Original Research Article

In-Silico Structural Annotation of Methylthioadenosine Nucleosidase Protein Zm00014a_031618 in Maize (*Zea mays L.*)

ABSTRACT

Aims: The aim of this study was *In-Silico* structural annotation of an amino acid sequence of Methylthioadenosine Nucleosidase Protein Zm00014a_031618 in Maize (*Zea mays*) retrieved from NCBI with the accession number PWZ58979.

Study design: The use of *In-Silico* studies for the structural annotation of Methylthioadenosine Nucleosidase protein.

Place and Duration of Study: The research was conducted at the Bioinformatics Laboratory, Chevron Biotechnology Centre, Modibbo Adama University of Yola, Nigeria. Between June 2018 to July 2018.

Methodology: The Methylthioadenosine Nucleosidase protein was retrieved from NCBI, physical and chemical parameters was calculated using ExPASy - ProtParam tool, the server SOPMA was used for secondary structure analysis (helix, sheets, and coils) and I-TASSER was used to obtain the 3D structure.

Results: ExPASy - ProtParam tool computated the various physical and chemical parameters such as molecular weight (MW) 30117.97, total number of positively (+R) 27, negatively charged residues (-R) 30, theoretical isoelectric point (pI) 5.96, aliphatic index (AI) 103.67 and grand average hydropathy (GRAVY) 0.293. The SOPMA server was used for calculating the secondary structural features of protein sequences as Alpha helix 39.16%, Extended strand 14.69%, Beta turn 6.64% and Random coil 39.51%. I-Tasser was used for predicting the 3D structure where 2qttA from PDB was used as the template.

Conclusion: This study helped in understanding the structural analysis of the Methylthioadenosine Nucleosidase Protein Zm00014a_031618 in maize (Zea mays)

Keywords: [Methylthioadenosine Nucleosidase, 3D structure, I-Tasser, in silico, Maize genome and Intraspecific}

1. INTRODUCTION

Maize (Zea mays L.) Poaceae for more than hundreds years has been a subject of genetics studies [1]. It is one of the most extensively studied plant species in genetics and it is usually used as research model for genome evolution and genetic diversity [2,3]. The genome is made up of 10 chromosomes with its size approximately 2.3 to 2.7 Gb and it is a diploid plant [4,5,6,7,]. Just like other larger genome of plant species, the *Z. mays* genome is typically made up of nongenic or low-copy DNA that harbor single genes. The repetitive elements is highly responsible for the wide range of diversity within the species which includes ribosomal DNA (rDNA), transposable elements (TEs) and high-copy short-tandem repeats mostly present at the centromeres, telomeres, and heterochromatin knobs [8,9,10,11]. *Z. mays* plant has an extraordinary levels of genomic diversity, phenotypic [12] and transcriptomic [13,14,15]. Looking at the genomic level *Z. mays* exhibits a high level of INDEL Polymorphisms [16,17] and Single Nucleotide Polymorphisms [18]. Averagely the

frequency of single nucleotide polymorphism (SNP) between two maize inbreds is said to be approximately 1 substitution per 100 bases [19,20]. Recent studies using sequencing data have shown that maize genome exhibits rather variable levels of naturally occurring genetic diversity which depends on the lines involved in the comparison [21,22].
Intraspecific genome variation has been long attributed to changes in size of heterochromatic DNA outside coding sequences that expanded and contracted the chromosomes (98). Intraspecific variations which are approximately 38.8% from the average of 5.5 pg/2n nucleus have been reported in *Z. mays* [23,24,25,26,27,22]. *Z. mays* is known to have large amount of intraspecific sequence variation [19,18] in form of deletion/insertion and single nucleotide polymorphisms. The main mechanism which have effect in the generation of intraspecific genome diversity and in the evolution of the maize genome, segmental duplications and whole genome duplications (polyploidization), retrotransposition and DNA transposition, expansion/contraction of simple sequence repeats (SSRs) and single base mutations and translocation of genes or gene segments by transposons and capture [22,28]. Intraspecific allelic variation is mostly as a result of qualitative changes that change the nature of the gene products and quantitative changes which also alter the amount of the gene product produced. Quantitative changes in gene expression can be as a result of cis- or trans variations in gene regulation [29].
The present study focused on the *In Silico* Structural Annotation of Methylthioadenosine Nucleosidase Protein Zm00014a_031618 in maize (*Zea mays*).

2. MATERIALS AND METHODS

2.1 Sequence Retrieval

Amino acid sequence of Methylthioadenosine Nucleosidase Protein Zm00014a_031618 (*Zea mays*) was retrieved from NCBI database (www.ncbi.nlm.nih.gov/protein/1394916517) with the accession number PWZ58979.

2.2 Physiochemical Analysis

The physiochemical properties of Methylthioadenosine Nucleosidase protein such as molecular weight, atomic composition, amino acid composition, theoretical pI, instability index, aliphatic index, extinction coefficients and grand average of hydropathocity (GRAVY) was determined using ProtParam tool (web.expasy.org/cgi-bin/protparam/protparam) [30].

2.3 Secondary Structure Analysis

The server SOPMA was used for secondary structure analysis (helix, sheets, and coils) of the Methylthioadenosine Nucleosidase protein (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) [31].

2.4 FASTA Sequence

The FASTA sequence of the Methylthioadenosine Nucleosidase Protein Zm00014a_031618 (*Zea mays*) was retrieved from NCBI databases [32].

2.5 3D Structural Prediction

The 3D structures was predicted with the use of I-TASSER [33]

2.6 Binding Residue Prediction

The binding residue of Methylthioadenosine Nucleosidase Protein was predicted using COACH server (34)

3. RESULTS AND DISCUSSION

The present study focused on the *In Silico* Structural Annotation of an amino acid sequence of Methylthioadenosine Nucleosidase Protein Zm00014a_031618 in maize (*Zea mays*) from NCBI database with the accession number PWZ58979 and 286 amino acid sequences.

The results presented in table 1 showed the physicochemical characterization of Methylthioadenosine Nucleosidase Protein Zm00014a_031618 in maize (*Zea mays*) with 286 amino acid sequence using

Expasy’s ProtParam server. The Molecular weight (MW), total number of positively (+R), negatively charged residues (-R), theoretical isoelectric point (pI), extinction coefficient (EC), aliphatic index (AI) and grand average hydropathy (GRAVY) was computed.

Table 1: Physiochemical Features of the Hypothetical Protein

Molecular weight (Da)	pI	-R	+R	EC	II	AI	GRAVY
30117.97	5.96	30	27	13200	23.10	103.67	0.293

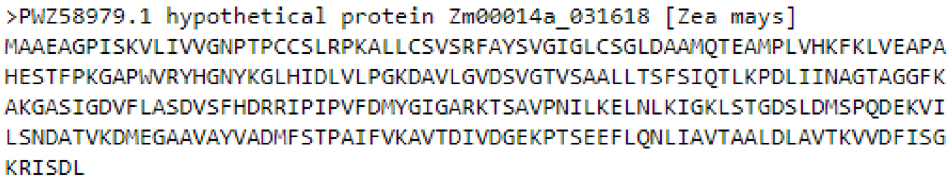
The results as presented in Table 2 showed the SOPMA which was used for calculating the secondary structural features of protein sequences such as Alpha helix, 310 helix, Pi helix, Beta bridge, Extended strand, Beta turn, Bend region, Random coil, Ambiguous states and Other states.

Table 2. Structural Features of the Methylthioadenosine Nucleosidase Protein

Parameter	% content	Parameter	% content
Alpha helix	39.16%	Beta turn	6.64%
310 helix	0.00%	Bend region	0.00%
Pi helix	0.00%	Random coil	39.51%
Beta bridge	0.00%	Ambiguous states	0.00%
Extended strand	14.69%	Other states	0.00%

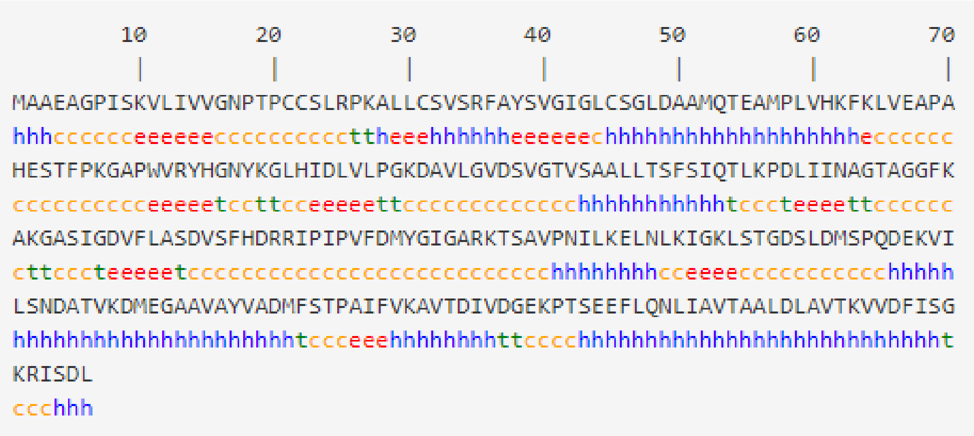
Table. 3. Top Five Models C-Scores from I-TASSER

Structural Models	**C-Scores**
1	**1.03**
2	**-2.32**
3	**-1.98**
4	**-2.65**
5	**-2.94**



```
>PWZ58979.1 hypothetical protein Zm00014a_031618 [Zea mays]
MAAEAGPISKVLIVVGNPTPCCSLRPKALLCSVSRFAYSVGIGLCSGLDAAMQTEAMPLVHKFKLVEAPA
HESTFFPKGAPWVRYHGNYKGLHIDLVLPGKDAVLGVDSVGTVSAALLTSFSIQTLKPDLIINAGTAGGFK
AKGASIGDVFLASDVSFHDRRIPIPVFDMYGIGARKTSAVPNILKELNLKIGKLSTGDSLDMSPQDEKVI
LSNDATVKDMEGAAVAYVADMFSTPAIFVKAVTDIVDGEKPTSEEFLQNLIAVTAALDLAVTKVVDFISG
KRISDL
```

Fig. 1. Fasta sequence of the Methylthioadenosine Nucleosidase Protein in maize Zm00014a_031618 (*Zea mays*)



```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MAAEAGPISKVLIVVGNPTPCCSLRPKALLCSVSRFAYSVGIGLCSGLDAAMQTEAMPLVHKFKLVEAPA
hhhcccccceeeeeeccccccccccttheeehhhhhheeeeeechhhhhhhhhhhhhhhhhhhecccccc
HESTFPKGAPWVRYHGNYKGLHIDLVLPGKDAVLGVDSVGTVSAALLTSFSIQTLKPDLIINAGTAGGFK
cccccccccceeeeetccttcceeeeettccccccccccccchhhhhhhhhhhtccctеeeettcccccc
AKGASIGDVFLASDVSFHDRRIPIPVFDMYGIGARKTSAVPNILKELNLKIGKLSTGDSLDMSPQDEKVI
cttccctееееetccccccccccccccccccccccccccchhhhhhhhcceeeeccccccccccchhhhh
LSNDATVKDMEGAAVAYVADMFSTPAIFVKAVTDIVDGEKPTSEEFLQNLIAVTAALDLAVTKVVDFISG
hhhhhhhhhhhhhhhhhhhhhtccceeehhhhhhhhttcccchhhhhhhhhhhhhhhhhhhhhhhhhhht
KRISDL
ccchhh
```

Fig. 2. Secondary Structure Prediction of Methylthioadenosine Nucleosidase Protein

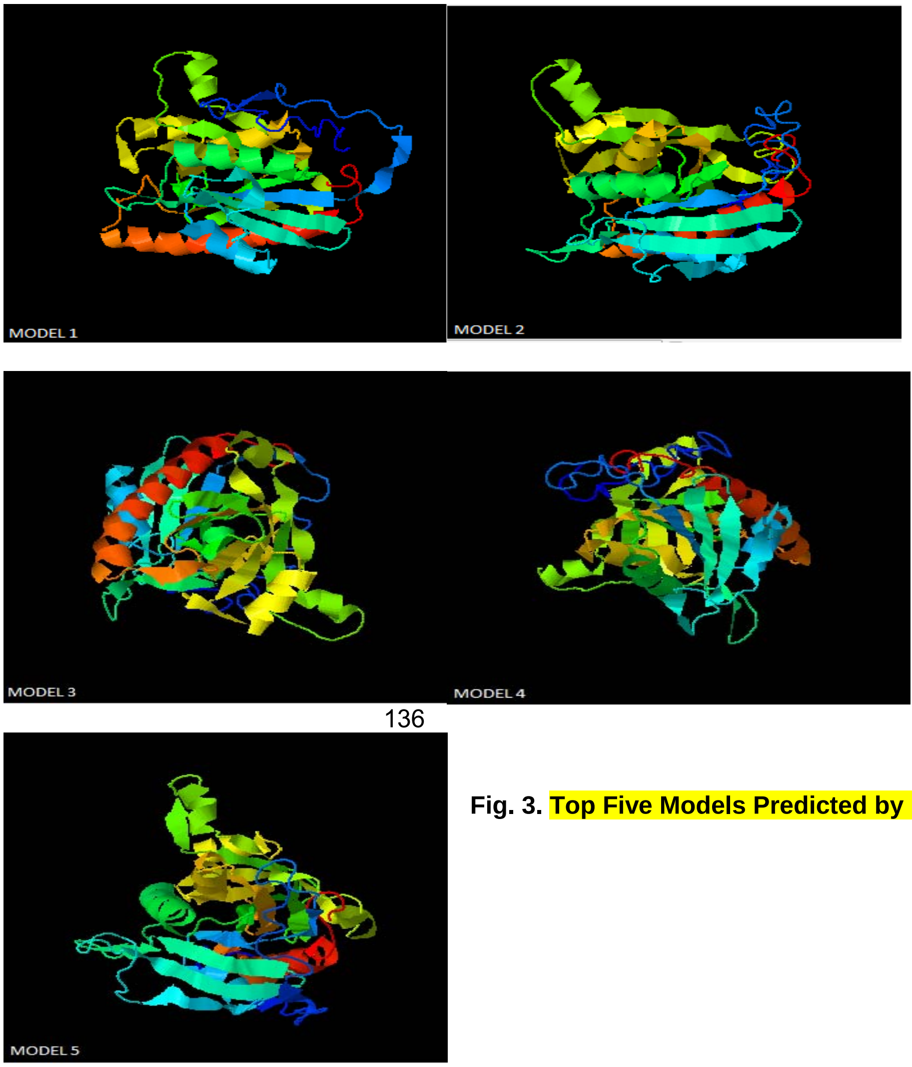


Fig. 3. Top Five Models Predicted by I-TASSER

Fig. 4. Structural Superposition of Methylthioadenosine Nucleosidase Protein with model 1 (2qttA from PDB)

ConCavity Results

Click to view	Rank	C-score	Predicted binding residues
●	1	0.59	50,51,52,55,109,133,134,135,136,137,200,201,218,219,220,221,243,244,246,250,256,260

Fig. 5. Predicted Binding Residues of Methylthioadenosine Nucleosidase Protein

The instability index (II) was computed to be 23.10 which make the Methylthioadenosine Nucleosidase protein classified as a stable protein because a protein whose instability index is less than 40 is said to be a stable protein [35]. The protein was predicted to have 286 amino acid sequences with several helices which is consistent with the ProtParam results present in Figure 1 this makes the protein more flexible for folding which is likely to increase the protein interaction. The sequence of Methylthioadenosine Nucleosidase protein was found to be rich in alanine. The proteins with very high AIn may show stability in a wide temperature range where lower AIn proteins are not thermal stable and show more flexibility. The amino acid sequences which had most in number are alanine [28] followed by leucine and valine [21], glycine and serine [23] and while the least is tryptophan (1). The Methylthioadenosine Nucleosidase protein had a total number of 30 negatively charged residues (Asp + Glu) and total number of 27 positively charged residues (Arg + Lys). The molecular formula of the protein was found as $C_{1354}H_{2181}N_{349}O_{401}S_{11}$. The GRAVY was shown to be 0.293GRAVY which shows a better interaction of protein and water is occurring in low GRAVY. [36]. The secondary structure of the Methylthioadenosine Nucleosidase protein was predicted by SOPMA server showed the random coil was the most predominant (39.51%), followed by alpha helix (39.16%), then extended strand (14.69%) and beta turn (6.64%) was the least. I-Tasser modeling server generated five models e PDB automatically. Model 1 with a C-score of 1.03 is best model because it has the highest score compare to the remaining four models. So the Methylthioadenosine Nucleosidase protein structure was compared with model 1 (2qttA from PDB) since it has the highest C score as the best model. Methylthioadenosine nucleosidase helps essentially in multiple metabolic pathways in plants [37].

4. CONCLUSION

This study has help in understanding the structural analysis of the Methylthioadenosine Nucleosidase Protein Zm00014a_031618 (*Zea mays*).

CONSENT

Not Applicable

ETHICAL APPROVAL

Not Applicable

REFERENCES

1. Coe E, East E. The birth of maize genetics. in maize handbook.volume ii: genetics and genomics. edited by Bennetzen JL, Hake S. New York,USA: Springer. 2009:3–15.

2. He E. “The origins of maize genetics,” Nature Reviews Genetics, L. 2001;2(11):898–905.

3. Lb L. “The future of maize,” in Handbook of maize: genetics and genomics, J. I. Bennetzen and S. hake, eds., Springer, Berlin, Germany. 2009;771–779.

4. Bennetzen JL. “Maize genome structure and evolution,” in handbook of maize: Genetics and genomics, J. L. Bennetzen and S. Hake, eds., Springer, Berlin, Germany. 2009;179–199

5. Schnable PS, Ware d, Fulton RS, *Stein JC, Wei F, Pasternak S*. et al. “The b73 maize genome: complexity, diversity, and dynamics,” Science. 2009;326(5956):1112–1115.

6. Rayburn AL, Biradar DB. Bullock DG, Mcmurphy IM. “Nuclear dna content in f1 hybrids of maize,” Heredity. 1993;70:294–300.

7. Zhou S, Wei F, Nguyen N. *Bechner* M, Potamousis K, Goldstein S. et al. “A single molecule scaffold for the maize genome,” Plos Genetics. 2009;5(11).

8. Mcclintock B. “The order of genes c, sh, and wx in zea mays with reference to a cytological known point on the chromosome,” Proceedings of The National Academy of Sciences of the United States of America. 1931;17(8):485–491.

9. Peacock WJ, Dennis ES, Rhoades, Pryor AJ. “Highly repeated dna sequence limited to knob heterochromatin in maize,” Proceedings of the National Academy of Sciences of the United States of America, 1981;78(7):4490–4494.

10. Ananiev EV, Phillips RL, Rines HW. “Chromosomespecific molecular organization of maize (zea mays l.) centromeric regions,” Proceedings of the National Academy of Sciences of the united States of America. 1998;95(22):13073–13078.

11. Morgante M. “Plant genome organisation and diversity: the year of the junk!,” Current Opinion in Biotechnology. 2006;17(2):168–173.

12. Buckler ES, Gaut BS, Mcmullen MD. Molecular and functional diversity of maize. Current Opinion in Plant Biology. 2006;9(2):172–6.

13. Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. The Plant Journal : for Cell and Molecular Biology. 2005;44(6):1054–64.

14. Stupar RM, Springer NM. Cis-transcriptional variation in maize inbred lines b73 and mo17 leads to additive expression patterns in the f1 hybrid. Genetics. 2006;173(4):2199–210.

15. Swanson-Wagner RA, Jia Y, Decook R, Borsuk LA, Nettleton D, Schnable PS. All possible modes of gene action are observed in a global comparison of gene expression in a maize f1 hybrid and its

inbred parents. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(18):6805–10.

16. Fu Y, Wen TJ, Ronin YI, Chen HD, Guo L, Mester DI, et al. Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. Genetics. 2006;174(3):1671–83.

17. Messing J, Dooner HK. Organization and variability of the maize genome. Current Opinion in Plant Biology. 2006;9(2):157–63.

18. Vroh BI, Mcmullen I, Sanchez-Villeda MD, Schroeder H, Gardiner S, Polacco J, et al. Single nucleotide polymorphisms and insertion–deletions for genetic markers and anchoring the maize fingerprint contig physical map. Crop Science. 2006;46(1):12.

19. Tenaillon MI, Sawkins MC, long AD, Gaut RL, Doebley JF, Gaut BS et al. "Patterns of DNA sequence polymorphism along chromosome 1 of maize (zea mays ssp. mays l.)," Proceedings of the National Academy of Sciences of the United States of America. 2001;98(16):9161–9166.

20. Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S. et al. "Snp frequency, haplotype structure and linkage disequilibrium in elitemaize inbred lines," Bmc Genetics. 2002;3(19)

21. Rafalski A, Ananiev E. "Genetic diversity, linkage disequilibrium and association mapping," in handbook of maize: genetics and genomics, J. L. Bennetzen and S. Hake, eds., Springer, Berlin, Germany. 2009;201–219

22. Rafalski A, Morgante M. "Corn and humans: recombination and linkage disequilibrium in two genomes of similar size," Trends in Genetics. 2004;20(2)103–111.

23. Rayburn AL, Price HJ, Smith JD, Gold JR. "c-Band heterochromatin and dna content in zea mays," American Journal of Botany. 1985;72(10):1610–1617.

24. Rayburn AL, "Flow cytometric assessment of nucleotide variability and its evolutionary implications," in Classical and Molecular Cytogenetic Analysis, W.J. Raup and B.S. Gill, eds, kansas Agricultural Experimental Station, Manhattan, kan, USA, 1994;110–115

25. Lee JH, Arumuganathan K, Kaeppler SM, Park S, Kim K, Chung Y. "Variability of chromosomal dna contents in maize (zea mays l.) inbred and hybrid lines," Planta, 2002;215(4)666–671.

26. Laurie DA, Bennet MD. "Nuclear dna content in the genera zea and sorghum. intergeneric, interspecific and intraspecific variation," Heredity.1985;55(3);307–313.

27. Biradar DP, Rayburn AL. "Heterosis and nuclear dna content in maize," Heredity. 1993;71(3)300–304.

28. Doebley J. "Molecular evidence for gene flow among zea species," Bioscience. 1990;40(6):443–448.

29. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. Nature. 2004;430:85–88.

30. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD. et al. Protein identification and analysis tools on the expasy server. The Proteomics Protocols Handbook. 2005;571-607

31. Geourjon C, Deleage G Sopma: Significant improvements in protein secondary structure prediction by prediction from multiple alignments. Comput Applic Bioci. 1995;11:681-684.

32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basiclocal alignment search tool. J Mol Biol. 1990;215:403—10

33. Yang Z. i-tasser: fully automated protein structure prediction in casp8. Proteins. 2009;77(9):100-113.

34. Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment, Bioinformatics. 2013;29:2588-2595

35. Guruprasad K. Reddy BV, Pandit MW. Protein Eng. 1990;4:155—61.

36. Ikai AJ. Thermo stability and aliphatic index of globular proteins. Journal of Biochemistry. 1980;88:1895-1898

37. Della RF, Porcelli M, Carteni-Farina M, Zappia V, Pegg AE. Escherichia coli S-adenosylhomocysteine/5′-methylthioadenosine nucleosidase. Purification, substrate specificity and mechanism of action. Biochem J. 1985;232:335–341