



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 17, 2016 – 01:57 PM EDT

PDB ID : 5LKD  
Title : Crystal structure of the Xi glutathione transferase ECM4 from *Saccharomyces cerevisiae* in complex with glutathione  
Authors : Schwartz, M.; Didierjean, C.; Hecker, A.; Girardet, J.M.; Morel-Rouhier, M.; Gelhaye, E.; Favier, F.  
Deposited on : 2016-07-22  
Resolution : 1.68 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027939  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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-20027939

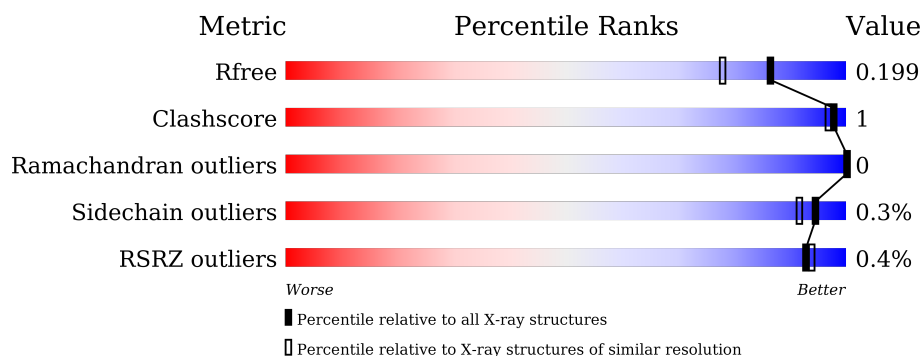
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.68 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))
$R_{free}$	91344	4802 (1.70-1.66)
Clashscore	102246	5317 (1.70-1.66)
Ramachandran outliers	100387	5225 (1.70-1.66)
Sidechain outliers	100360	5224 (1.70-1.66)
RSRZ outliers	91569	4813 (1.70-1.66)

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	378	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, orange 1%, yellow 1%, green 89%, grey 7%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>%</span> <span>89%</span> <span>7%</span> </div> </div>
1	B	378	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 85%, yellow 1%, grey 11%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>85%</span> <span>11%</span> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11990 atoms, of which 5579 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.

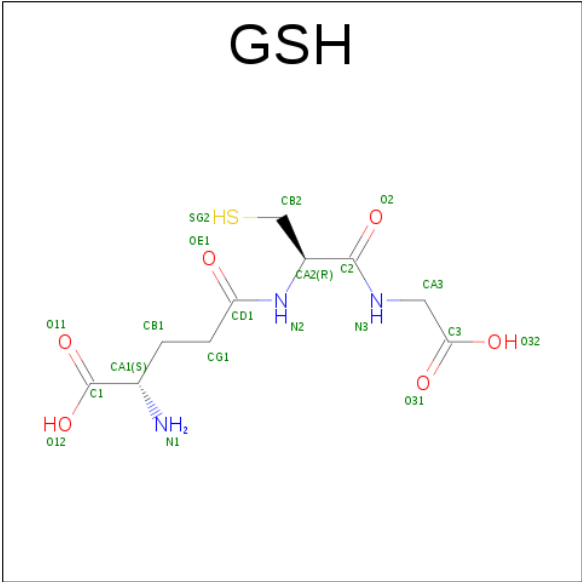
- Molecule 1 is a protein called Glutathione S-transferase omega-like 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	352	Total	C	H	N	O	S	263	0	0
			5757	1883	2833	499	538	4			
1	B	338	Total	C	H	N	O	S	174	0	0
			5527	1810	2718	479	516	4			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	371	LEU	-	expression tag	UNP P36156
A	372	GLU	-	expression tag	UNP P36156
A	373	HIS	-	expression tag	UNP P36156
A	374	HIS	-	expression tag	UNP P36156
A	375	HIS	-	expression tag	UNP P36156
A	376	HIS	-	expression tag	UNP P36156
A	377	HIS	-	expression tag	UNP P36156
A	378	HIS	-	expression tag	UNP P36156
B	371	LEU	-	expression tag	UNP P36156
B	372	GLU	-	expression tag	UNP P36156
B	373	HIS	-	expression tag	UNP P36156
B	374	HIS	-	expression tag	UNP P36156
B	375	HIS	-	expression tag	UNP P36156
B	376	HIS	-	expression tag	UNP P36156
B	377	HIS	-	expression tag	UNP P36156
B	378	HIS	-	expression tag	UNP P36156

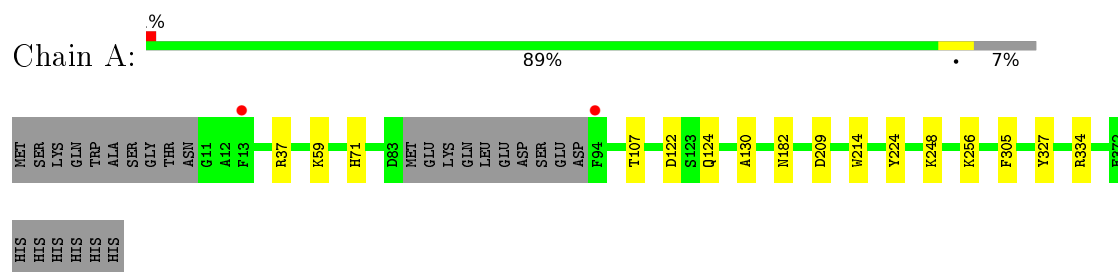
- Molecule 2 is GLUTATHIONE (three-letter code: GSH) (formula: C<sub>10</sub>H<sub>17</sub>N<sub>3</sub>O<sub>6</sub>S).



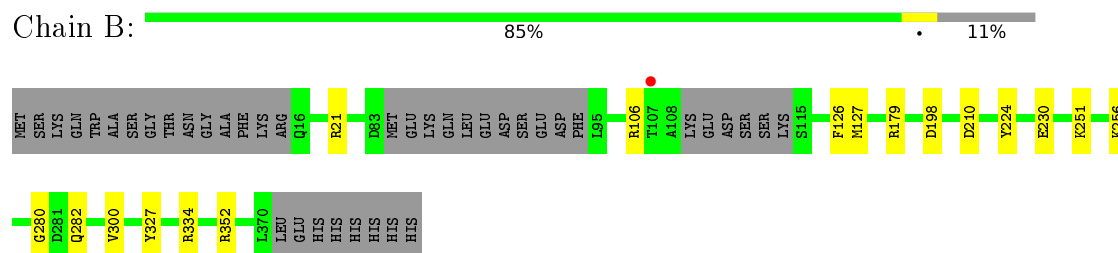
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: Glutathione S-transferase omega-like 2



- Molecule 1: Glutathione S-transferase omega-like 2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	55.44Å 82.44Å 80.58Å 90.00° 95.31° 90.00°	Depositor
Resolution (Å)	47.58 – 1.68 47.58 – 1.68	Depositor EDS
% Data completeness (in resolution range)	95.0 (47.58-1.68) 88.6 (47.58-1.68)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.18 (at 1.68Å)	Xtriage
Refinement program	PHENIX (1.10.1 _2155: ???)	Depositor
R, $R_{free}$	0.159 , 0.200 0.158 , 0.199	Depositor DCC
$R_{free}$ test set	3635 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	17.4	Xtriage
Anisotropy	0.791	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.43 , 55.0	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	11990	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.71% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.87	1/3008 (0.0%)	0.86	4/4080 (0.1%)
1	B	0.93	2/2890 (0.1%)	0.92	5/3924 (0.1%)
All	All	0.90	3/5898 (0.1%)	0.89	9/8004 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	300	VAL	CB-CG2	-8.43	1.35	1.52
1	B	224	TYR	CD1-CE1	-5.46	1.31	1.39
1	A	224	TYR	CE1-CZ	-5.03	1.32	1.38

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	106	ARG	NE-CZ-NH2	-7.68	116.46	120.30
1	B	179	ARG	NE-CZ-NH1	7.32	123.96	120.30
1	B	210	ASP	CB-CG-OD2	5.87	123.58	118.30
1	A	37	ARG	NE-CZ-NH2	-5.74	117.43	120.30
1	A	305	PHE	CB-CG-CD2	-5.34	117.06	120.80
1	A	209	ASP	CB-CG-OD1	5.20	122.98	118.30
1	B	21	ARG	NE-CZ-NH2	-5.19	117.71	120.30
1	B	198	ASP	CB-CG-OD1	5.13	122.92	118.30
1	A	214	TRP	CA-CB-CG	5.06	123.31	113.70

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	2924	2833	2832	5	0
1	B	2809	2718	2716	7	0
2	A	20	14	14	0	0
2	B	20	14	14	0	0
3	A	316	0	0	1	0
3	B	322	0	0	3	0
All	All	6411	5579	5576	12	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 1.

All (12) 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:A:248:LYS:NZ	3:A:501:HOH:O	2.24	0.64
1:B:352:ARG:NH2	3:B:501:HOH:O	2.32	0.63
1:B:256:LYS:NZ	1:B:280:GLY:O	2.32	0.60
1:B:230:GLU:CD	1:B:352:ARG:HH22	2.14	0.50
1:B:327:TYR:O	1:B:334:ARG:HB2	2.12	0.49
1:A:122:ASP:O	1:A:124:GLN:HG3	2.14	0.47
1:A:59:LYS:HE2	1:A:182:ASN:HA	1.99	0.43
1:B:126:PHE:CD2	1:B:127:MET:HG2	2.54	0.42
1:B:282:GLN:HG3	3:B:504:HOH:O	2.19	0.42
1:A:71:HIS:HA	1:A:130:ALA:HB2	2.02	0.41
1:A:327:TYR:O	1:A:334:ARG:HB2	2.21	0.41
1:B:251:LYS:NZ	3:B:521:HOH:O	2.53	0.41

There are no symmetry-related clashes.

## 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	348/378 (92%)	341 (98%)	7 (2%)	0	100	100
1	B	332/378 (88%)	326 (98%)	6 (2%)	0	100	100
All	All	680/756 (90%)	667 (98%)	13 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	318/342 (93%)	316 (99%)	2 (1%)	90	84
1	B	306/342 (90%)	306 (100%)	0	100	100
All	All	624/684 (91%)	622 (100%)	2 (0%)	94	92

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	107	THR
1	A	256	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

2 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	$\# Z  > 2$
2	GSH	A	401	-	13,19,19	2.42	3 (23%)	15,24,24	1.81	5 (33%)
2	GSH	B	401	1	13,19,19	2.07	3 (23%)	15,24,24	1.57	4 (26%)

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	401	-	-	0/18/24/24	0/0/0/0
2	GSH	B	401	1	-	0/18/24/24	0/0/0/0

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	401	GSH	O2-C2	-3.06	1.17	1.23
2	B	401	GSH	CD1-N2	3.91	1.42	1.34
2	A	401	GSH	CB2-CA2	4.44	1.57	1.52
2	A	401	GSH	CD1-N2	4.66	1.43	1.34
2	B	401	GSH	C2-N3	4.77	1.43	1.33
2	A	401	GSH	C2-N3	4.96	1.43	1.33

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	401	GSH	OE1-CD1-N2	-3.16	117.65	122.96

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	401	GSH	CB1-CG1-CD1	-2.99	106.43	113.26
2	B	401	GSH	CB1-CA1-N1	-2.78	102.70	110.54
2	A	401	GSH	CB1-CG1-CD1	-2.61	107.28	113.26
2	B	401	GSH	CB2-CA2-C2	-2.47	104.04	109.72
2	A	401	GSH	C2-CA2-N2	-2.09	105.37	111.28
2	A	401	GSH	CB2-CA2-N2	2.52	114.83	111.43
2	A	401	GSH	CG1-CD1-N2	2.98	120.86	115.85
2	B	401	GSH	CB2-CA2-N2	3.20	115.75	111.43

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	352/378 (93%)	-0.32	2 (0%) 90 92	14, 24, 41, 55	27 (7%)
1	B	338/378 (89%)	-0.33	1 (0%) 94 95	13, 22, 43, 55	18 (5%)
All	All	690/756 (91%)	-0.32	3 (0%) 93 94	13, 23, 42, 55	45 (6%)

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	94	PHE	3.6
1	B	107	THR	2.1
1	A	13	PHE	2.0

### 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	401	20/20	0.96	0.10	1.63	21,31,53,64	0
2	GSH	B	401	20/20	0.94	0.10	0.52	22,29,50,57	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.