WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Mar 18, 2024 – 06:04 PM EDT

PDB ID	:	2AAL
Title	:	Crystal Structures of the Wild-type, Mutant-P1A and Inactivated Malonate Semialdehyde Decarboxylase: A Structural Basis for the Decarboxylase and Hydratase Activities
Authors	:	Almrud, J.J.; Poelarends, G.J.; Johnson Jr., W.H.; Serrano, H.; Hackert, M.L.; Whitman, C.P.
Deposited on	:	2005-07-13
Resolution	:	1.65 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	:	2.36
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)	:	2.36

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 1.65 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.
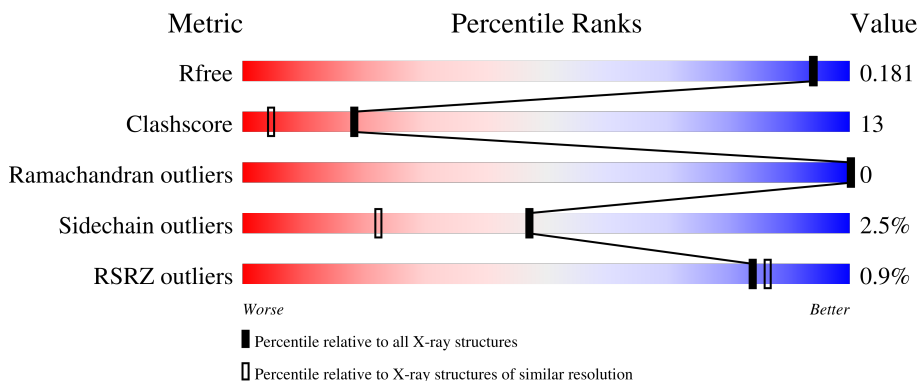
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1827 (1.66-1.66)
Clashscore	141614	1931 (1.66-1.66)
Ramachandran outliers	138981	1891 (1.66-1.66)
Sidechain outliers	138945	1891 (1.66-1.66)
RSRZ outliers	127900	1791 (1.66-1.66)

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 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	130	72% 24%
1	B	130	74% 21%
1	C	130	77% 15% 6%
1	D	130	78% 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	E	130	73% 22%
1	F	130	85% 10%

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
2	MLI	E	201	-	X	-	-
2	MLI	F	201	-	X	-	-

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 7105 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 Malonate semialdehyde decarboxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	129	Total	C	N	O	S	0	5	0
			1034	648	182	200	4			
1	B	130	Total	C	N	O	S	0	1	0
			1003	631	173	195	4			
1	C	129	Total	C	N	O	S	0	5	0
			1029	647	177	201	4			
1	D	129	Total	C	N	O	S	0	6	0
			1035	651	177	202	5			
1	E	129	Total	C	N	O	S	0	3	0
			1014	640	174	196	4			
1	F	129	Total	C	N	O	S	0	1	0
			998	628	172	193	5			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	7	LEU	ILE	conflict	UNP Q9EV83
A	130	GLY	-	expression tag	UNP Q9EV83
B	7	LEU	ILE	conflict	UNP Q9EV83
B	130	GLY	-	expression tag	UNP Q9EV83
C	7	LEU	ILE	conflict	UNP Q9EV83
C	130	GLY	-	expression tag	UNP Q9EV83
D	7	LEU	ILE	conflict	UNP Q9EV83
D	130	GLY	-	expression tag	UNP Q9EV83
E	7	LEU	ILE	conflict	UNP Q9EV83
E	130	GLY	-	expression tag	UNP Q9EV83
F	7	LEU	ILE	conflict	UNP Q9EV83
F	130	GLY	-	expression tag	UNP Q9EV83

- Molecule 2 is MALONATE ION (three-letter code: MLI) (formula: $C_3H_2O_4$).

MLI

O7

O⁻

C1   C2

C3

O9 O⁻

O6

O8

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 6	C 3	O 3	0	0
2	B	1	Total 6	C 3	O 3	0	0
2	C	1	Total 6	C 3	O 3	0	0
2	D	1	Total 6	C 3	O 3	0	0
2	E	1	Total 6	C 3	O 3	0	0
2	F	1	Total 6	C 3	O 3	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	148	Total 148	O 148	0	0
3	B	132	Total 132	O 132	0	0
3	C	161	Total 161	O 161	0	0
3	D	162	Total 162	O 162	0	0
3	E	174	Total 174	O 174	0	0
3	F	179	Total 179	O 179	0	0

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: Malonate semialdehyde decarboxylase

- Molecule 1: Malonate semialdehyde decarboxylase

- Molecule 1: Malonate semialdehyde decarboxylase
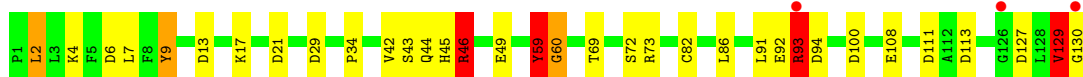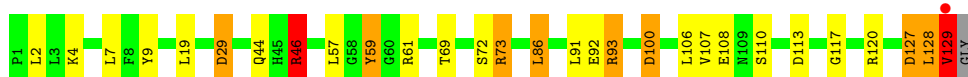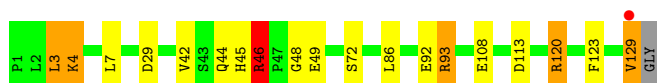

- Molecule 1: Malonate semialdehyde decarboxylase

- Molecule 1: Malonate semialdehyde decarboxylase

- Molecule 1: Malonate semialdehyde decarboxylase

Chain F:

4 Data and refinement statistics   ⓘ

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	51.89Å 51.89Å 219.08Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	72.55 – 1.65 73.03 – 1.60	Depositor EDS
% Data completeness (in resolution range)	99.0 (72.55-1.65) 98.8 (73.03-1.60)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	0.13	Depositor
$< I/\sigma(I) >$ [1]	2.50 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.184 , 0.239 0.188 , 0.181	Depositor DCC
R_{free} test set	4294 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	9.1	Xtriage
Anisotropy	0.027	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 17.5	EDS
L-test for twinning[2]	$< |L| > = 0.41, < L^2 > = 0.24$	Xtriage
Estimated twinning fraction	0.069 for -h,-k,l 0.237 for h,-h-k,-l 0.074 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7105	wwPDB-VP
Average B, all atoms (Å^2)	11.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.87% 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: MLI

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	1.32	6/1052 (0.6%)	1.47	16/1423 (1.1%)
1	B	1.42	13/1021 (1.3%)	1.58	19/1382 (1.4%)
1	C	1.52	10/1047 (1.0%)	1.57	18/1417 (1.3%)
1	D	1.56	13/1054 (1.2%)	1.34	19/1426 (1.3%)
1	E	1.64	14/1032 (1.4%)	1.72	23/1397 (1.6%)
1	F	1.33	8/1016 (0.8%)	1.58	11/1375 (0.8%)
All	All	1.47	64/6222 (1.0%)	1.55	106/8420 (1.3%)

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	2
1	C	0	3
1	F	0	1
All	All	0	6

The worst 5 of 64 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	39	TYR	CE1-CZ	-16.42	1.17	1.38
1	D	59	TYR	CE1-CZ	-15.58	1.18	1.38
1	C	129	VAL	CA-CB	-14.02	1.25	1.54
1	C	129	VAL	CB-CG1	-13.67	1.24	1.52
1	D	59	TYR	CG-CD2	-12.92	1.22	1.39

The worst 5 of 106 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	61	ARG	NE-CZ-NH1	33.53	137.07	120.30
1	F	3	LEU	O-C-N	-31.10	72.95	122.70
1	C	129	VAL	CB-CA-C	-24.75	64.37	111.40
1	B	93	ARG	NE-CZ-NH2	-21.31	109.65	120.30
1	B	46	ARG	NE-CZ-NH1	-19.91	110.35	120.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	59	TYR	Peptide
1	B	93	ARG	Sidechain
1	C	128	LEU	Mainchain,Peptide
1	C	59	TYR	Peptide
1	F	3	LEU	Mainchain

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	1034	0	1015	42	2
1	B	1003	0	984	43	3
1	C	1029	0	1008	23	2
1	D	1035	0	1013	28	0
1	E	1014	0	1000	17	0
1	F	998	0	979	18	0
2	A	6	0	2	0	0
2	B	6	0	2	0	0
2	C	6	0	2	0	0
2	D	6	0	2	1	0
2	E	6	0	2	0	0
2	F	6	0	2	0	0
3	A	148	0	0	16	3
3	B	132	0	0	15	2
3	C	161	0	0	7	6
3	D	162	0	0	10	3
3	E	174	0	0	8	4
3	F	179	0	0	7	3

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	7105	0	6011	162	14

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 13.

The worst 5 of 162 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:46:ARG:HH11	1:B:46:ARG:CG	1.16	1.35
1:E:4:LYS:HE2	3:E:363:HOH:O	1.25	1.31
1:A:46:ARG:NH2	1:B:9:TYR:OH	1.73	1.21
1:D:78:GLU:HG2	3:D:384:HOH:O	1.33	1.19
1:F:93:ARG:HD2	3:F:400:HOH:O	1.49	1.11

The worst 5 of 14 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:B:304:HOH:O	3:D:460:HOH:O[1_444]	1.18	1.02
1:B:93:ARG:NH2	1:C:93:ARG:CB[1_455]	1.67	0.53
1:A:93:ARG:O	1:B:93:ARG:NH1[1_665]	1.86	0.34
3:B:423:HOH:O	3:C:442:HOH:O[1_455]	1.89	0.31
1:C:46:ARG:NH2	3:D:385:HOH:O[1_544]	1.95	0.25

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	132/130 (102%)	128 (97%)	4 (3%)	0	100	100
1	B	129/130 (99%)	126 (98%)	3 (2%)	0	100	100

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	132/130 (102%)	128 (97%)	4 (3%)	0	100	100
1	D	133/130 (102%)	130 (98%)	3 (2%)	0	100	100
1	E	130/130 (100%)	125 (96%)	5 (4%)	0	100	100
1	F	128/130 (98%)	125 (98%)	3 (2%)	0	100	100
All	All	784/780 (100%)	762 (97%)	22 (3%)	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	111/106 (105%)	108 (97%)	3 (3%)	44	19
1	B	107/106 (101%)	105 (98%)	2 (2%)	57	34
1	C	111/106 (105%)	109 (98%)	2 (2%)	59	36
1	D	112/106 (106%)	109 (97%)	3 (3%)	44	19
1	E	109/106 (103%)	104 (95%)	5 (5%)	27	6
1	F	107/106 (101%)	105 (98%)	2 (2%)	57	34
All	All	657/636 (103%)	640 (97%)	17 (3%)	47	21

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	106	LEU
1	F	46	ARG
1	D	1	PRO
1	D	7	LEU
1	D	86	LEU

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

Mol	Chain	Res	Type
1	A	24	HIS
1	A	45	HIS
1	D	24	HIS
1	D	45	HIS

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 ⓘ

6 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	MLI	D	201	1	5,5,6	1.89	1 (20%)	5,5,7	2.85	2 (40%)
2	MLI	B	201	1	5,5,6	1.62	1 (20%)	5,5,7	3.32	2 (40%)
2	MLI	A	201	1	5,5,6	1.38	1 (20%)	5,5,7	3.08	2 (40%)
2	MLI	F	201	1	5,5,6	1.88	2 (40%)	5,5,7	3.14	2 (40%)
2	MLI	E	201	1	5,5,6	2.12	2 (40%)	5,5,7	3.85	2 (40%)
2	MLI	C	201	1	5,5,6	1.63	1 (20%)	5,5,7	3.46	2 (40%)

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	MLI	D	201	1	-	2/2/3/4	-
2	MLI	B	201	1	-	2/2/3/4	-
2	MLI	A	201	1	-	2/2/3/4	-
2	MLI	F	201	1	-	2/2/3/4	-
2	MLI	E	201	1	-	2/2/3/4	-
2	MLI	C	201	1	-	2/2/3/4	-

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	201	MLI	C1-C3	3.96	1.57	1.50
2	D	201	MLI	C1-C3	3.24	1.56	1.50
2	C	201	MLI	C1-C3	3.20	1.56	1.50
2	B	201	MLI	C1-C3	2.86	1.55	1.50
2	F	201	MLI	C1-C3	2.82	1.55	1.50

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	201	MLI	C2-C1-C3	-7.83	98.36	113.70
2	C	201	MLI	C2-C1-C3	-6.62	100.73	113.70
2	B	201	MLI	C2-C1-C3	-6.40	101.16	113.70
2	A	201	MLI	C2-C1-C3	-6.28	101.40	113.70
2	F	201	MLI	C2-C1-C3	-6.18	101.58	113.70

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	201	MLI	C2-C1-C3-O9
2	B	201	MLI	C2-C1-C3-O9
2	C	201	MLI	C2-C1-C3-O9
2	D	201	MLI	C2-C1-C3-O9
2	E	201	MLI	C2-C1-C3-O9

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	201	MLI	1	0

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	129/130 (99%)	-0.50	1 (0%)	86	88	6, 9, 15, 19	0
1	B	130/130 (100%)	-0.42	3 (2%)	60	61	6, 9, 15, 25	0
1	C	129/130 (99%)	-0.46	1 (0%)	86	88	6, 9, 15, 19	0
1	D	129/130 (99%)	-0.51	0	100	100	6, 9, 15, 19	0
1	E	129/130 (99%)	-0.50	1 (0%)	86	88	6, 9, 15, 19	0
1	F	129/130 (99%)	-0.53	1 (0%)	86	88	6, 9, 15, 19	0
All	All	775/780 (99%)	-0.49	7 (0%)	84	86	6, 9, 15, 25	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	129	VAL	5.2
1	B	130	GLY	3.9
1	F	129	VAL	2.7
1	B	93	ARG	2.5
1	B	126	GLY	2.5

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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($Å^2$)	Q<0.9
2	MLI	B	201	6/7	0.93	0.10	11,12,13,15	0
2	MLI	D	201	6/7	0.95	0.09	11,13,13,15	0
2	MLI	C	201	6/7	0.96	0.06	11,12,14,15	0
2	MLI	E	201	6/7	0.96	0.07	11,13,13,15	0
2	MLI	A	201	6/7	0.97	0.07	11,12,14,15	0
2	MLI	F	201	6/7	0.98	0.08	11,12,13,15	0

6.5 Other polymers ⓘ

There are no such residues in this entry.