



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 01:35 am GMT

PDB ID : 5I2G  
Title : 1,2-propanediol Dehydration in Roseburia inulinivorans; Structural Basis for Substrate and Enantiomer Selectivity  
Authors : LaMattina, J.W.; Reitzer, P.; Kapoor, S.; Galzerani, F.; Koch, D.J.; Gouvea, I.E.; Lanzilotta, W.N.  
Deposited on : 2016-02-08  
Resolution : 2.35 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	:	trunk28620
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)	:	recalc28949

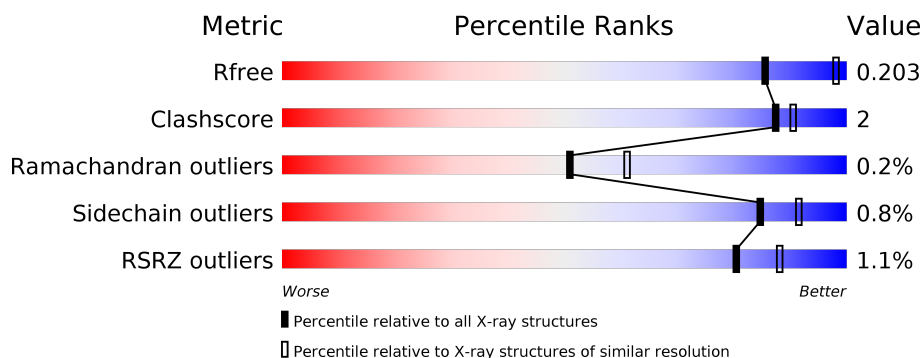
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.35 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	1522 (2.38-2.34)
Clashscore	112137	1626 (2.38-2.34)
Ramachandran outliers	110173	1605 (2.38-2.34)
Sidechain outliers	110143	1606 (2.38-2.34)
RSRZ outliers	101464	1528 (2.38-2.34)

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	843	<div> <div style="width: 100%; height: 10px; position: relative;"> <div style="position: absolute; top: -10px; left: 0; width: 100%; text-align: center;">%</div> <div style="position: absolute; top: 10px; left: 0; width: 100%; text-align: center;">88% 6% 6%</div> </div> </div>
1	B	843	<div> <div style="width: 100%; height: 10px; position: relative;"> <div style="position: absolute; top: -10px; left: 0; width: 100%; text-align: center;">%</div> <div style="position: absolute; top: 10px; left: 0; width: 100%; text-align: center;">88% 6% 6%</div> </div> </div>

## 2 Entry composition [i](#)

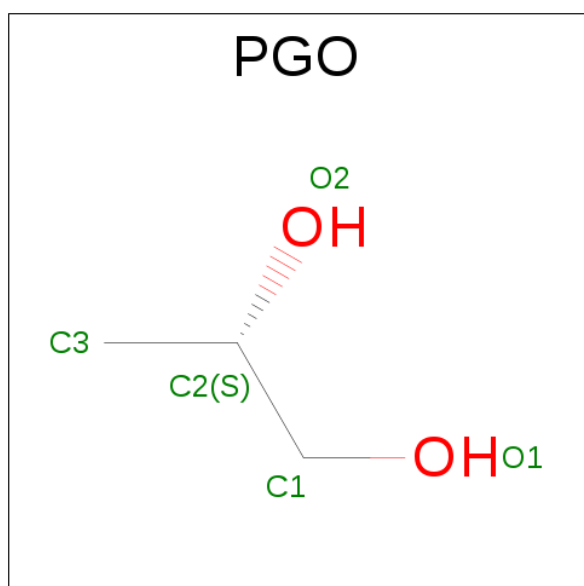
There are 3 unique types of molecules in this entry. The entry contains 12995 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Diol dehydratase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	796	Total	C	N	O	S	0	7	0
			6302	3989	1062	1208	43			
1	B	796	Total	C	N	O	S	0	8	0
			6314	3997	1063	1211	43			

- Molecule 2 is S-1,2-PROPANEDIOL (three-letter code: PGO) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			5	3	2		
2	B	1	Total	C	O	0	0
			5	3	2		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	199	Total 199	O 199	0	0
3	B	170	Total 170	O 170	0	0



- Molecule 1: Diol dehydratase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	204.33Å 83.25Å 138.26Å 90.00° 129.90° 90.00°	Depositor
Resolution (Å)	35.36 – 2.35 49.63 – 2.35	Depositor EDS
% Data completeness (in resolution range)	98.9 (35.36-2.35) 98.9 (49.63-2.35)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.61 (at 2.34Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.145 , 0.203 0.147 , 0.203	Depositor DCC
$R_{free}$ test set	2001 reflections (2.73%)	DCC
Wilson B-factor (Å <sup>2</sup> )	32.4	Xtriage
Anisotropy	0.603	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 47.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.014 for -h-2*k,l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	12995	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.29% 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: PGO

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	0.58	11/6449 (0.2%)	0.54	2/8723 (0.0%)
1	B	0.57	13/6461 (0.2%)	0.52	1/8740 (0.0%)
All	All	0.58	24/12910 (0.2%)	0.53	3/17463 (0.0%)

The worst 5 of 24 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	320	TRP	CE3-CZ3	-16.64	1.10	1.38
1	B	320	TRP	CE3-CZ3	-15.49	1.12	1.38
1	A	319	ILE	C-O	-11.32	1.01	1.23
1	B	320	TRP	CD2-CE2	-10.96	1.28	1.41
1	B	319	ILE	C-O	-10.74	1.02	1.23

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	321	VAL	CG1-CB-CG2	-6.26	100.89	110.90
1	A	321	VAL	CG1-CB-CG2	-5.67	101.82	110.90
1	A	320	TRP	CA-CB-CG	5.49	124.13	113.70

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	6302	0	6141	28	0
1	B	6314	0	6154	31	0
2	A	5	0	8	0	0
2	B	5	0	8	1	0
3	A	199	0	0	2	1
3	B	170	0	0	3	1
All	All	12995	0	12311	59	1

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 59 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:789:GLN:NE2	1:B:814:ARG:HH21	1.55	1.04
1:B:789:GLN:HE22	1:B:814:ARG:HH21	1.35	0.72
1:A:299:GLU:HG2	1:A:303:LYS:HE2	1.76	0.67
1:A:648:GLU:O	1:A:652:GLU:HG3	1.94	0.66
1:B:789:GLN:NE2	1:B:814:ARG:NH2	2.39	0.63

All (1) 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)
3:A:1040:HOH:O	3:B:1168:HOH:O[2_656]	2.07	0.13

## 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	799/843 (95%)	775 (97%)	22 (3%)	2 (0%)	44	53
1	B	800/843 (95%)	776 (97%)	23 (3%)	1 (0%)	55	66

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1599/1686 (95%)	1551 (97%)	45 (3%)	3 (0%)	51 61

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	696	VAL
1	B	696	VAL
1	A	815	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	669/699 (96%)	664 (99%)	5 (1%)	87 93
1	B	671/699 (96%)	665 (99%)	6 (1%)	82 90
All	All	1340/1398 (96%)	1329 (99%)	11 (1%)	85 91

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	357	ARG
1	B	304	GLU
1	B	390	SER
1	A	320	TRP
1	B	320	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	789	GLN
1	B	434	ASN
1	B	784	GLN
1	B	789	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 ⓘ

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	PGO	A	901	-	4,4,4	0.29	0	2,4,4	0.57	0
2	PGO	B	901	-	4,4,4	0.25	0	2,4,4	0.88	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	PGO	A	901	-	-	0/2/2/2	0/0/0/0
2	PGO	B	901	-	-	0/2/2/2	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	901	PGO	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	A	796/843 (94%)	-0.25	9 (1%) 80 88	21, 32, 55, 77	0
1	B	796/843 (94%)	-0.19	9 (1%) 80 88	22, 34, 56, 72	0
All	All	1592/1686 (94%)	-0.22	18 (1%) 80 88	21, 33, 55, 77	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	639	VAL	3.3
1	A	473	GLY	3.2
1	A	573	PHE	3.1
1	A	638	ASP	2.8
1	B	486	PHE	2.7

### 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	PGO	A	901	5/5	0.98	0.20	1.72	22,22,25,26	0
2	PGO	B	901	5/5	0.98	0.19	1.21	25,27,30,33	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.