WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 2, 2023 – 11:25 PM EDT

|		
PDB ID	:	1G67
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)

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.
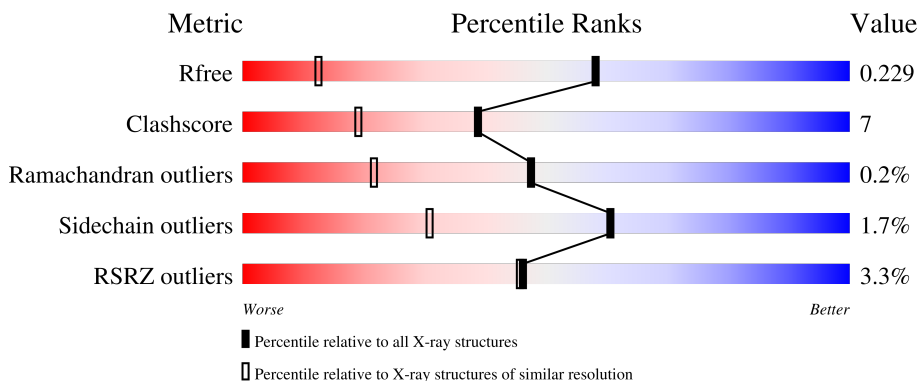
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1714 (1.40-1.40)
Clashscore	141614	1812 (1.40-1.40)
Ramachandran outliers	138981	1763 (1.40-1.40)
Sidechain outliers	138945	1762 (1.40-1.40)
RSRZ outliers	127900	1674 (1.40-1.40)

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 <=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	225	3% / 89% / 10% / •
1	B	225	4% / 84% / 16%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 3723 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.

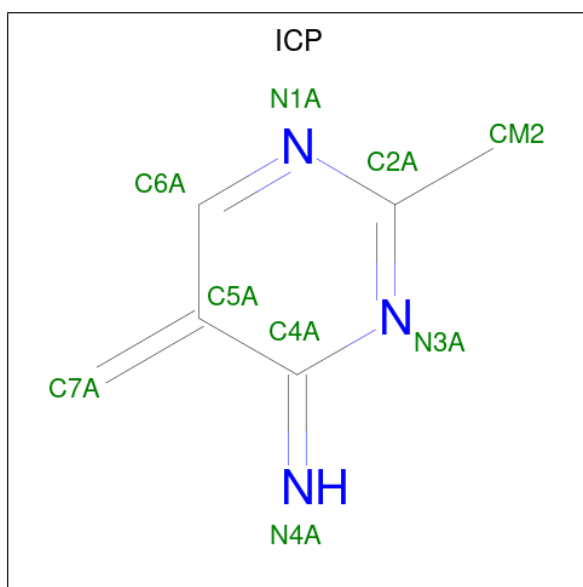| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	225	Total	C	N	O	S	0	3	0
			1688	1047	298	335	8			
1	B	225	Total	C	N	O	S	0	6	0
			1700	1057	303	332	8			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
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	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 (three-letter code: 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-METHYL-5-METHYLENE-5H-PYRIMIDIN-4-YLIDENEAMINE (three-letter code: ICP) (formula: $C_6H_7N_3$).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 9	C 6	N 3	0	0
3	B	1	Total 9	C 6	N 3	0	0

• Molecule 4 is PYROPHOSPHATE 2- (three-letter code: POP) (formula: $H_2O_7P_2$).
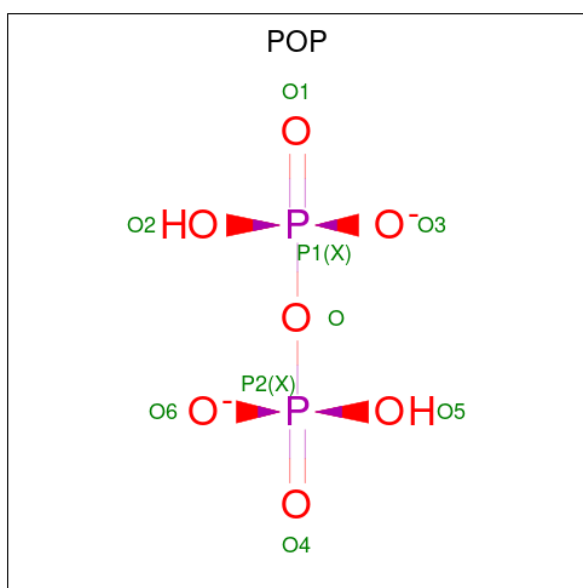
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 13	O 10	P 3	0	1
4	B	1	Total 13	O 10	P 3	0	1

- Molecule 5 is 4-METHYL-5-HYDROXYETHYLTHIAZOLE PHOSPHATE (three-letter code: TZP) (formula: $C_6H_{10}NO_4PS$).
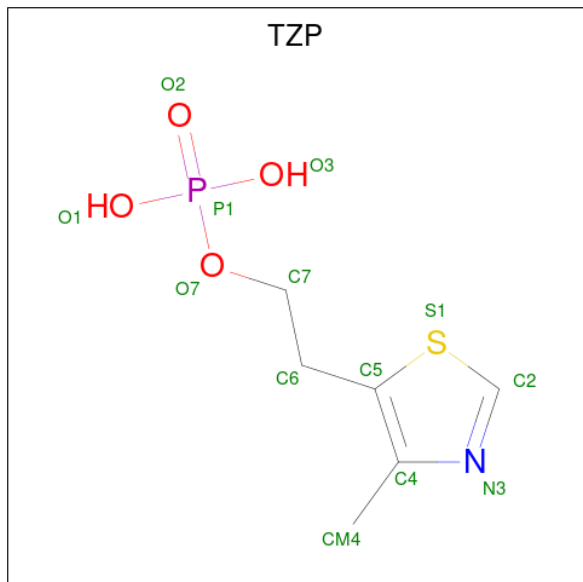
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
5	A	1	Total	C	N	O	P	S	0	0
			13	6	1	4	1	1		
5	B	1	Total	C	N	O	P	S	0	0
			13	6	1	4	1	1		
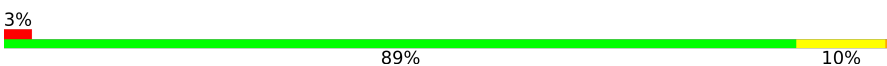
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	132	Total	O	0	0
			132	132		
6	B	131	Total	O	0	0
			131	131		

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

Chain A:

• Molecule 1: THIAMIN PHOSPHATE SYNTHASE

Chain B:

4 Data and refinement statistics (i)

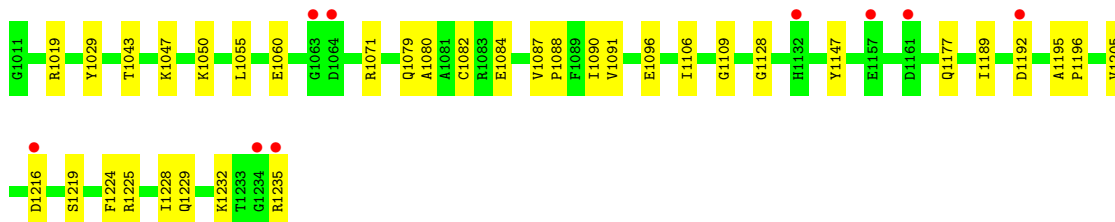| Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	76.98Å 76.98Å 140.15Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.99 – 1.40 42.99 – 1.40	Depositor EDS
% Data completeness (in resolution range)	77.0 (42.99-1.40) 77.1 (42.99-1.40)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.90 (at 1.40Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.216 , 0.240 0.208 , 0.229	Depositor DCC
R_{free} test set	3240 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	13.4	Xtriage
Anisotropy	0.035	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.43 , 53.1	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3723	wwPDB-VP
Average B, all atoms (Å^2)	16.0	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 37.36 % 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 4.3396e-04.*

---

[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 (i)

5.1 Standard geometry (i)

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

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	0.28	0/1721	0.55	0/2317
1	B	0.27	0/1750	0.55	0/2356
All	All	0.27	0/3471	0.55	0/4673

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	1688	0	1711	23	0
1	B	1700	0	1733	30	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	9	0	6	0	0
3	B	9	0	6	0	0
4	A	13	0	0	0	0
4	B	13	0	0	1	0
5	A	13	0	8	0	0
5	B	13	0	8	0	0
6	A	132	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	131	0	0	3	0
All	All	3723	0	3472	52	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 7.

The worst 5 of 52 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:114:ASN:ND2	1:A:116:LYS:HG2	1.64	1.11
1:B:1189:ILE:HG21	1:B:1205[B]:VAL:HG11	1.40	1.02
1:A:114:ASN:HD22	1:A:116:LYS:HG2	1.35	0.85
1:A:12:ILE:H	1:A:12:ILE:HD13	1.45	0.80
1:B:1189:ILE:HG21	1:B:1205[B]:VAL:CG1	2.13	0.79

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	226/225 (100%)	218 (96%)	7 (3%)	1 (0%)	34	12
1	B	229/225 (102%)	221 (96%)	8 (4%)	0	100	100
All	All	455/450 (101%)	439 (96%)	15 (3%)	1 (0%)	47	21

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	64	ASP

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	175/172 (102%)	171 (98%)	4 (2%)	50	18
1	B	178/172 (104%)	176 (99%)	2 (1%)	73	50
All	All	353/344 (103%)	347 (98%)	6 (2%)	60	31

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	235	ARG
1	B	1029	TYR
1	B	1192	ASP
1	A	29	TYR
1	A	12	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	229	GLN
1	B	1079	GLN
1	B	1229	GLN
1	B	1213	GLN
1	A	114	ASN

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 ⓘ

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	ICP	A	2001	-	4,9,9	3.56	2 (50%)	4,12,12	2.63	2 (50%)
5	TZP	A	2005	-	9,13,13	1.34	1 (11%)	8,18,18	1.28	0
4	POP	A	2003[A]	2	6,8,8	1.77	1 (16%)	13,13,13	1.64	3 (23%)
5	TZP	B	2006	-	9,13,13	1.32	1 (11%)	8,18,18	1.35	1 (12%)
4	POP	B	2004[B]	2	6,8,8	1.82	1 (16%)	13,13,13	1.66	3 (23%)
4	POP	B	2004[A]	2	6,8,8	1.78	1 (16%)	13,13,13	1.62	3 (23%)
4	POP	A	2003[B]	2	6,8,8	1.80	1 (16%)	13,13,13	1.66	3 (23%)
3	ICP	B	2002	-	4,9,9	3.56	2 (50%)	4,12,12	2.66	2 (50%)

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	ICP	A	2001	-	-	-	0/1/1/1
5	TZP	A	2005	-	-	0/6/7/7	0/1/1/1
4	POP	A	2003[A]	2	-	1/6/6/6	-
5	TZP	B	2006	-	-	0/6/7/7	0/1/1/1
4	POP	B	2004[B]	2	-	0/6/6/6	-
4	POP	B	2004[A]	2	-	1/6/6/6	-
4	POP	A	2003[B]	2	-	0/6/6/6	-
3	ICP	B	2002	-	-	-	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	2001	ICP	C7A-C5A	6.40	1.50	1.33
3	B	2002	ICP	C7A-C5A	6.39	1.50	1.33
5	A	2005	TZP	C4-C5	-2.98	1.36	1.42
5	B	2006	TZP	C4-C5	-2.97	1.36	1.42
3	A	2001	ICP	C6A-N1A	2.63	1.37	1.32

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	2002	ICP	N1A-C2A-N3A	-4.31	123.04	126.92
3	A	2001	ICP	N1A-C2A-N3A	-4.25	123.09	126.92
4	B	2004[B]	POP	O3-P1-O2	2.97	119.00	107.64
4	A	2003[B]	POP	O3-P1-O2	2.96	118.95	107.64
4	A	2003[A]	POP	O3-P1-O2	2.91	118.78	107.64

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	2004[A]	POP	P1-O-P2-O4
4	A	2003[A]	POP	P1-O-P2-O4

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	2004[B]	POP	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 ⓘ

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, 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(Å²)	Q<0.9
1	A	225/225 (100%)	0.28	6 (2%)	54	54	8, 14, 25, 35	1 (0%)
1	B	225/225 (100%)	0.28	9 (4%)	38	39	8, 14, 25, 54	0
All	All	450/450 (100%)	0.28	15 (3%)	46	46	8, 14, 25, 54	1 (0%)

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1063	GLY	6.4
1	A	63	GLY	4.4
1	A	160	LYS	4.4
1	B	1235	ARG	4.2
1	A	64	ASP	3.4

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, 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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\mathring{A}^2$)	Q<0.9
3	ICP	B	2002	9/9	0.88	0.12	13,14,17,17	0
3	ICP	A	2001	9/9	0.89	0.12	10,14,17,20	0
4	POP	B	2004[A]	9/9	0.94	0.11	5,12,16,18	4
4	POP	B	2004[B]	9/9	0.94	0.11	5,12,16,17	4
4	POP	A	2003[A]	9/9	0.96	0.09	9,12,14,16	4
4	POP	A	2003[B]	9/9	0.96	0.09	9,12,14,16	4
5	TZP	A	2005	13/13	0.97	0.09	8,10,18,18	0
5	TZP	B	2006	13/13	0.97	0.09	10,11,16,19	0
2	MG	A	2007	1/1	1.00	0.05	11,11,11,11	0
2	MG	B	2008	1/1	1.00	0.07	12,12,12,12	0

6.5 Other polymers ⓘ

There are no such residues in this entry.