



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 24, 2020 – 05:06 am BST

PDB ID : 3KAJ  
Title : Apoenzyme structure of homogluthathione synthetase from Glycine max in open conformation  
Authors : Galant, A.; Arkus, K.A.J.; Zubieta, C.; Cahoon, R.E.; Jez, J.M.  
Deposited on : 2009-10-19  
Resolution : 2.00 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

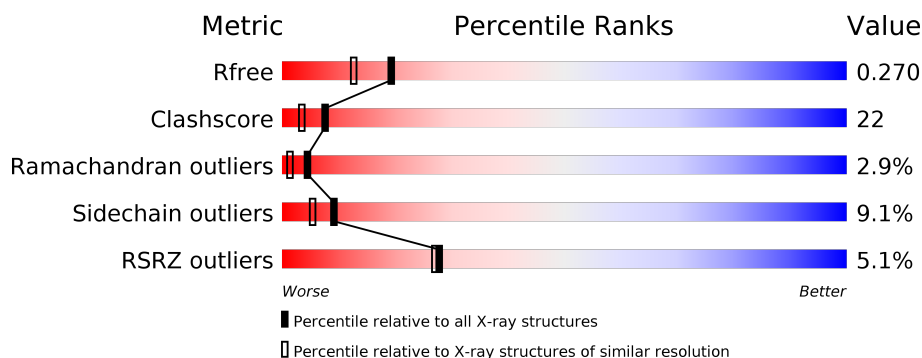
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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	A	499	<div> <div>6%</div> <div> <div></div> <div>66%</div> <div>18%</div> <div>6%</div> <div>9%</div> </div> </div>
1	B	499	<div> <div>4%</div> <div> <div></div> <div>64%</div> <div>19%</div> <div>5%</div> <div>11%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7754 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 Homoglutathione synthetase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	452	Total	C	N	O	S	0	13	0
			3696	2347	638	695	16			
1	B	442	Total	C	N	O	S	0	9	0
			3589	2280	626	667	16			

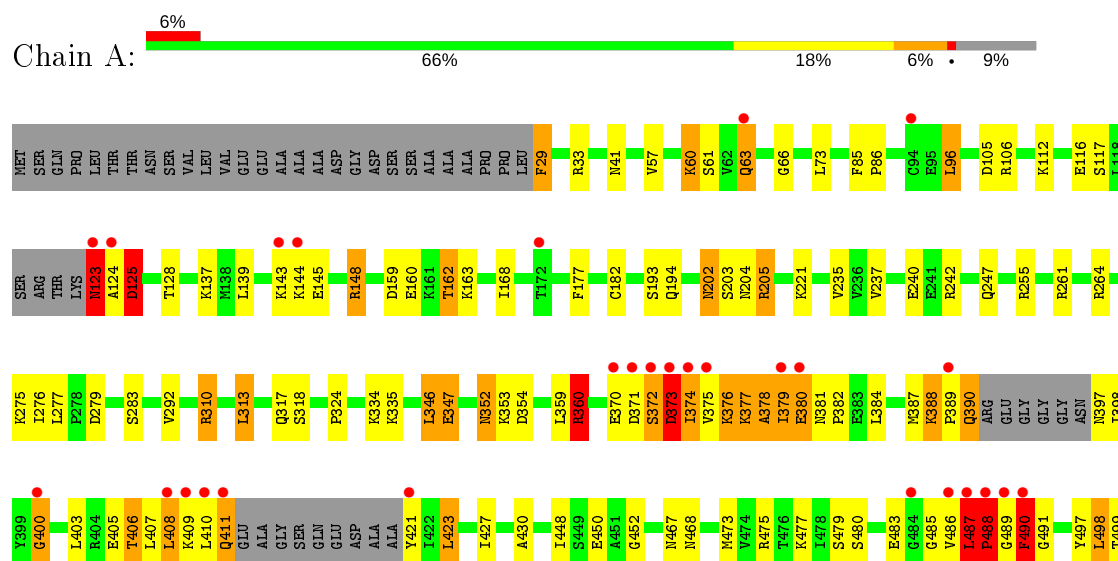
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	249	Total	O	0	0
			249	249		
2	B	220	Total	O	0	0
			220	220		

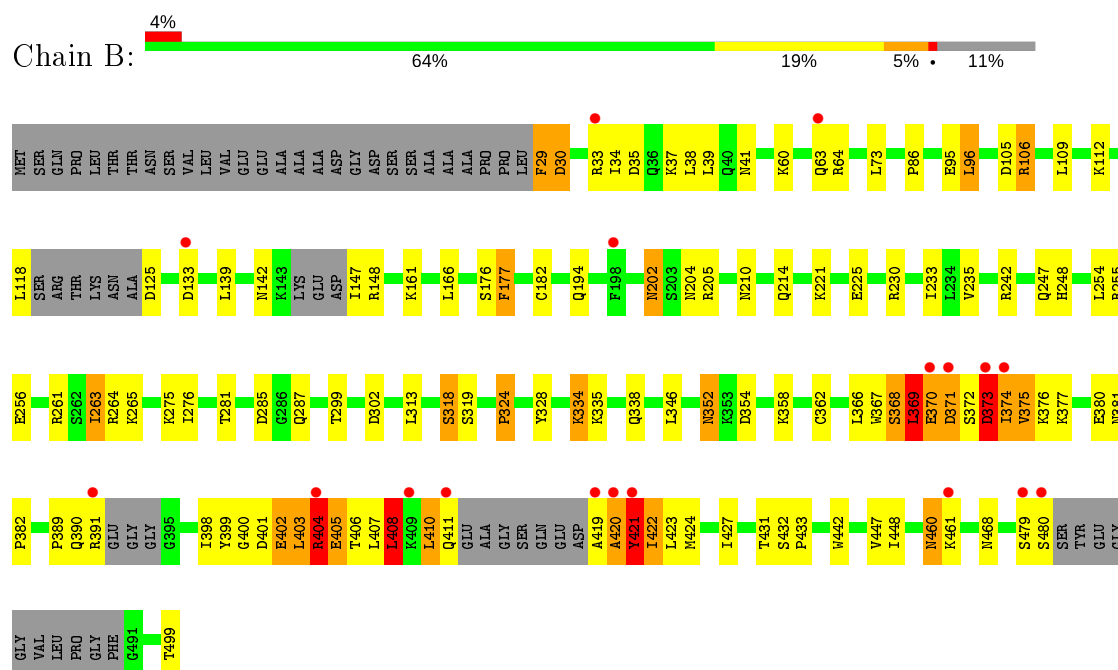
### 3 Residue-property plots [i](#)

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: Homogluthathione synthetase



#### • Molecule 1: Homogluthathione synthetase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.96Å 80.55Å 90.00Å 90.00° 96.94° 90.00°	Depositor
Resolution (Å)	28.38 – 2.00 28.37 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (28.38-2.00) 97.3 (28.37-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.26 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.198 , 0.269 0.203 , 0.270	Depositor DCC
$R_{free}$ test set	3031 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.9	Xtriage
Anisotropy	0.093	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 47.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7754	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	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 21.42 % 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 7.0353e-03. 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

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	0.99	5/3764 (0.1%)	1.10	20/5084 (0.4%)
1	B	1.03	5/3652 (0.1%)	1.09	18/4928 (0.4%)
All	All	1.01	10/7416 (0.1%)	1.09	38/10012 (0.4%)

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	1

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	410	LEU	CB-CG	-16.25	1.05	1.52
1	B	402	GLU	C-N	11.02	1.59	1.34
1	B	408	LEU	CB-CG	10.72	1.83	1.52
1	A	182	CYS	CB-SG	-10.04	1.65	1.82
1	B	182	CYS	CB-SG	-9.56	1.66	1.82

The worst 5 of 38 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	410	LEU	CA-CB-CG	20.67	162.85	115.30
1	B	410	LEU	CB-CG-CD1	16.53	139.11	111.00
1	A	123	ASN	O-C-N	-13.29	101.43	122.70
1	A	148	ARG	NE-CZ-NH2	-12.71	113.94	120.30
1	A	310	ARG	NE-CZ-NH2	-12.06	114.27	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	468	ASN	Peptide

## 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	3696	0	3718	144	1
1	B	3589	0	3633	182	1
2	A	249	0	0	12	1
2	B	220	0	0	16	0
All	All	7754	0	7351	323	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 22.

The worst 5 of 323 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:408:LEU:CG	1:B:408:LEU:CB	1.83	1.53
1:B:367:TRP:HB3	1:B:374:ILE:CD1	1.60	1.32
1:B:369:LEU:O	1:B:369:LEU:HD13	1.38	1.20
1:A:475:ARG:HD3	1:A:490:PHE:CE1	1.77	1.18
1:B:404[B]:ARG:CD	1:B:407:LEU:HD12	1.73	1.17

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)
1:A:354:ASP:OD1	1:A:483:GLU:O[2_555]	1.84	0.36
1:B:404[A]:ARG:NH1	2:A:554:HOH:O[2_546]	2.16	0.04

## 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	457/499 (92%)	418 (92%)	25 (6%)	14 (3%)	4	1
1	B	439/499 (88%)	411 (94%)	14 (3%)	14 (3%)	4	1
All	All	896/998 (90%)	829 (92%)	39 (4%)	28 (3%)	4	1

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	370	GLU
1	A	374	ILE
1	A	408	LEU
1	A	487	LEU
1	A	490	PHE

### 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	411/431 (95%)	375 (91%)	36 (9%)	10	6
1	B	398/431 (92%)	359 (90%)	39 (10%)	8	4
All	All	809/862 (94%)	734 (91%)	75 (9%)	9	5

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	486	VAL

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	B	96	LEU
1	B	405	GLU
1	A	487	LEU
1	B	35	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	115	GLN
1	B	142	ASN
1	B	337	GLN
1	A	468	ASN
1	B	352	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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	452/499 (90%)	0.05	28 (6%)	20 19	12, 21, 49, 69	1 (0%)
1	B	442/499 (88%)	0.10	18 (4%)	37 36	14, 25, 44, 59	0
All	All	894/998 (89%)	0.08	46 (5%)	28 27	12, 23, 45, 69	1 (0%)

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	374	ILE	19.0
1	A	374	ILE	7.0
1	A	373	ASP	6.9
1	B	419	ALA	6.1
1	B	421	TYR	5.9

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.