick, J.J.; Begley, T.P.; Ealick, S.E.
Deposited on : 2000-11-03
Resolution : 1.40 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 ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

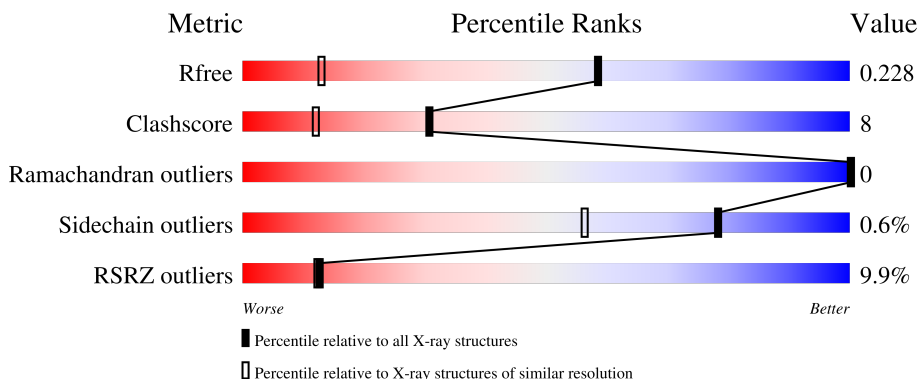
1 Overall quality at a glance

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}	180053	2563 (1.40-1.40)
Clashscore	190562	2660 (1.40-1.40)
Ramachandran outliers	187476	2611 (1.40-1.40)
Sidechain outliers	187428	2610 (1.40-1.40)
RSRZ outliers	180081	2561 (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 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 $\leq 5\%$. 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	227	10% 79% 20%
1	B	227	10% 89% 10%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3780 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 THIAMIN PHOSPHATE SYNTHASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	226	Total	C	N	O	S	0	4	0
			1700	1054	303	335	8			
1	B	227	Total	C	N	O	S	0	7	1
			1709	1060	307	334	8			

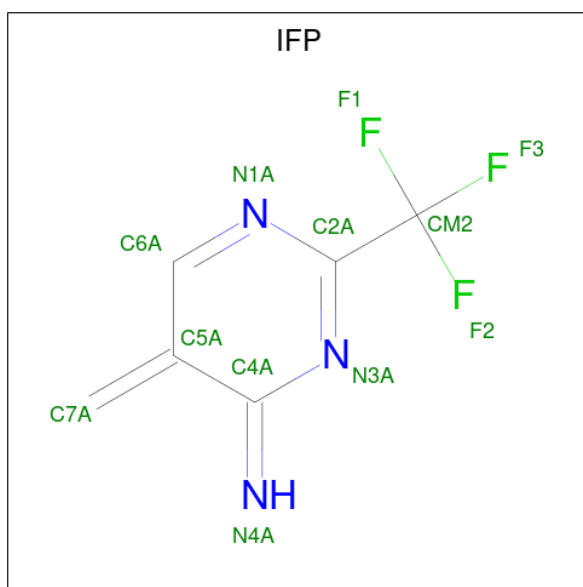
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	HIS	-	cloning artifact	UNP P39594
A	10	HIS	-	cloning artifact	UNP P39594
A	11	GLY	-	cloning artifact	UNP P39594
A	12	ILE	-	cloning artifact	UNP P39594
A	13	ARG	-	cloning artifact	UNP P39594
A	130	ALA	SER	engineered mutation	UNP P39594
B	1009	HIS	-	cloning artifact	UNP P39594
B	1010	HIS	-	cloning artifact	UNP P39594
B	1011	GLY	-	cloning artifact	UNP P39594
B	1012	ILE	-	cloning artifact	UNP P39594
B	1013	ARG	-	cloning artifact	UNP P39594
B	1130	ALA	SER	engineered mutation	UNP P39594

- Molecule 2 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

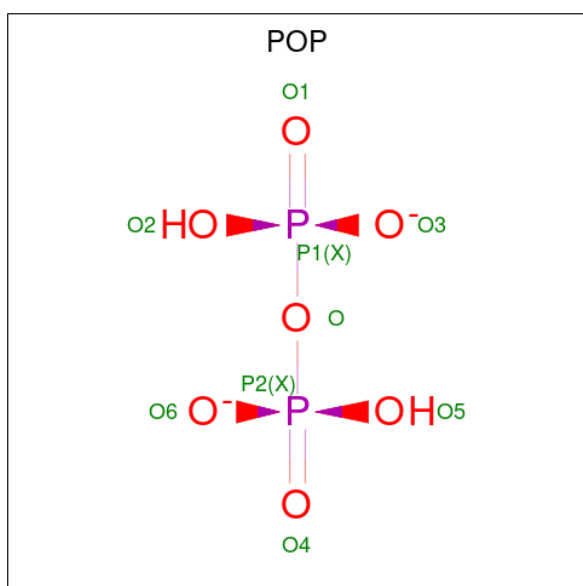
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Mg	0	0
			1	1		
2	B	1	Total	Mg	0	0
			1	1		

- Molecule 3 is 2-TRIFLUOROMETHYL-5-METHYLENE-5H-PYRIMIDIN-4-YLIDENEA MINE (CCD ID: IFP) (formula: C₆H₄F₃N₃).



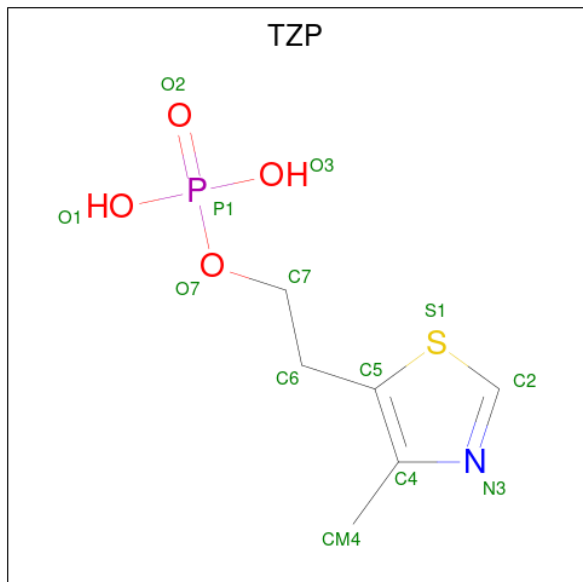
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	F	N	0	0
			12	6	3	3		
3	B	1	Total	C	F	N	0	0
			12	6	3	3		

- Molecule 4 is PYROPHOSPHATE 2- (CCD ID: POP) (formula: $\text{H}_2\text{O}_7\text{P}_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	O	P	0	1
			13	10	3		
4	B	1	Total	O	P	0	1
			13	10	3		

- Molecule 5 is 4-METHYL-5-HYDROXYETHYLTHIAZOLE PHOSPHATE (CCD ID: TZP) (formula: C₆H₁₀NO₄PS).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
5	A	1	Total	C	N	O	P	S	0	0
			13	6	1	4	1	1		
5	B	1	Total	C	N	O	P	S	0	0
			13	6	1	4	1	1		

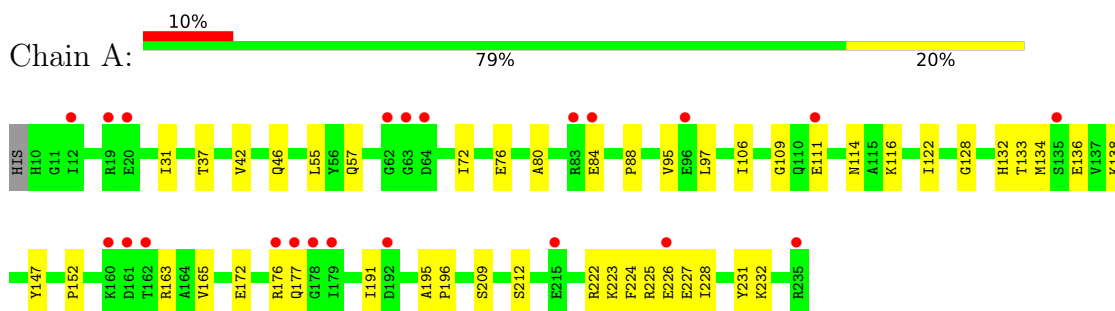
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	144	Total	O	0	0
			144	144		
6	B	149	Total	O	0	0
			149	149		

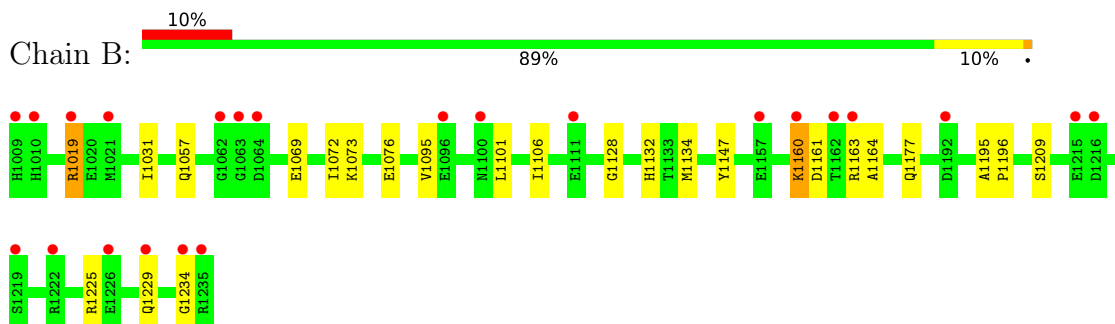
3 Residue-property plots

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: THIAMIN PHOSPHATE SYNTHASE



• Molecule 1: THIAMIN PHOSPHATE SYNTHASE



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	76.65Å 76.65Å 140.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.86 – 1.40 42.86 – 1.40	Depositor EDS
% Data completeness (in resolution range)	93.6 (42.86-1.40) 93.7 (42.86-1.40)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.66 (at 1.40Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.215 , 0.237 0.208 , 0.228	Depositor DCC
R_{free} test set	3938 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	11.6	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 39.0	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	0.94	EDS
Total number of atoms	3780	wwPDB-VP
Average B, all atoms (Å ²)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 40.77 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.6051e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: IFP, MG, POP, TZP

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	0.35	0/1739	0.80	2/2341 (0.1%)
1	B	0.38	1/1767 (0.1%)	0.82	2/2380 (0.1%)
All	All	0.36	1/3506 (0.0%)	0.81	4/4721 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1234	GLY	C-N	-5.70	1.25	1.33

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	209	SER	N-CA-C	7.18	119.97	111.71
1	B	1209	SER	N-CA-C	6.73	118.62	111.28
1	B	1164	ALA	N-CA-C	5.28	117.69	110.24
1	A	122	ILE	N-CA-C	5.14	117.92	112.83

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	1700	0	1718	35	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	1709	0	1729	22	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	12	0	3	0	0
3	B	12	0	3	1	0
4	A	13	0	0	1	0
4	B	13	0	0	0	0
5	A	13	0	8	0	0
5	B	13	0	8	1	0
6	A	144	0	0	1	0
6	B	149	0	0	0	0
All	All	3780	0	3469	58	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 8.

The worst 5 of 58 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:1019:ARG:HH11	1:B:1019:ARG:HB2	1.35	0.90
1:A:116:LYS:HB2	1:A:116:LYS:HZ2	1.39	0.86
1:B:1160:LYS:HE3	1:B:1160:LYS:H	1.51	0.75
1:B:1019:ARG:HB2	1:B:1019:ARG:NH1	2.05	0.71
1:B:1160:LYS:HE3	1:B:1160:LYS:N	2.14	0.62

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	228/227 (100%)	223 (98%)	5 (2%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	232/227 (102%)	227 (98%)	5 (2%)	0	100	100
All	All	460/454 (101%)	450 (98%)	10 (2%)	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	177/174 (102%)	177 (100%)	0	100	100
1	B	180/174 (103%)	178 (99%)	2 (1%)	65	37
All	All	357/348 (103%)	355 (99%)	2 (1%)	78	56

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1019	ARG
1	B	1160	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	1213	GLN
1	B	1100	ASN
1	A	229	GLN
1	A	199	GLN
1	B	1079	GLN

5.3.3 RNA ⓘ

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 2 are monoatomic - leaving 8 for Mogul analysis.

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	IFP	A	2001	-	9,12,12	2.55	2 (22%)	11,18,18	2.70	3 (27%)
4	POP	A	2003[A]	2	6,8,8	1.68	2 (33%)	12,13,13	1.04	0
4	POP	B	2004[B]	2	6,8,8	1.73	2 (33%)	12,13,13	1.10	0
5	TZP	A	2005	-	12,13,13	1.49	2 (16%)	15,18,18	1.60	2 (13%)
4	POP	B	2004[A]	2	6,8,8	1.74	3 (50%)	12,13,13	1.01	0
4	POP	A	2003[B]	2	6,8,8	1.57	0	12,13,13	1.08	0
5	TZP	B	2006	-	12,13,13	1.57	2 (16%)	15,18,18	1.60	2 (13%)
3	IFP	B	2002	-	9,12,12	2.65	3 (33%)	11,18,18	2.75	3 (27%)

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	IFP	A	2001	-	-	0/6/19/19	0/1/1/1
4	POP	A	2003[A]	2	-	0/6/6/6	-
4	POP	B	2004[B]	2	-	0/6/6/6	-
5	TZP	A	2005	-	-	0/7/7/7	0/1/1/1
4	POP	B	2004[A]	2	-	0/6/6/6	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	POP	A	2003[B]	2	-	2/6/6/6	-
5	TZP	B	2006	-	-	0/7/7/7	0/1/1/1
3	IFP	B	2002	-	-	0/6/19/19	0/1/1/1

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	2002	IFP	C7A-C5A	7.09	1.49	1.33
3	A	2001	IFP	C7A-C5A	6.89	1.49	1.33
5	B	2006	TZP	C4-N3	-3.42	1.33	1.38
5	A	2005	TZP	C4-N3	-3.05	1.33	1.38
5	B	2006	TZP	C2-S1	-2.97	1.67	1.72

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	2002	IFP	C5A-C6A-N1A	-6.11	121.81	126.20
3	A	2001	IFP	C5A-C6A-N1A	-5.92	121.95	126.20
3	A	2001	IFP	CM2-C2A-N3A	5.63	120.03	114.11
3	B	2002	IFP	CM2-C2A-N3A	5.60	120.00	114.11
5	A	2005	TZP	S1-C2-N3	-3.76	112.85	116.08

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	2003[B]	POP	P1-O-P2-O5
4	A	2003[B]	POP	P1-O-P2-O4

There are no ring outliers.

3 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	2003[B]	POP	1	0
5	B	2006	TZP	1	0
3	B	2002	IFP	1	0

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, 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	226/227 (99%)	0.68	22 (9%) 13 13	6, 12, 24, 31	4 (1%)
1	B	227/227 (100%)	0.53	23 (10%) 12 12	7, 12, 23, 36	7 (3%)
All	All	453/454 (99%)	0.61	45 (9%) 13 12	6, 12, 24, 36	11 (2%)

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1157	GLU	4.8
1	B	1064	ASP	4.6
1	B	1234	GLY	4.0
1	B	1062	GLY	4.0
1	B	1063	GLY	3.9

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 oligosaccharides 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($\text{\AA}^2$)	Q<0.9
4	POP	A	2003[A]	9/9	0.93	0.14	9,10,14,15	4
4	POP	A	2003[B]	9/9	0.93	0.14	6,9,11,11	4
4	POP	B	2004[A]	9/9	0.94	0.13	9,10,13,14	4
4	POP	B	2004[B]	9/9	0.94	0.13	9,10,11,12	4
3	IFP	A	2001	12/12	0.95	0.07	8,10,12,16	0
3	IFP	B	2002	12/12	0.96	0.07	9,11,13,17	0
2	MG	A	2007	1/1	0.97	0.07	9,9,9,9	0
5	TZP	A	2005	13/13	0.98	0.06	6,8,14,14	0
5	TZP	B	2006	13/13	0.98	0.06	7,8,14,15	0
2	MG	B	2008	1/1	0.99	0.05	9,9,9,9	0

6.5 Other polymers [i](#)

There are no such residues in this entry.