



# wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 03:50 PM GMT

PDB ID : 4KGM  
Title : Bacterial tRNA(HIS) Guanylyltransferase (Thg1)-Like Protein in complex with ATP  
Authors : Hyde, S.J.; Rao, B.S.; Eckenroth, B.E.; Jackman, J.E.; Doublet, S.  
Deposited on : 2013-04-29  
Resolution : 2.36 Å (reported)

This is a wwPDB 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

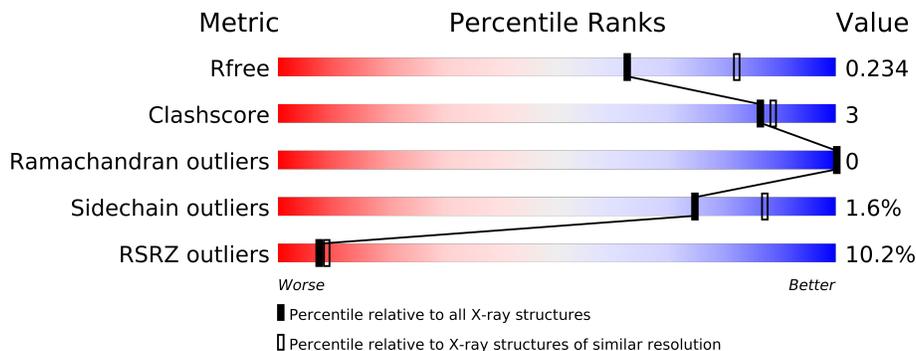
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.36 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))
$R_{free}$	66092	3327 (2.40-2.32)
Clashscore	79885	1064 (2.38-2.34)
Ramachandran outliers	78287	1048 (2.38-2.34)
Sidechain outliers	78261	1049 (2.38-2.34)
RSRZ outliers	66119	3330 (2.40-2.32)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	245	
1	B	245	
1	C	245	
1	D	245	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	ATP	B	301	-	X
2	ATP	D	301	-	X
4	MG	B	303	-	X
5	GOL	A	305	-	X

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Type</b>	<b>Chain</b>	<b>Res</b>	<b>Geometry</b>	<b>Electron density</b>
5	GOL	B	304	-	X
5	GOL	C	305	-	X
5	GOL	D	304	-	X

## 2 Entry composition i

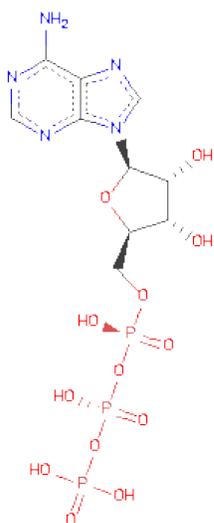
There are 6 unique types of molecules in this entry. The entry contains 7903 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Thg1-like uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	230	Total	C	N	O	S	0	1	0
			1812	1151	313	337	11			
1	B	228	Total	C	N	O	S	0	0	0
			1835	1170	312	342	11			
1	C	229	Total	C	N	O	S	0	0	0
			1845	1175	315	344	11			
1	D	224	Total	C	N	O	S	0	0	0
			1781	1136	306	329	10			

- Molecule 2 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula:  $C_{10}H_{16}N_5O_{13}P_3$ ).



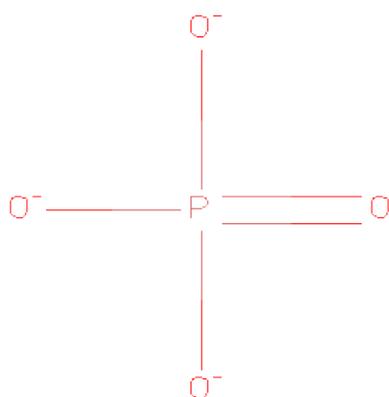
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
			Total	C	N	O			P
2	A	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
2	B	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	C	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
2	D	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	P	0	0
			5	4	1		
3	A	1	Total	O	P	0	0
			5	4	1		
3	C	1	Total	O	P	0	0
			5	4	1		
3	C	1	Total	O	P	0	0
			5	4	1		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

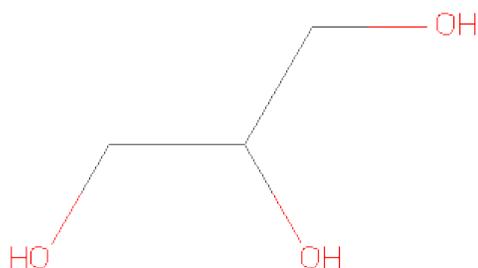
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	Mg	0	0
			2	2		
4	A	2	Total	Mg	0	0
			2	2		
4	D	2	Total	Mg	0	0
			2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	2	Total	Mg	0	0
			2	2		

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			6	3	3		
5	B	1	Total	C	O	0	0
			6	3	3		
5	C	1	Total	C	O	0	0
			6	3	3		
5	D	1	Total	C	O	0	0
			6	3	3		

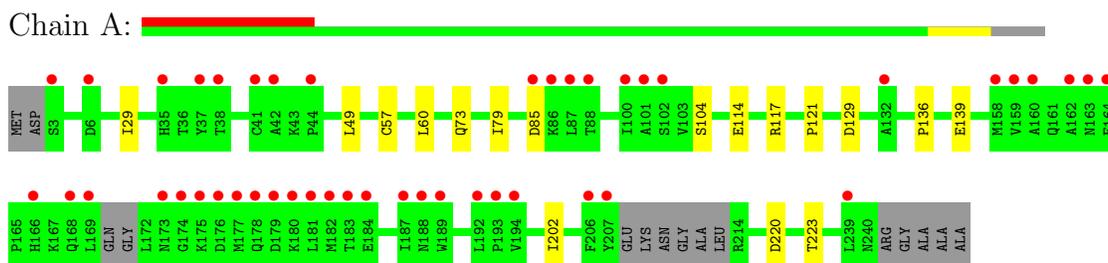
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	76	Total	O	0	0
			76	76		
6	B	92	Total	O	0	0
			92	92		
6	C	143	Total	O	0	0
			143	143		
6	D	143	Total	O	0	0
			143	143		

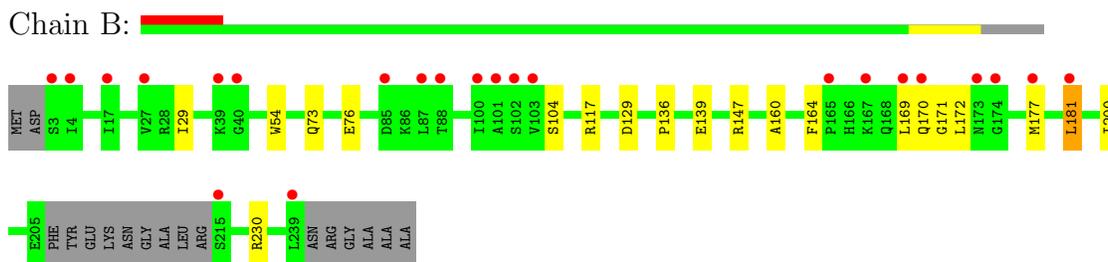
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

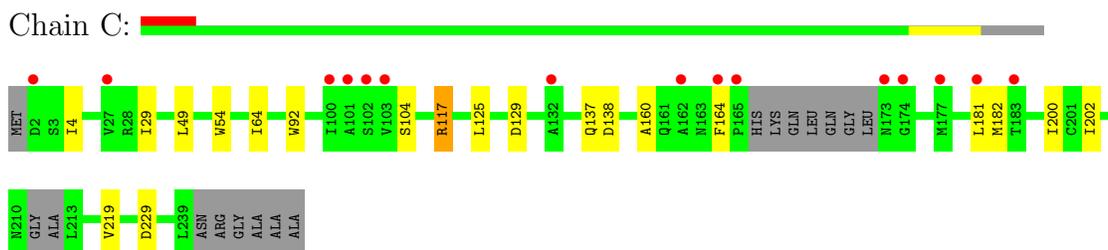
- Molecule 1: Thg1-like uncharacterized protein



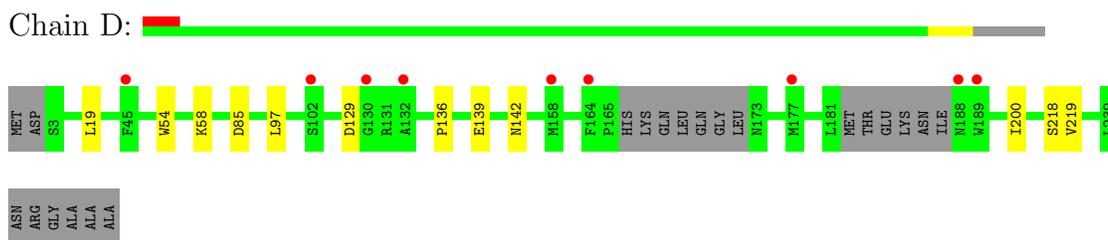
- Molecule 1: Thg1-like uncharacterized protein



- Molecule 1: Thg1-like uncharacterized protein



- Molecule 1: Thg1-like uncharacterized protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.16Å 218.34Å 126.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.70 – 2.36 36.70 – 2.36	Depositor EDS
% Data completeness (in resolution range)	99.5 (36.70-2.36) 99.5 (36.70-2.36)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.23 (at 2.37Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.1.1168)	Depositor
R, $R_{free}$	0.199 , 0.232 0.200 , 0.234	Depositor DCC
$R_{free}$ test set	5681 reflections (10.12%)	DCC
Wilson B-factor (Å <sup>2</sup> )	42.3	Xtrriage
Anisotropy	0.369	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 23.4	EDS
Estimated twinning fraction	No twinning to report.	Xtrriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Outliers	0 of 56158 reflections	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7903	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 27.53 % 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 2.1639e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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

## 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, MG, PO4, ATP

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	0.21	0/1857	0.37	0/2522
1	B	0.22	0/1877	0.39	0/2541
1	C	0.23	0/1887	0.39	0/2553
1	D	0.23	0/1824	0.38	0/2475
All	All	0.22	0/7445	0.38	0/10091

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1812	0	1672	8	0
1	B	1835	0	1769	13	0
1	C	1845	0	1744	12	0
1	D	1781	0	1677	6	0
2	A	31	0	12	1	0
2	B	31	0	12	0	0
2	C	31	0	12	1	0
2	D	31	0	12	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	10	0	0	0	0
3	C	10	0	0	0	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
4	C	2	0	0	0	0
4	D	2	0	0	0	0
5	A	6	0	8	0	0
5	B	6	0	8	0	0
5	C	6	0	8	1	0
5	D	6	0	8	0	0
6	A	76	0	0	0	0
6	B	92	0	0	1	0
6	C	143	0	0	2	2
6	D	143	0	0	0	1
All	All	7903	0	6942	37	3

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 3.

The worst 5 of 37 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:164:PHE:HB2	1:B:169:LEU:HD21	1.83	0.60
1:A:29:ILE:HD11	1:A:104:SER:HB3	1.83	0.59
1:C:29:ILE:HD11	1:C:104:SER:HB3	1.88	0.54
1:B:169:LEU:HD12	1:B:177:MET:HG2	1.90	0.54
1:A:60:LEU:HD13	1:A:79:ILE:HD11	1.89	0.54

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
6:D:452:HOH:O	6:D:453:HOH:O[3_756]	2.04	0.16
6:C:442:HOH:O	6:C:444:HOH:O[3_756]	2.14	0.06
6:C:442:HOH:O	6:C:445:HOH:O[3_756]	2.14	0.06

## 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	225/245 (92%)	218 (97%)	7 (3%)	0	100	100
1	B	224/245 (91%)	221 (99%)	3 (1%)	0	100	100
1	C	223/245 (91%)	217 (97%)	6 (3%)	0	100	100
1	D	218/245 (89%)	212 (97%)	6 (3%)	0	100	100
All	All	890/980 (91%)	868 (98%)	22 (2%)	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	181/214 (85%)	178 (98%)	3 (2%)	73	86
1	B	193/214 (90%)	190 (98%)	3 (2%)	75	87
1	C	190/214 (89%)	186 (98%)	4 (2%)	66	82
1	D	182/214 (85%)	180 (99%)	2 (1%)	84	93
All	All	746/856 (87%)	734 (98%)	12 (2%)	75	87

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	181	LEU
1	C	117	ARG
1	C	202	ILE
1	B	129	ASP
1	C	182	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	190	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	154	ASN
1	C	190	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

### 5.6 Ligand geometry ⓘ

Of 20 ligands modelled in this entry, 8 are monoatomic - leaving 12 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
2	ATP	A	301	4	33,33,33	1.01	2 (6%)	52,52,52	1.77	9 (17%)
3	PO4	A	302	-	4,4,4	0.29	0	6,6,6	0.31	0
5	GOL	A	305	-	5,5,5	0.31	0	5,5,5	0.38	0
3	PO4	A	306	-	4,4,4	0.28	0	6,6,6	0.31	0
2	ATP	B	301	4	33,33,33	1.02	2 (6%)	52,52,52	1.80	9 (17%)
5	GOL	B	304	-	5,5,5	0.33	0	5,5,5	0.30	0
2	ATP	C	301	4	33,33,33	1.02	2 (6%)	52,52,52	1.86	9 (17%)
3	PO4	C	302	-	4,4,4	0.29	0	6,6,6	0.31	0
5	GOL	C	305	-	5,5,5	0.33	0	5,5,5	0.28	0
3	PO4	C	306	-	4,4,4	0.27	0	6,6,6	0.31	0
2	ATP	D	301	4	33,33,33	1.02	2 (6%)	52,52,52	1.82	9 (17%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
5	GOL	D	304	-	5,5,5	0.36	0	5,5,5	0.32	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
2	ATP	A	301	4	-	0/22/38/38	0/1/3/3
3	PO4	A	302	-	-	0/0/0/0	0/0/0/0
5	GOL	A	305	-	-	0/4/4/4	0/0/0/0
3	PO4	A	306	-	-	0/0/0/0	0/0/0/0
2	ATP	B	301	4	-	0/22/38/38	0/1/3/3
5	GOL	B	304	-	-	0/4/4/4	0/0/0/0
2	ATP	C	301	4	-	0/22/38/38	0/1/3/3
3	PO4	C	302	-	-	0/0/0/0	0/0/0/0
5	GOL	C	305	-	-	0/4/4/4	0/0/0/0
3	PO4	C	306	-	-	0/0/0/0	0/0/0/0
2	ATP	D	301	4	-	0/22/38/38	0/1/3/3
5	GOL	D	304	-	-	0/4/4/4	0/0/0/0

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	301	ATP	C5-C4	3.17	1.47	1.40
2	B	301	ATP	C5-C4	3.13	1.47	1.40
2	A	301	ATP	C5-C4	3.11	1.47	1.40
2	D	301	ATP	C5-C4	3.10	1.47	1.40
2	B	301	ATP	C4-N9	-2.55	1.34	1.37

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	301	ATP	N3-C2-N1	-6.12	123.59	128.71
2	D	301	ATP	N3-C2-N1	-6.06	123.64	128.71
2	B	301	ATP	N3-C2-N1	-6.01	123.68	128.71
2	A	301	ATP	N3-C2-N1	-5.88	123.80	128.71
2	D	301	ATP	N3-C4-N9	5.59	135.53	125.43

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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	230/245 (93%)	0.84	46 (20%) 2 1	15, 40, 74, 82	0
1	B	228/245 (93%)	0.46	23 (10%) 7 9	15, 33, 55, 81	0
1	C	229/245 (93%)	0.38	15 (6%) 18 20	4, 23, 62, 73	0
1	D	224/245 (91%)	0.28	9 (4%) 36 40	5, 22, 59, 73	0
All	All	911/980 (92%)	0.49	93 (10%) 7 9	4, 30, 65, 82	0

The worst 5 of 93 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	170	GLN	5.1
1	A	239	LEU	5.0
1	A	187	ILE	5.0
1	C	181	LEU	4.7
1	A	162	ALA	4.6

### 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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	GOL	A	305	6/6	0.40	21.08	32,34,37,39	6
5	GOL	B	304	6/6	0.30	13.15	38,39,43,43	0
5	GOL	C	305	6/6	0.30	6.99	14,19,20,24	6
5	GOL	D	304	6/6	0.23	5.68	14,19,21,22	6
2	ATP	D	301	31/31	0.29	3.04	24,35,41,41	31
2	ATP	B	301	31/31	0.24	2.68	28,37,47,52	31
4	MG	B	303	1/1	0.21	2.08	42,42,42,42	0
2	ATP	C	301	31/31	0.19	1.79	23,37,48,56	13
3	PO4	C	306	5/5	0.18	0.36	36,38,51,52	0
2	ATP	A	301	31/31	0.23	0.26	35,42,49,60	31
3	PO4	C	302	5/5	0.18	0.22	37,43,51,55	0
3	PO4	A	306	5/5	0.16	-0.33	43,44,55,66	0
3	PO4	A	302	5/5	0.15	-1.08	43,47,57,66	0
4	MG	C	303	1/1	0.11	-1.11	25,25,25,25	0
4	MG	C	304	1/1	0.14	-1.12	34,34,34,34	0
4	MG	A	303	1/1	0.06	-1.97	40,40,40,40	0
4	MG	D	302	1/1	0.09	-2.14	26,26,26,26	0
4	MG	B	302	1/1	0.07	-2.58	37,37,37,37	0
4	MG	A	304	1/1	0.05	-3.29	38,38,38,38	0
4	MG	D	303	1/1	0.09	-3.42	40,40,40,40	0

## 6.5 Other polymers

There are no such residues in this entry.