wwPDB X-ray Structure Validation Summary Report (i)

Dec 15, 2016 – 04:29 PM EST

PDB ID	:	5LSU
Title	:	Structure of the Epigenetic Oncogene MMSET and inhibition by N-Alkyl Sinefungin Derivatives
Authors	:	Tisi, D.; Pathuri, P.; Heightman, T.
Deposited on	:	2016-09-05
Resolution	:	2.14 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
http://wwpdb.org/validation/2016/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:

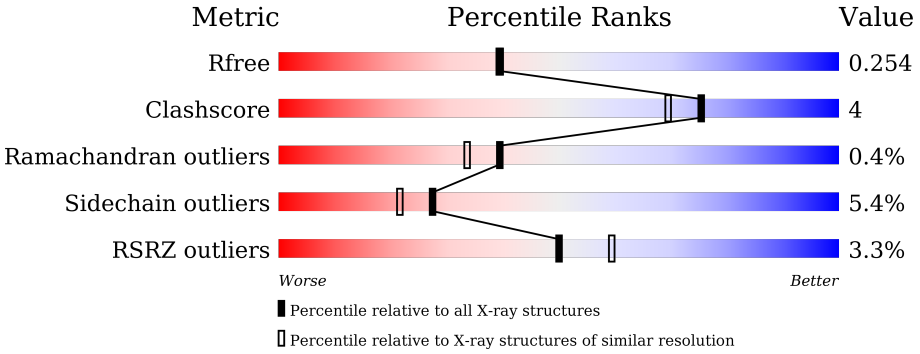
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

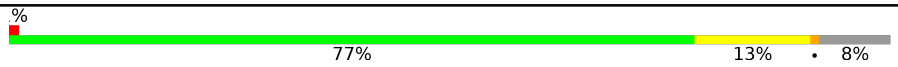
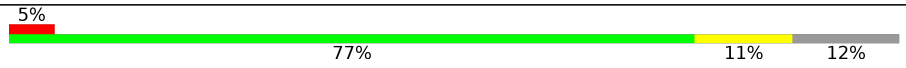
The reported resolution of this entry is 2.14 Å.

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	1693 (2.16-2.12)
Clashscore	102246	1824 (2.16-2.12)
Ramachandran outliers	100387	1798 (2.16-2.12)
Sidechain outliers	100360	1798 (2.16-2.12)
RSRZ outliers	91569	1699 (2.16-2.12)

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 >=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 <=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	252	
1	B	252	

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	ZN	B	1301	-	-	-	X

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4212 atoms, of which 46 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 Histone-lysine N-methyltransferase NSD2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	231	1825	1140	321	343	21	0	1	0
1	B	222	1753	1100	310	321	22	0	2	0

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	952	MET	-	initiating methionine	UNP O96028
A	953	GLY	-	expression tag	UNP O96028
A	954	SER	-	expression tag	UNP O96028
A	955	SER	-	expression tag	UNP O96028
A	956	HIS	-	expression tag	UNP O96028
A	957	HIS	-	expression tag	UNP O96028
A	958	HIS	-	expression tag	UNP O96028
A	959	HIS	-	expression tag	UNP O96028
A	960	HIS	-	expression tag	UNP O96028
A	961	HIS	-	expression tag	UNP O96028
A	962	SER	-	expression tag	UNP O96028
A	963	SER	-	expression tag	UNP O96028
A	964	GLY	-	expression tag	UNP O96028
A	965	LEU	-	expression tag	UNP O96028
A	966	VAL	-	expression tag	UNP O96028
A	967	PRO	-	expression tag	UNP O96028
A	968	ARG	-	expression tag	UNP O96028
A	969	GLY	-	expression tag	UNP O96028
A	970	SER	-	expression tag	UNP O96028
A	971	HIS	-	expression tag	UNP O96028
A	972	MET	-	expression tag	UNP O96028
A	975	LEU	GLN	engineered mutation	UNP O96028
A	978	LEU	ALA	engineered mutation	UNP O96028
A	1071	LEU	ASP	engineered mutation	UNP O96028
A	1072	GLN	GLY	engineered mutation	UNP O96028

Continued on next page...

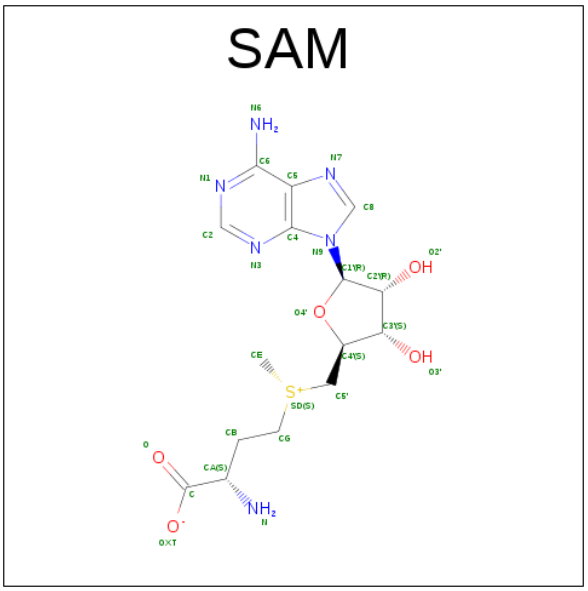
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1073	ARG	LYS	engineered mutation	UNP O96028
B	952	MET	-	initiating methionine	UNP O96028
B	953	GLY	-	expression tag	UNP O96028
B	954	SER	-	expression tag	UNP O96028
B	955	SER	-	expression tag	UNP O96028
B	956	HIS	-	expression tag	UNP O96028
B	957	HIS	-	expression tag	UNP O96028
B	958	HIS	-	expression tag	UNP O96028
B	959	HIS	-	expression tag	UNP O96028
B	960	HIS	-	expression tag	UNP O96028
B	961	HIS	-	expression tag	UNP O96028
B	962	SER	-	expression tag	UNP O96028
B	963	SER	-	expression tag	UNP O96028
B	964	GLY	-	expression tag	UNP O96028
B	965	LEU	-	expression tag	UNP O96028
B	966	VAL	-	expression tag	UNP O96028
B	967	PRO	-	expression tag	UNP O96028
B	968	ARG	-	expression tag	UNP O96028
B	969	GLY	-	expression tag	UNP O96028
B	970	SER	-	expression tag	UNP O96028
B	971	HIS	-	expression tag	UNP O96028
B	972	MET	-	expression tag	UNP O96028
B	975	LEU	GLN	engineered mutation	UNP O96028
B	978	LEU	ALA	engineered mutation	UNP O96028
B	1071	LEU	ASP	engineered mutation	UNP O96028
B	1072	GLN	GLY	engineered mutation	UNP O96028
B	1073	ARG	LYS	engineered mutation	UNP O96028

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
2	B	3	3	3	0	0
2	A	3	3	3	0	0

- Molecule 3 is S-ADENOSYLMETHIONINE (three-letter code: SAM) (formula: $C_{15}H_{22}N_6O_5S$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	H	N	O	S		
3	A	1	50	15	23	6	5	1	0	0
			Total	C	H	N	O	S		
3	B	1	50	15	23	6	5	1	0	0

- Molecule 4 is water.

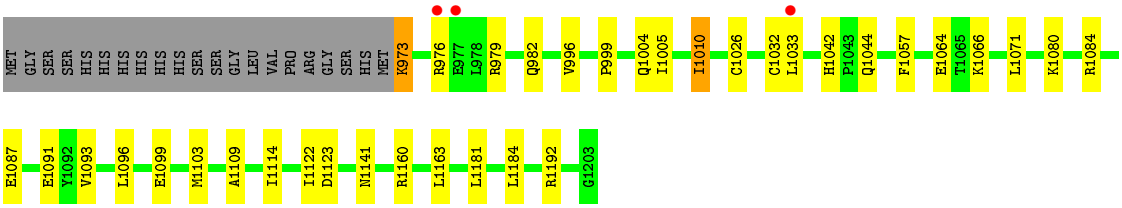
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	280	280	280	0	0
			Total	O		
4	B	248	248	248	0	0

3 Residue-property plots (i)

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.

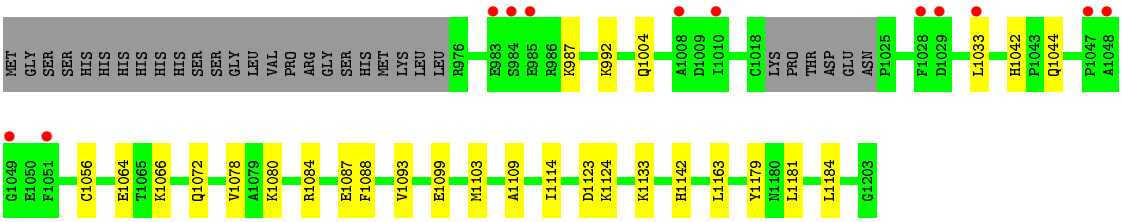
- Molecule 1: Histone-lysine N-methyltransferase NSD2

Chain A:



- Molecule 1: Histone-lysine N-methyltransferase NSD2

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	94.58Å 63.11Å 81.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.81 – 2.14 61.56 – 2.14	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.81-2.14) 99.9 (61.56-2.14)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.36 (at 2.14Å)	Xtriage
Refinement program	BUSTER 2.11.7	Depositor
R, R_{free}	0.171 , 0.238 0.184 , 0.254	Depositor DCC
R_{free} test set	1384 reflections (5.32%)	DCC
Wilson B-factor (Å^2)	35.0	Xtriage
Anisotropy	0.678	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 66.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4212	wwPDB-VP
Average B, all atoms (Å^2)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.52% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, SAM

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	0.60	0/1868	0.70	0/2521
1	B	0.61	0/1797	0.71	0/2421
All	All	0.61	0/3665	0.70	0/4942

There are no bond length outliers.

There are no bond angle outliers.

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	1825	0	1746	20	0
1	B	1753	0	1692	11	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
3	A	27	23	22	3	0
3	B	27	23	22	0	0
4	A	280	0	0	0	0
4	B	248	0	0	0	0
All	All	4166	46	3482	28	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 4.

The worst 5 of 28 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:999:PRO:HB3	1:A:1005[A]:ILE:HD11	1.65	0.79
1:A:1071:LEU:HD13	1:B:1124:LYS:HG2	1.74	0.69
1:A:1071:LEU:CD1	1:B:1124:LYS:HG2	2.25	0.67
1:A:1141:ASN:HD22	3:A:1304:SAM:C8	2.17	0.58
1:A:1042:HIS:HD2	1:A:1044:GLN:H	1.52	0.56

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	230/252 (91%)	222 (96%)	6 (3%)	2 (1%)	21 12
1	B	220/252 (87%)	211 (96%)	9 (4%)	0	100 100
All	All	450/504 (89%)	433 (96%)	15 (3%)	2 (0%)	39 33

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	982	GLN
1	A	1010	ILE

5.3.2 Protein sidechains (i)

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	200/224 (89%)	189 (94%)	11 (6%)	27	21
1	B	191/224 (85%)	181 (95%)	10 (5%)	29	23
All	All	391/448 (87%)	370 (95%)	21 (5%)	27	21

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1181	LEU
1	B	987	LYS
1	B	1123	ASP
1	A	1163	LEU
1	B	1163	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1141	ASN
1	B	1072	GLN
1	B	1042	HIS
1	A	1072	GLN
1	B	1004	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)

Of 8 ligands modelled in this entry, 6 are monoatomic - leaving 2 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	$\#\|Z\| > 2$
3	SAM	A	1304	-	23,29,29	0.78	1 (4%)	15,42,42	0.65	0
3	SAM	B	1304	-	23,29,29	0.60	0	15,42,42	0.79	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
3	SAM	A	1304	-	-	0/8/33/33	0/3/3/3
3	SAM	B	1304	-	-	0/8/33/33	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1304	SAM	O4'-C1'	2.50	1.44	1.41

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 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1304	SAM	3	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, 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(Å^2)	Q<0.9
1	A	231/252 (91%)	-0.23	3 (1%)	79	84	30, 48, 94, 128	0
1	B	222/252 (88%)	0.01	12 (5%)	29	38	25, 47, 105, 116	0
All	All	453/504 (89%)	-0.11	15 (3%)	50	59	25, 48, 104, 128	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	983	GLU	6.3
1	B	1008	ALA	6.3
1	B	1051	PHE	5.8
1	B	1029	ASP	3.8
1	B	1048	ALA	3.7

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, 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	LLDF	B-factors(Å^2)	Q<0.9
2	ZN	B	1301	1/1	1.00	0.11	2.27	32,32,32,32	0
2	ZN	A	1301	1/1	1.00	0.12	0.77	42,42,42,42	0
3	SAM	A	1304	27/27	0.95	0.10	-0.24	28,44,51,54	0
3	SAM	B	1304	27/27	0.98	0.09	-0.32	24,31,39,42	0
2	ZN	A	1302	1/1	0.99	0.06	-1.04	54,54,54,54	0
2	ZN	B	1303	1/1	0.91	0.05	-1.59	85,85,85,85	0
2	ZN	B	1302	1/1	0.98	0.07	-1.71	81,81,81,81	0
2	ZN	A	1303	1/1	0.99	0.05	-2.67	48,48,48,48	0

6.5 Other polymers (i)

There are no such residues in this entry.