



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 09:26 PM GMT

PDB ID : 1P0Y  
Title : Crystal structure of the SET domain of LSM1 bound to MeLysine and AdoHcy  
Authors : Trievel, R.C.; Flynn, E.M.; Houtz, R.L.; Hurley, J.H.  
Deposited on : 2003-04-11  
Resolution : 2.55 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

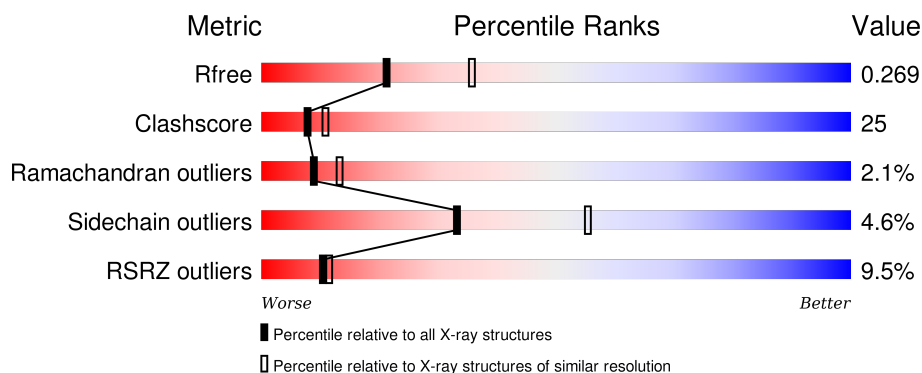
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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))
$R_{free}$	91344	4549 (2.58-2.50)
Clashscore	102246	5292 (2.58-2.50)
Ramachandran outliers	100387	5194 (2.58-2.50)
Sidechain outliers	100360	5196 (2.58-2.50)
RSRZ outliers	91569	4561 (2.58-2.50)

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	444	<div> <div>13%</div> <div>54%</div> <div>38%</div> <div>5%</div> <div>.</div> </div>
1	B	444	<div> <div>10%</div> <div>58%</div> <div>38%</div> <div>.</div> <div>..</div> </div>
1	C	444	<div> <div>5%</div> <div>60%</div> <div>34%</div> <div>.</div> <div>..</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11308 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 Ribulose-1,5 bisphosphate carboxylase/oxygenase large subunit N-methyltransferase, chloroplast.

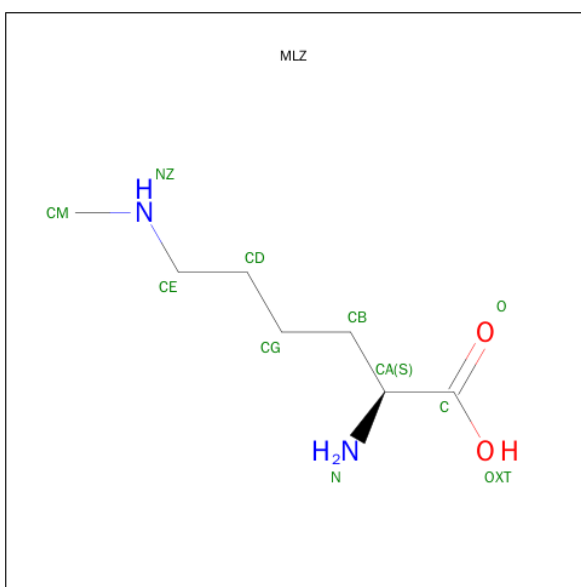
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	430	Total	C	N	O	S	0	0	0
			3464	2217	573	667	7			
1	B	441	Total	C	N	O	S	0	0	0
			3549	2275	586	681	7			
1	C	439	Total	C	N	O	S	0	0	0
			3532	2265	583	677	7			

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	45	MET	-	INITIATING MET	UNP Q43088
A	483	GLU	-	ENGINEERED	UNP Q43088
A	484	ASN	-	ENGINEERED	UNP Q43088
A	485	LEU	-	ENGINEERED	UNP Q43088
A	486	TYR	-	ENGINEERED	UNP Q43088
A	487	PHE	-	ENGINEERED	UNP Q43088
A	488	GLN	-	ENGINEERED	UNP Q43088
B	45	MET	-	INITIATING MET	UNP Q43088
B	483	GLU	-	ENGINEERED	UNP Q43088
B	484	ASN	-	ENGINEERED	UNP Q43088
B	485	LEU	-	ENGINEERED	UNP Q43088
B	486	TYR	-	ENGINEERED	UNP Q43088
B	487	PHE	-	ENGINEERED	UNP Q43088
B	488	GLN	-	ENGINEERED	UNP Q43088
C	45	MET	-	INITIATING MET	UNP Q43088
C	483	GLU	-	ENGINEERED	UNP Q43088
C	484	ASN	-	ENGINEERED	UNP Q43088
C	485	LEU	-	ENGINEERED	UNP Q43088
C	486	TYR	-	ENGINEERED	UNP Q43088
C	487	PHE	-	ENGINEERED	UNP Q43088
C	488	GLN	-	ENGINEERED	UNP Q43088

- SAH
- 
- The image displays the chemical structure of S-adenosylmethionine (SAH). It features a 5'-methylthioadenosine (MTA) moiety linked to a ribose sugar, which is in turn linked to an adenosine moiety. The MTA part consists of a methionine side chain (CH<sub>2</sub>-CH<sub>2</sub>-CH(CH<sub>3</sub>)-NH<sub>2</sub>) attached to a sulfur atom (S), which is bonded to a 5'-methylthioadenosine (MTA) group. The ribose sugar is a five-membered ring with hydroxyl groups at the 2' and 3' positions. The adenosine part consists of an adenine base (a purine ring system) attached to a ribose sugar. The structure is labeled with atom names and numbers, and the overall molecule is identified as SAH.

- Molecule 3 is N-METHYL-LYSINE (three-letter code: MLZ) (formula:  $\text{C}_7\text{H}_{16}\text{N}_2\text{O}_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	C	3	Total	C	N	O	0	0
			33	21	6	6		

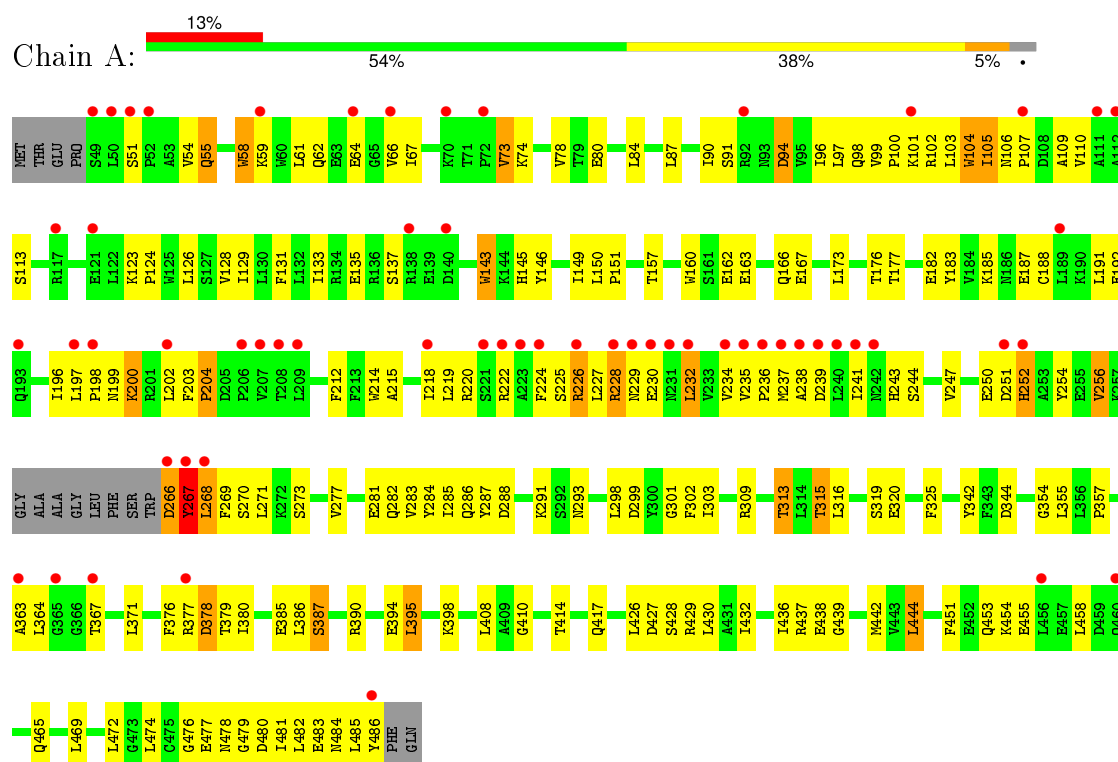
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	221	Total	O	0	0
			221	221		
4	B	210	Total	O	0	0
			210	210		
4	C	221	Total	O	0	0
			221	221		

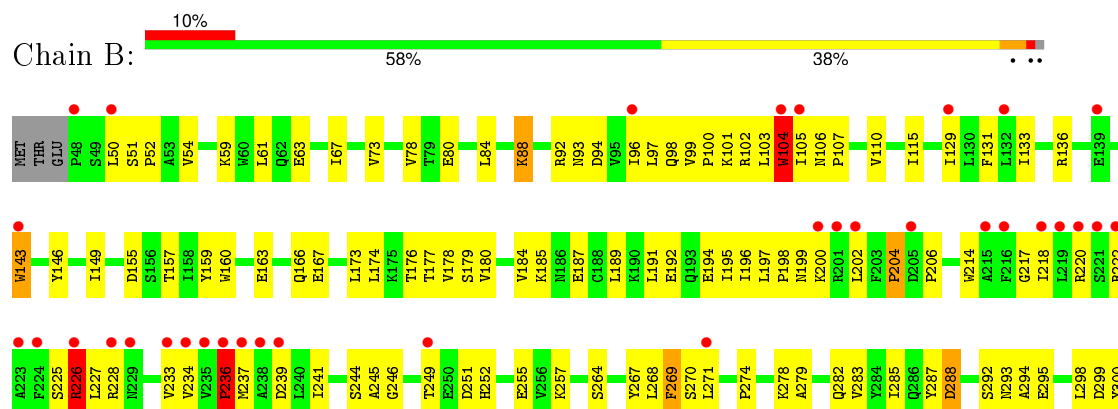
### 3 Residue-property plots

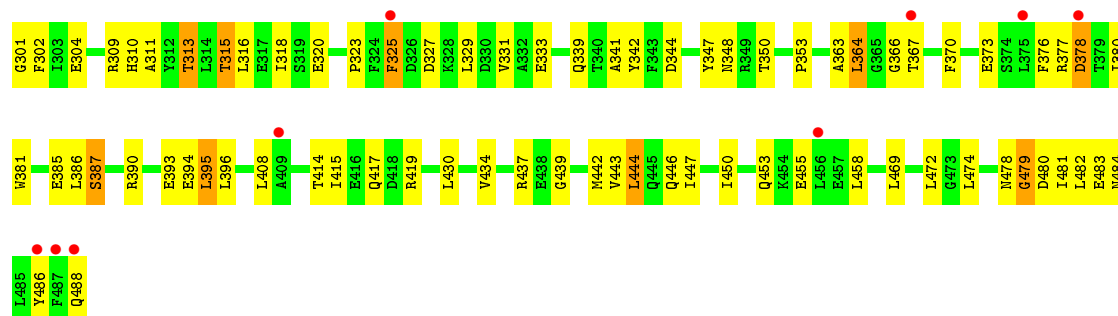
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Ribulose-1,5 biphosphate carboxylase/oxygenase large subunit N-methyltransferase, chloroplast

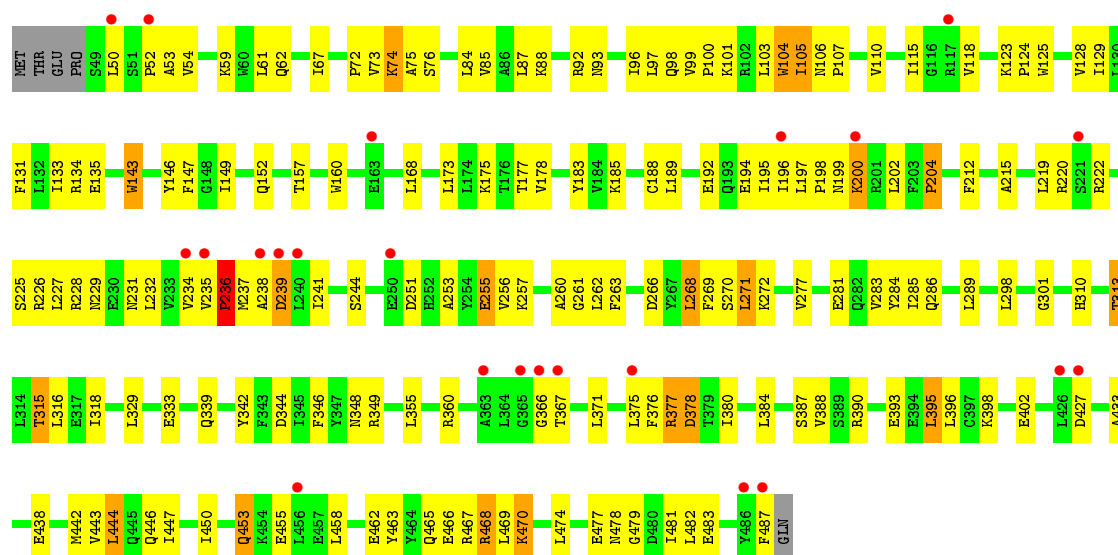


- Molecule 1: Ribulose-1,5 biphosphate carboxylase/oxygenase large subunit N-methyltransferase, chloroplast





- Molecule 1: Ribulose-1,5 biphosphate carboxylase/oxygenase large subunit N-methyltransferase, chloroplast



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	130.41Å 153.19Å 266.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.65 – 2.55 29.65 – 2.55	Depositor EDS
% Data completeness (in resolution range)	85.1 (29.65-2.55) 85.2 (29.65-2.55)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.04	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.82 (at 2.54Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.228 , 0.269 0.228 , 0.269	Depositor DCC
$R_{free}$ test set	3761 reflections (5.08%)	DCC
Wilson B-factor (Å <sup>2</sup> )	51.8	Xtriage
Anisotropy	0.825	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 50.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 83829 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11308	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.74% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: MLZ, SAH

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	0.38	0/3534	0.60	1/4793 (0.0%)
1	B	0.39	0/3625	0.61	2/4917 (0.0%)
1	C	0.38	0/3607	0.60	0/4894
All	All	0.38	0/10766	0.60	3/14604 (0.0%)

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	1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	267	TYR	N-CA-C	-5.37	96.51	111.00
1	B	269	PHE	N-CA-C	-5.24	96.84	111.00
1	B	104	TRP	CA-CB-CG	5.01	123.22	113.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	267	TYR	Sidechain

## 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	3464	0	3432	212	0
1	B	3549	0	3509	168	0
1	C	3532	0	3493	174	0
2	A	26	0	19	0	0
2	B	26	0	19	1	0
2	C	26	0	19	1	0
3	C	33	0	45	5	0
4	A	221	0	0	22	0
4	B	210	0	0	15	0
4	C	221	0	0	23	0
All	All	11308	0	10536	530	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 25.

The worst 5 of 530 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:395:LEU:HD22	1:C:469:LEU:HD12	1.42	1.01
1:A:97:LEU:HD22	1:A:237:MET:HG3	1.42	0.98
1:A:98:GLN:HB3	1:A:268:LEU:HG	1.48	0.95
1:B:367:THR:HA	4:B:803:HOH:O	1.65	0.95
1:C:74:LYS:HG2	1:C:87:LEU:HD21	1.51	0.92

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	426/444 (96%)	389 (91%)	27 (6%)	10 (2%)	8	11
1	B	439/444 (99%)	398 (91%)	30 (7%)	11 (2%)	7	10
1	C	437/444 (98%)	402 (92%)	29 (7%)	6 (1%)	14	23
All	All	1302/1332 (98%)	1189 (91%)	86 (7%)	27 (2%)	9	13

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	377	ARG
1	C	377	ARG
1	A	204	PRO
1	A	252	HIS
1	A	377	ARG

### 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	380/390 (97%)	360 (95%)	20 (5%)	28	48
1	B	387/390 (99%)	373 (96%)	14 (4%)	42	67
1	C	385/390 (99%)	366 (95%)	19 (5%)	31	52
All	All	1152/1170 (98%)	1099 (95%)	53 (5%)	33	55

5 of 53 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	104	TRP
1	B	378	ASP
1	C	444	LEU
1	B	143	TRP
1	B	236	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such

sidechains are listed below:

Mol	Chain	Res	Type
1	B	348	ASN
1	B	484	ASN
1	C	446	GLN
1	B	412	HIS
1	B	460	GLN

### 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](#)

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	SAH	A	800	-	20,28,28	0.88	0	19,40,40	0.93	0
2	SAH	B	801	-	20,28,28	1.00	1 (5%)	19,40,40	0.88	0
2	SAH	C	802	-	20,28,28	0.90	1 (5%)	19,40,40	0.85	0
3	MLZ	C	900	-	7,10,10	0.48	0	5,11,11	0.90	0
3	MLZ	C	901	-	7,10,10	0.52	0	5,11,11	0.44	0
3	MLZ	C	902	-	7,10,10	0.48	0	5,11,11	0.62	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	SAH	A	800	-	-	0/7/31/31	0/3/3/3
2	SAH	B	801	-	-	0/7/31/31	0/3/3/3
2	SAH	C	802	-	-	0/7/31/31	0/3/3/3
3	MLZ	C	900	-	-	0/6/10/10	0/0/0/0
3	MLZ	C	901	-	-	0/6/10/10	0/0/0/0
3	MLZ	C	902	-	-	0/6/10/10	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	802	SAH	C2-N3	2.21	1.36	1.32
2	B	801	SAH	C2-N3	2.45	1.36	1.32

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	801	SAH	1	0
2	C	802	SAH	1	0
3	C	900	MLZ	2	0
3	C	901	MLZ	2	0
3	C	902	MLZ	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	430/444 (96%)	0.66	59 (13%) 4 4	46, 70, 120, 130	0
1	B	441/444 (99%)	0.47	43 (9%) 10 11	41, 64, 101, 123	0
1	C	439/444 (98%)	0.33	23 (5%) 31 36	45, 66, 97, 113	0
All	All	1310/1332 (98%)	0.48	125 (9%) 10 11	41, 67, 108, 130	0

The worst 5 of 125 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	267	TYR	7.5
1	C	487	PHE	6.4
1	B	229	ASN	6.1
1	A	239	ASP	6.1
1	B	228	ARG	6.0

### 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
3	MLZ	C	902	11/11	0.90	0.32	1.21	46,50,54,55	0
3	MLZ	C	901	11/11	0.91	0.28	0.50	52,53,56,57	0
3	MLZ	C	900	11/11	0.80	0.38	0.37	74,75,79,80	0
2	SAH	C	802	26/26	0.94	0.22	0.24	53,61,64,64	0
2	SAH	A	800	26/26	0.92	0.28	0.23	66,69,76,77	0
2	SAH	B	801	26/26	0.95	0.20	-0.25	44,47,50,52	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.