



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:41 PM GMT

PDB ID : 4FSX  
Title : crystal structure of Se-substituted Zea mays ZMET2 in complex with SAH  
Authors : Du, J.; Patel, D.J.  
Deposited on : 2012-06-27  
Resolution : 3.20 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

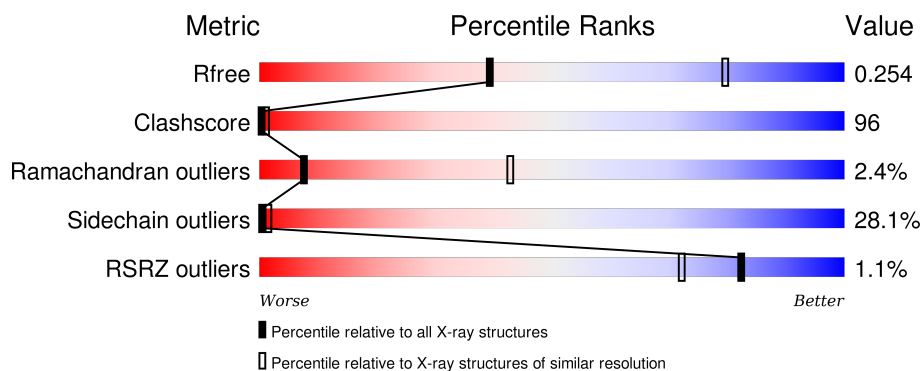
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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	784	<div> <div></div> <div> <div>21%</div> <div>47%</div> <div>17%</div> <div>•</div> <div>13%</div> </div> </div>
1	B	784	<div> <div></div> <div> <div>19%</div> <div>48%</div> <div>18%</div> <div>•</div> <div>14%</div> </div> </div>

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	SAH	A	1000	-	-	-	X

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 10631 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.

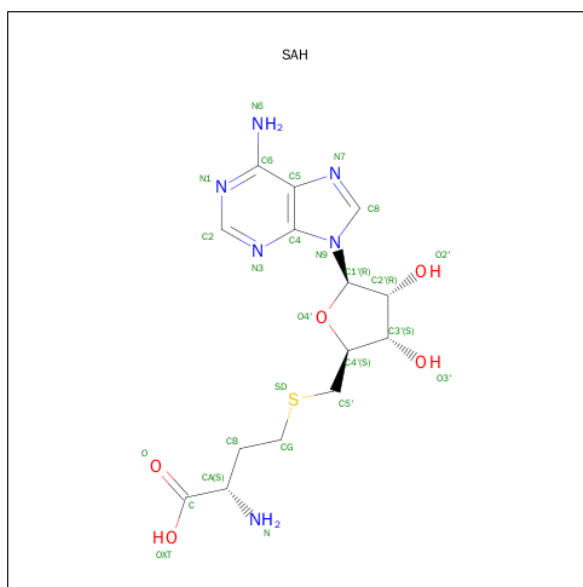
- Molecule 1 is a protein called DNA (cytosine-5)-methyltransferase 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	680	Total	C	N	O	S	Se	0	0	0
			5331	3401	904	993	18	15			
1	B	675	Total	C	N	O	S	Se	0	0	0
			5248	3346	893	976	18	15			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	129	SER	-	EXPRESSION TAG	UNP Q9AXT8
B	129	SER	-	EXPRESSION TAG	UNP Q9AXT8

- Molecule 2 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula:  $C_{14}H_{20}N_6O_5S$ ).



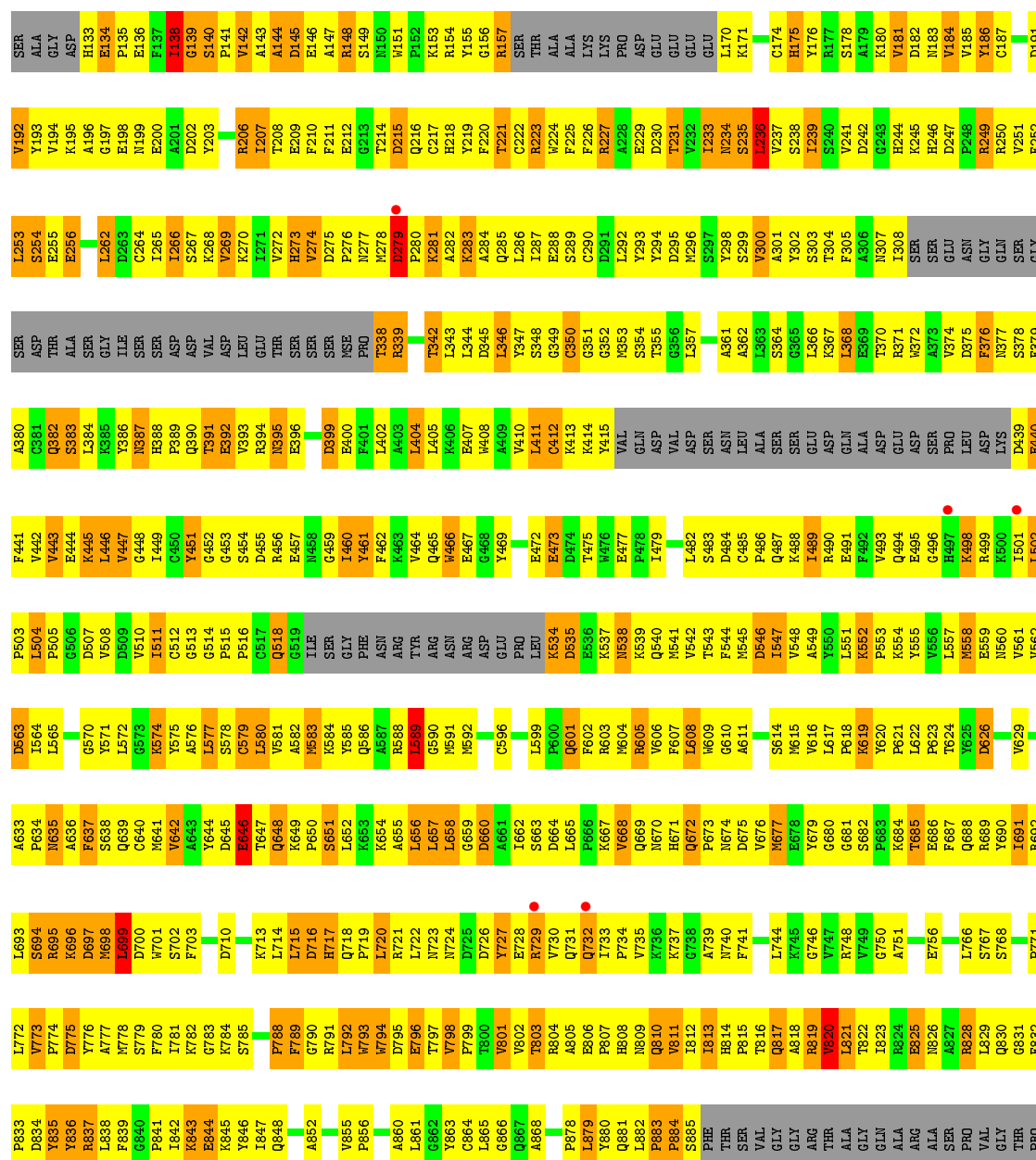
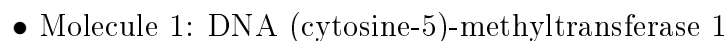
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			26	14	6	5	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	B	1	Total	C	N	O	S	0	0
			26	14	6	5	1		





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.86Å 88.95Å 113.49Å 93.47° 95.53° 110.41°	Depositor
Resolution (Å)	40.25 – 3.20 48.87 – 3.19	Depositor EDS
% Data completeness (in resolution range)	97.6 (40.25-3.20) 95.6 (48.87-3.19)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	0.17	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.07 (at 3.19Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.7.1_743)	Depositor
R, $R_{free}$	0.240 , 0.263 0.233 , 0.254	Depositor DCC
$R_{free}$ test set	1937 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	59.9	Xtriage
Anisotropy	0.609	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 68.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 38702 reflections	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	10631	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 4.11% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.64	0/5445	0.79	5/7358 (0.1%)
1	B	0.63	1/5362 (0.0%)	0.82	12/7252 (0.2%)
All	All	0.63	1/10807 (0.0%)	0.80	17/14610 (0.1%)

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	3
1	B	0	1
All	All	0	4

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	579	CYS	CB-SG	-5.50	1.72	1.81

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	535	ASP	CB-CG-OD1	-9.04	110.17	118.30
1	A	632	GLY	N-CA-C	-8.60	91.61	113.10
1	B	535	ASP	N-CA-C	8.56	134.12	111.00
1	B	146	GLU	N-CA-C	-6.83	92.56	111.00
1	B	746	GLY	N-CA-C	-6.81	96.08	113.10

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	145	ASP	Peptide
1	A	455	ASP	Peptide
1	A	507	ASP	Peptide
1	B	504	LEU	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	5331	0	5166	987	0
1	B	5248	0	5023	1014	0
2	A	26	0	19	6	0
2	B	26	0	19	6	0
All	All	10631	0	10227	1998	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 96.

The worst 5 of 1998 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:695:ARG:HG2	1:B:835:TYR:CD1	1.50	1.46
1:B:656:LEU:C	1:B:657:LEU:HD12	1.33	1.45
1:A:134:GLU:HB3	1:A:135:PRO:CD	1.43	1.38
1:B:883:PRO:CG	1:B:884:PRO:HD2	1.53	1.37
1:B:695:ARG:HG2	1:B:835:TYR:CE1	1.64	1.32

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	670/784 (86%)	591 (88%)	62 (9%)	17 (2%)	7	41
1	B	665/784 (85%)	607 (91%)	43 (6%)	15 (2%)	8	44
All	All	1335/1568 (85%)	1198 (90%)	105 (8%)	32 (2%)	7	43

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	142	VAL
1	A	634	PRO
1	A	646	GLU
1	B	140	SER
1	B	279	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	564/652 (86%)	413 (73%)	151 (27%)	0	2
1	B	544/652 (83%)	384 (71%)	160 (29%)	0	1
All	All	1108/1304 (85%)	797 (72%)	311 (28%)	0	1

5 of 311 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	813	ILE
1	B	234	ASN
1	B	732	GLN
1	A	839	PHE
1	B	149	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	814	HIS
1	B	183	ASN
1	B	810	GLN
1	A	817	GLN
1	B	175	HIS

### 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](#)

2 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	1000	-	20,28,28	1.60	4 (20%)	19,40,40	3.59	4 (21%)
2	SAH	B	1000	-	20,28,28	1.60	4 (20%)	19,40,40	3.56	4 (21%)

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	1000	-	-	0/7/31/31	0/3/3/3
2	SAH	B	1000	-	-	0/7/31/31	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1000	SAH	C5-C4	-2.96	1.33	1.40
2	A	1000	SAH	C5-C4	-2.95	1.33	1.40
2	A	1000	SAH	O4'-C1'	2.15	1.43	1.41
2	B	1000	SAH	O4'-C1'	2.20	1.44	1.41
2	A	1000	SAH	C2-N1	3.19	1.40	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1000	SAH	N3-C2-N1	-14.09	118.11	128.89
2	B	1000	SAH	N3-C2-N1	-14.00	118.18	128.89
2	A	1000	SAH	C4'-O4'-C1'	-5.32	103.87	109.72
2	B	1000	SAH	C4'-O4'-C1'	-5.27	103.93	109.72
2	A	1000	SAH	C2'-C1'-N9	-2.14	111.02	114.29

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1000	SAH	6	0
2	B	1000	SAH	6	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	665/784 (84%)	-0.15	9 (1%) 78 65	21, 62, 129, 200	0
1	B	660/784 (84%)	-0.24	5 (0%) 87 80	17, 69, 134, 285	0
All	All	1325/1568 (84%)	-0.19	14 (1%) 82 72	17, 65, 132, 285	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	759	PRO	3.9
1	A	746	GLY	3.3
1	A	484	ASP	3.2
1	A	445	LYS	2.7
1	B	729	ARG	2.6

### 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
2	SAH	A	1000	26/26	0.91	0.28	3.48	42,53,71,112	0
2	SAH	B	1000	26/26	0.95	0.19	-0.16	31,55,87,93	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.