



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 11:21 am BST

PDB ID : 2BE3  
Title : Structure of a GTP Pyrophosphokinase Family Protein from *Streptococcus pneumoniae*  
Authors : Cuff, M.E.; Hatzos, C.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2005-10-21  
Resolution : 2.40 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](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 : 2.11  
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) : 2.11

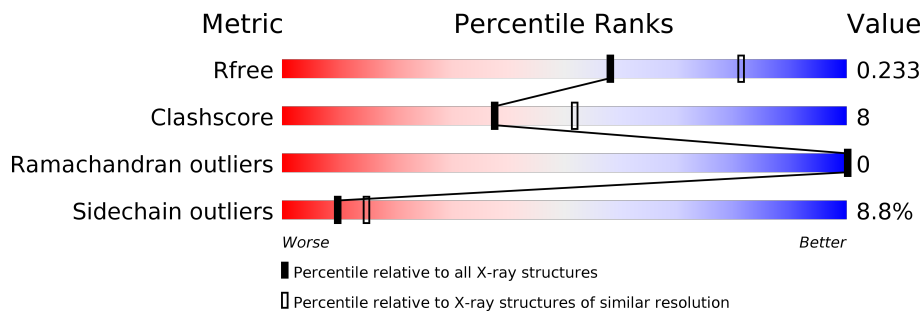
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)

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 on 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	226	 75% 13% • 10%
1	B	226	 63% 16% • 16%

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3557 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 GTP pyrophosphokinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	203	1702	1078	300	318	6	0	2	0
1	B	189	1584	1006	280	293	5	0	1	0

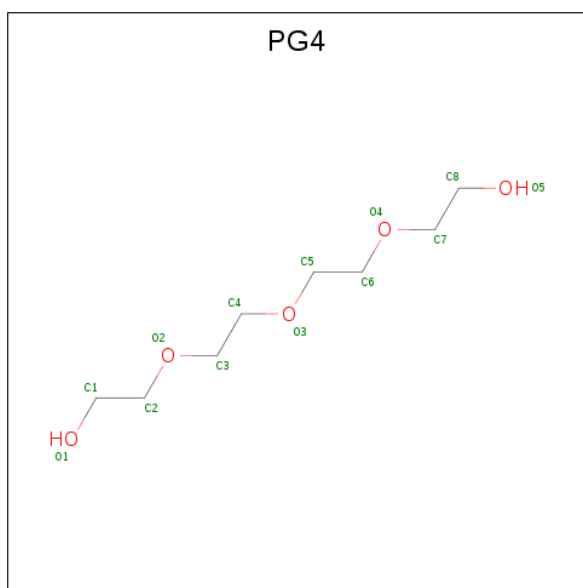
There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	CLONING ARTIFACT	UNP Q97QV1
A	-1	ASN	-	CLONING ARTIFACT	UNP Q97QV1
A	0	ALA	-	CLONING ARTIFACT	UNP Q97QV1
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
A	56	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
A	79	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
A	99	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
A	147	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
A	187	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
B	-2	SER	-	CLONING ARTIFACT	UNP Q97QV1
B	-1	ASN	-	CLONING ARTIFACT	UNP Q97QV1
B	0	ALA	-	CLONING ARTIFACT	UNP Q97QV1
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
B	56	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
B	79	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
B	99	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
B	147	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1
B	187	MSE	MET	MODIFIED RESIDUE	UNP Q97QV1

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Cl	0	0
			1	1		

- Molecule 3 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula:  $C_8H_{18}O_5$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			13	8	5		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is water.

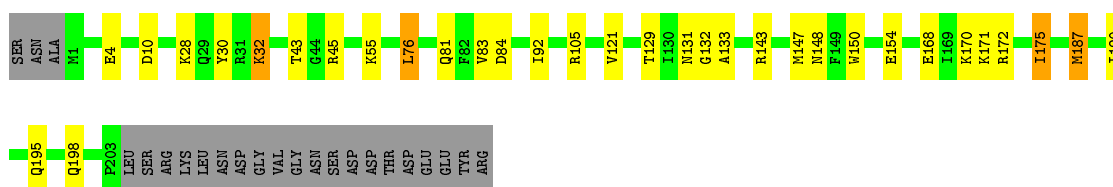
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
5	A	142	Total 142	O 142	0	0
5	B	109	Total 109	O 109	0	0

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

These plots are drawn for all protein, RNA and DNA 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. 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. 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.

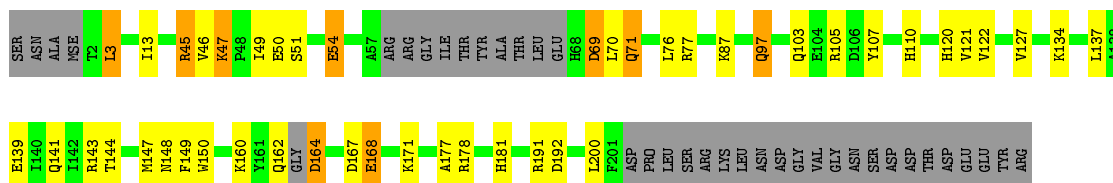
- Molecule 1: GTP pyrophosphokinase

Chain A: 



- Molecule 1: GTP pyrophosphokinase

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	104.84Å 104.84Å 170.56Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.00 – 2.40 38.60 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.5 (38.00-2.40) 98.2 (38.60-2.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.88 (at 2.29Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.182 , 0.234 0.181 , 0.233	Depositor DCC
$R_{free}$ test set	1130 reflections (4.55%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	45.9	Xtrriage
Anisotropy	0.234	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 45.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	3557	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	73.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.97% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, PG4, CL

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	0.73	0/1727	0.76	2/2317 (0.1%)
1	B	0.68	0/1605	0.76	2/2149 (0.1%)
All	All	0.70	0/3332	0.76	4/4466 (0.1%)

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	B	0	1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	76	LEU	CA-CB-CG	-6.69	99.92	115.30
1	B	3	LEU	CA-CB-CG	5.81	128.67	115.30
1	B	178	ARG	NE-CZ-NH2	-5.51	117.55	120.30
1	A	84	ASP	CB-CG-OD1	5.17	122.95	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	69	ASP	Peptide



## 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	1702	0	1709	24	0
1	B	1584	0	1593	28	0
2	A	1	0	0	0	0
3	A	13	0	18	1	0
4	A	6	0	8	0	0
5	A	142	0	0	4	0
5	B	109	0	0	4	0
All	All	3557	0	3328	51	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 51 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:181:HIS:CE1	5:B:329:HOH:O	2.24	0.91
1:B:181:HIS:ND1	5:B:329:HOH:O	2.04	0.90
1:B:120:HIS:CD2	1:B:141:GLN:HG2	2.16	0.80
1:A:187:MSE:HA	1:A:187:MSE:HE3	1.67	0.76
1:B:71:GLN:HG2	1:B:122:VAL:HG21	1.68	0.74

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	203/226 (90%)	201 (99%)	2 (1%)	0	100	100
1	B	184/226 (81%)	176 (96%)	8 (4%)	0	100	100
All	All	387/452 (86%)	377 (97%)	10 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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	183/195 (94%)	173 (94%)	10 (6%)	21	35
1	B	171/195 (88%)	149 (87%)	22 (13%)	4	5
All	All	354/390 (91%)	322 (91%)	32 (9%)	10	14

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	50	GLU
1	B	71	GLN
1	B	164	ASP
1	B	70	LEU
1	B	76	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	198	GLN
1	B	71	GLN
1	B	148	ASN
1	A	182	GLN
1	B	131	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 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	PG4	A	303	-	12,12,12	0.50	0	11,11,11	0.34	0
4	GOL	A	302	-	5,5,5	0.41	0	5,5,5	0.46	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	PG4	A	303	-	-	5/10/10/10	-
4	GOL	A	302	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	303	PG4	O3-C5-C6-O4
3	A	303	PG4	O1-C1-C2-O2
4	A	302	GOL	C1-C2-C3-O3
4	A	302	GOL	O2-C2-C3-O3
3	A	303	PG4	C8-C7-O4-C6

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	303	PG4	1	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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.