



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 22, 2024 – 02:33 PM EDT

PDB ID : 6K79  
Title : Glycerol kinase form *Thermococcus kodakarensis*, complex structure with substrate.  
Authors : Koga, Y.; Angkawidjaja, C.; Matsumura, H.; Hokao, R.  
Deposited on : 2019-06-06  
Resolution : 2.19 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 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 : 4.02b-467  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 1.20.1  
EDS : 2.37.1  
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) : 2.37.1

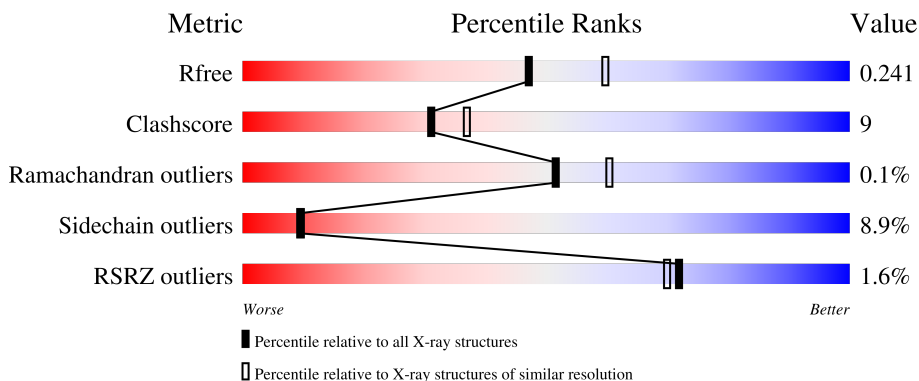
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.19 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	B	497	 78% 17% ..
2	A	497	 79% 17% ..
2	C	497	 76% 20% ..
2	D	497	 77% 17% ..

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:

<b>Mol</b>	<b>Type</b>	<b>Chain</b>	<b>Res</b>	<b>Chirality</b>	<b>Geometry</b>	<b>Clashes</b>	<b>Electron density</b>
4	GOL	A	502	-	-	X	-
4	GOL	D	502	-	-	X	-

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 16106 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 Glycerol kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	490	3895	2490	657	737	11	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	145	LEU	VAL	conflict	UNP O93623
B	247	GLU	GLN	conflict	UNP O93623
B	271	GLU	LYS	conflict	UNP O93623

- Molecule 2 is a protein called Glycerol kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	490	3894	2489	658	736	11	0	0	0
2	D	491	3902	2494	659	737	12	0	0	0
2	A	490	3894	2489	658	736	11	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	271	GLU	LYS	conflict	UNP O93623
D	271	GLU	LYS	conflict	UNP O93623
A	271	GLU	LYS	conflict	UNP O93623

- Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>4</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			10	6	4		
3	C	1	Total	C	O	0	0
			10	6	4		
3	D	1	Total	C	O	0	0
			10	6	4		
3	A	1	Total	C	O	0	0
			10	6	4		

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



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

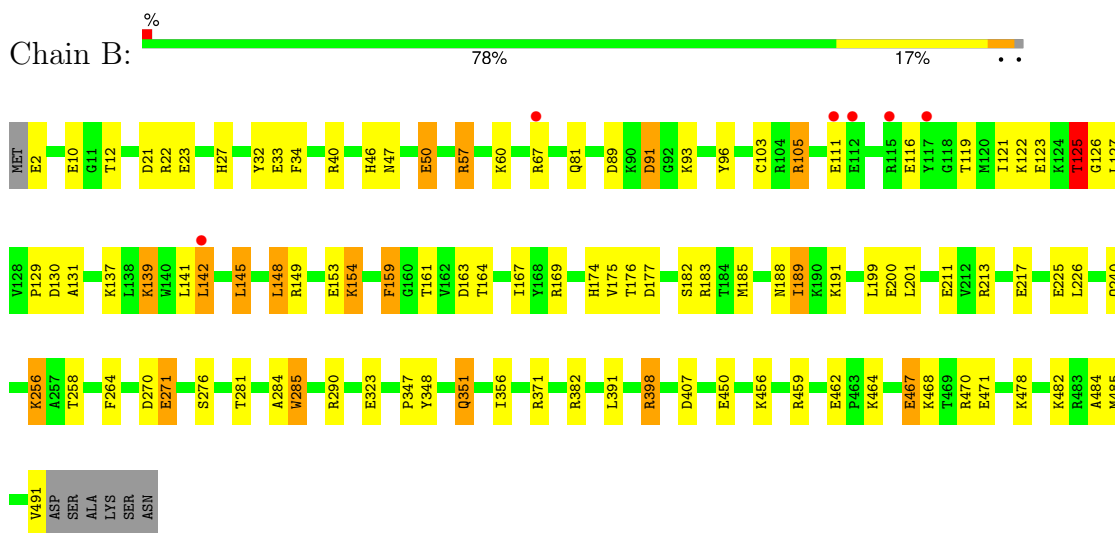
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	152	Total	O	0	0
			152	152		
5	C	86	Total	O	0	0
			86	86		
5	D	103	Total	O	0	0
			103	103		
5	A	116	Total	O	0	0
			116	116		

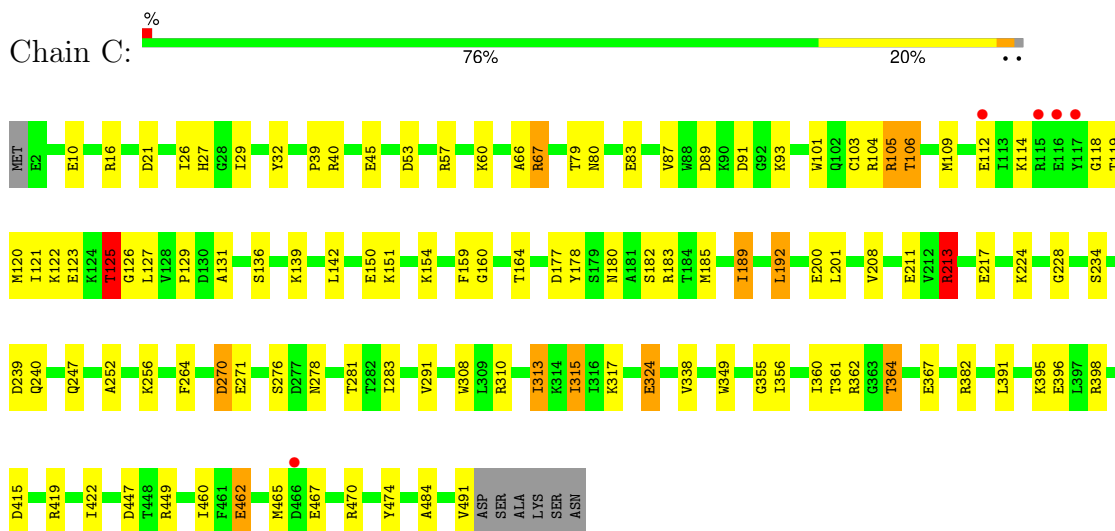
### 3 Residue-property plots [i](#)

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.

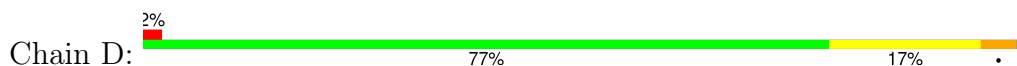
- Molecule 1: Glycerol kinase

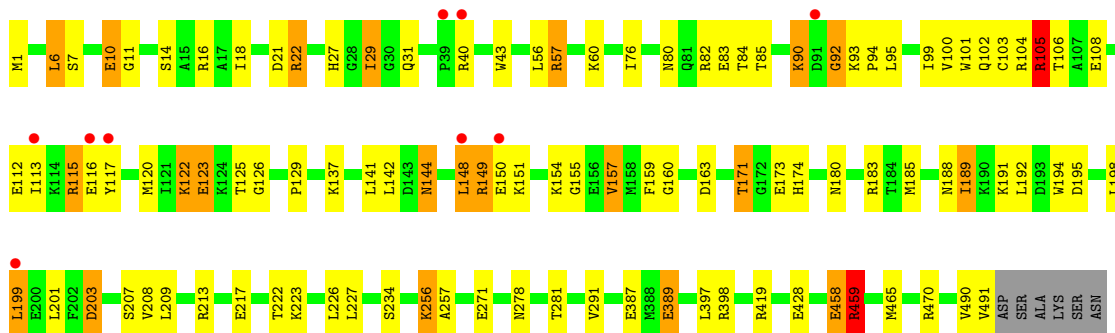


- Molecule 2: Glycerol kinase

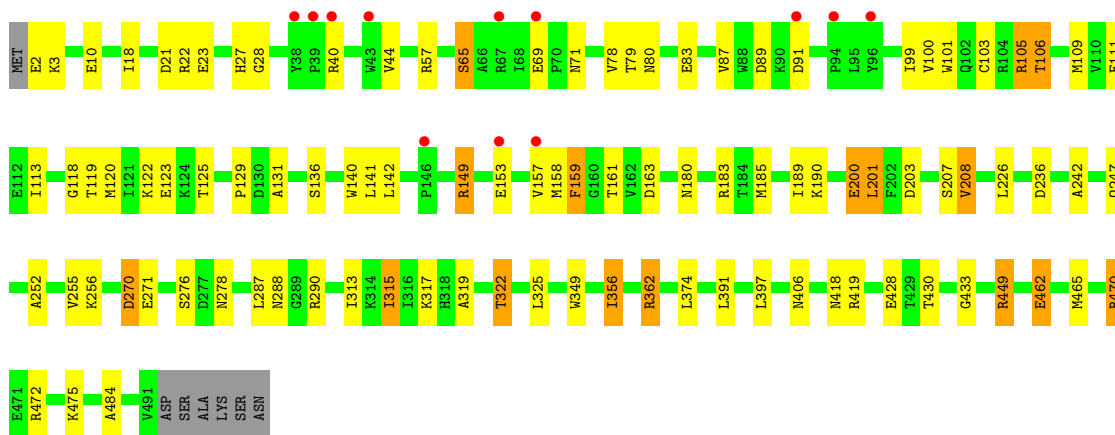
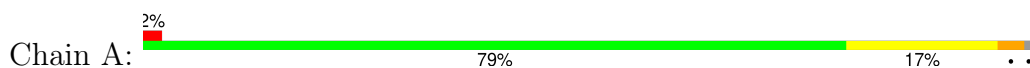


- Molecule 2: Glycerol kinase





• Molecule 2: Glycerol kinase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.04Å 72.13Å 115.99Å 103.43° 94.96° 99.48°	Depositor
Resolution (Å)	46.20 – 2.19 46.20 – 2.19	Depositor EDS
% Data completeness (in resolution range)	98.6 (46.20-2.19) 98.6 (46.20-2.19)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.64 (at 2.18Å)	Xtrriage
Refinement program	REFMAC 5.8.0222	Depositor
R, $R_{free}$	0.185 , 0.238 0.190 , 0.241	Depositor DCC
$R_{free}$ test set	5226 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.2	Xtrriage
Anisotropy	0.044	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 42.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	16106	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.56% 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, PGE

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	B	0.62	1/3977 (0.0%)	0.77	3/5388 (0.1%)
2	A	0.56	0/3976	0.72	0/5387
2	C	0.53	0/3976	0.71	5/5387 (0.1%)
2	D	0.53	0/3984	0.72	0/5397
All	All	0.56	1/15913 (0.0%)	0.73	8/21559 (0.0%)

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	8
2	A	0	7
2	C	0	5
2	D	0	8
All	All	0	28

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	471	GLU	CD-OE2	5.35	1.31	1.25

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	105	ARG	CB-CA-C	-8.54	93.32	110.40
1	B	371	ARG	NE-CZ-NH2	-6.96	116.82	120.30
1	B	217	GLU	CB-CA-C	-5.74	98.92	110.40
2	C	213	ARG	NE-CZ-NH2	-5.67	117.47	120.30

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	16	ARG	CB-CA-C	-5.59	99.22	110.40

There are no chirality outliers.

5 of 28 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	105	ARG	Sidechain
1	B	154	LYS	Peptide
1	B	183	ARG	Sidechain
1	B	40	ARG	Sidechain
1	B	67	ARG	Sidechain

## 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	B	3895	0	3864	68	0
2	A	3894	0	3864	72	0
2	C	3894	0	3864	66	0
2	D	3902	0	3876	78	0
3	A	10	0	14	1	0
3	B	10	0	14	1	0
3	C	10	0	14	1	0
3	D	10	0	14	0	0
4	A	6	0	8	5	0
4	B	6	0	8	0	0
4	C	6	0	8	0	0
4	D	6	0	8	6	0
5	A	116	0	0	6	0
5	B	152	0	0	5	0
5	C	86	0	0	2	0
5	D	103	0	0	3	0
All	All	16106	0	15556	283	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 9.

The worst 5 of 283 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)
2:A:288:ASN:N	5:A:601:HOH:O	1.77	1.14
2:D:459:ARG:HG3	2:D:459:ARG:HH11	1.15	1.11
1:B:121:ILE:HG21	1:B:185:MET:CE	1.83	1.08
2:D:82:ARG:HE	4:D:502:GOL:H31	1.08	1.07
1:B:121:ILE:CG2	1:B:185:MET:HE2	1.89	1.02

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	B	488/497 (98%)	473 (97%)	15 (3%)	0	100	100
2	A	488/497 (98%)	470 (96%)	18 (4%)	0	100	100
2	C	488/497 (98%)	464 (95%)	24 (5%)	0	100	100
2	D	489/497 (98%)	465 (95%)	23 (5%)	1 (0%)	47	55
All	All	1953/1988 (98%)	1872 (96%)	80 (4%)	1 (0%)	51	60

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	92	GLY

#### 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	B	403/409 (98%)	368 (91%)	35 (9%)	10	10
2	A	403/409 (98%)	369 (92%)	34 (8%)	11	11
2	C	403/409 (98%)	371 (92%)	32 (8%)	12	12
2	D	404/409 (99%)	361 (89%)	43 (11%)	6	6
All	All	1613/1636 (99%)	1469 (91%)	144 (9%)	9	9

5 of 144 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	57	ARG
2	A	475	LYS
2	A	91	ASP
2	A	201	LEU
2	C	125	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	247	GLN
2	A	80	ASN
2	A	278	ASN
2	A	240	GLN
2	C	80	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 monosaccharides in this entry.

## 5.6 Ligand geometry

8 ligands are modelled in this entry.

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	PGE	B	501	-	9,9,9	0.39	0	8,8,8	0.51	0
4	GOL	B	502	-	5,5,5	0.36	0	5,5,5	0.87	0
4	GOL	D	502	-	5,5,5	0.63	0	5,5,5	1.04	0
3	PGE	D	501	-	9,9,9	0.57	0	8,8,8	0.67	0
3	PGE	C	501	-	9,9,9	0.41	0	8,8,8	0.56	0
4	GOL	C	502	-	5,5,5	0.19	0	5,5,5	0.79	0
4	GOL	A	502	-	5,5,5	0.73	0	5,5,5	1.62	1 (20%)
3	PGE	A	501	-	9,9,9	0.48	0	8,8,8	0.58	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	PGE	B	501	-	-	3/7/7/7	-
4	GOL	B	502	-	-	2/4/4/4	-
4	GOL	D	502	-	-	2/4/4/4	-
3	PGE	D	501	-	-	3/7/7/7	-
3	PGE	C	501	-	-	3/7/7/7	-
4	GOL	C	502	-	-	2/4/4/4	-
4	GOL	A	502	-	-	0/4/4/4	-
3	PGE	A	501	-	-	4/7/7/7	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	502	GOL	O2-C2-C1	2.52	119.60	109.18

There are no chirality outliers.

5 of 19 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	502	GOL	C1-C2-C3-O3
3	C	501	PGE	O1-C1-C2-O2
3	B	501	PGE	O3-C5-C6-O4
3	D	501	PGE	O3-C5-C6-O4
3	B	501	PGE	O2-C3-C4-O3

There are no ring outliers.

5 monomers are involved in 14 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	501	PGE	1	0
4	D	502	GOL	6	0
3	C	501	PGE	1	0
4	A	502	GOL	5	0
3	A	501	PGE	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 [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	B	490/497 (98%)	-0.24	6 (1%) 79 77	15, 27, 52, 77	0
2	A	490/497 (98%)	-0.06	12 (2%) 59 56	15, 31, 63, 88	0
2	C	490/497 (98%)	-0.14	5 (1%) 82 81	19, 32, 56, 84	0
2	D	491/497 (98%)	-0.03	9 (1%) 68 66	17, 32, 65, 112	0
All	All	1961/1988 (98%)	-0.12	32 (1%) 72 70	15, 31, 60, 112	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	117	TYR	8.7
2	A	40	ARG	3.4
2	D	91	ASP	3.4
2	D	40	ARG	3.3
2	C	115	ARG	3.2

### 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 monosaccharides 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, 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	B-factors( $\text{\AA}^2$ )	Q<0.9
4	GOL	D	502	6/6	0.81	0.34	46,52,61,61	0
4	GOL	C	502	6/6	0.90	0.23	30,35,39,56	0
3	PGE	D	501	10/10	0.91	0.19	42,45,46,47	0
4	GOL	A	502	6/6	0.91	0.26	36,45,49,57	0
3	PGE	C	501	10/10	0.92	0.12	38,42,47,48	0
3	PGE	B	501	10/10	0.92	0.13	34,38,40,41	0
3	PGE	A	501	10/10	0.92	0.16	33,37,41,46	0
4	GOL	B	502	6/6	0.93	0.22	28,32,44,58	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.