



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 1, 2017 – 12:34 AM EDT

PDB ID : 1KJ8  
Title : Crystal Structure of PurT-Encoded Glycinamide Ribonucleotide Transformylase in Complex with Mg-ATP and GAR  
Authors : Thoden, J.B.; Firestine, S.M.; Benkovic, S.J.; Holden, H.M.  
Deposited on : unknown  
Resolution : 1.60 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

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030345

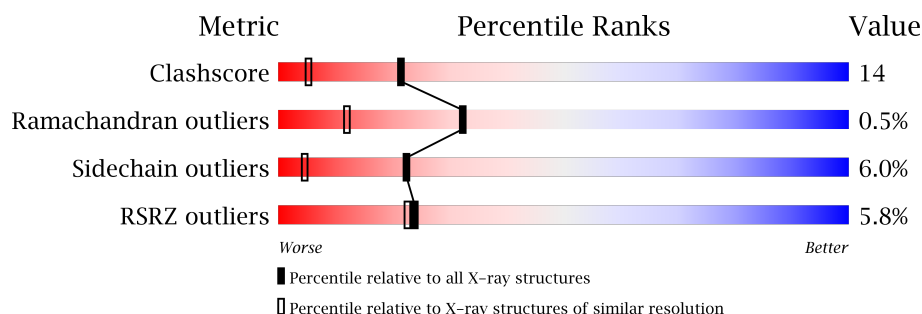
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.60 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))
Clashscore	112137	2967 (1.60-1.60)
Ramachandran outliers	110173	2887 (1.60-1.60)
Sidechain outliers	110143	2886 (1.60-1.60)
RSRZ outliers	101464	2714 (1.60-1.60)

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. 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	391	
1	B	391	

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	NA	B	395	-	-	-	X
8	EDO	B	399	-	-	X	X

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 6896 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 phosphoribosylglycinamide formyltransferase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	386	Total	C	N	O	S	0	6	0
			2971	1874	528	556	13			
1	B	384	Total	C	N	O	S	0	3	0
			2946	1857	523	552	14			

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Na	0	0
			1	1		
3	A	1	Total	Na	0	0
			1	1		

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

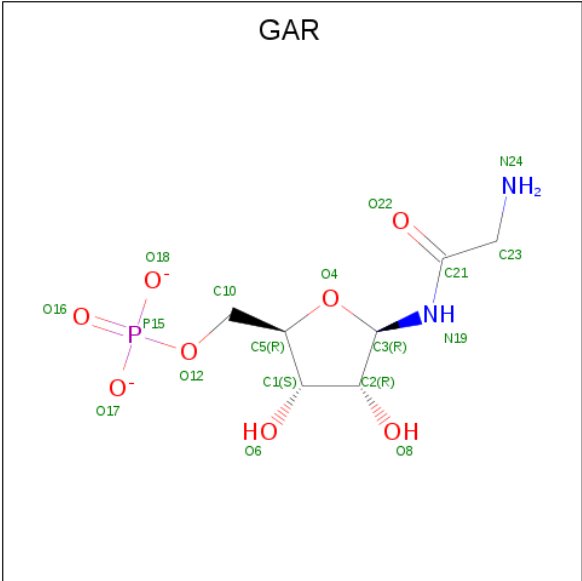
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Cl	0	0
			1	1		
4	A	3	Total	Cl	0	0
			3	3		

- Molecule 5 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: C<sub>10</sub>H<sub>16</sub>N<sub>5</sub>O<sub>13</sub>P<sub>3</sub>).



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

- Molecule 6 is GLYCINAMIDE RIBONUCLEOTIDE (three-letter code: GAR) (formula: C<sub>7</sub>H<sub>13</sub>N<sub>2</sub>O<sub>8</sub>P).



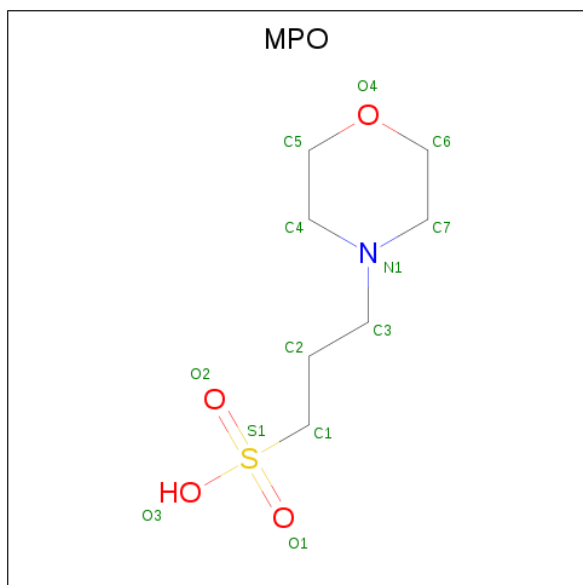
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	A	1	Total	C	N	O	P	0	0
			18	7	2	8	1		

Continued on next page...

*Continued from previous page...*

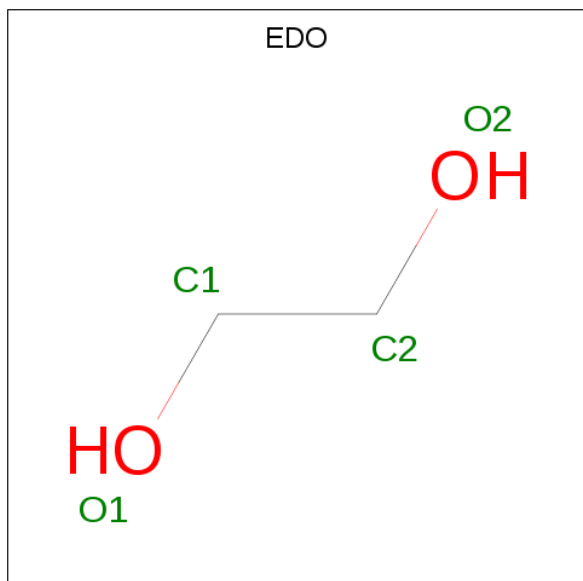
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	B	1	Total	C	N	O	P	0	0
			18	7	2	8	1		

- Molecule 7 is 3[N-MORPHOLINO]PROPANE SULFONIC ACID (three-letter code: MPO) (formula:  $C_7H_{15}NO_4S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	A	1	Total	C	N	O	S	0	0
			13	7	1	4	1		

- Molecule 8 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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

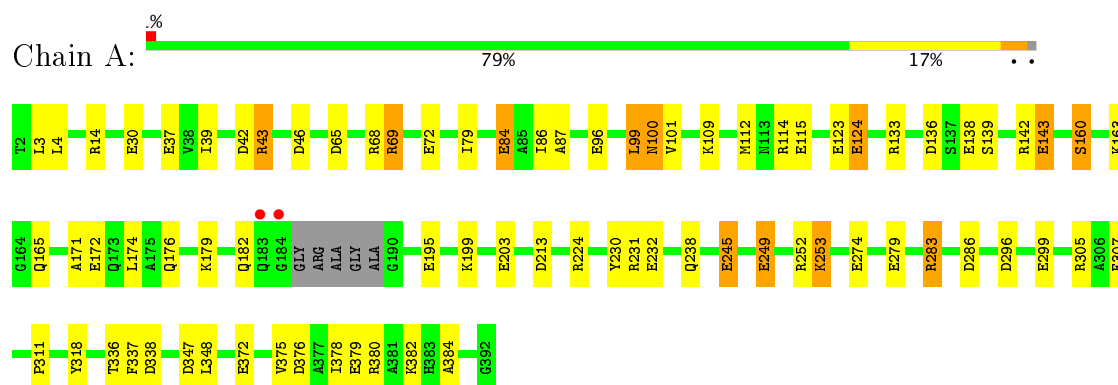
- Molecule 9 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	A	502	Total O 502 502	0	0
9	B	344	Total O 344 344	0	0

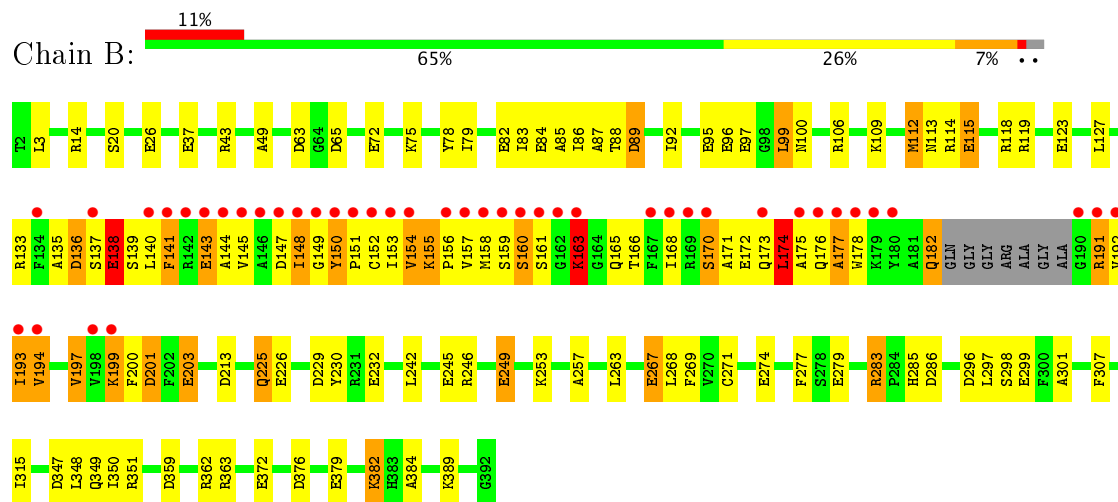
### 3 Residue-property plots [i](#)

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 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: phosphoribosylglycinamide formyltransferase 2



- Molecule 1: phosphoribosylglycinamide formyltransferase 2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	62.10 Å   179.30 Å   75.80 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	30.00 – 1.60 69.82 – 1.54	Depositor EDS
% Data completeness (in resolution range)	97.0 (30.00-1.60) 92.4 (69.82-1.54)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.34 (at 1.54 Å)	Xtriage
Refinement program	TNT	Depositor
R, $R_{free}$	0.181   ,   0.228 0.183   ,   (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	14.6	Xtriage
Anisotropy	0.221	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33   ,   105.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	6896	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.28% 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: MG, CL, NA, GAR, EDO, ATP, MPO

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	1.01	21/3044 (0.7%)	1.25	30/4123 (0.7%)
1	B	1.02	21/3007 (0.7%)	1.30	31/4075 (0.8%)
All	All	1.01	42/6051 (0.7%)	1.27	61/8198 (0.7%)

The worst 5 of 42 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	245	GLU	CD-OE2	8.12	1.34	1.25
1	B	249	GLU	CD-OE2	8.11	1.34	1.25
1	A	172	GLU	CD-OE2	7.81	1.34	1.25
1	A	72	GLU	CD-OE2	7.68	1.34	1.25
1	B	226	GLU	CD-OE2	7.24	1.33	1.25

The worst 5 of 61 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	363	ARG	NE-CZ-NH1	9.11	124.85	120.30
1	A	252[A]	ARG	NE-CZ-NH1	8.78	124.69	120.30
1	A	252[B]	ARG	NE-CZ-NH1	8.78	124.69	120.30
1	B	286	ASP	CB-CG-OD1	8.11	125.60	118.30
1	A	133	ARG	NE-CZ-NH2	-8.02	116.29	120.30

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	2971	0	3001	37	0
1	B	2946	0	2970	127	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	3	0	0	1	0
4	B	1	0	0	0	0
5	A	31	0	12	0	0
5	B	31	0	12	1	0
6	A	18	0	13	0	0
6	B	18	0	13	0	0
7	A	13	0	15	0	0
8	A	4	0	6	2	0
8	B	8	0	12	10	0
9	A	502	0	0	8	1
9	B	344	0	0	9	1
All	All	6896	0	6054	164	2

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 14.

The worst 5 of 164 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:156:PRO:HG2	1:B:159:SER:HB2	1.33	1.09
1:B:174:LEU:HD23	1:B:174:LEU:H	1.24	1.00
1:B:141:PHE:CE1	1:B:145:VAL:HG21	2.01	0.95
1:B:297:LEU:HA	8:B:399:EDO:H11	1.52	0.92
1:B:136:ASP:HB3	1:B:191:ARG:NH1	1.88	0.87

All (2) 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	Interatomic distance (Å)	Clash overlap (Å)
9:B:477:HOH:O	9:B:477:HOH:O[2_655]	2.10	0.10
9:A:521:HOH:O	9:A:755:HOH:O[4_556]	2.17	0.03

## 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	388/391 (99%)	376 (97%)	12 (3%)	0	100	100
1	B	383/391 (98%)	365 (95%)	14 (4%)	4 (1%)	18	4
All	All	771/782 (99%)	741 (96%)	26 (3%)	4 (0%)	32	12

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	163	LYS
1	B	174	LEU
1	B	177	ALA
1	B	197	VAL

### 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	316/311 (102%)	306 (97%)	10 (3%)	44	17
1	B	312/311 (100%)	285 (91%)	27 (9%)	12	1
All	All	628/622 (101%)	591 (94%)	37 (6%)	22	5

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	143	GLU
1	B	160	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	263	LEU
1	B	148	ILE
1	B	154	VAL

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	248	GLN
1	B	176	GLN
1	B	225	GLN
1	A	238	GLN
1	B	182	GLN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 18 ligands modelled in this entry, 10 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	$\# Z  > 2$
5	ATP	A	1	2	27,33,33	0.89	0	25,52,52	2.14	4 (16%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
6	GAR	A	399	-	17,18,18	0.74	0	21,26,26	1.17	3 (14%)
7	MPO	A	400	-	13,13,13	1.99	3 (23%)	16,17,17	1.90	3 (18%)
8	EDO	A	401	-	3,3,3	0.59	0	2,2,2	0.05	0
5	ATP	B	397	2	27,33,33	0.91	2 (7%)	25,52,52	1.73	6 (24%)
6	GAR	B	398	-	17,18,18	0.72	1 (5%)	21,26,26	0.98	1 (4%)
8	EDO	B	399	-	3,3,3	0.55	0	2,2,2	0.18	0
8	EDO	B	400	-	3,3,3	0.32	0	2,2,2	0.41	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
5	ATP	A	1	2	-	0/18/38/38	0/3/3/3
6	GAR	A	399	-	-	0/12/28/28	0/1/1/1
7	MPO	A	400	-	-	0/7/15/15	0/1/1/1
8	EDO	A	401	-	-	0/1/1/1	0/0/0/0
5	ATP	B	397	2	-	0/18/38/38	0/3/3/3
6	GAR	B	398	-	-	0/12/28/28	0/1/1/1
8	EDO	B	399	-	-	0/1/1/1	0/0/0/0
8	EDO	B	400	-	-	0/1/1/1	0/0/0/0

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	397	ATP	O4'-C1'	-2.41	1.37	1.41
6	B	398	GAR	O4-C3	-2.01	1.39	1.43
7	A	400	MPO	O1-S1	2.05	1.51	1.45
5	B	397	ATP	C2-N1	2.55	1.38	1.33
7	A	400	MPO	C3-N1	4.48	1.57	1.47

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1	ATP	N3-C2-N1	-7.49	122.33	128.86
5	B	397	ATP	N3-C2-N1	-4.42	125.01	128.86
5	A	1	ATP	C5-C6-N1	-2.97	110.72	119.70
5	B	397	ATP	C2'-C3'-C4'	-2.42	97.91	102.62
5	B	397	ATP	C5-C6-N1	-2.36	112.57	119.70

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	A	401	EDO	2	0
5	B	397	ATP	1	0
8	B	399	EDO	10	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 [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	386/391 (98%)	-0.72	2 (0%) 90 91	9, 16, 40, 91	0
1	B	384/391 (98%)	0.00	43 (11%) 6 5	10, 19, 85, 99	0
All	All	770/782 (98%)	-0.36	45 (5%) 24 22	9, 17, 71, 99	0

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	175	ALA	11.3
1	B	150	TYR	9.6
1	B	159	SER	9.1
1	B	141	PHE	8.5
1	B	160	SER	8.3

### 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	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
8	EDO	B	399	4/4	0.72	0.25	12.68	45,56,61,64	0
3	NA	B	395	1/1	0.96	0.09	2.72	23,23,23,23	0
3	NA	A	395	1/1	0.96	0.08	1.51	29,29,29,29	0
7	MPO	A	400	13/13	0.98	0.06	0.93	17,24,31,32	0
5	ATP	B	397	31/31	0.50	0.16	-0.26	0,39,92,97	0
6	GAR	A	399	18/18	0.99	0.06	-0.29	11,15,17,20	0
6	GAR	B	398	18/18	0.99	0.05	-0.38	10,14,21,29	0
4	CL	A	397	1/1	0.98	0.07	-0.60	25,25,25,25	0
5	ATP	A	1	31/31	0.99	0.05	-0.83	8,14,19,23	0
2	MG	A	393	1/1	0.99	0.04	-2.71	16,16,16,16	0
4	CL	A	398	1/1	0.97	0.07	-	34,34,34,34	0
2	MG	B	394	1/1	0.91	0.09	-	39,39,39,39	0
4	CL	B	396	1/1	0.97	0.07	-	35,35,35,35	0
2	MG	A	394	1/1	1.00	0.04	-	12,12,12,12	0
2	MG	B	393	1/1	0.93	0.15	-	51,51,51,51	0
8	EDO	B	400	4/4	0.84	0.17	-	33,36,55,58	0
8	EDO	A	401	4/4	0.81	0.18	-	47,60,60,62	0
4	CL	A	396	1/1	0.99	0.04	-	21,21,21,21	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.