



Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 05:33 PM GMT

PDB ID : 3SMT
Title : Crystal structure of human SET domain-containing protein3
Authors : Dong, A.; Zeng, H.; Walker, J.R.; Loppnau, P.; Bountra, C.; Weigelt, J.; Arrowsmith, C.H.; Edwards, A.M.; Min, J.; Wu, H.; Structural Genomics Consortium (SGC)
Deposited on : 2011-06-28
Resolution : 2.04 Å(reported)

This is a full wwPDB validation 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/ValidationPDFNotes.html>

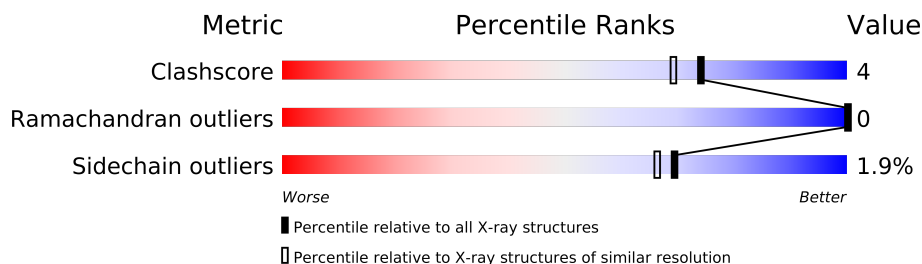
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : **FAILED**
Percentile statistics : 21963
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.04 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))
Clashscore	79885	1577 (2.06-2.02)
Ramachandran outliers	78287	1565 (2.06-2.02)
Sidechain outliers	78261	1565 (2.06-2.02)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	497	

2 Entry composition i

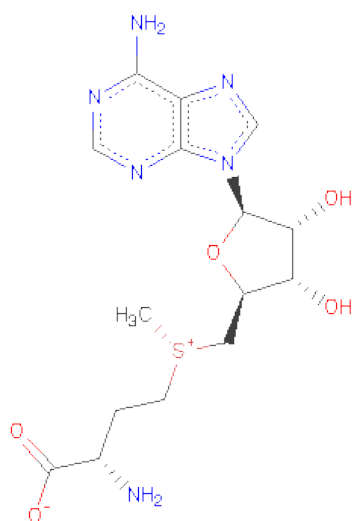
There are 7 unique types of molecules in this entry. The entry contains 4067 atoms, of which 0 are hydrogen and 0 are deuterium.

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 setd3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	465	3707	2367	613	710	7	10	0	8	0

- Molecule 2 is S-ADENOSYLMETHIONINE (three-letter code: SAM) (formula: C₁₅H₂₂N₆O₅S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	27	15	6	5	1	0	0

- Molecule 3 is ARSENIC (three-letter code: ARS) (formula: As).

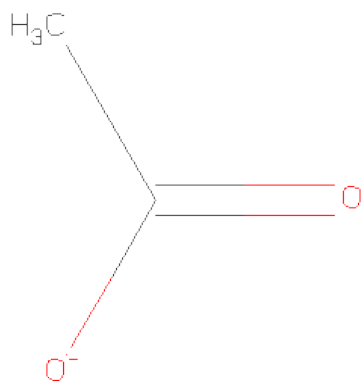
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	As	0	0
			2	2		

- Molecule 4 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: C₈H₁₈O₅).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			7	4	3		

- Molecule 5 is ACETATE ION (three-letter code: ACT) (formula: $\text{C}_2\text{H}_3\text{O}_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			4	2	2		
5	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 6 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	11	Total	X	0	0
			11	11		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	301	Total	O	0	4
			305	305		

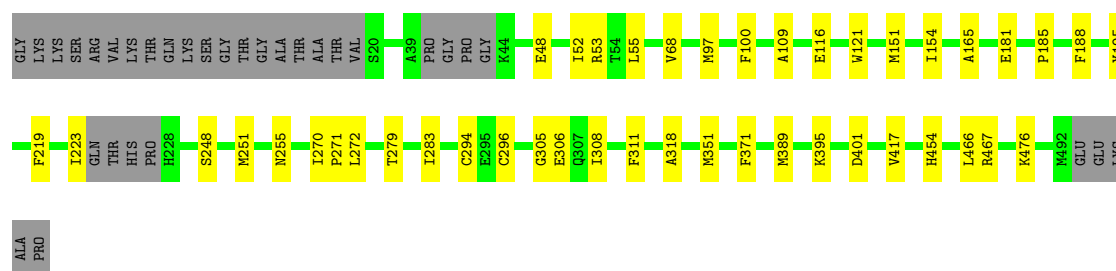
3 Residue-property plots

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.

Note EDS failed to run properly.

- Molecule 1: Histone-lysine N-methyltransferase setd3

Chain A: 



4 Data and refinement statistics i

EDS failed to run properly - this section will therefore be incomplete.

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	113.86Å 113.86Å 83.57Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.30 – 2.04	Depositor
% Data completeness (in resolution range)	(Not available) (49.30-2.04)	Depositor
R_{merge}	0.06	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.75 (at 2.03Å)	Xtriage
Refinement program	BUSTER 2.8.0	Depositor
R, R_{free}	0.208 , 0.259	Depositor
Wilson B-factor (Å ²)	36.9	Xtriage
Anisotropy	0.128	Xtriage
Estimated twinning fraction	0.022 for -h,-k,l	Xtriage
L-test for twinning	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtriage
Outliers	0 of 40151 reflections	Xtriage
Total number of atoms	4067	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.75% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: UNX, PG4, ARS, SAM, ACT

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	0.51	0/3780	0.63	0/5117

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3707	0	3458	30	0
2	A	27	0	22	0	0
3	A	2	0	0	6	0
4	A	7	0	9	4	0
5	A	8	0	6	0	0
6	A	11	0	0	0	0
7	A	305	0	0	1	0
All	All	4067	0	3495	30	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 4.

All (30) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:296:CYS:SG	3:A:498:ARS:AS	2.53	1.27
1:A:294[A]:CYS:SG	3:A:499:ARS:AS	2.73	1.07
1:A:294[B]:CYS:HB3	3:A:499:ARS:AS	2.45	0.77
1:A:283:ILE:HD11	1:A:311:PHE:H	1.60	0.67
1:A:306:GLU:HA	4:A:4001:PG4:H71	1.75	0.66
1:A:389:MSE:HE2	1:A:417:VAL:HG21	1.79	0.65
1:A:351:MSE:HE1	1:A:395:LYS:HG3	1.79	0.64
1:A:255:ASN:HB2	1:A:270:ILE:CD1	2.28	0.63
1:A:151:MSE:SE	1:A:154:ILE:HD13	2.49	0.62
1:A:68:VAL:HG21	1:A:165:ALA:HA	1.91	0.53
1:A:116:GLU:HG2	7:A:788:HOH:O	2.10	0.52
1:A:296:CYS:CB	3:A:498:ARS:AS	3.19	0.50
1:A:305:GLY:O	4:A:4001:PG4:H61	2.12	0.49
1:A:109:ALA:HB2	1:A:308:ILE:HD11	1.93	0.49
1:A:195:TYR:O	1:A:467:ARG:HD3	2.13	0.47
1:A:255:ASN:HB2	1:A:270:ILE:HD11	1.97	0.45
1:A:219:PHE:O	1:A:223:ILE:HG12	2.17	0.45
1:A:351:MSE:HE1	1:A:395:LYS:CG	2.44	0.45
1:A:255:ASN:HB2	1:A:270:ILE:HD13	1.96	0.44
1:A:248:SER:HA	1:A:251:MSE:HG2	1.99	0.44
1:A:296:CYS:HB3	3:A:498:ARS:AS	2.78	0.43
1:A:271:PRO:O	1:A:272:LEU:HB2	2.18	0.43
1:A:100:PHE:HE1	4:A:4001:PG4:H52	1.83	0.42
1:A:294[B]:CYS:CB	3:A:499:ARS:AS	3.24	0.42
1:A:371:PHE:O	1:A:476:LYS:HD3	2.19	0.42
1:A:121:TRP:HA	1:A:294[A]:CYS:O	2.20	0.42
1:A:100:PHE:CE1	4:A:4001:PG4:H52	2.55	0.42
1:A:318[B]:ALA:HA	1:A:466:LEU:HD11	2.00	0.41
1:A:185:PRO:HA	1:A:188:PHE:CD1	2.56	0.41
1:A:48:GLU:O	1:A:52:ILE:HG22	2.21	0.40

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	467/497 (94%)	457 (98%)	10 (2%)	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	377/427 (88%)	369 (98%)	8 (2%)	66	62

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	53	ARG
1	A	55	LEU
1	A	97	MSE
1	A	181[A]	GLU
1	A	181[B]	GLU
1	A	279	THR
1	A	401	ASP
1	A	454	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	333	HIS

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 17 ligands modelled in this entry, 11 are unknown and 2 are monoatomic - leaving 4 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
2	SAM	A	1000	-	26,29,29	1.35	5 (19%)	38,42,42	2.17	11 (28%)
5	ACT	A	1001	-	1,3,3	0.17	0	0,3,3	0.00	-
4	PG4	A	4001	-	6,6,12	1.02	0	5,5,11	0.57	0
5	ACT	A	500	-	1,3,3	2.19	1 (100%)	0,3,3	0.00	-

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	SAM	A	1000	-	-	0/13/33/33	0/1/3/3
5	ACT	A	1001	-	-	0/0/0/0	0/0/0/0
4	PG4	A	4001	-	-	0/4/4/10	0/0/0/0
5	ACT	A	500	-	-	0/0/0/0	0/0/0/0

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1000	SAM	C2'-C1'	3.72	1.58	1.53
2	A	1000	SAM	O2'-C2'	2.53	1.49	1.43
2	A	1000	SAM	C8-N9	2.41	1.40	1.36
2	A	1000	SAM	CG-CB	2.21	1.55	1.52
5	A	500	ACT	CH3-C	2.19	1.51	1.48
2	A	1000	SAM	C8-N7	-2.13	1.30	1.34

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1000	SAM	N3-C2-N1	-8.60	121.52	128.71
2	A	1000	SAM	N3-C4-N9	3.48	131.72	125.43
2	A	1000	SAM	C4'-O4'-C1'	3.11	113.13	109.75
2	A	1000	SAM	C5'-C4'-C3'	-3.10	109.05	116.32
2	A	1000	SAM	CB-CA-C	-2.99	106.92	111.44
2	A	1000	SAM	C2'-C1'-N9	2.79	120.43	113.27
2	A	1000	SAM	C2-N3-C4	2.31	120.58	114.01
2	A	1000	SAM	C4-C5-N7	-2.29	107.56	109.52
2	A	1000	SAM	C1'-N9-C4	-2.28	122.70	126.64
2	A	1000	SAM	N7-C8-N9	-2.21	108.09	114.36
2	A	1000	SAM	C8-N7-C5	2.06	109.96	103.58

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 ⓘ

EDS failed to run properly - this section will therefore be empty.

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

EDS failed to run properly - this section will therefore be empty.

6.3 Carbohydrates ⓘ

EDS failed to run properly - this section will therefore be empty.

6.4 Ligands ⓘ

EDS failed to run properly - this section will therefore be empty.

6.5 Other polymers ⓘ

EDS failed to run properly - this section will therefore be empty.