



**International Journal of Biology, Pharmacy
and Allied Sciences (IJBPAS)**

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

STRUCTURE AND FUNCTION ANALYSIS OF *ADENOSINE DEAMINASE* FROM *MYCOBACTERIUM TUBERCULOSIS* (MTADA)

DWIVEDI AK*, TIWARI S AND SINGH N

School of Biotechnology Gautam Buddha University Greater Noida, India

*Corresponding Author: E Mail: avaneeshkt@gmail.com

Received 25th May 2019; Revised 24th June 2019; Accepted 24th July 2019; Available online 1st Dec. 2019

<https://doi.org/10.31032/IJBPAS/2019/8.12.4884>

ABSTRACT

Adenosine to Inosine conversion is one of the most common modifications found in RNA world, a process known as A-to-I RNA editing. Adenosine deaminase that acts on RNA (ADA) is the main enzyme responsible for this A-to-I RNA editing and essential for RNA maturation. We have studied the structure prediction of Adenosine deaminase enzyme from *Mycobacterium tuberculosis* (MtADA). The structure prediction was done using homology modeling methods. *Mus musculus* ADA with sequence similarity of 47.6% with MtADA were used as template for model generation. The predicted model was validated using online servers and tools. The structural architecture and domain organization of MtADA was analyzed using various rapid computational approaches. The active site of MtADA was predicted containing residues Met309 & Ser310. The molecular dynamic simulation study was performed with tRNA. The predicted structure could prove valuable for advance studies and structure based drug designing against *Mycobacterium* to combat infections.

Keywords: Homology modeling, Computational tools, Protein interaction, RNA modification, Molecular Dynamics Simulation, Structure-function relationship

1. INTRODUCTION

RNA editing is a crucial co/post-transcriptional modification process in which an alteration in the

primary nucleotide sequences occurs via enzymatic reactions [1]. Methylation, acetylation, deamination and trans-glycosylation are some of

the crucial modifications that occur on RNAs [2]. Evolutionary studies have demonstrated that RNA modification enzymes are one of the most conserved classes among the bacteria, archaea and eukaryotes [3]. RNA editing is a dynamic mechanism which can affect RNA splicing, nuclear-cytoplasmic transport, mRNA stability, translation and regulation of genes by RNA interference [4]. The malfunction of the editing machinery is associated with different abnormalities like ageing, cancer, autoimmune, neurological or cardiovascular disorders [3].

Among all the forms of RNA modifications, hydrolytic deamination of adenosine (A) to inosine (I) is the most common RNA editing event [5]. A-to-I RNA editing is catalyzed by members of the ADA family, which bind to double stranded RNA (dsRNA) substrates [6]. Inosine is widespread among various types of RNAs including ribosomal RNA (rRNA), messenger RNA (mRNA), transfer RNA (tRNA), microRNA (miRNA) and long non-coding RNA (lncRNA) that can appear in different locations in these RNAs [2]. The functional consequence of the inosine modification depends on both the nucleotide position modified and the type of RNA [2]. Because inosine is recognized as guanosine (G) by splicing and translation machineries due to their similar characteristic therefore, it can affect the stability, biogenesis and target recognition properties of RNAs [6].

ADA is a metallo-dependent hydrolase that catalyzes the irreversible hydrolytic deamination of adenosine/ deoxyadenosine to inosine /deoxyinosine and releases ammonia [7]. Within the bacteria, one of the apparent roles of this enzyme is purine metabolism and scavenging ammonia.

Tuberculosis (TB) a most concerning problem in developing countries is caused by *Mycobacterium tuberculosis* a pathogenic bacteria. TB is hard to cure, and responsible for more than a million deaths each year [8]. The situation has further worsened by the emergence of extensively drug-resistant tuberculosis (XDR- TB). Thus, there is an urgent need to look into a new target site in bacterium and develop new anti-tubercular drugs for effective treatment of this disease. As ADA is an essential enzyme for proper cellular functioning therefore, it seems very promising to defeat TB. Three-dimensional (3D) structure determination is the most important necessity of any advanced proteomic analysis for drug designing. The present study deals with detailed tertiary structure derived by homology modeling, validation of structure, functional characterization and molecular dynamics simulation study. The study also dealt with active site prediction, role in different metabolic pathway and protein-protein interaction study of MtADA which can act as potential therapeutic target for controlling TB infections.

2. MATERIAL AND METHODS

Protein Sequence of MtADA

The full-length MtADA of 365 amino acid sequence information was obtained from a protein database of NCBI program in fasta format. (Accession no. AIR16106.1).

Structure Prediction & Validation of MtADA

The template selection of MtADA model was generated by HMM-HMM correlation using Hhpred server for structure prediction [9]. The predicted structure file of MtADA was retrieved on accomplishment of 3-D modeling in PDB format. Met1 to Met361 residues were modeled based on the available templates. Chimera [10] was used for visualizing the structure.

The predicted 3-D model of MtADA was validated with the help of PROCHECK [11]. The G-factors is a measurement of property for unusual or out of the ordinary structure. The values of G factor below -0.5 is unusual while below -1.0 is highly unusual structure. The main-chain bond angles and bond-lengths were compared with ideal values derived from small-molecule data. Structure refined using different restraints showed actually great deviations from normality.

The problem of error recognition between experimental and theoretical models was of great importance of structural biology. The protein structure problem ProSA [12] is a user friendly

web server used for protein structure validation. It calculates overall quality of a given 3D molecule in terms plot of local quality score points which requires atomic coordinates of the model to be evaluated. ProSA-web server calculates C-alpha atoms of the input structure for validation.

Our ProQ [13] result indicates acceptable quality of constructed structure. Here, LG and MaxSub scores are taken into consideration for quality estimation of modeled protein structure. Different ranges of LGscore shows quality of predicted model. LGscore>1.5 indicates fairly good model, LGscore>2.5 shows very good model, and LGscore>4 shows extremely good model. Values of MaxSub>0.1 indicates fairly good model, MaxSub>0.5 indicates very good model and MaxSub>0.8 indicates extremely good model.

Verify3D [14] program accesses the quality of the modeled structure and validates model on the basis of amino acid scores. To pass the criteria, at least 80% of the amino acids of given sequence should score more than 0.2 in the 3D/1D profile. Profile 3-D utilized to determine the stereo-chemical aspects and local structure similarity.

The superimposition of the predicted structure on the template structure was done by using SuperPose tool [15]. The RMSD values of the protein were calculated. Additionally, the general stereo-chemical nature of the protein was analyzed whether all amino acid residues are in allowed or disallowed regions.

Further evaluation of 3D structure of MtADA done by using VADAR (Volume, Area, Dihedral Angle Reporter) [16]. VADAR have a group of more than 15 algorithms and programs for assessing and analyzing 3D coordinates of a given protein. The validation of results done by comparing pre-published data and visual examinations. It uses homology modeling or threading methods to determine protein structure and quality & quantity assessment.

Structure Annotation & Domain Identification

ProFunc Server identifies the biochemical function of a given protein from its 3D structure [17]. ProFunc applies number of methods, including residue conservation, fold matching, surface cleft analysis, and functional 3D templates, to identify active site and possible homologues in the PDB.

The Domain Identification was performed with Dihedral Alignment server (DIAL) [18]. It recognizes the domains in the predicted modeled of MtADA. DIAL is an online server for domain identification of given proteins by clustering secondary structure and substructures. Each cluster represents a structural domain and thus, measurements of compactness of clusters were calculated for each domain organization. Disjoint factor values >1.0 provide satisfactory structural solutions of structural domain organization in the given structure.

Structure Function Relationship

ProFunc and Sequence Annotated by Structure servers (SAS) [19] were used to analyze the predicted and validated structure of MtADA for its functional relationship. The sequence information and modeled structure of MtADA suggests their functional relationship. ScanProsite [20] was applied for establishing functional relationship of MtADA. ScanProsite scans proteins for best matches against motifs available in Prosite motifs including or excluding the ones with a high probability of occurrence. Prosite determines the function of a protein from its nucleotide sequence. It's a database of important biological sites and patterns, which computationally searches for the reliable protein family of a given sequence. The role of protein motif detection and function establishment is essential tool of sequence analysis. It works on technique based on weight matrices (also known as profiles) to detect a proteins or domains.

Protein-Protein Interaction Analysis

MtADA protein interacting partners and their functions were identified using STRING Database [21]. Protein jointly contributes to a shared function and proteomic connections such as gene fusion, co- occurrences, curated databases and text mining were examined by this database. Using this approach, the functional, biological, pathways and molecular characterization of MtADA was done.

Binding studies of MtADA

Meta-server way to ligand binding site prediction

is COACH [22]. Initiate from the structure of target proteins, COACH creates a complementary ligand binding site predictions employing two comparative methods, TM-SITE and S-SITE, which acknowledges the ligand-binding templates from the BioLiP protein function database by confining precise substructure and sequence profile comparisons. C-score represent the confidence score of predicted binding site in a given protein structure. C-score values lies in 0 to 1. Reliability of the prediction of binding site is directly proportional to the C-score. The cluster size indicates the number of templates taken in the experiment. In TM-align, RMSD score is the root mean square deviation between residues that are structurally aligned. BioLiP database was also used to predict the binding site in given protein.

Active Site Prediction of MtADA

The Modeled 3D structure of MtADA [PMID PM0081199] was used for active site prediction. MtADA protein preparation and active site identification was done in Maestro (Schrödinger) [23] using Protein Preparation Wizard. Active site prediction was done by SiteMap module. Site finding algorithm was used for grid generation of size 1Å on whole protein surface. These site points are connected together close to the surface of protein, where potential ligand binding sites are identified.

Phylogenetic evaluation

Evolutionary information was inferred using

Neighbor-joining (NJ) method of phylogenetic tree analysis. Phylogenetics shows the relation between organisms, gene, and proteins of different species. It reveals the evolutionary history and relationship between species. E- Value is the criteria to sort out the organisms which will help to plot the evolutionary tree with the help of different bioinformatics applications.

We aligned different homologous sequences for Adenosine deaminase from different species for generating a phylogenetic tree. Protein Blast program was used to align the sequences and to find out closely related sequences with higher e-value score. Phylogenetic tree constructed with the help of Clustal Omega [24]. The aligned file was further used as input for Phylip program [25]. By using UPGMA and NJ method, we analyzed the information about the branches and roots attached with it. Phylip provides unrooted and rooted tree and draws a tree based on distance matrix. Multiple Sequence alignment of amino acid sequences of different species of Adenosine deaminase was done using Clustal Omega. substitution matrix was used for alignment. To compute evolutionary distances, p-distance method was used. Total 6 amino acid sequences were used in complete analysis.

Molecular Dynamics Simulation studies

The top ranked model was selected for energy minimization to remove steric clashes between side chains and to analyze the biophysical

behavior in the dynamic system using GROMACS 4.5.6. [26] The Root Mean Square Deviation (RMSD) of the protein and template was calculated by superimposing the structures using Chimera.

Biophysical behavior was assessed by simulating the homology models in water and ionic environments. It is a widely used method to predict the stability of apo MtADA and MtADA-substrate complexes. MD simulations of MtADA and its complexes were performed with the GROMACS 4.5.6 package using the amber99sb force-field. The topology was generated using pdb2gmx modules of GROMACS. The model was placed in a cubic box using the editconf module. The protein was centered in the box with 1.0 nm distance from the edges. Further the system was solvated in simple point charge water model (SPC216) using solvate module. The genion module was used to add 1 Na⁺ ions on the MtADA for neutralization. Additionally, each dodecahedron box was energy minimized with the steepest descent algorithm to attain a reasonable geometry and solvent orientation. The systems were equilibrated for about 100ns under two step ensemble processes (NVT and NPT) with position restrained dynamics. Finally the systems were submitted to molecular dynamics simulation for about 100 ns to observe stability of MtADA and with its substrate. The resulting trajectories were analyzed using g_hbond, g_rmsf, g_rms and

trajconv utilities of GROMACS. All graphical presentation was prepared using Origin 6.0 [27].

3. RESULTS AND DISCUSSION

Protein sequence of MtADA

The MtADA is a 365 amino acid enzyme with a molecular weight of 39.7kDa. A total of 48(13.1%) amino acids have proton donor property i.e., Arginine 27(7.4%), Histidine 15(4.1%) and Lysine 6(1.6%) whereas 49(13.4%) amino acids are as proton acceptor i.e., Aspartic Acid 30(8.2%) residues and Glutamic Acid 19(5.2%) residues. Nature of protein seems hydrophobic as it contains 204(60.7%) residues as non-polar. The buried side of enzyme consists of Alanine 52(14.2%), Glycine 25(6.8%), Valine 26(7.1%), Leucine 34(9.3%), Isoleucine 19(5.2%), Proline 16(4.4%), Methionine 10(2.7%), Phenylalanine 20(5.5%) and Tryptophan 2(0.5%). The hydrophilic portion of enzyme contains 6(1.6%) Cysteine, 6(1.6%) Tyrosine, 19(5.2%) Serine, 19(5.2%) Threonine, 5(1.4%) Asparagine and 9(2.5%) Glutamine. A total of exposed non-polar residues are 64(17.5%).

3D Structure Prediction and Validation

The 3D model of MtADA was generated through homology modeling [Figure 1]. It was observed that MtADA showed maximum similarity with the crystal structure of the *M. musculus* ADA (PDB id: 1ADD_A) and *Human* ADA (PDB id: 3IAR_A) which were selected as template with sequence similarity of

47.6% & 45.1% and E-value of 4.7E-50 & 1E-33 respectively. The structure contains 8 beta sheets (Pro14 to His21, Asp140 to Val146, Leu149 to Met153, Gly176 to Ala180, Phe205 to Ala209, Arg229 to His 232, Leu264 to Leu266 and Val295 to Val297) whereas 16 Alpha helices and seven 3_{10} helices.

Since the accurate prediction and validation of the structure is itself a complex task, here we have tried to validate predicted model via different approaches. The constructed structure fulfilled all the validation criteria from PROCHECK. A total of 118 protein structures were taken with less than 2.0 Å and <20 R factor scores for validating structure. Over 90% residues must be in favored regions for a good quality model. According to Ramachandran plot, 90.1% residues of ϕ/ψ angles were in favored regions, 6.5% residues in allowed region, 1.6% residues in the less allowed region and 1.9% residues in outlier space [Figure 2(a)]. The value for G-factor was calculated -0.10, indicated that the structure was unusual.

The ProSA z-score shows overall model quality. The values were plotted in a plot that contains the z-scores of all experimentally determined protein chains in coordinate file of given protein. The groups of structures from different sources i.e., X-ray and NMR were represented by light and dark blue colors and used to check whether the z-score of the given

protein was within the range of scores typically found for native proteins of similar size. ProSA Z-score of modeled structure was -8.24 [Figure 2(b)], indicating that constructed structure occupies same regions as X-Ray analyzed native structure. From the structural analysis with Anolea we have concluded that 31.7% of amino acids have high energy. We observed that energy of all the residues were largely negative except few regions. The ProSA Residue score for all the residues in the predicted 3D model were negative [Figure 2(c)]. The RMSD of two aligned structures was 2.16Å, indicating that the constructed structure was accurately modeled. In addition, we deposited our constructed model in Protein Modeling Database with PMDB PM0081237.

Our ProQ results indicated acceptable quality of constructed structure. Here, LG and MaxSub score is taken into consideration for quality estimation of modeled protein structure with LG score 5.422 and MaxSub 0.548. Score of our predicted model came under the category of Extremely Good Model.

The quality of final structure was further validated by Verify-3D with a score of 90.30% as the residues have averaged 3D-1D score \geq 0.2 [Figure 3 (a)]. Profile-3D with Quality score of 146.66 and 3D-1D averaged score of 0.68 [Figure 3 (b)]. The results of superpose tool shows overall RMSD score of 2.71 (alpha

carbon have 2.46 and backbone had 2.36) with chain A of 1ADD.pdb using needle algorithm. During alignment 26.7% residues had identity, 44.2% had similarity and 12.2% gaps were found with an alignment score of 340.

The statistical results of VADAR using atomic radii from Eisenberg, showed that the MtADA consist of 54% (196) helix, 15% (55) beta, 30% (110) coil and 19% (72) turns in structure. The mean h_bond distance calculated was 2.2 ± 0.4 , mean h_bond energy -1.6 ± 1.1 and 86% (313) residues were with h_bond. The total volume were observed to be 46605.2 \AA^3 , where as expected volume was 47617.2 \AA^3 . The 3D profile quality indexed in [Table 1a & 1b] [Figure 4].

Structure Annotation & Domain Identification

The wiring diagram was generated by using SAS server [Figure 5]. The Profunc server revealed the gene ontology. The cell components were intracellular in nature, biological processes were cellular and metabolic and biological functions were metal and cation binding. InterPro scan for sequence motifs, 4 motifs matched in scan against PROSITE, TIGRFAM, PROFILES, PRINTS, PFam-A and PRODOM motifs were aden_deam: adenosine deaminase, Adenosine/AMP deaminase) and Adenosine deaminase [add]. Sequence search was done

with existing PDB entries, 66 matching sequences found by FASTA search. BLAST search was performed against Uniprot database, 50 matching sequences were found by BLAST search. 3D functional template search was performed for active site template of enzyme, 13 significant hits out of 3518 enzyme active site templates were found.

The Profunc server results for the secondary structure of MtADA of 365 amino acids of which 49 amino acids (13.4%) involve in the formation of strand, 168 amino acids (46.0%) in alpha helix, 14 amino acids (3.8%) in 3–10 helix. The presence of 1 beta sheet, 7 beta-alpha-beta motifs, 3 beta bulges, 8 strands, 21 helices, 32 helix-helix interactions, 31 beta turns and 3 gamma turns. Additionally, domains were identified using 3D coordinates. The first domain related to Alpha Beta class in which both the helix and strands were present in 15 to 55% and 10 to 45% respectively. In this domain 117 residues were detected from Met1 to Arg27, Pro81 to Ile144 and Asn298 to Gly323. The second domain belonging to few secondary structures type involved 53 amino acid residues. The observed residues ranged from Pro28 to Thr80, the last domain was related to Alpha Beta class consisting of 195 residues and ranged from Thr145 to Asn297 and Trp324 to Glu365.

On the basis of percentage residue occupancy in

helical, loop & strand regions, the domains in DIAL server were categorized into 4 different types, all alpha containing (>60% helix), all beta containing (>60% strands), alpha beta containing (15-55% helix or loop) and few secondary structure containing type (10-45% helix or loop). The functional sites and motifs found in 4 domains characterized by DIAL server are tabulated in [Table 2].

Structure and Function Relationship

The MtADA had 13 hits by 4 distinct catalytic sub units with a high probability of occurrence, identified by ScanProsite. Thr9 to Arg11 and Thr145 to Arg147 had Protein kinase C phosphorylation sites, Thr25 to Phe30, Gly141 to Val146, Gly181 to His186 and Gly275 to Ile280 had N-myristoylation sites, Thr30 to Asp33, Thr46 to Asp49, Ser118 to Asp121, Ser217 to Glu220 and Ser279 to Glu282 had Casein kinase II phosphorylation sites, while Arg190 to Met197 and Arg315 to Gly322 had Tyrosine kinase phosphorylation sites [Table 3]. Hence these results gave an in-depth knowledge based on the function related to the predicted structure of MtADA in cellular processes.

Protein-protein Interaction Analysis

The MtADA Protein interacting partners were identified using STRING database [Figure 6]. MtADA have interactions with protein families responsible for nucleoside phosphorylase and

dehydrogenase. InterPro Domain Architecture (IDA) tool of STRING database concluded the presence of PNP/MTAP phosphorylase, Inosine-5-monophosphate dehydrogenase and IMP dehydrogenase/ GMP reductase functions. MtADA may have transferase activity, dehydrogenase activity, deaminase activity and phosphorylase activity [Table 4].

Binding studies of MtADA

The predictions are linked with outcomes from various methods which include COFACTOR, FINDSITE and ConCavity to create a final ligand binding site predictions. The consensus binding residues were predicted at 17 different positions in protein with Deoxycytosine (DCF) as ligand taking 1A4I_B.pdb as a template with C-score of 0.66. With TM-SITE search in our protein MtADA, Zinc, Iron and Cobalt act as cofactor with predicted binding site residues at 19th, 21st, 178th and 300th positions and protein-ligand binding site structure were predicted with C-score of 0.52 and a cluster size of 99 [Figure 7(a)]. The S-SITE search shows that predicted binding site residues at 21 different positions were important for ligand binding in which 10 residues were involved in Zinc binding. In another program COFACTOR results, predicted binding site residues for Zinc as ligand at 19th, 21st, 208th and 300th residues of protein with C-score of 0.65 with 1a4IA.pdb as template,

TM-score of 0.888 and RMSD value of 1.41. In program FINDSITE, 20 different positions in protein were involved in binding site with a C-score of 0.80 and cluster size of 32. The program ConCavity, the predicted binding residues were 25 in number with a C-score of 0.36 [Figure 7(b)]. The ligand used to predict binding site was Deoxycoformycin (DCF) using cofactor.

Active Site prediction

SiteMap program was used to establish possible active sites of MtADA. Five active sites were found on the basis of site-score [Figure 8]. Site score depends on properties as number of site points, hydrophilic score and enclosure score. The highest site-score which was of Site1 (1.032) an active site [Table 5].

Phylogenetic Analysis

To draw a phylogenetic tree we had selected the adenosine deaminase protein sequences from the different species where their structural information was available. On the basis of e-value scores, we selected 7 different Adenosine deaminase proteins from different species i.e., Human, *Mus musculus*, *Pseudomonas*, *Plasmodium*, *Burkholderia* and *Arthrobacter*. These sequences were again realigned for tree generation with a score. The alignment file was

further used as input in Phylip program to get internal and external nodes [Table 6]. Which showed the 3 different groups had been plotted 1st is of Human ADA and *Mus musculus* ADA which shows the valuable similarity to our query sequence. The Second is of *Pseudomonas*, *Burkholderia* and *Arthrobacter* species and last group consists of *Plasmodium* and Human Growth factor ADA. These three were diverged by the evolution time factor. [Figure 9].

Molecular Dynamics Simulation studies

RMSD was evaluated for 100 ns of simulation for both complex, average RMSD for MtADA bonded with cofactor (Zn) and with Substrate (tRNA loop) was 0.41 and 0.179 nm. Last 25ns stable trajectory was taken for further analysis to reduce computational cost. Average RMSF for MtADA bonded with cofactor (Zn) and with Substrate (tRNA loop) was 0.114 and 0.06nm respectively. MtADA formed 9 hydrogen bonds with substrate and 308 hydrogen bonds with water molecules present in the dodecahedron box. Hence, we had concluded that the predicted model was stable and could be exploited for drug design experiments [Figure 10].

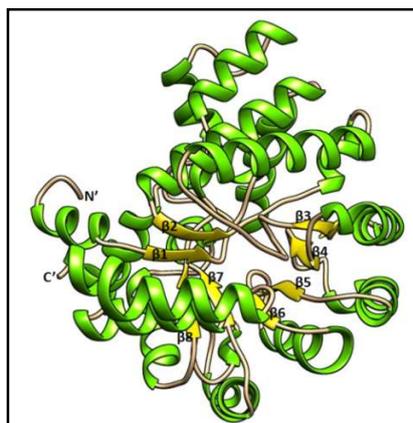


Figure 1: Three dimensional structure of MtADA. The structure consists of 8 beta sheets and 16 Alpha helices and seven 3_{10} helices

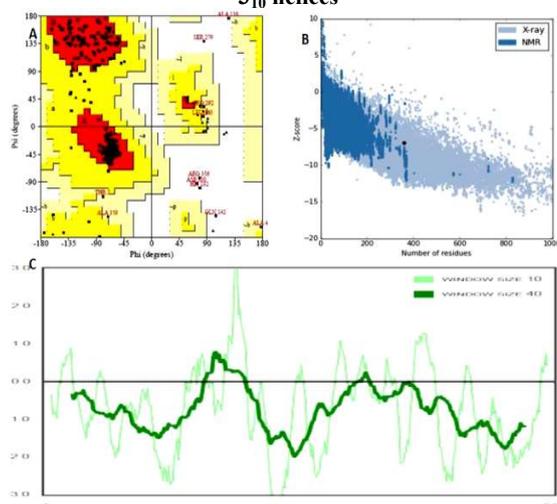


Figure 2: Structure validation of MtADA (a) Ramachandran Plot generated using Procheck server. In Ramachandran plot, 90.1% residues of ϕ/ψ angles were in favored regions, 6.5% residues in allowed region, 1.6% residues in the less allowed region and 1.9% residues in outlier space. (b) Structure comparison using ProQ Server. The groups of structures from X-ray and NMR are represented by light and dark blue colors. The z-score of the MtADA is within the range of scores typically found for native proteins of similar size. ProSA z-score of modeled structure was -8.24. (c) The ProSA Residue score for all the residues in the predicted 3D model were largely negative except few regions.

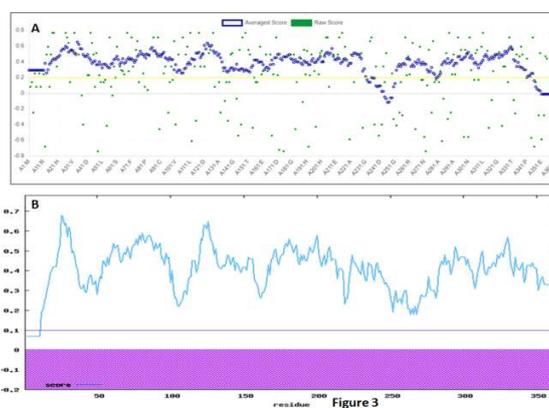


Figure 3: The quality of final structure has been validated by (a) Verify-3D with average 3D-1D score ≥ 0.2 . (b) Profile-3D have 3D-1D average score of 0.68

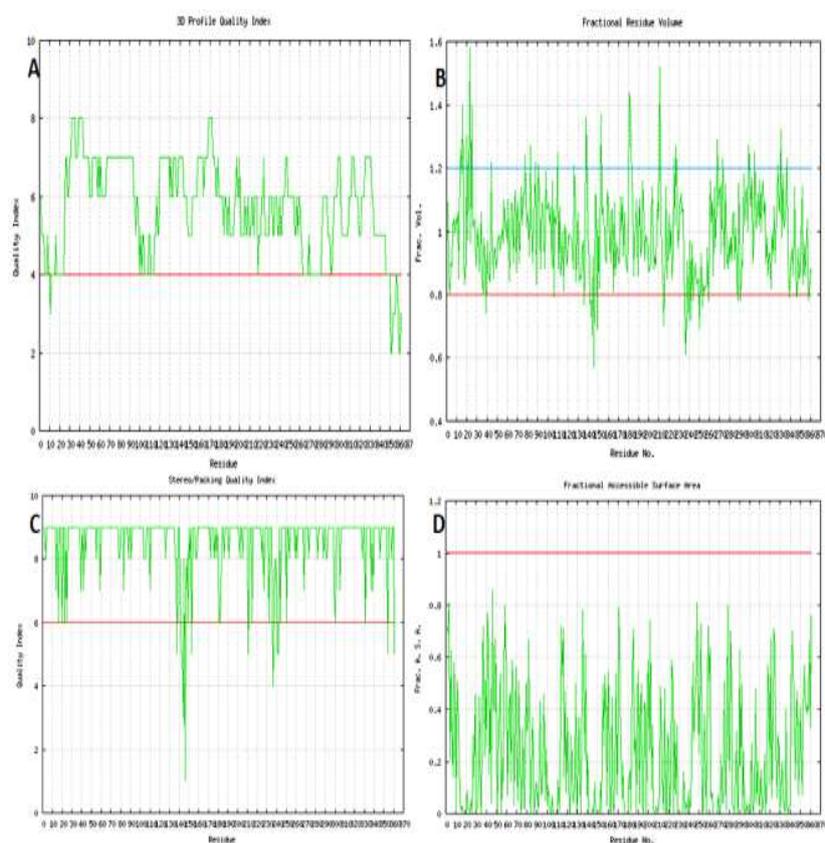


Figure 4: Sterio-chemical quality estimation of MtADA using VADAR Server (a) 3D Profile Quality Index, (b) Fractional Residue Volume, (c) Stereo Quality Index and (d) Fractional Accessible Surface Area data were plotted

Table 1(a): Dihedral Angles found in modeled structure using VADAR (Volume, Area, Dihedral Angle Reporter) Server

DIHEDRAL ANGLES		
Statistics	Observed (in \AA^2)	Expected (in \AA^2)
Mean Helix Phi	-64.7 sd=6.6	-65.3 sd=11.9
Mean Helix Psi	-39.5 sd=13.9	-39.4 sd=25.5
res with Gauche+ Chi	129 (47 %)	147 (55%)
res with Gauche- Chi	43 (15 %)	53 (20%)
res with Trans Chi	97 (36 %)	67 (25%)
Mean Chi Gauche+	-65.9 sd=9.3	-66.7 sd=15.0
Mean Chi Gauche-	62.0 sd=10.1	64.1 sd=15.7
Mean Chi Trans	171.2 sd=7.4	168.6 sd=16.8
Std. dev of chi pooled	8.77	15.70
Mean Omega (omega>90)	-178.1 sd=5.2	180.0 sd=5.8
# res with (omega<90)	1 (0 %)	-

Table 1(b): Accessible surface area calculated in modeled structure using VADAR (Volume, Area, Dihedral Angle Reporter) Server

ACCESSIBLE SURFACE AREA (ASA)		
Statistics	Observed (in Å ²)	Expected (in Å ²)
Total ASA	14178.6	14155.1
ASA of backbone	1714.2	-
ASA of side chains	12464.3	-
ASA of C	8674.6	-
ASA of N	1326.8	-
ASA of N+	876.5	-
ASA of O	2430.9	-
ASA of O-	747.1	-
ASA of S	122.7	-
Exposed non-polar ASA	8797.3	8648.9
Exposed polar ASA	3757.7	2835.7
Exposed charged ASA	1623.6	2693.9
Side exposed nonpolar ASA	7592.5	-
Side exposed polar ASA	2287.1	-
Side exposed Charged ASA	1556.0	-
Fraction nonpolar ASA	0.62 ± 0.03	0.61
Fraction polar ASA	0.27 ± 0.05	0.20
Fraction charged ASA	0.11 ± 0.05	0.19
Mean residue ASA	39.3 sd=43.6	-
Mean frac ASA	0.2 sd=0.2	-
% side ASA hydrophobic	35.52	-

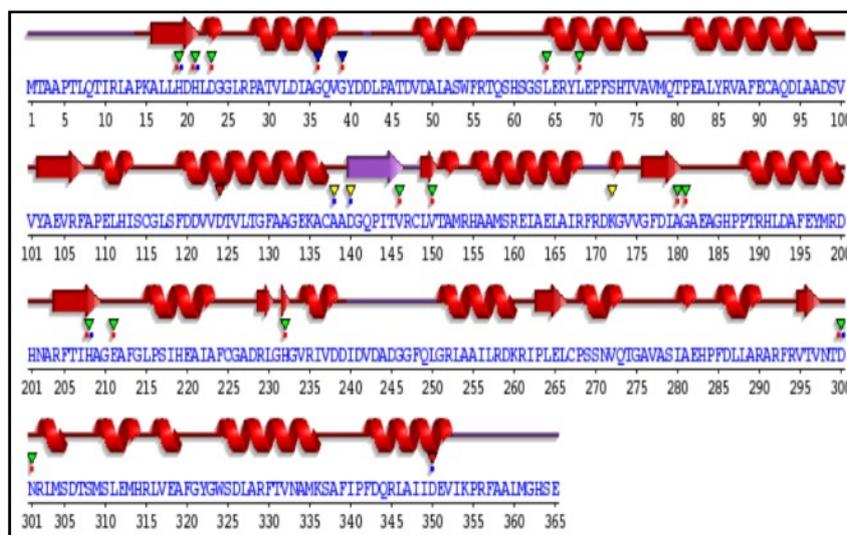


Figure 5: Wiring diagram of the MtADA secondary structure by SAS Server showing predicted active sites and binding sites of metals & ligands

Table 2: Domain Identification Algorithm (DIAL) Server shows conserved regions mapped on domain sequence along with its homologs.

Chain	Domain	No. of Residues	Domain Class	Functional Sites	Motifs	Conserved Residues
A	1	232 (1-75 and 204-360)	Few Secondary Structures	Protein Kinase C Phosphorylation		His80, His91, Cys97, His104, Cys139, His155, His186, Trp196
				9-11	TiR	
				74-76	TvR	
				Casein Kinase II Phosphorylation		
				30-33	TvID	
				46-49	TdvD	
				89-92	SihE	
				151-154	SiaE	
				Tyrosine kinase Phosphorylation		
				187-194	Rlv.EafgY	
				N-myristoylation		
25-30	GLrpAT					
147-152	GAvaSI					
A	2	128 (76-203)	Alpha beta class	Protein Kinase C Phosphorylation		Tyr10, Arg11, Phe14, Cys16, Glu18, Asp19, Asp23, Tyr27, Arg31, Phe32, Pro34, His37, Ile38, Cys40, Phe44, Asp45, Asp46, Asp49, Phe55, Cys62, Asp65, Glu67, Pro68, Arg72, Cys73, Met78, Arg79, His80, Met83, Arg85, Arg93, Phe94, Arg95, Phe102.
				70-72	TvR	
				Casein Kinase II Phosphorylation		
				43-46	SfdD	
				Tyrosine kinase Phosphorylation		
				115-122	Rhl.DafeY	
				N-myristoylation		
				66-71	GQpiTV	
106-111	GAcaGH					
B	3	307 (1-36, 73-225 and 244-361)	Few Secondary Structures	Protein Kinase C Phosphorylation		Cys55, His76, Cys79, Cys101, Cys112, His119, His150, His155, Phe159, Tyr161, Met162, Arg163, His165, Phe169, His172, Phe177, Pro180, Ile182, His183, Ile186, Phe188, Glu189, Asp191, Phe194, Glu195, Arg198, Ile202, Arg204, Pro209, Cys213, Pro214, His229, Pro230, Phe231, Arg236, Arg238, Phe239, Arg240, Asp246, Arg248, Met250, Met255, Met259, His260, Phe266, Trp270
				9-11	TiR	
				109-111	TvR	
				Casein Kinase II Phosphorylation		
				30-33	TvID	
				82-85	SfdD	
				181-184	SihE	
				225-228	SiaE	
				Tyrosine kinase Phosphorylation		
				154-161	Rhl.DafeY	
				261-268	Rhl.EafgY	
				N-myristoylation		
				25-30	GLrpAT	
				105-110	GQpiTV	
145-150	GAcaGH					
221-226	GAvaSI					

Table 3: List of ScanProsite Hits. Motifs mapped on domain along with its modifications

S. No.	Motifs	Site	Position	Modification
1	Protein kinase C phosphorylation site	TiR	9-11	Threonine to Phosphothreonine
		TvR	145 - 147	Threonine to Phosphothreonine
2	N-myristoylation site	GLrpAT	25 - 30	
		GQpiTV	141- 146	
		GAcaGH	181- 186	
		GAvaSI	275-280	
3	Casein kinase II phosphorylation site	TvID	30 - 33	Threonine to Phosphothreonine
		TdvD	46 - 49	Threonine to Phosphothreonine
		SfdD	118 - 121	Serine to Phosphoserine
		SihE	217 - 220	Serine to Phosphoserine
		SiaE	279 - 282	Serine to Phosphoserine
4	Tyrosine kinase phosphorylation site 1	RhlDafeY	190 - 197	
		RlvEafgY	315 - 322	

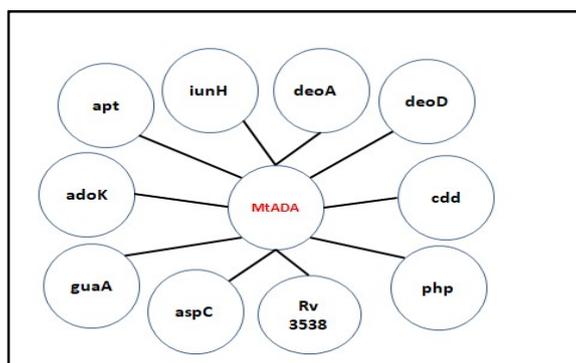


Figure 6: Protein-Protein Interactions analysis of MtADA using STRING database. Ten proteins with known structure were found interacting

Table 4: MtADA interacting proteins and their role in biological process, molecular function, pathways and features using STRING Databases

BIOLOGICAL PROCESS				
#term ID	term description	Observed/ Background gene count	false discovery rate	matching proteins in network
GO:0009116	nucleoside metabolic process	7 of 44	1.38e-09	add,adoK,apt,cdd,deoA,deoD,guaA
GO:0042278	purine nucleoside metabolic process	5 of 22	1.75e-07	add,adoK,apt,deoD,guaA
GO:0042455	ribonucleoside biosynthetic process	5 of 24	1.93e-07	add,adoK,apt,cdd,guaA
GO:0009119	ribonucleoside metabolic process	5 of 33	4.60e-07	add,adoK,apt,cdd,guaA
GO:0043094	cellular metabolic compound salvage	4 of 12	8.62e-07	add,adoK,apt,cdd
GO:0043094	cellular metabolic compound salvage	4 of 12	8.62e-07	add,adoK,apt,cdd
GO:0009112	nucleobase metabolic process	4 of 13	9.00e-07	add,apt,cdd,deoA
GO:0046129	purine ribonucleoside biosynthetic process	4 of 13	9.00e-07	add,adoK,apt,guaA
GO:0046128	purine ribonucleoside metabolic process	4 of 21	3.38e-06	add,adoK,apt,guaA
GO:0009164	nucleoside catabolic process	3 of 4	6.75e-06	add,cdd,deoD
MOLECULAR FUNCTION				
GO:0016763	transferase activity, transferring pentosyl groups	3 of 21	0.0019	apt,deoA,deoD
GO:0019239	deaminase activity	2 of 4	0.0032	add,cdd
GO:0016814	hydrolase activity, acting on carbon-nitrogen bonds, in cyclic amidines	2 of 9	0.0078	add,cdd
GO:0003824	catalytic activity	9 of 1310	0.0147	add,adoK,apt,aspC,cdd,deoA,deoD,guaA,php
GO:0016740	transferase activity	5 of 440	0.0445	adoK,apt,aspC,deoA,deoD
KEGG PATHWAYS				
mtu00230	Purine metabolism	6 of 70	1.88e-07	add,adoK,apt,deoD,guaA,iunH
mtu01100	Metabolic pathways	9 of 596	8.99e-06	add,adoK,apt,aspC,cdd,deoA,deoD,guaA,iunH
mtu00983	Drug metabolism - other enzymes	3 of 11	2.16e-05	cdd,deoA,guaA
mtu00240	Pyrimidine metabolism	3 of 41	0.00056	cdd,deoA,deoD
mtu00760	Nicotinate and nicotinamide metabolism	2 of 19	0.0033	deoD,iunH
UNIPROT KEYWORDS				
KW-0660	Purine salvage	2 of 4	0.0019	adoK,apt
KW-0328	Glycosyltransferase	3 of 52	0.0035	apt,deoA,deoD
KW-0862	Zinc	3 of 92	0.0115	add,cdd,php
KW-0808	Transferase	5 of 476	0.0279	adoK,apt,aspC,deoA,deoD
INTERPRO PROTEIN DOMAINS AND FEATURES				
IPR032466	Metal-dependent hydrolase	2,15	0.0215	add,php

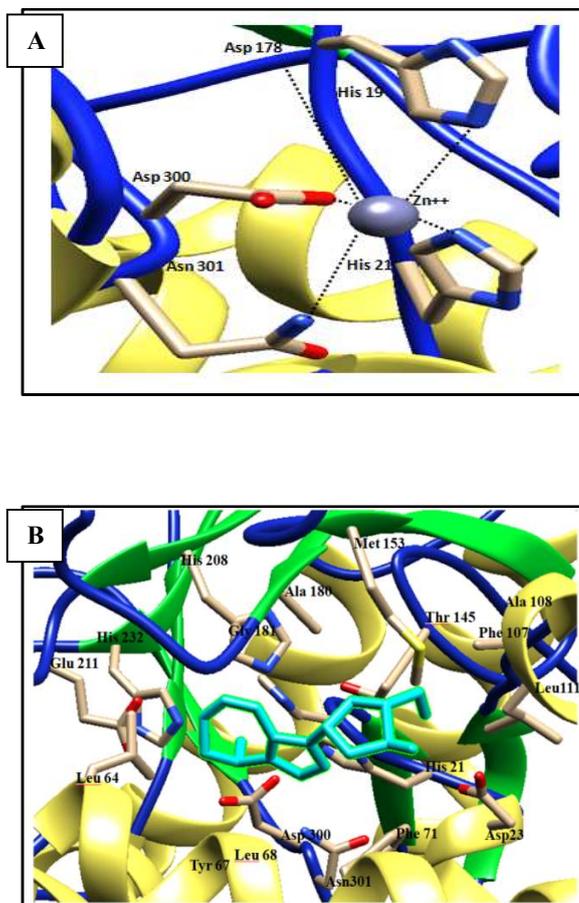


Figure 7: Binding studies using COACH server (a) MtADA in complex to Zinc showing residues participate in binding site using TM site program. (b) MtADA in complex to Deoxycformycin (DCF) showing residues participate in binding site using COFACTOR program

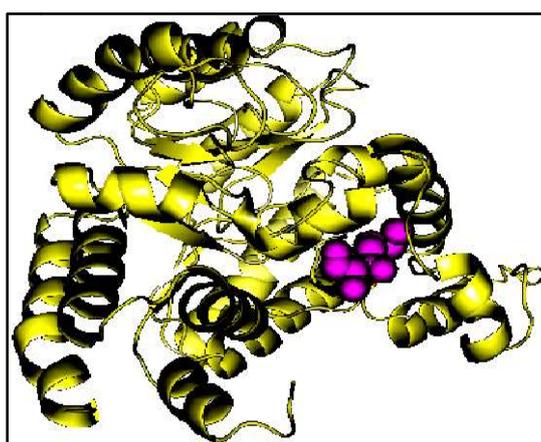


Figure 8: Active sites of MtADA using SiteMap program. Five different active sites were computed and best one is colored megenta in structure (Met309 & Ser310)

Table 5: Active Site prediction of MtADA using SiteMap

S.No.	Title Site score	Size	Sitescore
1.	Sitemap_ada_site_1	132	1.032
2.	Sitemap_ada_site_2	67	0.851
3.	Sitemap_ada_site_3	48	0.745
4.	Sitemap_ada_site_4	37	0.677
5.	Sitemap_ada_site_5	32	0.812

Table 6: Sequence alignment and template comparisons

	Organism	Percent Coverage	Percent Identity
1	<i>Mycobacterium tuberculosis</i>	100	100
2	<i>Mus musculus</i>	91	29.52
3	Human	93.2	28.82
4	<i>Plasmodium vivax</i>	90.7	19.34
5	Human Growth factor	88.2	19.25
6	<i>Pseudomonas aeruginosa</i>	87.7	25.31
7	<i>Burkholderia ambifaria</i>	91.2	22.82
8	<i>Anthrobacter aureescens</i>	91.8	25.37

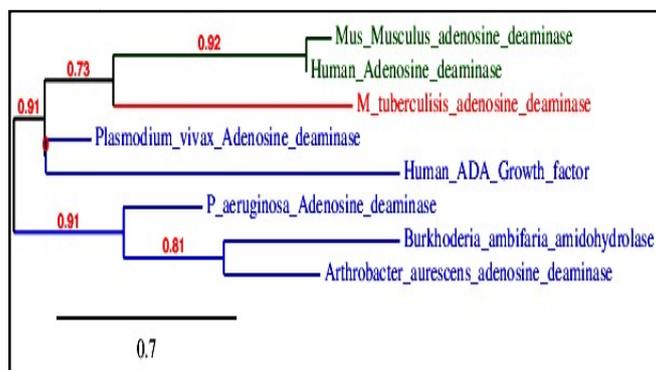
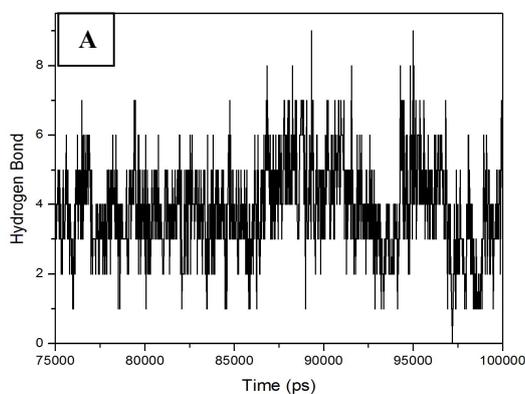


Figure 9: Phylogenetic association of Adenosine deaminase from the ADA gene from different organisms whose structure is already known along with good alignment scores. The evolutionary history was inferred using Clustal omega (Neighbor-Joining method). Branches correspond to the distance to other ADA



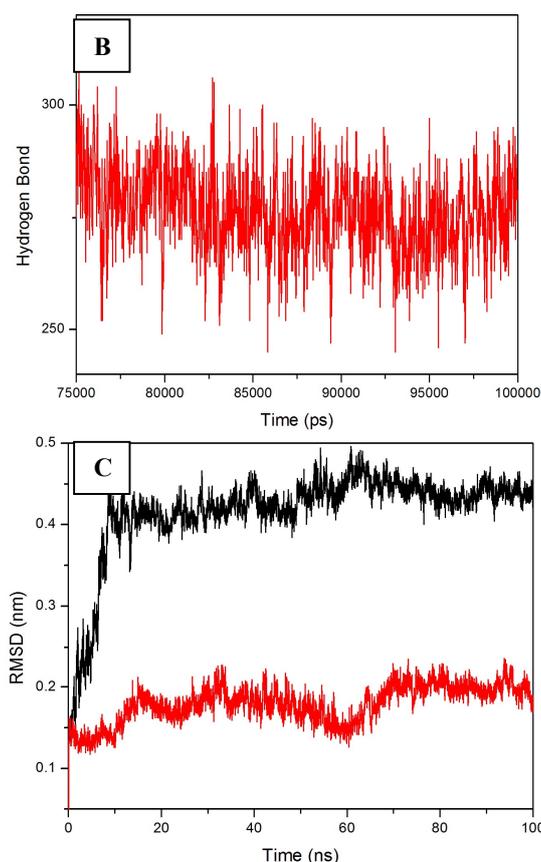


Figure 10: Molecular dynamics simulation studies of MtADA alone and in complex to tRNA. (a) The values of Zinc bound MtADA with no. of h_bond interaction formed in respect to time (black). (b) The value of MtADA docked with tRNA showing no. of h_bond interaction (red). The last 100ns data was taken to check the stability. (c) The RMSD values of both complexes plotted with respect to time

4. CONCLUSION

The current study reported the modeled structure of MtADA. The functional, roles of MtADA were determined. Predicted 3D model of MtADA was observed to accurate for performing molecular docking studies in context of drug development. The active site of MtADA was predicted and found to have amino acid residues Met309 & Ser310. The predicted structure is being used for designing of selective and potent inhibitors of MtADA in search of lead molecules for development of therapeutic molecules against tuberculosis

infections.

ACKNOWLEDGEMENT

The authors acknowledge financial assistance from Indian Council of Medical Research (ICMR) New Delhi. AKD acknowledge the SRF support from ICMR.

5. REFERENCES

- [1] Singh, M. Dysregulated A to I RNA editing and non-coding RNAs in neurodegeneration. *Frontiers in genetics*, (2013) 3, 326.
- [2] Wang, Y., Zheng, Y., & Beal, P. A. Adenosine deaminases that act on RNA

- (ADARs). In *The Enzymes* (2017) (Vol. 41, pp. 215-268). Academic Press.
- [3] Gatsiou, A., Vlachogiannis, N., Lunella, F. F., Sachse, M., & Stellos, K. Adenosine-to-Inosine RNA editing in health and disease. *Antioxidants and Redox Signaling*, (2017) (ja).
- [4] Witkin, K. L., Hanlon, S. E., Strasburger, J. A., Coffin, J. M., Jaffrey, S. R., Howcroft, T. K., & Read-Connole, E. RNA editing, epitranscriptomics, and processing in cancer progression. *Cancer biology & therapy*, (2015) 16(1), 21-27.
- [5] Picardi, E., D'Erchia, A. M., Gallo, A., Montalvo, A., & Pesole, G. Uncovering RNA editing sites in long non-coding RNAs. *Frontiers in bioengineering and biotechnology*, (2014) 2, 64.
- [6] Kim, D., Hwang, J. K., & Pan, J. G. Escherichia coli ASKA Clone Library Harboring tRNA-Specific Adenosine Deaminase (tadA) Reveals Resistance towards Xanthorhizol. *Molecules*, (2015) 20(9), 16290-16305.
- [7] Larson, E. T., Deng, W., Krumm, B. E., Napuli, A., Mueller, N., Van Voorhis, W. C., & Luft, J. Structures of substrate-and inhibitor-bound adenosine deaminase from a human malaria parasite show a dramatic conformational change and shed light on drug selectivity. *Journal of molecular biology*, (2008)381(4), 975-988.
- [8] Kyu, H. H., Maddison, E. R., Henry, N. J., Mumford, J. E., Barber, R., Shields, C., & Wang, H.. The global burden of tuberculosis: results from the Global Burden of Disease Study 2015. *The Lancet Infectious Diseases*, (2018)18(3), 261-284.
- [9] Söding, J., Biegert, A., & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research*, (2005) 33(suppl_2), W244-W248.
- [10] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, (2004) 25(13), 1605-1612.
- [11] Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, (1993)26(2), 283-291.
- [12] Wiederstein, M., & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of

- proteins. *Nucleic acids research*, (2007) 35(suppl_2), W407-W410.
- [13] Wallner, B., & Elofsson, A. Can correct protein models be identified? *Protein science*, (2003)12(5), 1073-1086.
- [14] Eisenberg, D., Lüthy, R., & Bowie, J. U.. [20] VERIFY3D: assessment of protein models with three-dimensional profiles. In *Methods in enzymology* (1997) (Vol. 277, pp. 396-404). Academic Press.
- [15] Maiti, R., Van Domselaar, G. H., Zhang, H., & Wishart, D. S. SuperPose: a simple server for sophisticated structural superposition. *Nucleic acids research*, (2004)32 (suppl_2), W590-W594.
- [16] Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D., & Wishart, D. S. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic acids research*, (2003) 31(13), 3316-3319.
- [17] Laskowski, R. A., Watson, J. D., & Thornton, J. M. ProFunc: a server for predicting protein function from 3D structure. *Nucleic acids research*, (2005)33(suppl_2), W89-W93.
- [18] Pugalenth, G., Archunan, G., & Sowdhamini, R. DIAL: a web-based server for the automatic identification of structural domains in proteins. *Nucleic acids research*, (2005) 33(suppl_2), W130-W132.
- [19] Milburn, D., Laskowski, R. A., & Thornton, J. M. Sequences annotated by structure: a tool to facilitate the use of structural information in sequence analysis. *Protein engineering*, (1998) 11(10), 855-859.
- [20] De Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., & Hulo, N.. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research*, (2006) 34(suppl_2), W362-W365.
- [21] Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., & Bork, P. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research*, (2008) 37(suppl_1), D412-D416.
- [22] Yang, J., Roy, A., & Zhang, Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, (2013) 29(20), 2588-2595.
- [23] <http://www.schrodinger.com/citations#Maestro>.
- [24] Sievers, Fabian, and Desmond G. Higgins. "Clustal Omega for making accurate alignments of many protein sequences." *Protein Science* (2018) 27.1: 135-145.

- [25] Baum, Bernard R. "PHYMLIP: phylogeny inference package. Version 3.2. Joel Felsenstein." *Q Rev Biol* (1989)64.4: 539-541.
- [26] Vermaas, Josh V., *et al.* "TopoGromacs: Automated topology conversion from CHARMM to GROMACS within VMD." (2016): 1112-1116.
- [27] Lindberg, Max. "Fluorescent fusion proteins as probes to characterize tau fibril polymorphism". (2019).