WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 20, 2023 – 06:30 AM EDT

PDB ID	:	2GQ3
Title	:	mycobacterium tuberculosis malate synthase in complex with magnesium, malate, and coenzyme A
Authors	:	Anstrom, D.M.; Remington, S.J.
Deposited on	:	2006-04-19
Resolution	:	2.30 Å(reported)

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

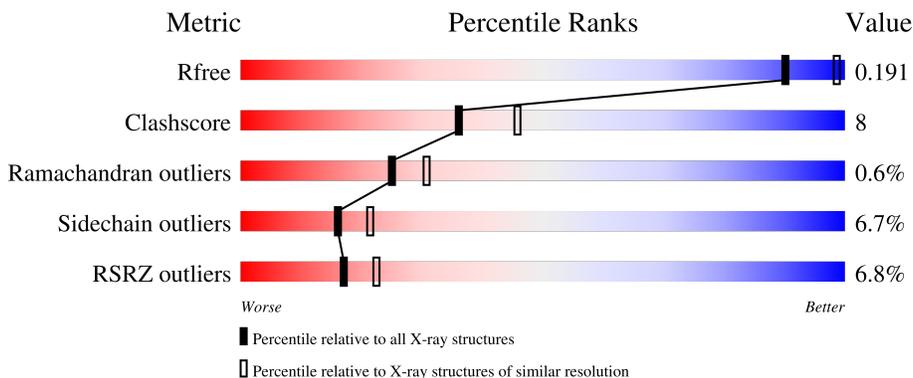|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.30 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	Percentile Ranks	Value
Rfree		0.191
Clashscore		8
Ramachandran outliers		0.6%
Sidechain outliers		6.7%
RSRZ outliers		6.8%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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% 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	729	2% / 83% / 14% / ••
1	B	729	11% / 73% / 21% / ••

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	MLT	A	900	X	-	-	-
3	MLT	B	901	X	-	-	-

2　Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 11633 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.

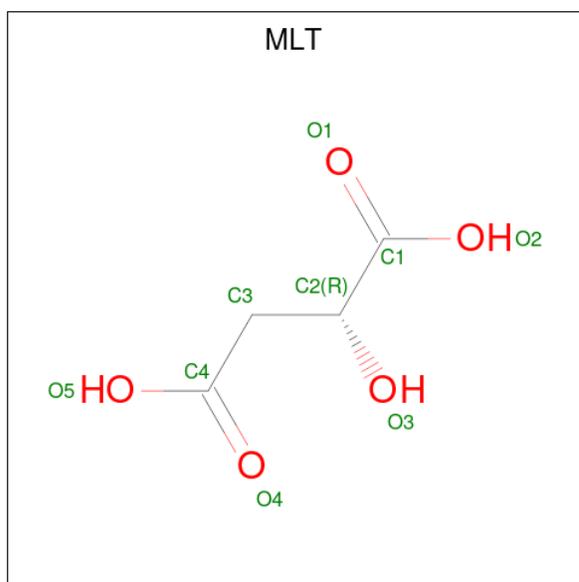| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	720	Total 5487	C 3444	N 964	O 1057	S 22	0	0	0
1	B	714	Total 5429	C 3411	N 953	O 1043	S 22	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	cloning artifact	UNP P0A5J4
A	0	SER	-	cloning artifact	UNP P0A5J4
B	-1	GLY	-	cloning artifact	UNP P0A5J4
B	0	SER	-	cloning artifact	UNP P0A5J4

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total 2	Mg 2	0	0
2	B	2	Total 2	Mg 2	0	0

- Molecule 3 is D-MALATE (three-letter code: MLT) (formula: $C_4H_6O_5$).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 9	C 4	O 5	0	0
3	B	1	Total 9	C 4	O 5	0	0

- Molecule 4 is COENZYME A (three-letter code: COA) (formula: $C_{21}H_{36}N_7O_{16}P_3S$).
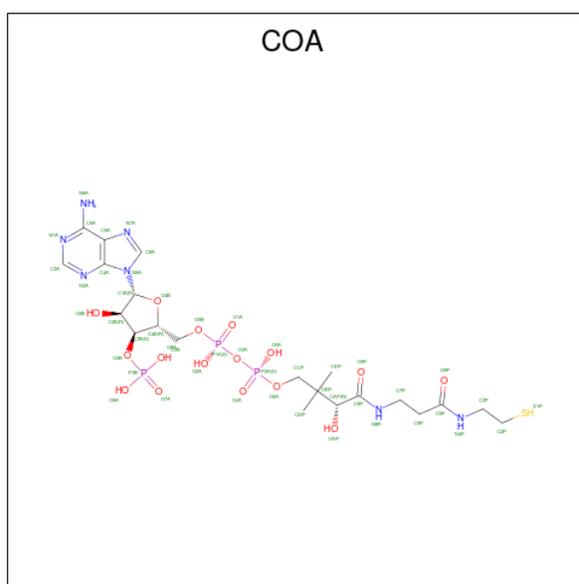
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
4	A	1	Total 48	C 21	N 7	O 16	P 3	S 1	0	0

- Molecule 5 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID

(three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).
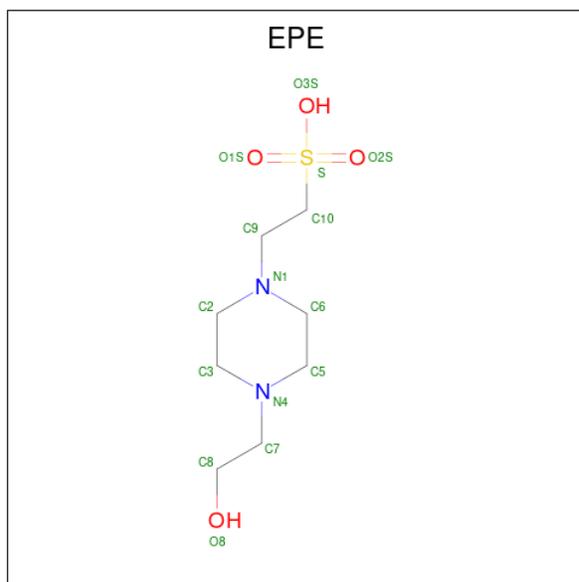
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	S	0	0
			15	8	2	4	1		
5	B	1	Total	C	N	O	S	0	0
			15	8	2	4	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	448	Total	O	0	0
			448	448		
6	B	169	Total	O	0	0
			169	169		

3 Residue-property plots ⓘ

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 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: Malate synthase G

Chain A:

• Molecule 1: Malate synthase G

Chain B:

N600 G607 Y608 V609 V610 R611 W612 G618 C619 S620 K621 V622 P623 D624 I625 H626 D627 V628 A629 L630 M631 R634 R638 A645 N646 W647 L648 D657 V658 R659 A660 S661 R664 M665 A666 P667 L668 V669 D670 R671 A674 G675 D676 V677 A678 Y679 R680 P681 M682 A683 P684 N685

F686 D687 D688 S689 I690 A691 N707 T710 E711 H715 R716 R717 R718 R719 R724 A725 A726 GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	120.38Å 120.38Å 238.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.09 – 2.30 49.08 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.4 (49.09-2.30) 99.4 (49.08-2.30)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.97 (at 2.29Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.178 , 0.244 0.181 , 0.191	Depositor DCC
R_{free} test set	7746 reflections (9.91%)	wwPDB-VP
Wilson B-factor (Å^2)	36.4	Xtriage
Anisotropy	0.020	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 44.1	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	11633	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.39% 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: MLT, EPE, MG, COA

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	1.17	8/5591 (0.1%)	1.07	18/7605 (0.2%)
1	B	0.98	8/5532 (0.1%)	0.96	12/7530 (0.2%)
All	All	1.08	16/11123 (0.1%)	1.02	30/15135 (0.2%)

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	A	0	2
1	B	0	4
All	All	0	6

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	14	ARG	CZ-NH2	14.31	1.51	1.33
1	B	686	PHE	CG-CD1	6.75	1.48	1.38
1	B	14	ARG	NE-CZ	6.36	1.41	1.33
1	B	686	PHE	CE2-CZ	6.33	1.49	1.37
1	A	478	VAL	CB-CG1	6.24	1.66	1.52

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	638	ARG	NE-CZ-NH1	14.83	127.71	120.30
1	A	638	ARG	NE-CZ-NH2	-12.60	114.00	120.30
1	B	14	ARG	NE-CZ-NH1	-12.57	114.02	120.30
1	A	312	ARG	NE-CZ-NH2	-10.37	115.12	120.30

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	638	ARG	NE-CZ-NH1	9.59	125.09	120.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	2	THR	Peptide
1	A	206	ASP	Peptide
1	B	183	SER	Peptide
1	B	586	GLU	Peptide
1	B	674	ALA	Peptide

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	5487	0	5407	78	0
1	B	5429	0	5345	100	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	9	0	3	0	0
3	B	9	0	3	0	0
4	A	48	0	31	20	0
5	A	15	0	17	1	0
5	B	15	0	17	1	0
6	A	448	0	0	1	0
6	B	169	0	0	2	0
All	All	11633	0	10823	175	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 8.

The worst 5 of 175 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:206:ASP:HB2	1:A:207:LYS:HB2	1.33	1.11

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:11:ARG:HG2	1:B:11:ARG:HH21	1.13	1.06
1:A:631:MET:SD	4:A:800:COA:H62	2.04	0.97
1:B:204:LEU:HB3	1:B:205:PRO:HD2	1.55	0.88
1:A:206:ASP:HB2	1:A:207:LYS:CB	2.04	0.86

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	714/729 (98%)	701 (98%)	12 (2%)	1 (0%)	51	64
1	B	708/729 (97%)	670 (95%)	31 (4%)	7 (1%)	15	17
All	All	1422/1458 (98%)	1371 (96%)	43 (3%)	8 (1%)	25	31

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	207	LYS
1	B	586	GLU
1	B	588	ALA
1	B	207	LYS
1	B	29	ASP

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	575/586 (98%)	545 (95%)	30 (5%)	23	32
1	B	567/586 (97%)	521 (92%)	46 (8%)	11	15
All	All	1142/1172 (97%)	1066 (93%)	76 (7%)	16	21

5 of 76 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	380	ASP
1	B	659	ARG
1	B	390	THR
1	B	576	GLU
1	B	724	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	21	ASN
1	B	340	ASN
1	B	673	ASN
1	B	532	GLN
1	B	672	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 ⓘ

Of 9 ligands modelled in this entry, 4 are monoatomic - leaving 5 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	EPE	A	5000	-	15,15,15	1.34	2 (13%)	18,20,20	2.27	8 (44%)
4	COA	A	800	-	41,50,50	1.82	2 (4%)	52,75,75	2.18	14 (26%)
3	MLT	A	900	2	8,8,8	1.34	1 (12%)	10,10,10	2.26	4 (40%)
5	EPE	B	6000	-	15,15,15	0.76	1 (6%)	18,20,20	2.47	7 (38%)
3	MLT	B	901	2	8,8,8	1.31	0	10,10,10	1.55	2 (20%)

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	EPE	A	5000	-	-	2/9/19/19	0/1/1/1
4	COA	A	800	-	-	16/44/64/64	0/3/3/3
3	MLT	A	900	2	1/1/3/3	2/8/8/8	-
5	EPE	B	6000	-	-	3/9/19/19	0/1/1/1
3	MLT	B	901	2	1/1/3/3	2/8/8/8	-

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	800	COA	O9P-C9P	9.89	1.43	1.23
5	A	5000	EPE	C10-S	3.55	1.82	1.77
4	A	800	COA	C2A-N3A	3.38	1.37	1.32
5	A	5000	EPE	O1S-S	2.61	1.52	1.45
3	A	900	MLT	O2-C1	-2.55	1.22	1.30

The worst 5 of 35 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	800	COA	C7P-N8P-C9P	8.53	137.81	122.59
5	B	6000	EPE	O1S-S-C10	7.11	115.48	106.92

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	800	COA	N3A-C2A-N1A	-5.50	120.08	128.68
4	A	800	COA	C6P-C7P-N8P	4.20	120.38	111.90
5	A	5000	EPE	C7-N4-C3	4.18	121.92	111.23

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	900	MLT	C2
3	B	901	MLT	C2

5 of 25 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	900	MLT	C1-C2-C3-C4
3	A	900	MLT	O3-C2-C3-C4
3	B	901	MLT	C1-C2-C3-C4
3	B	901	MLT	O3-C2-C3-C4
4	A	800	COA	P1A-O3A-P2A-O6A

There are no ring outliers.

3 monomers are involved in 22 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	5000	EPE	1	0
4	A	800	COA	20	0
5	B	6000	EPE	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand COA A 800

Bond lengths

Bond angles

Torsions

Rings
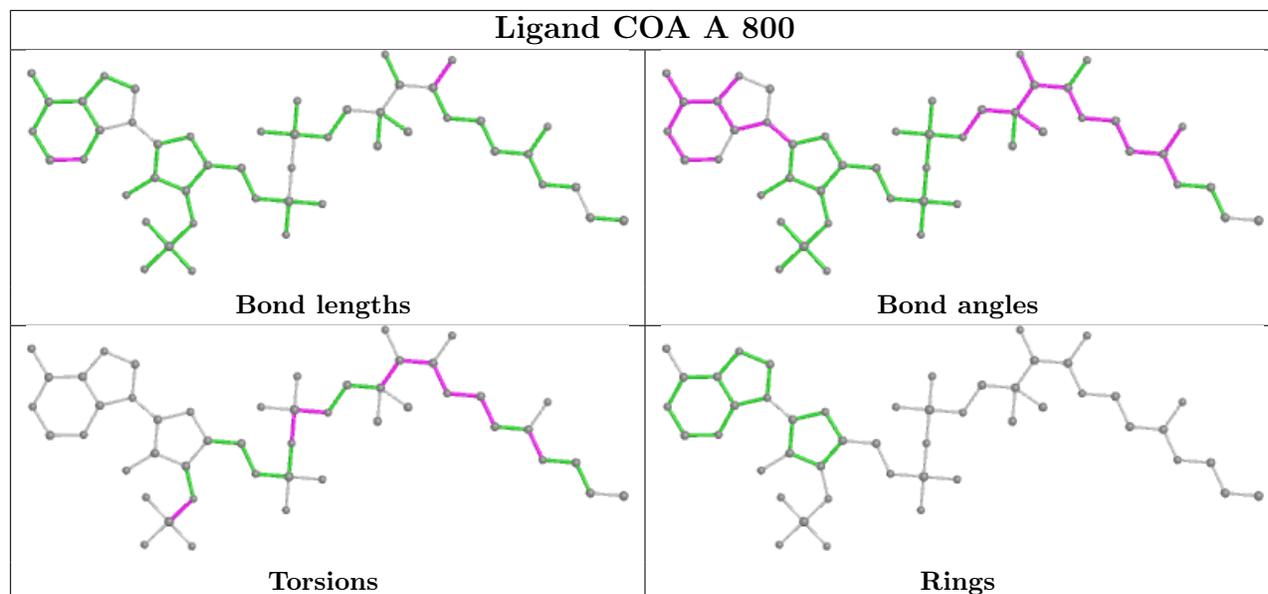
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 95th 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(Å²)	Q<0.9
1	A	720/729 (98%)	0.19	18 (2%)	57	64	31, 38, 52, 75	1 (0%)
1	B	714/729 (97%)	0.57	80 (11%)	5	7	28, 39, 50, 64	0
All	All	1434/1458 (98%)	0.38	98 (6%)	17	22	28, 38, 50, 75	1 (0%)

The worst 5 of 98 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	205	PRO	6.4
1	B	247	SER	5.1
1	B	204	LEU	4.9
1	B	251	THR	4.6
1	B	245	PRO	4.4

6.2 Non-standard residues in protein, DNA, RNA chains (i)

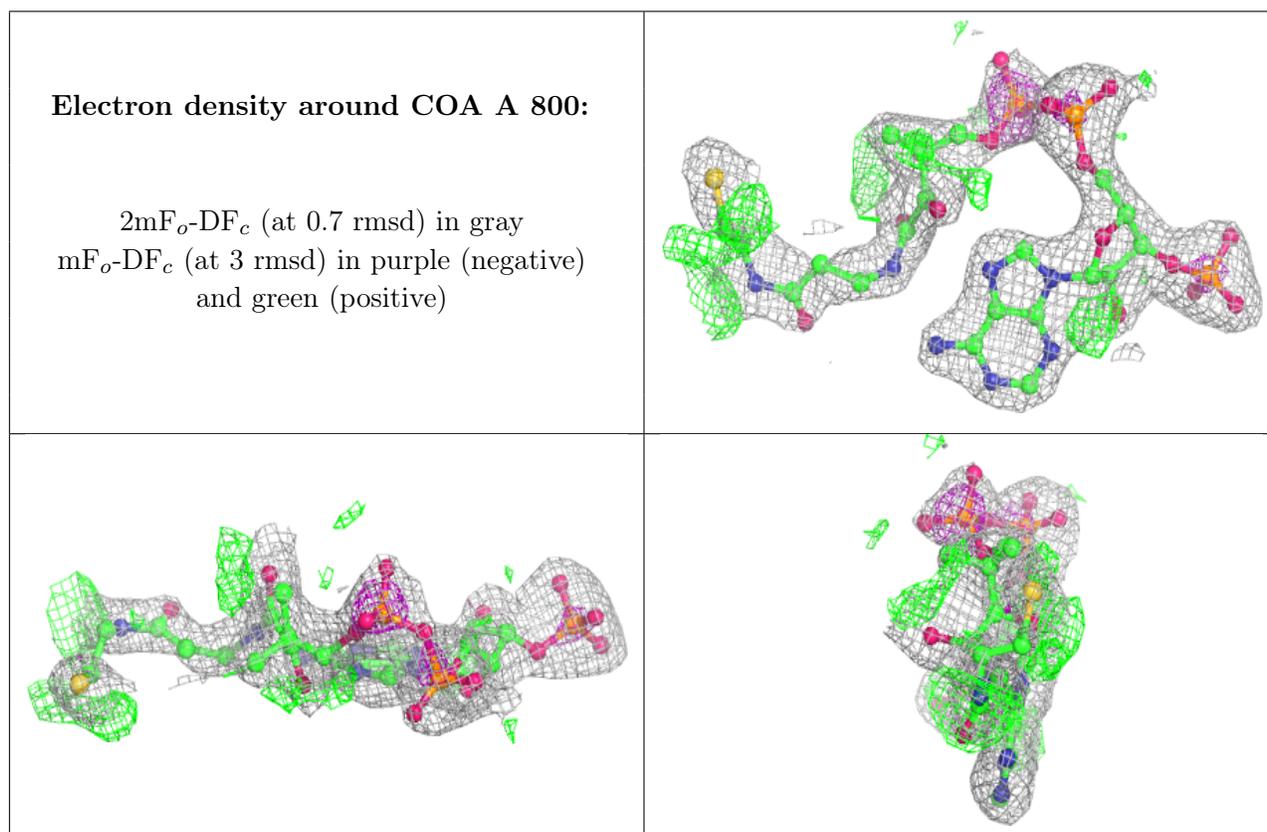There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no monosaccharides 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. The B-factors column lists the minimum, median, 95th 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	B-factors(Å²)	Q<0.9
4	COA	A	800	48/48	0.90	0.20	23,41,52,56	48
2	MG	B	1003	1/1	0.93	0.29	76,76,76,76	0
5	EPE	A	5000	15/15	0.93	0.13	52,55,61,62	0
5	EPE	B	6000	15/15	0.94	0.24	80,93,97,97	0
3	MLT	B	901	9/9	0.95	0.08	29,37,49,53	0
2	MG	A	1002	1/1	0.95	0.24	40,40,40,40	0
2	MG	B	1001	1/1	0.97	0.04	37,37,37,37	0
3	MLT	A	900	9/9	0.97	0.12	19,23,43,49	0
2	MG	A	1000	1/1	0.98	0.08	24,24,24,24	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around COA A 800:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.