wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2021 – 02:52 PM EDT

PDB ID	:	1D8C
Title	:	MALATE SYNTHASE G COMPLEXED WITH MAGNESIUM AND GLYOXYLATE
Authors	:	Howard, B.R.; Endrizzi, J.A.; Remington, S.J.
Deposited on	:	1999-10-22
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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

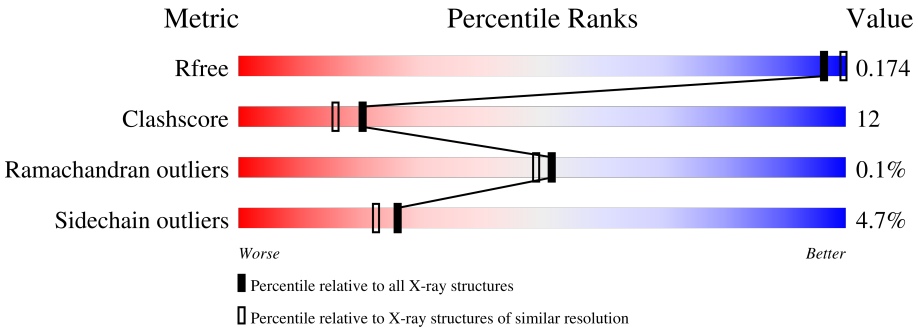
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

1 Overall quality at a glance (i)

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)

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 >=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 <=5%

Mol	Chain	Length	Quality of chain
1	A	723	66% 28%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5719 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.

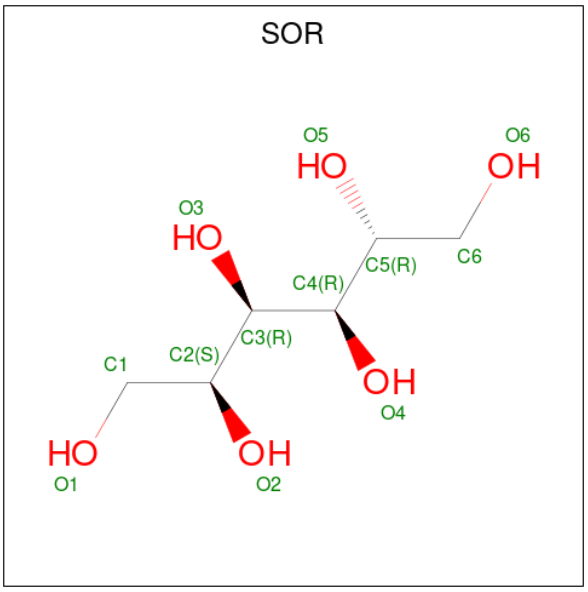
- Molecule 1 is a protein called MALATE SYNTHASE G.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	709	5331	3365	930	1009	6	21	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	conflict	UNP P37330
A	2	ALA	SER	engineered mutation	UNP P37330
A	122	MSE	MET	modified residue	UNP P37330
A	154	MSE	MET	modified residue	UNP P37330
A	294	MSE	MET	modified residue	UNP P37330
A	302	MSE	MET	modified residue	UNP P37330
A	344	MSE	MET	modified residue	UNP P37330
A	366	MSE	MET	modified residue	UNP P37330
A	393	MSE	MET	modified residue	UNP P37330
A	412	MSE	MET	modified residue	UNP P37330
A	415	MSE	MET	modified residue	UNP P37330
A	422	MSE	MET	modified residue	UNP P37330
A	425	MSE	MET	modified residue	UNP P37330
A	461	MSE	MET	modified residue	UNP P37330
A	465	MSE	MET	modified residue	UNP P37330
A	470	MSE	MET	modified residue	UNP P37330
A	476	MSE	MET	modified residue	UNP P37330
A	508	MSE	MET	modified residue	UNP P37330
A	511	MSE	MET	modified residue	UNP P37330
A	515	MSE	MET	modified residue	UNP P37330
A	518	MSE	MET	modified residue	UNP P37330
A	629	MSE	MET	modified residue	UNP P37330
A	663	MSE	MET	modified residue	UNP P37330
A	680	MSE	MET	modified residue	UNP P37330

- Molecule 2 is sorbitol (three-letter code: SOR) (formula: $C_6H_{14}O_6$).

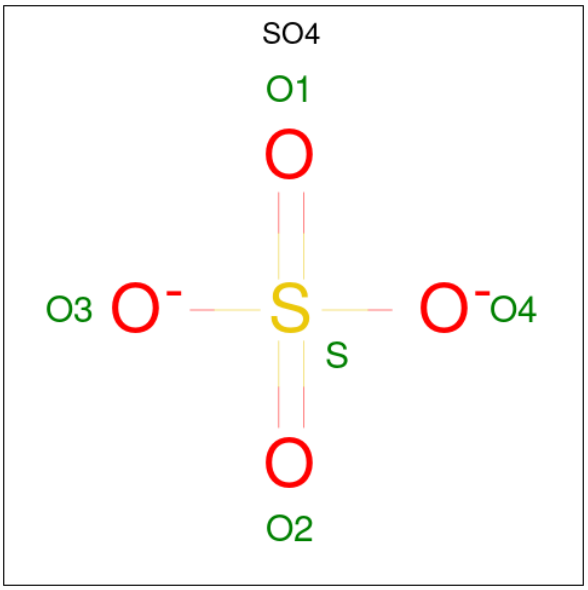


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 12 6 6	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

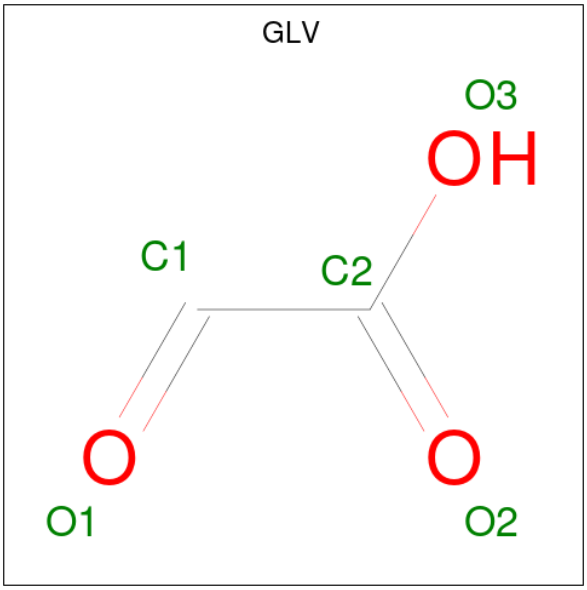
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Mg 1 1	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0

- Molecule 5 is GLYOXYLIC ACID (three-letter code: GLV) (formula: $C_2H_2O_3$).



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

- Molecule 6 is water.

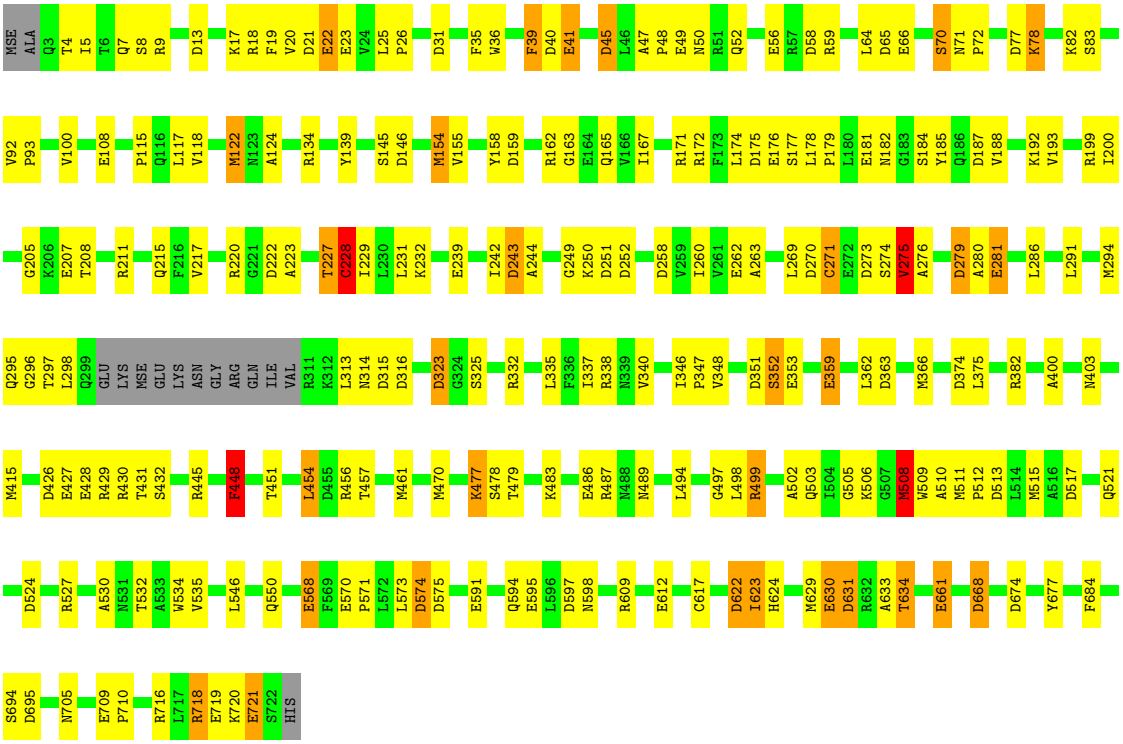
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	345	Total O 345 345	0	0

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. 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. 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: MALATE SYNTHASE G

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	73.80Å 88.70Å 109.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00 19.90 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.6 (20.00-2.00) 99.5 (19.90-2.00)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.60 (at 2.01Å)	Xtriage
Refinement program	TNT	Depositor
R, R_{free}	(Not available) , (Not available) 0.168 , 0.174	Depositor DCC
R_{free} test set	2522 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	25.6	Xtriage
Anisotropy	0.608	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 114.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	5719	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: MG, GLV, SO4, SOR

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	1.02	24/5416 (0.4%)	1.51	101/7345 (1.4%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	56	GLU	CD-OE2	8.08	1.34	1.25
1	A	262	GLU	CD-OE2	7.75	1.34	1.25
1	A	630	GLU	CD-OE2	7.45	1.33	1.25
1	A	486	GLU	CD-OE2	6.82	1.33	1.25
1	A	41	GLU	CD-OE2	6.63	1.32	1.25

The worst 5 of 101 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	9	ARG	NE-CZ-NH1	11.18	125.89	120.30
1	A	382	ARG	NE-CZ-NH1	10.69	125.65	120.30
1	A	382	ARG	NE-CZ-NH2	-9.79	115.41	120.30
1	A	243	ASP	CB-CG-OD2	-9.14	110.07	118.30
1	A	175	ASP	CB-CG-OD1	9.12	126.51	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	5331	0	5096	130	0
2	A	12	0	14	0	0
3	A	1	0	0	0	0
4	A	25	0	0	1	0
5	A	5	0	1	0	0
6	A	345	0	0	9	0
All	All	5719	0	5111	130	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 12.

The worst 5 of 130 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:346:ILE:HG13	1:A:347:PRO:HD2	1.40	1.03
1:A:122:MSE:HE2	1:A:276:ALA:H	1.23	1.00
1:A:508:MSE:HG3	1:A:534:TRP:HB3	1.51	0.91
1:A:7:GLN:HG2	1:A:36:TRP:HB3	1.53	0.91
1:A:508:MSE:HG3	1:A:534:TRP:CB	2.07	0.85

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	705/723 (98%)	684 (97%)	20 (3%)	1 (0%)	51 49

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	155	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	531/576 (92%)	506 (95%)	25 (5%)	26 22

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	281	GLU
1	A	430	ARG
1	A	718	ARG
1	A	352	SER
1	A	448	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	418	ASN
1	A	441	GLN
1	A	705	ASN
1	A	521	GLN
1	A	295	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 1 is monoatomic - leaving 7 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	$\#\|Z\| > 2$
5	GLV	A	2000	3	1,4,4	0.65	0	0,4,4	-	-
4	SO4	A	6000	-	4,4,4	0.92	0	6,6,6	0.39	0
4	SO4	A	5000	-	4,4,4	0.91	0	6,6,6	0.21	0
2	SOR	A	4000	-	11,11,11	0.40	0	14,14,14	1.12	1 (7%)
4	SO4	A	8000	-	4,4,4	1.11	0	6,6,6	0.39	0
4	SO4	A	9000	-	4,4,4	0.86	0	6,6,6	0.16	0
4	SO4	A	7000	-	4,4,4	1.71	1 (25%)	6,6,6	0.33	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
5	GLV	A	2000	3	-	0/0/2/2	-
2	SOR	A	4000	-	-	2/16/16/16	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	7000	SO4	O1-S	2.73	1.60	1.46

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	4000	SOR	C5-C4-C3	-2.02	109.31	112.47

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	4000	SOR	O3-C3-C4-O4
2	A	4000	SOR	C2-C3-C4-O4

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	6000	SO4	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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.