



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 7, 2017 – 10:56 PM EDT

PDB ID : 5F05  
Title : Crystal structure of glutathione transferase F5 from Populus trichocarpa  
Authors : Didierjean, C.; Rouhier, N.; Pegéot, H.; Gense, F.  
Deposited on : unknown  
Resolution : 1.70 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-20029824  
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-20029824

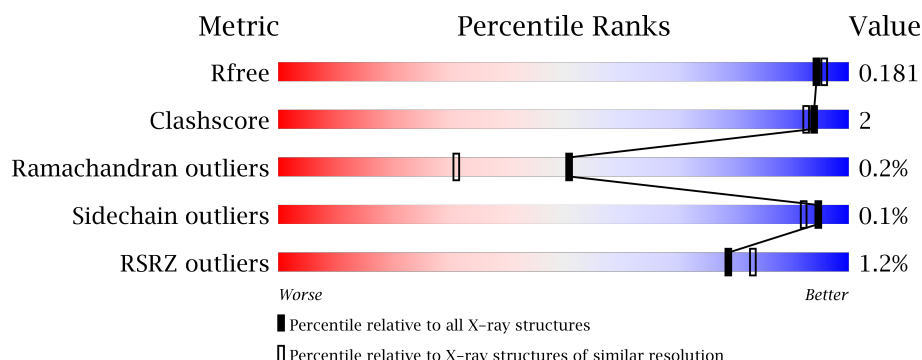
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.70 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))
$R_{free}$	100719	3453 (1.70-1.70)
Clashscore	112137	3876 (1.70-1.70)
Ramachandran outliers	110173	3815 (1.70-1.70)
Sidechain outliers	110143	3815 (1.70-1.70)
RSRZ outliers	101464	3491 (1.70-1.70)

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	212	<div> <div>2%</div> <div>94%</div> <div>5%</div> </div>
1	B	212	<div> <div>96%</div> </div>
1	C	212	<div> <div>2%</div> <div>97%</div> <div>1%</div> </div>
1	D	212	<div> <div>94%</div> <div>5%</div> </div>

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GSH	A	301	-	-	-	X
3	GOL	A	302	-	-	-	X
3	GOL	B	304	-	-	-	X
3	GOL	D	304	-	-	-	X
6	PG4	B	303	-	-	-	X
7	PGE	D	303	-	-	-	X

## 2 Entry composition [i](#)

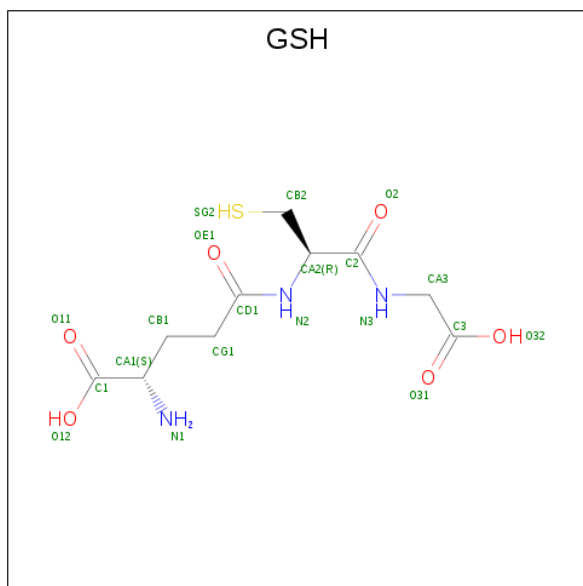
There are 8 unique types of molecules in this entry. The entry contains 14499 atoms, of which 6816 are hydrogens and 0 are deuteriums.

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 Phi class glutathione transferase GSTF5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	210	Total	C	H	N	O	0	4	0
			3268	1051	1633	277	307			
1	B	210	Total	C	H	N	O	0	8	0
			3307	1062	1654	280	311			
1	C	211	Total	C	H	N	O	0	6	0
			3303	1060	1652	280	311			
1	D	211	Total	C	H	N	O	0	10	0
			3331	1068	1665	282	316			

- Molecule 2 is GLUTATHIONE (three-letter code: GSH) (formula: C<sub>10</sub>H<sub>17</sub>N<sub>3</sub>O<sub>6</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	0	0
			35	10	15	3	6		
2	B	1	Total	C	H	N	O	0	0
			35	10	15	3	6		

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	C	1	Total	C	H	N	O	S	0	0
			35	10	15	3	6	1		
2	D	1	Total	C	H	N	O	S	0	0
			35	10	15	3	6	1		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).

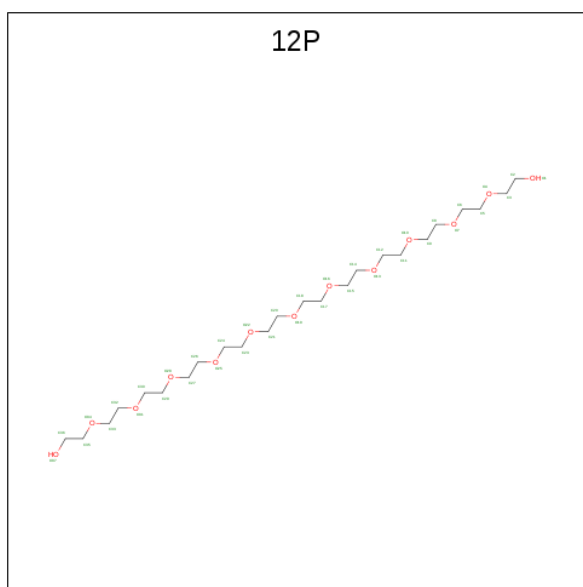


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			11	3	5	3		
3	B	1	Total	C	H	O	0	0
			11	3	5	3		
3	C	1	Total	C	H	O	0	0
			11	3	5	3		
3	D	1	Total	C	H	O	0	0
			11	3	5	3		

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

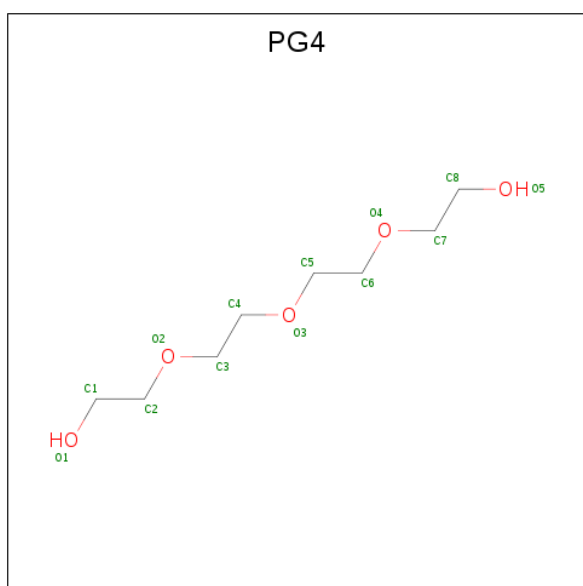
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		
4	A	1	Total	Mg	0	0
			1	1		

- Molecule 5 is DODECAETHYLENE GLYCOL (three-letter code: 12P) (formula:  $C_{24}H_{50}O_{13}$ ).



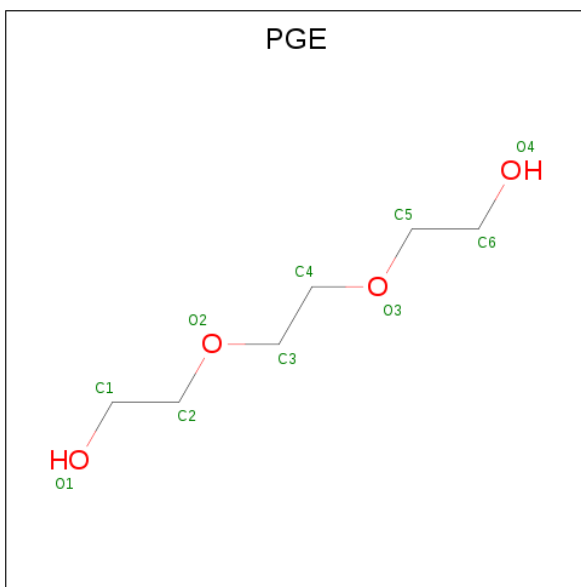
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	B	1	Total	C	H	O	0	0
			87	24	50	13		
5	D	1	Total	C	H	O	0	0
			87	24	50	13		

- Molecule 6 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula:  $C_8H_{18}O_5$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	B	1	Total	C	H	O	0	0
			31	8	18	5		

- Molecule 7 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula:  $C_6H_{14}O_4$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	D	1	Total	C	H	O	0	0
			24	6	14	4		

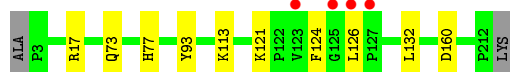
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	220	Total	O	0	0
			220	220		
8	B	209	Total	O	0	0
			209	209		
8	C	243	Total	O	0	0
			243	243		
8	D	203	Total	O	0	0
			203	203		

### 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.

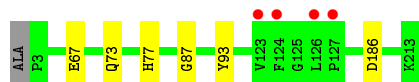
- Molecule 1: Phi class glutathione transferase GSTF5



- Molecule 1: Phi class glutathione transferase GSTF5



- Molecule 1: Phi class glutathione transferase GSTF5



- Molecule 1: Phi class glutathione transferase GSTF5





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	54.46Å 85.95Å 88.52Å 90.00° 97.93° 90.00°	Depositor
Resolution (Å)	43.35 – 1.70 45.69 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.5 (43.35-1.70) 99.6 (45.69-1.70)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.82 (at 1.70Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.144 , 0.181 0.147 , 0.181	Depositor DCC
$R_{free}$ test set	4424 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	11.4	Xtriage
Anisotropy	0.818	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.43 , 55.1	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	14499	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	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 41.80 % 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 2.2377e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, MG, PGE, 12P, PG4, GSH

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	0.69	0/1694	0.73	2/2310 (0.1%)
1	B	0.72	0/1736	0.75	0/2364
1	C	0.70	0/1718	0.71	1/2340 (0.0%)
1	D	0.70	0/1752	0.72	0/2384
All	All	0.70	0/6900	0.73	3/9398 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	17	ARG	NE-CZ-NH1	6.49	123.55	120.30
1	C	186	ASP	CB-CG-OD1	5.18	122.97	118.30
1	A	160	ASP	CB-CG-OD1	5.07	122.86	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts

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	1635	1633	1617	6	0
1	B	1653	1654	1608	6	0
1	C	1651	1652	1624	3	0
1	D	1666	1665	1605	8	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	20	15	15	0	0
2	B	20	15	15	0	0
2	C	20	15	15	0	0
2	D	20	15	15	0	0
3	A	6	5	8	0	0
3	B	6	5	8	0	0
3	C	6	5	8	0	0
3	D	6	5	8	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	B	37	50	50	3	0
5	D	37	50	50	4	0
6	B	13	18	18	1	0
7	D	10	14	14	2	0
8	A	220	0	0	0	0
8	B	209	0	0	0	0
8	C	243	0	0	0	0
8	D	203	0	0	0	0
All	All	7683	6816	6678	22	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 2.

The worst 5 of 22 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:5[A]:LYS:NZ	1:B:30:GLU:OE1	2.36	0.57
1:B:87:GLY:HA3	5:B:301:12P:H322	1.90	0.53
1:D:62:ASP:H	7:D:303:PGE:H6	1.75	0.52
1:A:113:LYS:CE	1:A:132:LEU:HD21	2.40	0.51
1:D:117:GLU:HA	1:D:121:LYS:HD3	1.95	0.49

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	212/212 (100%)	208 (98%)	3 (1%)	1 (0%)	32	15
1	B	216/212 (102%)	214 (99%)	2 (1%)	0	100	100
1	C	215/212 (101%)	210 (98%)	4 (2%)	1 (0%)	32	15
1	D	219/212 (103%)	217 (99%)	2 (1%)	0	100	100
All	All	862/848 (102%)	849 (98%)	11 (1%)	2 (0%)	51	31

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	124	PHE
1	C	67	GLU

### 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	177/175 (101%)	177 (100%)	0	100	100
1	B	182/175 (104%)	182 (100%)	0	100	100
1	C	180/175 (103%)	180 (100%)	0	100	100
1	D	185/175 (106%)	184 (100%)	1 (0%)	91	87
All	All	724/700 (103%)	723 (100%)	1 (0%)	94	92

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	12	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	77	HIS
1	A	85	GLN
1	A	100	GLN
1	B	77	HIS
1	B	78	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 14 ligands modelled in this entry, 2 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	GSH	A	301	-	11,19,19	0.96	0	14,24,24	1.32	2 (14%)
3	GOL	A	302	-	5,5,5	0.50	0	5,5,5	0.64	0
5	12P	B	301	-	36,36,36	0.51	0	35,35,35	0.68	0
2	GSH	B	302	-	11,19,19	0.80	0	14,24,24	1.34	2 (14%)
6	PG4	B	303	-	12,12,12	0.63	0	11,11,11	0.37	0
3	GOL	B	304	-	5,5,5	0.41	0	5,5,5	0.54	0
2	GSH	C	301	-	11,19,19	1.11	1 (9%)	14,24,24	0.89	0
3	GOL	C	302	-	5,5,5	0.50	0	5,5,5	1.07	1 (20%)
5	12P	D	301	-	36,36,36	0.60	0	35,35,35	0.67	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	GSH	D	302	-	11,19,19	0.73	0	14,24,24	0.85	0
7	PGE	D	303	-	9,9,9	0.24	0	8,8,8	0.84	0
3	GOL	D	304	-	5,5,5	0.30	0	5,5,5	0.45	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	GSH	A	301	-	-	0/18/24/24	0/0/0/0
3	GOL	A	302	-	-	0/4/4/4	0/0/0/0
5	12P	B	301	-	-	0/34/34/34	0/0/0/0
2	GSH	B	302	-	-	0/18/24/24	0/0/0/0
6	PG4	B	303	-	-	0/10/10/10	0/0/0/0
3	GOL	B	304	-	-	0/4/4/4	0/0/0/0
2	GSH	C	301	-	-	0/18/24/24	0/0/0/0
3	GOL	C	302	-	-	0/4/4/4	0/0/0/0
5	12P	D	301	-	-	0/34/34/34	0/0/0/0
2	GSH	D	302	-	-	0/18/24/24	0/0/0/0
7	PGE	D	303	-	-	0/7/7/7	0/0/0/0
3	GOL	D	304	-	-	0/4/4/4	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	301	GSH	CB2-CA2	2.02	1.55	1.53

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	302	GSH	CA2-CB2-SG2	-3.34	110.25	114.15
2	A	301	GSH	OE1-CD1-N2	-2.47	118.71	122.97
3	C	302	GOL	O1-C1-C2	-2.32	98.40	110.07
2	A	301	GSH	O2-C2-CA2	-2.23	115.62	120.43
2	B	302	GSH	CB1-CG1-CD1	-2.07	108.49	113.18

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	301	12P	3	0
6	B	303	PG4	1	0
5	D	301	12P	4	0
7	D	303	PGE	2	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	210/212 (99%)	-0.55	4 (1%) 67 72	5, 11, 30, 49	0
1	B	210/212 (99%)	-0.58	0 100 100	5, 12, 28, 41	0
1	C	211/212 (99%)	-0.54	4 (1%) 67 72	5, 11, 31, 55	0
1	D	211/212 (99%)	-0.58	2 (0%) 84 87	5, 11, 30, 41	0
All	All	842/848 (99%)	-0.56	10 (1%) 79 83	5, 11, 30, 55	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	124	PHE	4.9
1	C	126	LEU	4.2
1	A	126	LEU	3.6
1	C	123	VAL	3.3
1	C	127	PRO	3.1

### 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
2	GSH	A	301	20/20	0.95	0.14	6.70	9,20,34,47	0
6	PG4	B	303	13/13	0.86	0.13	4.33	22,38,49,51	0
3	GOL	A	302	6/6	0.92	0.08	3.38	19,38,47,47	0
3	GOL	B	304	6/6	0.94	0.10	3.07	16,22,31,38	0
3	GOL	D	304	6/6	0.95	0.11	2.24	16,22,40,42	0
7	PGE	D	303	10/10	0.92	0.10	2.06	19,38,49,51	0
2	GSH	C	301	20/20	0.98	0.09	1.57	7,13,24,28	0
2	GSH	B	302	20/20	0.98	0.10	1.43	6,14,27,31	0
5	12P	D	301	37/37	0.89	0.14	1.40	18,40,63,69	0
5	12P	B	301	37/37	0.87	0.12	1.12	13,36,60,69	0
2	GSH	D	302	20/20	0.98	0.07	0.69	7,12,28,35	0
3	GOL	C	302	6/6	0.95	0.08	0.59	16,35,43,51	0
4	MG	B	305	1/1	0.98	0.12	-	22,22,22,22	0
4	MG	A	303	1/1	0.98	0.11	-	20,20,20,20	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.