OPEN ACCESS
**Haya: The Saudi Journal of Life Sciences**
Abbreviated Key Title: Haya Saudi J Life Sci
ISSN 2415-623X (Print) |ISSN 2415-6221 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com/sjls

**Original Research Article**

# In Silico Analysis of Molecular Interaction of EhSir2a with its Interacting Proteins from Human Pathogen *Entamoeba histolytica*

Pinaki Biswas[1], Somasri Dam[2*]

[1]Department of Microbiology, The University of Burdwan, Burdwan-713104, West Bengal, India
[2]Assistant Professor, Department of Microbiology, The University of Burdwan, Burdwan-713104, West Bengal, India

**DOI:** 10.36348/sjls.2020.v05i08.002 | **Received:** 30.07.2020 | **Accepted:** 07.08.2020 | **Published:** 22.08.2020

***Corresponding author:** Somasri Dam

## Abstract

The human pathogen, *Entamoeba histolytica* contains four Sir2 homologs in its genome. We have designed the 3D tertiary structure of EhSir2a and its interacting partners by comparative homology modeling and studied their interaction by molecular docking. Sir2 proteins are known to interact with their substrates through the deacetylase domain present in its C-terminus. Interestingly, EhSir2a contains a unique Zn finger domain at its N terminus and this is not present in any known Sir2 protein. This study shows that EhSir2a may interact with this N-terminal residue also. It interacts with its substrate, elongation factor EhEF2 through this zinc-finger domain. The interaction sites are different for alpha-tubulin homologs, the other substrates of EhSir2a identified by yeast two-hybrid library screening. The coordinate files of the best-modeled structure for EhSir2a and its interacting proteins were processed for protein-protein docking using ClusPro v2.0 and HawkDock server. Tyr432, Asn422, Arg408 residues of α-tubulin are essential for interaction with EhSir2a. Here we report that molecular interactions of EhSir2a with its interacting partners are not restricted to the conserved $NAD^+$ dependent deacetylase domain; it may also involve the N-terminal residue.

**Keywords:** *Entamoeba histolytica*, Sirtuin, Homology modelling, Molecular docking, Gromacs.



## INTRODUCTION

The protozoan parasite *Entamoeba histolytica* is an etiological agent of amoebiasis in humans. Amoebiasis was recorded as the third leading cause of death from parasitic infection worldwide, with its greatest impact on the people of developing countries [1]. Approximately 50 million people worldwide suffer from invasive amoebic infection each year, resulting in 40-100 thousand deaths annually [2, 3]. Unicellular protozoa like *E. histolytica* is the early evolutionary divergent of eukaryotic lineage. The mode of nutrition in many such protozoa is parasitic which makes them medically significant. Thus, *E. histolytica* being a protozoan parasite may serve as a model organism to study not only their cellular biology and evolution but also to understand the mechanism of pathogenesis and identification of potential drug targets. *E. histolytica* is different from other pathogenic protozoa by having an unusual mode of cell cycle events [4-8]. The parasite is prone to oxidative stress due to the lack of most of the antioxidant defense mechanisms such as glutathione peroxidase, glutathione reductase, and catalase enzymes [9, 10]. Identification of different essential proteins in *E. histolytica* may enlighten us with their physiological role and related interactome. Sirtuins are $NAD^+$-dependent class III histone deacetylases which are well conserved and widely distributed among all domains of life from archaea to eukarya [11]. Deacetylation of substrates like histone and other non-histone proteins require $NAD^+$ as a co-substrate and nicotinamide and O-acetyl-ADP-ribose are released as metabolites [12]. Increasing cellular concentration of nicotinamide may feedback-inhibit sirtuin activity by non-competitive binding [13, 14]. Likewise, the other metabolite, O-acetyl-ADP-ribose has also been reported as a signaling molecule [15-18]. Sirtuin may provide a direct link between metabolic programming and the control of gene expression. Previous study of sirtuins in protozoan parasites like, *Entamoeba histolytica* [19], *Plasmodium falciparum* [20, 21], *Leishmania donovani* [22, 23], *Leishmania infantum* [24-28], *Leishmania amazonensis* [29], *Eimeria tenella* [30], *Cryptosporidium parvum* [31], *Cryptosporidium hominis* [32], *Trypanosoma cruzi* [33, 34], *Trypanosoma brucei* [35] and *Giardia lambia* [36, 37] reflect their versatile functionality in cellular processes [38-40].

EhSir2a regulates microtubule assembly during cell cycle progression by tubulin deacetylation. The interactors of EhSir2a include two alpha-tubulin homologs (EhAT1, Acc. no. XP_650067.1 and EhAT2, Acc. no. XP_653419.1), elongation factor 2 (EhEF2, Acc. no. XP_651009.2), putative proteasome beta subunit (EhPBST3, Acc no. XP_655858.1), translation initiation factor (EhMIF4G, Acc. no. XP_654481.1), phospholipase B (EhPLB1, Acc. no. XP_654113.1) and serine/threonine phosphatase [19]. In the present study, we have constructed a three-dimensional structural model of EhSir2a and its interactors (EhAT1, EhAT2, EhEF2, and EhPBST3). The molecular interactions between these proteins have been reported here.

## MATERIAL AND METHODS

### EhSir2a interacting proteins

EhSir2a interacting proteins were searched using the FASTA sequence of EhSir2a as a query in the latest version of the STRING v11.0 database at a confidence level of 0.4. Interaction of EhSir2a with its protein partners was investigated based on text mining, co-expression, experimental data, known metabolic, and signal transduction pathways [41, 42].

### Homology modeling and structural analysis

The amino acid sequence of EhSir2a (Acc. Id: XP_657434.1) and its interactors (EhAT1, Acc. Id: XP_650067.1; EhAT2, Acc. Id: XP_653419.1; EhEF2, Acc. Id: XP_651009.2; EhPBST3, Acc. Id: XP_655858.1) were retrieved from NCBI protein database in FASTA format. The interactors were obtained by yeast two-hybrid screening against the cDNA library and also by STRING analysis. For homology modeling, the SWISS-MODEL server is used to search a suitable template in SWISS-MODEL Template Library for evolutionary related structures matching the target sequence [43]. Among all the templates, the best one was determined by an optimal combination of high coverage, sequence similarity, sequence identity, and low resolution. Homology based modeling was done with MODELLER v9.20 [44], and Galaxy TBM followed by Galaxy Refine [45]. The quality and the integrity of the predicted models were evaluated by QMEAN scoring function [46], Ramachandran plot (RAMPAGE) [47], ResProx (Resolution-by-proxy) [48], ProQ (Protein quality prediction) [49], ProSA (Protein Structure Analysis) [50] and SAVES v5 (Verify 3D, ERRAT, PROVE, PROCHECK, WHATCHECK). All the protein structures were visually represented with the help of PyMOL. Protein structures were analyzed by the PDBsum tool [51].

### Molecular docking of EhSir2a and its interactors

Molecular Docking study was performed by ClusPro (v2.0), a blind rigid PIPER docking program based on fast Fourier transformation to generate low energy interaction conformations of a protein-protein complex using the pairwise docking potentials. Stable interaction complex was refined by filtering and clustering of docked confrontation using pairwise RMSD subsequently stabilization using Monte–Carlo simulations. ClusPro presents the protein-protein interaction in four different modes: Balanced, Electrostatic-favoured, Hydrophobic-favored, and VdW+Elec favored. Protein-protein interactions with the highest members in cluster formation of each interaction mode were selected [52-54]. HawkDock Server treats a comparatively smaller protein as a flexible and larger protein as a stationary receptor. ATTRACT docking algorithm, HawkRank scoring function with MM/GBSA free energy decomposition analysis was employed to predict the binding free energy and decompose the free energy contributions to the binding free energy of a protein-protein complex in per residue. The best ten models of interacting proteins were re-ranked by MM/GBSA (Molecular Mechanics energies combined with the Generalized Born and Surface Area continuum solvation) calculation [55, 56]. All protein-protein interactions were represented diagrammatically using the LigPlot program [57]. All scoring functions and free energy were calculated as described in respective references.

### Prediction of the functional and biological role of EhSir2a

COFACTOR server was used to predict the functional and biological significance of protein molecules by analyzing its structure, sequence, and protein-protein interaction. 3D structural model file EhSir2a in PDB format was submitted and threaded through the BioLiP protein function database on local and as well as global scale to identify functional sites and homologies. Functional insights, including Gene Ontology (GO), Enzyme Commission (EC), and ligand-binding sites are predicted from the best templates depending on functional homology. For GO, it acquires data by sequence, sequence-profile alignments from UniProt-GOA, and by protein-protein interaction from STRING [58, 59].

### Molecular Dynamics Simulation

Molecular dynamics (MD) simulation of Ehsir2a was performed individually using GROMACS 5.1.4 [60]. The geometry of each complex was regularized using the GROMOS96 54a7 force field [61]. All the structures were placed in the center of a cubic box with a minimum distance of 1 nm between the protein and the wall of the box on all sides and the box was solvated using SPC/E water. The initial charge of the system was neutralized by adding counter ions of $Na^+$ and $Cl^-$. The salt concentration was set to 150 mM to mimic the physiological ion concentration. To remove any steric clashes, all the molecular systems were energy minimized using the steepest descent algorithm. The system was equilibrated under NVT (constant Number of particles, Volume, and Temperature) conditions for 100 ps at 300 K using the Berendsen thermostat and followed by NPT (constant

Number of particles, Pressure, and Temperature) condition pressure was equilibrated for 100 ps to 1 atm using the Berendsen barostat. During both the equilibrations, all the heavy atoms of the proteins were position restrained with a force constant of 1000 kJ $mol^{-1}nm^{-2}$. $E_{pot}$ should be negative, and on the order of $10^5$-$10^6$, depending on the system size and number of water molecules. Energy minimization gave us a reasonable starting structure, in terms of geometry and solvent orientation. Electrostatic interactions were calculated using Particle Mesh Ewald (PME) formalism. Equilibration of the solvent and ions around protein was done to begin molecular dynamics. A position restraining force can permit the movement of the heavy atoms of the protein except hydrogen atom, but only after overcoming a substantial energy penalty. The utility of position restraints is that they allow us to equilibrate our solvent around our protein, without the added variable of structural changes in the protein. Finally, 10 ns MD simulations were performed for each complex to analyze the stability of the system. The trajectory produced in MD simulation was analyzed using gmx rms, gmx rmsf, gmx gyrate, and gmx hbond of GROMACS utilities to obtain the root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), the radius of gyration ($R_g$) and the number of H-bond s formed in EhSir2a. The differences in the energies like kinetic, potential, total, and pressure, and temperature were computed as a function of simulation time to check whether the systems obey NVT or NPT ensemble throughout the simulation. The trajectories were analyzed using the tools from GROMACS distribution. All the graphs were generated using the Grace tool (https://plasma-gate.weizmann.ac.il/Grace/).

# RESULTS

### Predicted interactors of EhSir2a

EhSir2a interacts with various proteins as found from STRING v11.0, to execute different types of molecular actions such as activation, inhibition, binding, phenotype, catalysis, post-translational modification, and gene expression. The interactors are predicted to be involved in the DNA repair pathway, chromosome structure maintenance, genome segregation, chromatin dynamics, transcription, and stress response. Some uncharacterized hypothetical proteins are also found as substrates of EhSir2a.

### Zinc finger domain of EhSir2a

Bioinformatic analysis of EhSir2a showed that it has a unique N-terminal sequence of 134 amino acid residues in which a Zn-finger (ZnF) domain of 62 residues is embedded. The catalytic deacetylase domain is 240 amino acids long in EhSir2a. N-terminal part of EhSir2a is predominated by positively charged residues (Arginine and Lysine) and also unique as it is absent in other EhSir2 homologs. Zn-finger domain in EhSir2a is structured by five α-helices and one β-sheet but the crystallographic data showed the presence of two β-sheets and one α-helix in classical zinc-finger domain [62, 63] which indicates that ZnF in EhSir2a belongs to non-classical type where the position of cysteine/histidine combinations are very much different from currently approved 30 types of ZnF by the HUGO Gene Nomenclature Committee [62, 64].

### Homology modeling

Homology modeling of EhSir2a and its interactors were built by using suitable templates and their quality was evaluated to determine the best stable model (Table-1).

**Table-1: Template, quality estimation and evaluation of selected model of EhSir2a and its interactors using RAMPAGE: Assessment of the Ramachandran Plot, ProQ (Protein Quality Prediction), ResProx (Resolution-by-proxy), QMEAN4 score and ProSA**

| Parameters | | | Name of Proteins | | | | |
|---|---|---|---|---|---|---|---|
| | | | EhSir2a | EhAT1 | EhAT2 | EhEF2 | EhPBST3 |
| Name of the template | | | 4RMH | 5UBQ, 5IYZ | 5UBQ, 5IYZ | 3J7P | 5LE5, 5T0H |
| Quality Estimation Tools | RAMPAGE (Ramachandran plot) | Number of residues in favoured region | 99.5% | 98.5% | 98.0% | 97.0% | 96.0% |
| | | Number of residues in allowed region | 0.5% | 0.7% | 1.3% | 2.3% | 4.0% |
| | | Number of residues in outlier region | 0.0% | 0.9% | 0.7% | 0.7% | 0.0% |
| | ProQ | LG score | 2.914 | 4.566 | 4.509 | 4.772 | 3.960 |
| | | MaxSub | 0.197 | 0.353 | 0.365 | 0.279 | 0.302 |
| | ResProx Value | | 1.766 | 1.656 | 1.803 | 1.946 | 1.853 |
| | QMEAN4 value | | -1.93 | -1.98 | -2.08 | -0.76 | -0.26 |
| | All Atom | | -1.14 | -0.30 | -0.15 | -0.43 | 0.37 |
| | Cβ | | -1.43 | -1.32 | -0.90 | -1.31 | -1.15 |
| | Solvation | | 0.93 | 0.76 | 1.07 | 1.22 | 0.38 |
| | Torsion | | -1.88 | -2.07 | -2.36 | -0.91 | -0.19 |
| | ProSA (Overall model quality in Z-score) | | -5.72 | -10.03 | -10.17 | -12.66 | -8.78 |

All the PDB coordinate data in each of the modeled protein structures have been predicted with atomic resolution below 2Å whereas, a typical crystallographic model based on 2.0 Å data has a coordinate error of less than 0.2 Å. Resolution below 2 Å suggests that the structural quality of protein models was built with fewer systematic errors (such as missing or misplaced atoms) and with a less average coordinate error. Predicted QMEAN Z-score of modeled protein structures suggests that the 'degree of nativeness' or calculated secondary structure matched with the expected values from a representative set of high-resolution experimental structures. LG score and MaxSub values predicted by ProQ for the modeled proteins are near or above 3.0 and 0.1 proposed that structural quality is satisfactory for the theoretical model. Predicted Ramachandran plot of modeled protein structures showed above 95% of the amino acid residues fall into the favoured region, which indicates that theoretical conformations of protein have very less amount of steric hindrance or clashes between atoms with an acceptable range of phi and psi angles. ProSA predicts that the overall structural quality of the modeled proteins has a similar score as native proteins of the same residue obtained from X-ray analysis, NMR spectroscopy, and theoretical calculations. The best stable structured protein model of EhSir2a, EhAT1, EhAT2, EhEF2 and EhPBST3 were deposited in Protein Model DataBase (https://bioinformatics.cineca.it/PMDB/) as EhSir2a.pdb (PMDB id: PM0081615), EhAT1 (PMDB id: PM0082260), EhAT2 (PMDB id: PM0082261), EhEF2 (PMDB id: PM0082262) and EhPBST3 (PMDB id: PM0082264).

**Structural analysis of EhSir2a and its interactors**

The tertiary structure of modeled EhSir2a showed the presence of 6 parallel and 3 antiparallel β-sheets, 3 β-α-β units, 1 β-hairpin, 3 β-bulges (one parallel classic, one antiparallel G1 and one antiparallel special), 24 α-helices, 23 helix-helix interactions and 34 β-turns (Figure-1).

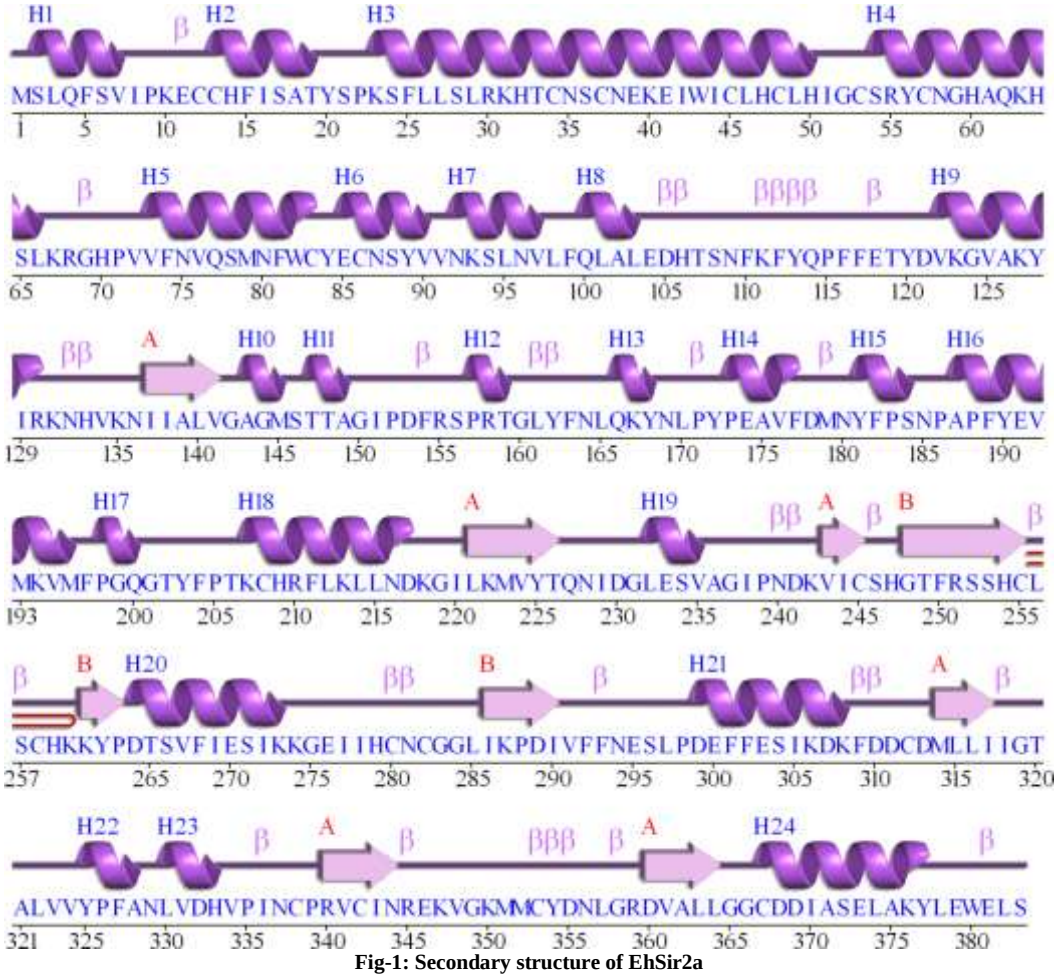
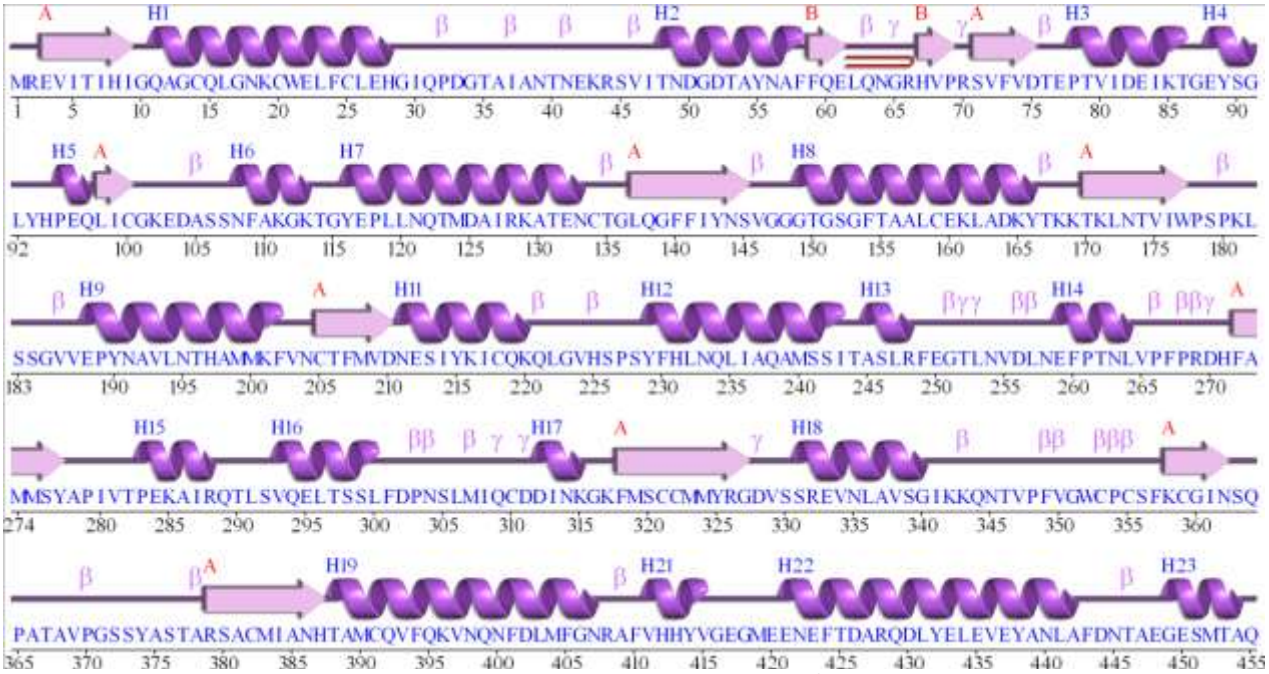


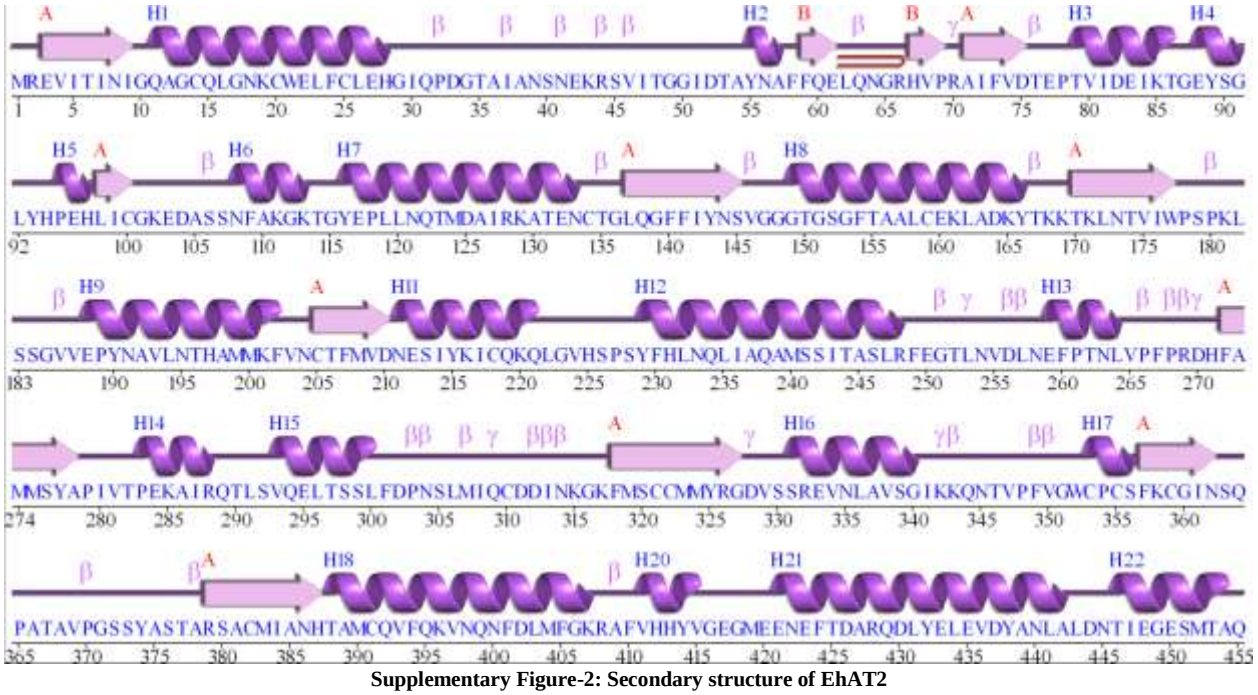

**Fig-1: Secondary structure of EhSir2a**

Theoretically designed structure of EhAT1 have 23 α-helices, 12 β-strands (10 mixed type and 2 antiparallel type), 5 β-α-β units, 1 β-hairpin, 1 ψ-loop, 2 β-bulges (one parallel classic and one antiparallel classic type), 26 helix-helix interactions, 33 β-turns, 9 γ-turns (one classic type and 8 inverse type) (Supplementary Figure-1).



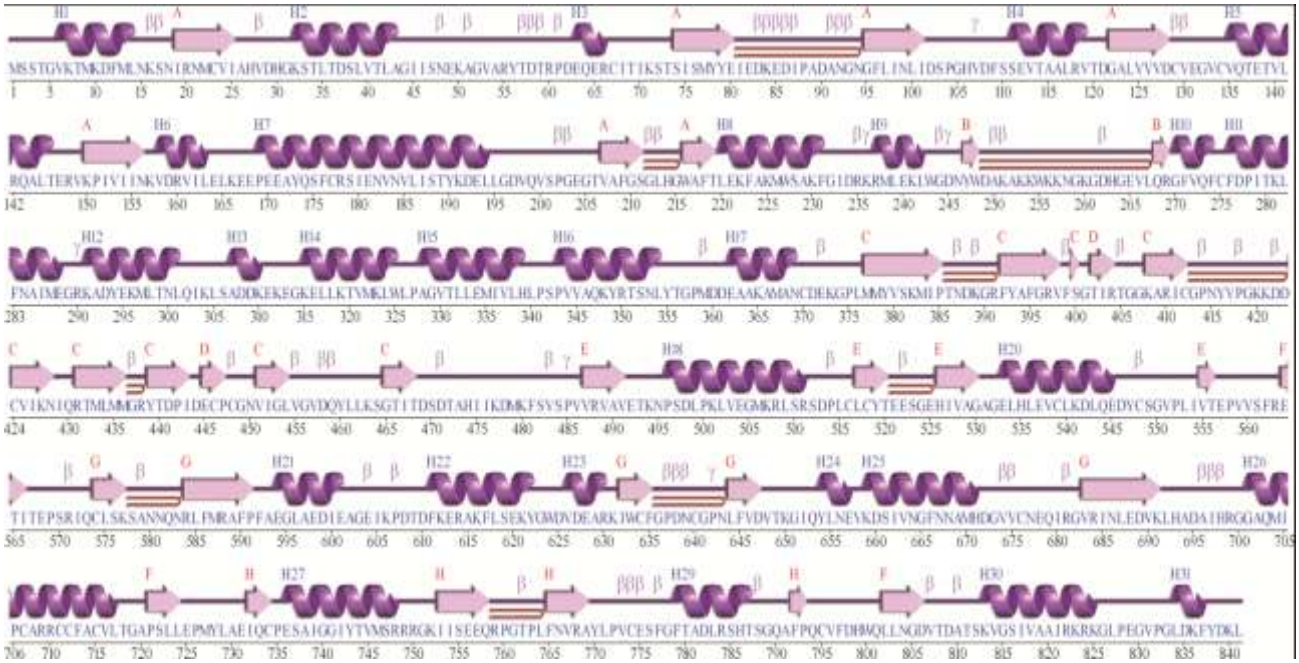

**Supplementary Figure-1: Secondary structure of EhAT1**

Structurally, EhAT2 is composed of 22 α-helices, 12 β-strands (10 mixed type and 2 antiparallel type), 5β-α-β units, 1 β-hairpin, 1 ψ-loop, 2 β-bulges (one parallel classic and one antiparallel classic type), 20 helix-helix interactions, 31 β-turns, 8 γ-turns (two classic type and 6 inverse type) (Supplementary Figure-2).



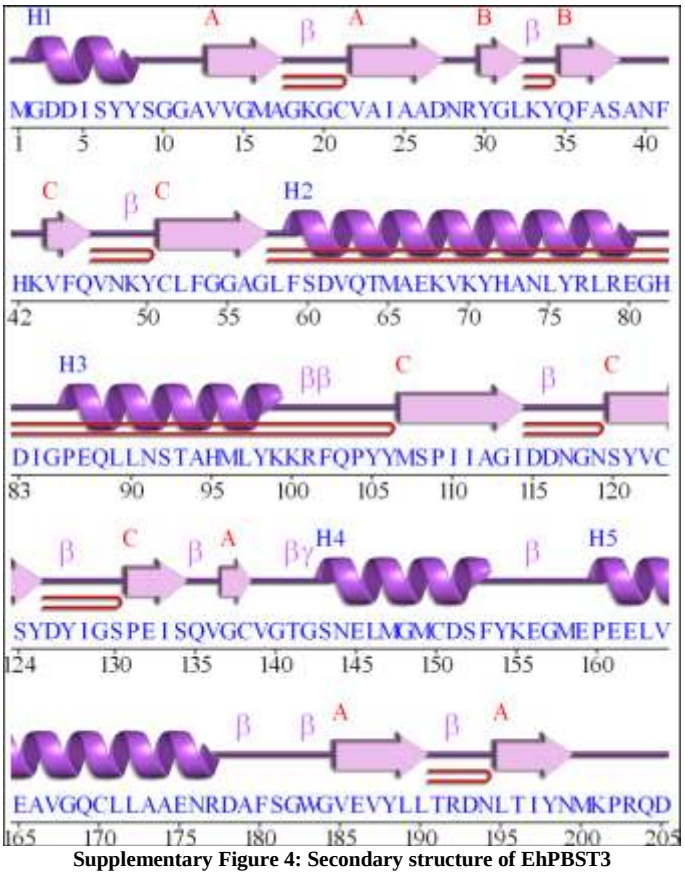

**Supplementary Figure-2: Secondary structure of EhAT2**

EhEF2 protein structure is contributed by 31 α-helices, 35 β-strands (12 mixed type and 23 antiparallel type), 3 β-α-β units, 10 β-hairpins, 1 ψ-loop, 11 β-bulges (one special antiparallel type, four G1 antiparallel type and six classic antiparallel type), 33 helix-helix interactions, 68 β-turns and 8 γ-turns (two classic type and six inverse type) (Supplementary Figure-3).



**Supplementary Figure-3: Secondary structure of EhEF2**

Structurally, EhPBST3 has 5 α-helices, 7 β-hairpins, 3 β-bulges (one classic antiparallel and two antiparallel G1 type), 12 antiparallel β-strands, 2 helix-helix interactions, 13 β-turns and 1 inverse type γ-turn (Supplementary Figure-4).




**Supplementary Figure 4: Secondary structure of EhPBST3**

**Molecular interaction of EhSir2a with its interactors**

Molecular Docking study was executed using the ClusPro v2.0 protein-protein molecular docking program. Among the four types of interaction modes, a balanced form of interaction was visualized by PyMOL (Table-2).

**Table-2: Molecular Docking of EhSir2a with its interactors using ClusPro v2.0**

| | ClusPro v2.0 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Balanced | | | Electrostatic-favored | | | Hydrophobic-favored | | | VdW+Elec | | |
| | Members | Representative | Weighted Score | Members | Representative | Weighted Score | Members | Representative | Weighted Score | Members | Representative | Weighted Score |
| EhSir2a + EhAT1 | 66 | Center | -848.6 | 71 | Center | -924.9 | 140 | Center | -1248.2 | 130 | Center | -246.2 |
| | | Lowest Energy | -985.2 | | Lowest Energy | -1093.4 | | Lowest Energy | -1248.2 | | Lowest Energy | -301.4 |
| EhSir2a + EhAT2 | 48 | Center | -986.2 | 49 | Center | -1086.4 | 61 | Center | -1034.3 | 126 | Center | -256.0 |
| | | Lowest Energy | -1083.7 | | Lowest Energy | -1086.4 | | Lowest Energy | -1217.7 | | Lowest Energy | -306.1 |
| EhSir2a + EhEF2 | 58 | Center | -901.0 | 51 | Center | -965.2 | 145 | Center | -994.8 | 58 | Center | -269.4 |
| | | Lowest Energy | -1088.4 | | Lowest Energy | -1184.2 | | Lowest Energy | -1368.4 | | Lowest Energy | -308.2 |
| EhSir2a + EhPBST3 | 89 | Center | -859.5 | 62 | Center | -892.3 | 119 | Center | -1155.3 | 238 | Center | -270.2 |
| | | Lowest Energy | -1044.8 | | Lowest Energy | -1059.2 | | Lowest Energy | -1387.9 | | Lowest Energy | -282.9 |

We further analyzed the docked conformation to find out the binding mode of interaction between the EhSir2a-EhSir2a interactor complex. Although the hydrophobic interactions are greatly responsible but comparative analysis of protein-protein interaction study with the HawkDock server showed hydrogen bonds are also necessary to stabilize the interaction between the proteins. Interaction study of EhSir2a with EhAT1, EhAT2, EhEF2, and EhPBST3 was carried out by HawkDock evaluated lowest free energy of binding were -54.46 Kcal.mol$^{-1}$, -47.21 Kcal.mol$^{-1}$, -41.99 Kcal.mol$^{-1}$ and -46.38 Kcal.mol$^{-1}$ respectively (Table-3).

**Table-3: Structural prediction and analysis of protein-protein complex of EhSir2a and its interactors using HawkDock server**

| | HawkDock Server | | No. of H-bonds | Residues involved in H-bond interaction | | Bond length (Å) | Residues involved in hydrophobic contacts | |
|---|---|---|---|---|---|---|---|---|
| | HawkDock Score | MM/GBSA Binding free energy of complex | | EhSir2a (B) | EhSir2a interactor (A) | | EhSir2a (B) | EhSir2a interactor (A) |
| EhSir2a + EhAT1 | -4861.48 | -54.46 kcal/mol | 9 | (Tyr203)O | (Arg269)NH1 | 3.01 | Phe204, Leu214, Lys213, Pro205, Val235, Asn240, Ile238, Ser234, Lys261, Pro263, Asp264, Arg251 | Lys201, Phe267, Phe405, Arg408, Arg428, Asp430, Tyr432, Gly449, Glu450, Met452, Ala454 |
| | | | | (Glu381)OE2 | (Asn422)ND2 | 2.81 | | |
| | | | | (Glu381)O | (Thr425)OG1 | 2.97 | | |
| | | | | (Leu382)O | (Thr425)OG1 | 1.53 | | |
| | | | | (Ser383)O | (Gln429)NE2 | 3.30 | | |
| | | | | (Arg210)NH1 | (Glu433)OE2 | 2.63 | | |
| | | | | (Arg210)NH2 | (Glu433)OE2 | 2.45 | | |
| | | | | (Thr265)N | (Thr453)O | 2.98 | | |
| | | | | (Thr265)N | (Gln455)O | 2.85 | | |
| EhSir2a + EhAT2 | -5059.60 | -47.21 kcal/mol | 5 | (Asn92)OD1 | (Lys315)NZ | 2.70 | Lys93, Asn96, Val97, Thr107, Leu101, His106, Thr33, Lys111, Ser108, Phe112, Phe99, Asn80, Leu98 | Asn314, Lys317, Gln395, Gln399, Phe405, Gln429, Tyr432, Glu435, Val436, Ala439, Asn440, Ala442, Leu443 |
| | | | | (Glu41)OE2 | (Arg408)NH1 | 3.13 | | |
| | | | | (Asn109)OD1 | (Asn398)ND2 | 2.29 | | |
| | | | | (Cys37) SG | (Asn422)OD1 | 2.28 | | |
| | | | | (Asn109)O | (Arg428)NH1 | 2.11 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| EhSir2a + EhEF2 | -4707.73 | -41.99 kcal/mol | 7 | (Lys350)NZ | (Glu168)OE1 | 3.08 | Pro71, His70, Leu364, Ala18, Tyr20, Ser21, Leu98, Val97, Ser94, Val91, Asn87, Tyr84, Met79 | Val130, Glu131, Leu165, Gly264, Ser737, Ile739, Gly740, Gly741, Tyr743, Thr744, His786, Thr787 |
| | | | | (Lys347)NZ | (Lys166)O | 3.29 | | |
| | | | | (Lys347)NZ | (Glu168)OE1 | 2.30 | | |
| | | | | (Asn356)ND2 | (Glu171)OE2 | 2.25 | | |
| | | | | (Thr19)O | (Lys260)NZ | 2.21 | | |
| | | | | (Cys83)SG | (Ser785)O | 3.16 | | |
| | | | | (Asn80)OD1 | (Ser788)OG | 2.31 | | |
| EhSir2a + EhPBST3 | -5597.87 | -46.38 kcal/mol | 5 | (Asn356)ND2 | (Asp151)OD2 | 3.10 | Lys350, Phe25, Lys111, Pro22, Cys353, Phe74, Val72, Val73, Val76, Cys12, Ala18, Tyr20, Thr19, Phe15, Phe99 | Met147, Ser143, Ser130, Arg177, Ile128, Ile133, Leu90, Asp178, Glu87, Tyr7, Phe180, Lys33, Asn91 |
| | | | | (Met352)CE | (Asn144)ND2 | 3.18 | | |
| | | | | (Lys10)LZ | (Asp4)OD1 | 2.94 | | |
| | | | | (Pro71)O | (Gln135)NE2 | 2.51 | | |
| | | | | (His14)O | (Tyr34)OH | 2.92 | | |

Alpha-tubulin conserved domain of both EhAT1 and EhAT2 have following nucleotide-binding sites at R2, Q11, A12, Q15, E77, P78, T79, S106, S107, N108, K111, G148, G149, T150, G151, I176, V186, E188, N204, N211, Y229, N233. Interaction analysis of both α-tubulin proteins with EhSir2a showed that hydroxyl amino acids and acidic amino acids are essential for the formation of H-bonds. Tyr432, Asn422, Arg408 residues of EhAT1, and EhAT2 are common in interaction with Ehsir2a. This implies that Tyr432, Asn422, Arg408 of α-tubulin are essential for interaction with EhSir2a (Figure-2, Supplementary Figure 5-8). The molecular interaction of EhSir2a with its interacting proteins is shown in Figure-3.

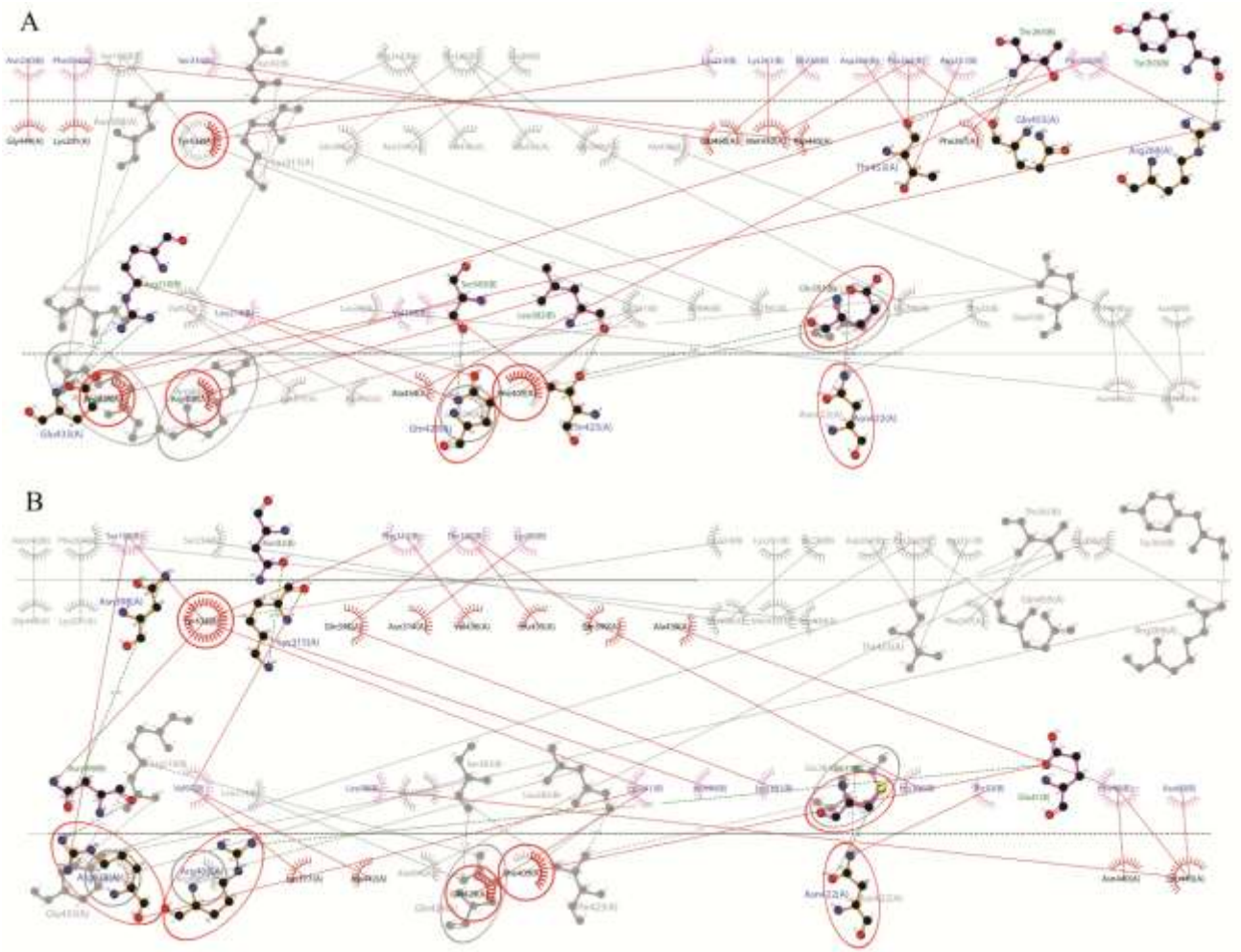
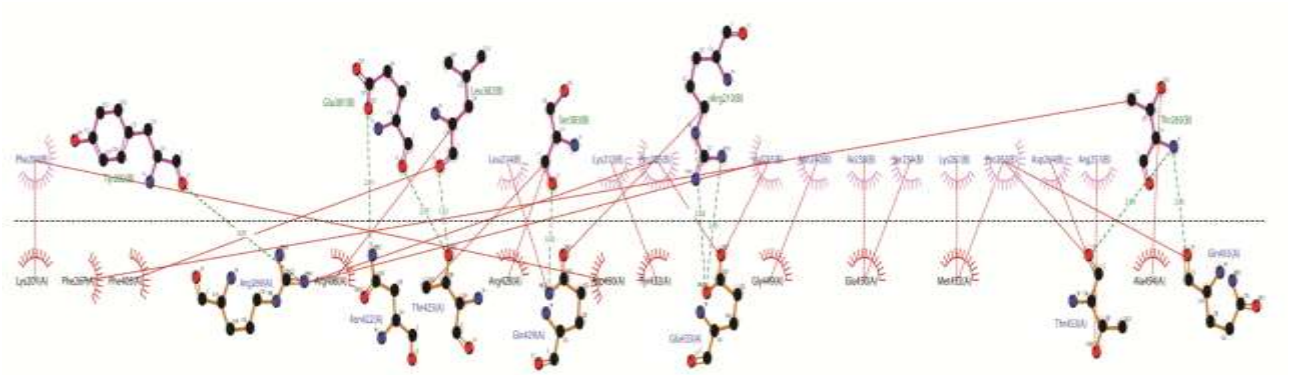
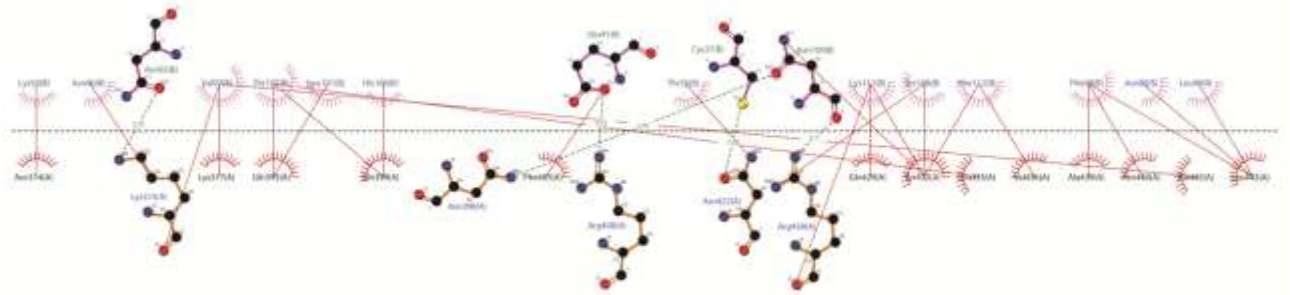


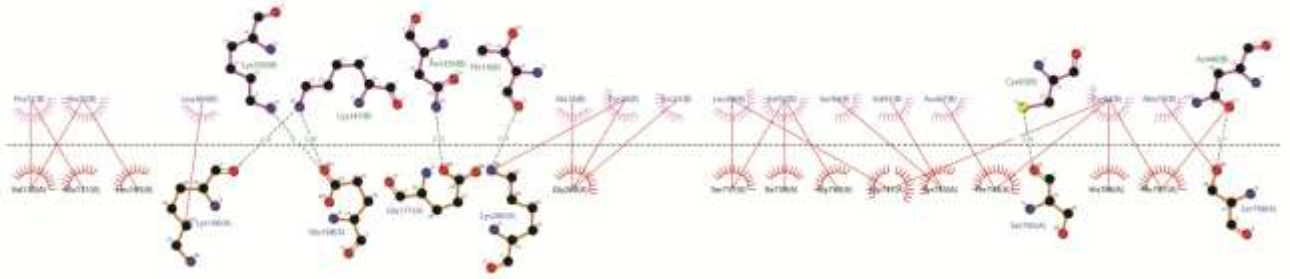

**Fig-2: Merged representation of molecular interactions between EhSir2a and two isoforms of α-tubulin, EhAT1 and EhAT2. (A) Interaction with EhAT1 is actively highlighted (B) Interaction with EhAT2 is actively highlighted. Encircled amino acids are common residues involved in interaction with EhSir2a with both the isoform of Eh α-tubulin**



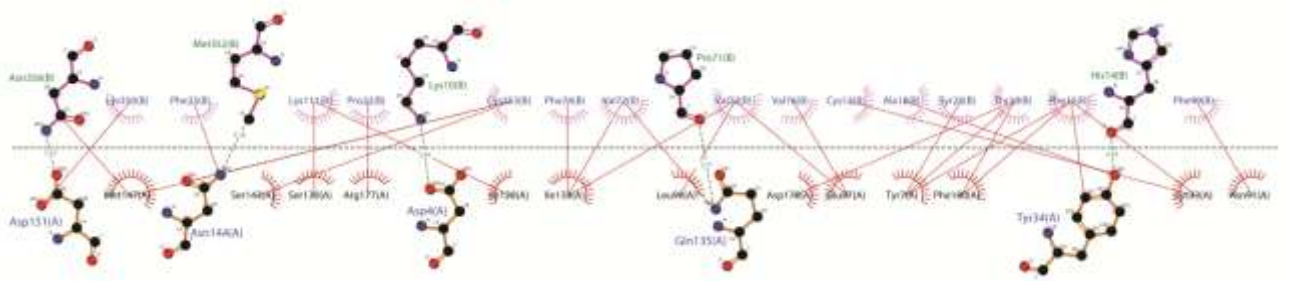
**Supplementary Figure-5: Schematic representation of the molecular interactions between EhSir2a and EhAT1. Amino acids of EhSir2a (B) (pink coloured arc) involved in hydrophobic contacts with amino acids of EhAT1 (A) (red coloured arc) are indicated by an arc with spokes radiating towards each other. The atoms of amino acids involved in hydrophobic interaction have small red spokes on them. The horizontal black dash line represents the interface. The green coloured dash line indicates the hydrogen bond between corresponding atoms with its bond length**



**Supplementary Figure-6: Schematic representation of the molecular interactions between EhSir2a and EhAT2. Amino acids of EhSir2a (B) (pink coloured arc) involved in hydrophobic contacts with amino acids of EhAT2 (A) (red coloured arc) are indicated by an arc with spokes radiating towards each other. The atoms of amino acids involved in hydrophobic interaction have small red spokes on them. The horizontal black dash line represents the interface. The green coloured dash line indicates the hydrogen bond between corresponding atoms with its bond length**



**Supplementary Figure-7: Schematic representation of the molecular interactions between EhSir2a and EhEF2. Amino acids of EhSir2a (B) (pink coloured arc) involved in hydrophobic contacts with amino acids of EhEF2 (A) (red coloured arc) are indicated by an arc with spokes radiating towards each other. The atoms of amino acids involved in hydrophobic interaction have small red spokes on them. The horizontal black dash line represents the interface. The green coloured dash line indicates the hydrogen bond between corresponding atoms with its bond length**



**Supplementary Figure-8: Schematic representation of the molecular interactions between EhSir2a and EhPBST3. Amino acids of EhSir2a (B) (pink coloured arc) involved in hydrophobic contacts with amino acids of EhPBST3 (A) (red coloured arc) are indicated by an arc with spokes radiating towards each other. The atoms of amino acids involved in hydrophobic interaction have small red spokes on them. The horizontal black dash line represents the interface. The green coloured dash line indicates the hydrogen bond between corresponding atoms with its bond length**

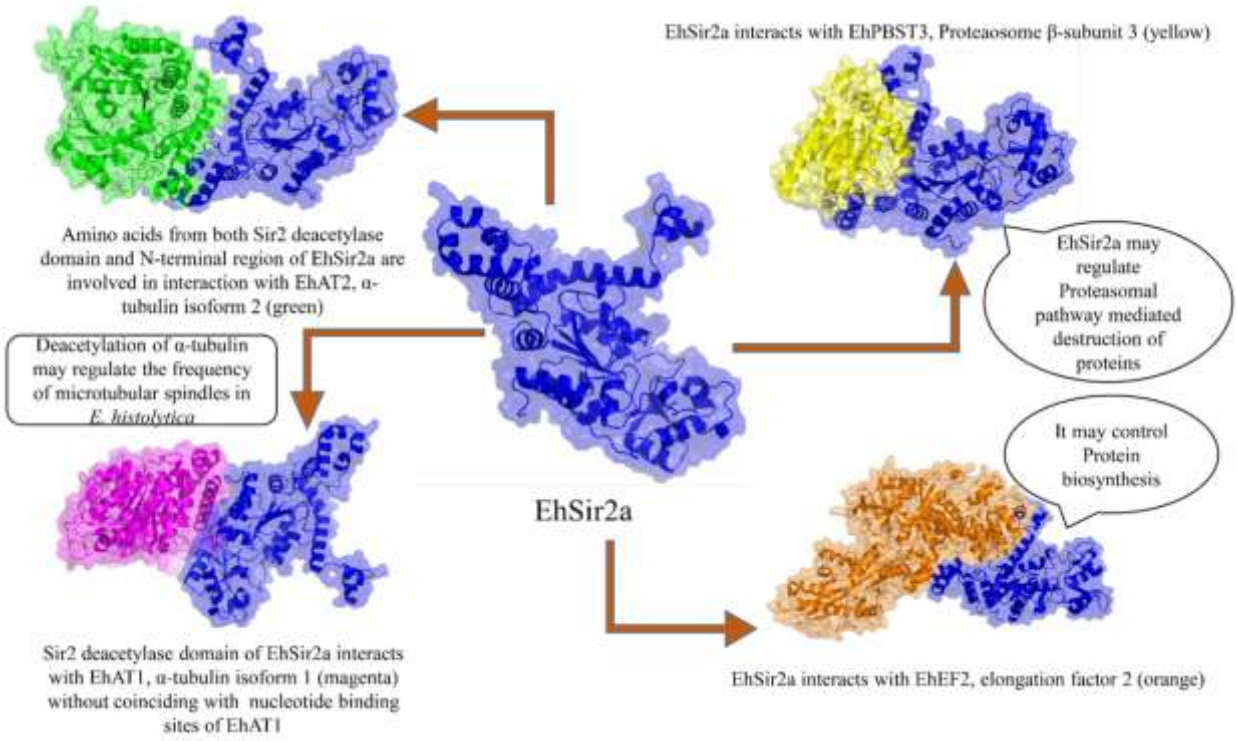



**Fig-3: Schematic representation of EhSir2a with its interactors.**

## MD simulation of EhSir2a

Energy minimization (EM) of EhSir2a protein was conducted using SPC/E water in a cubic box with GROMOS96 54a7 force field at 300 K. Initially, the system had a non-zero total charge 2.999998 but it was neutralized by adding three $Cl^-$ ions and the potential energy was 2.40702 x$10^5$ kJ $mol^{-1}$. EM was achieved in 499 steps by steepest descent method with a final minimized energy of -4.32909 x $10^5$ kJ $mol^{-1}$ with an average value of -4.01012 x $10^5$ kJ $mol^{-1}$ (Figure-4).

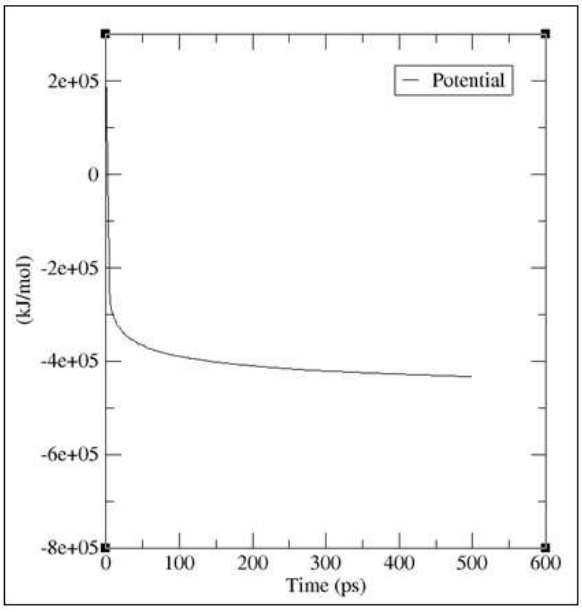
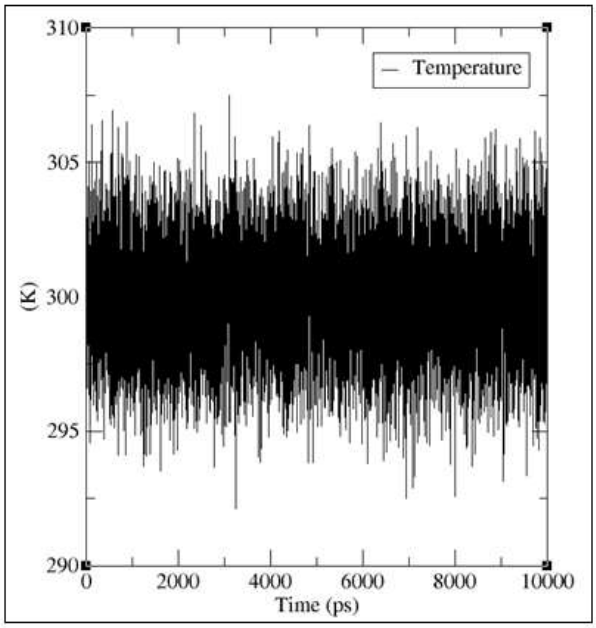


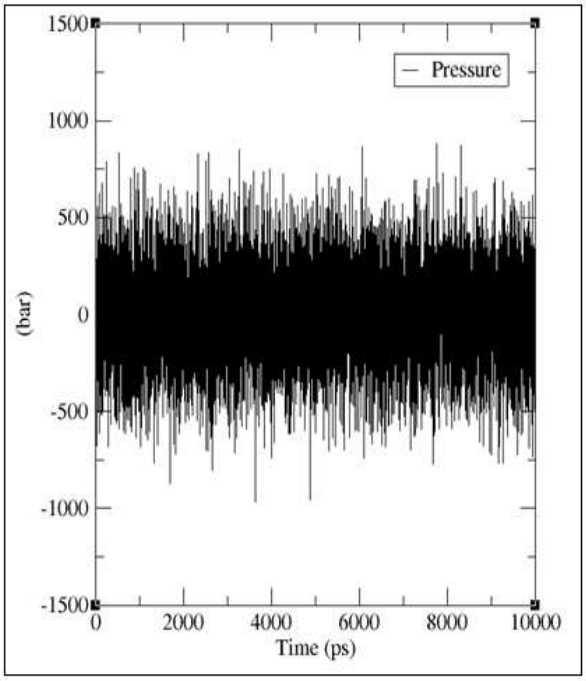

**Fig-4: Energy minimization curve of EhSir2a model structure by steepest descent method showed a decline in potential energy and remained constant after 499 steps with the potential energy of -4.32909 x $10^5$ KJ.$mol^{-1}$**

This value is significant for a satisfactory EM. After the NVT ensemble, the average temperature of this energy minimized protein in the system was recorded as 299.971 K, which is very much close to 300 K states that molecular simulation didn't collapse because of the flexibility of this protein model (Supplementary Figure-9).

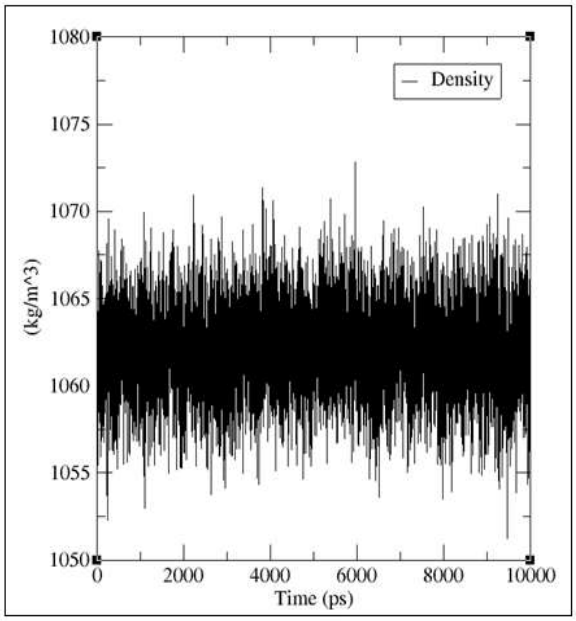



**Supplementary Figure-9: Temperature curve of EhSir2a model structure showed a rise in system temperature, becoming stable at 300 K with a running average of 299.971 K. Significant at P < 0.001, i.e., the system is well equilibrated in terms of temperature**

After the NPT ensemble, the average pressure and density were recorded as 1.64519 bar and 1062.07 Kg $m^{-3}$ (Supplementary Figure 10, 11), respectively which suggest that the system reached the equilibrium and stabilized in terms of density and pressure within nanosecond time scale.

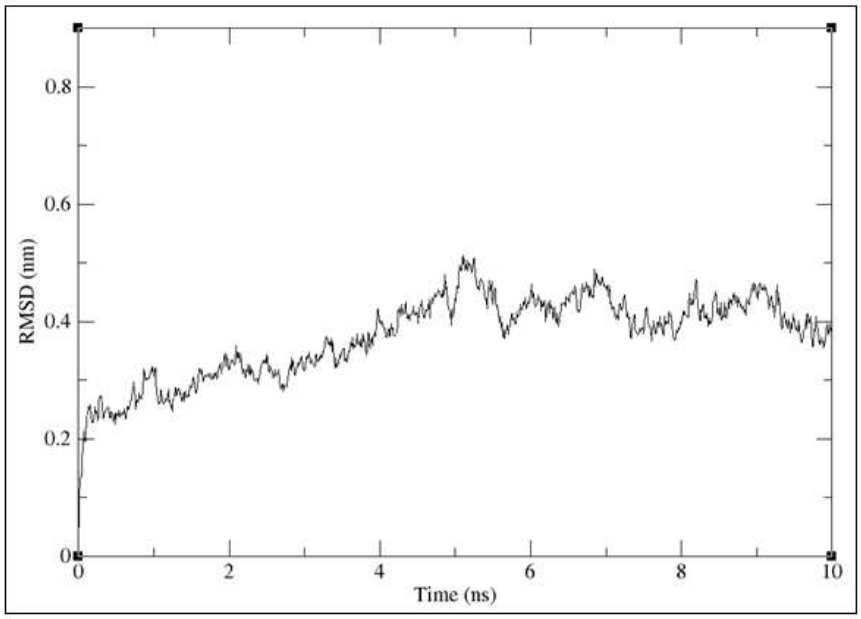



**Supplementary Figure-10: Pressure curve of EhSir2a model structure at 300 K showed an average pressure of 1.64519 bar**



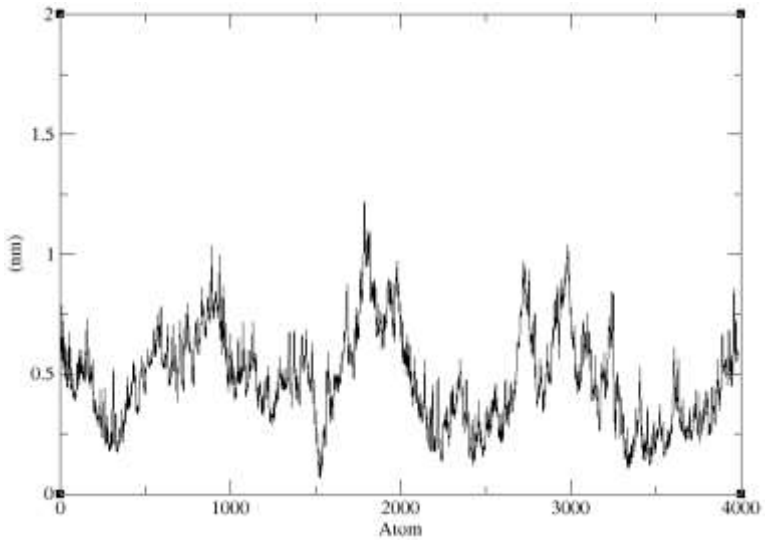

**Supplementary Figure-11: Density curve of EhSir2a model structure at 300 K showed an average density of 1062.07 kg $m^{-3}$**

The structural stability of the designed model of EhSir2a was run through 10 ns of MD simulation followed by the analysis of RMSD, RMSF, $R_g$, H-bonds to obtain a better picture of structural properties in aqueous condition having physiological salt concentration. MD trajectories of EhSir2a used to analyze the RMSD of the protein backbone atoms as a function of time (Supplementary Figure-12).




**Supplementary Figure-12: RMSD of EhSir2a backbone structure over 10 ns**

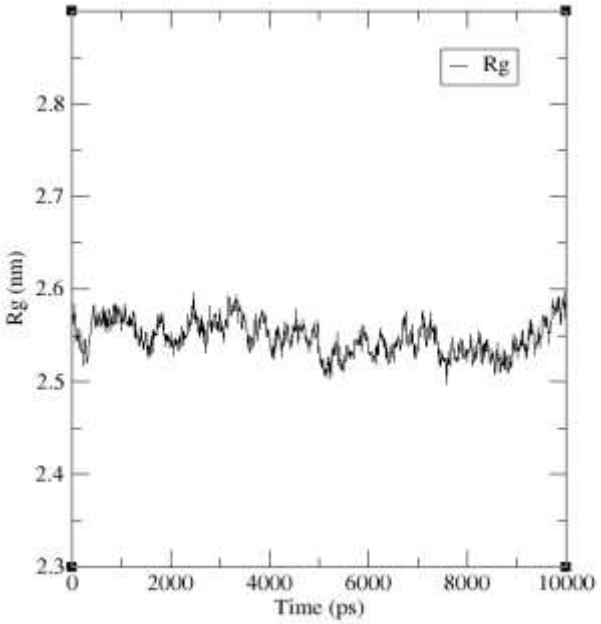
From this figure, it can be observed that the RMSD value within 0.1-0.5 during the 0-10 ns time scale. Initially, the RMSD value is increased up to 0.5 within 5 ns but after that, no further increment in RMSD value suggests that EhSir2a reaches an equilibrium state. The RMSF study of EhSir2a MD simulation is used to analyze the flexibility of the backbone structure (Supplementary Figure 13).




**Supplementary Figure-13: RMSF showed the fluctuation of each atoms in amino acid residue of EhSir2a**

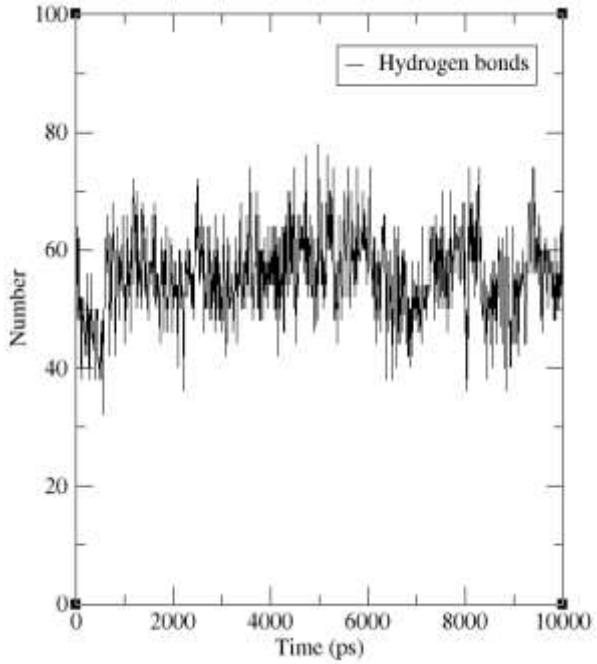
The RMSF of the atoms in residues in the active site of the ZnF domain and Sir2 domain EhSir2a show more fluctuation during simulation from its average position, indicate the flexibility and accessibility of the region for its interactor proteins. The low RMSF value indicates limited movements during simulation to its average position. This suggests that these atoms belong to the amino acid residues which are rigid due to chemical bonds. From the $R_g$ plot, we can determine the compaction level in EhSir2a. The $R_g$ value of EhSir2a varies between 2.5-2.6 nm which reveals the stability of EhSir2a designed structure in simulated biological conditions over a 10 ns time scale (Supplementary Figure-14).

a

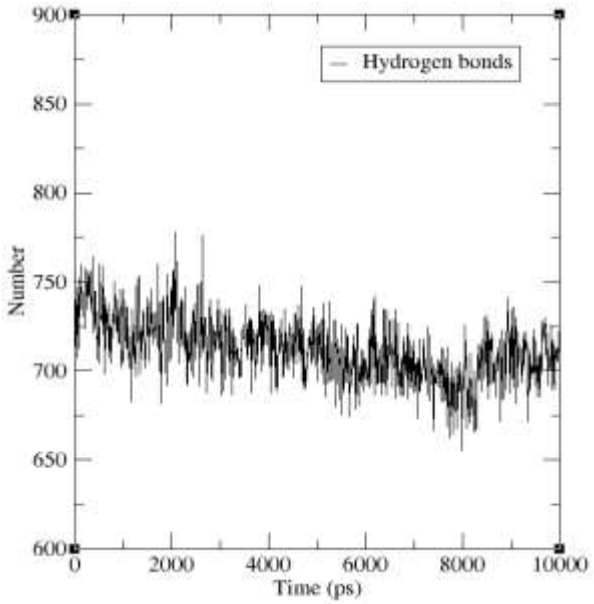



**Supplementary Figure-14: The radius of gyration of EhSir2a over 10,000 ps time frame showed the compactness of structure**

In a 10 ns simulation time, hydrogen bonds with both intermolecular level (Supplementary Figure-15) and solvent molecule (Supplementary Figure-16) suggest its stability of protein structural conformation.




**Supplementary Figure-15: Number of hydrogen bonds in EhSir2a at intermolecular level fluctuate initially but achieved equilibrium within nanosecond time scale**




**Supplementary Figure-16: Number of hydrogen bonds in EhSir2a with solvent molecule gradually decrease with time and reached equilibrium within nanosecond time scale**

## DISCUSSION

From a previous study, we have identified the interactors of EhSir2a by yeast two-hybrid genetic screening against the cDNA library of *E. histolytica* and identified their probable biological role in this parasite [19]. This study shows the detailed molecular interactions occurring between EhSir2a and its interacting proteins that are identified from the experimental method or by STRING analysis. Few interactors are not found by STRING analysis even at a confidence level of 0.4 but identified by cDNA library screening. This may be because *E. histolytica* being a protozoan parasite behaves differently than model organisms. Zinc finger domain-containing proteins are extremely abundant in eukaryotic genomes. It has versatile functions which include DNA binding, RNA packaging, transcriptional activation, regulation of apoptosis, protein folding and assembly, protein-protein, and protein-lipid interactions [65-70]. The Zn finger domain of EhSir2a is different from the other 30 types of zinc-finger domain found in higher eukaryotes. According to NCBI conserve domain search analysis, EhSir2a-ZnF is closely related to Zn-finger in ubiquitin-hydrolases (zf-UBP, Acc. Id: pfam02148). The Zn finger domain present at the N-terminal part of EhSir2a is probably responsible for its interaction with chromatin and other proteins. The RMSF of the atoms in the amino acid residues present in the active site of the ZnF domain and Sir2 domain of EhSir2a show more fluctuation during the simulation as compared to its average position. This indicates the flexibility and accessibility of the region for its interactor proteins. The low RMSF value indicates limited movements during simulation concerning its average position suggesting these atoms belong to the amino acid residues which are rigid. We have studied the molecular interaction between EhSir2a with its interactor EhAT1 and found that interacting site is different from the nucleotide-binding sites of EhAT1. Both EhAT1 and EhAT2 interact with the deacetylase domain of EhSir2a. The interaction of EhAT2 is different in the sense that it interacts with the N terminus of EhSir2a also, where EhAT1 has not any.

## CONCLUSION

The molecular interaction between EhSir2a and EhAT1 with the low free energy of binding suggests the stable interaction between them in the intercellular environment. Several proteins with diverse functionality predicted to be the interactors of EhSir2a indicates that it is involved in different cellular functions. The presence of positive amino acids in the zinc finger domain in EhSir2a may interact with specific sites on DNA and again, this may be the reason for interaction with an elongation factor, EhEF2. Various intercellular localization of EhSir2a and stable interaction with proteasome beta type3 domain-containing protein, EhPBST3 suggests its role in controlling the degradation of non-lysosomal protein in cytosol and nucleus. The intercellular concentration of regulatory protein works as a sensor of cellular cell cycle phase and metabolic state. The concentration of these regulatory proteins orchestrates with a function of time by degradation via the proteasomal pathway. From this study, it can be concluded that EhSir2a, a sirtuin homolog from the human parasite may have some novel functions that are not associated with sirtuins from other organisms.

## DECLARATIONS

**Ethics approval and consent to participate:** Not applicable

**Consent for publication:** Not applicable

**Availability of data and material:** Not applicable

**Competing interests:** The authors declare that they have no conflict of interest.

**Funding:** DAE-BRNS project (37 (1)/20/15/2014-BRNS/808), 2014-2017

## ACKNOWLEDGEMENTS

This study was supported by DAE-BRNS project (37 (1)/20/15/2014-BRNS/808), 2014-2017 to SD and PB was supported by State-funded fellowship, The University of Burdwan, West Bengal, India.

## REFERENCES

1. Haque, R. (2007). Human intestinal parasites. *Journal of health, population, and nutrition*, *25*(4), 387-391.
2. Petri Jr, W. A., Haque, R., Lyerly, D., & Vines, R. R. (2000). Estimating the impact of amebiasis on health. *Parasitology Today*, *16*(8), 320-321.

3. Amoebiasis. (1997). Releve epidemiologique hebdomadaire. 72(14):97-99.
4. Das, S., & Lohia, A. (2002). Delinking of S phase and cytokinesis in the protozoan parasite Entamoeba histolytica. *Cellular Microbiology*, *4*(1), 55-60.
5. Mukherjee, C., Majumder, S., & Lohia, A. (2009). Inter-cellular variation in DNA content of Entamoeba histolytica originates from temporal and spatial uncoupling of cytokinesis from the nuclear cycle. *PLoS Negl Trop Dis*, *3*(4), e409.
6. Lohia, A., Mukherjee, C., Majumder, S., & Dastidar, P. G. (2007). Genome re-duplication and irregular segregation occur during the cell cycle of Entamoeba histolytica. *Bioscience reports*, *27*(6), 373-384.
7. Lohia, A. (2003). The cell cycle of Entamoeba histolytica. *Molecular and cellular biochemistry*, *253*(1-2), 217-222.
8. Banerjee, S., Das, S., & Lohia, A. (2002). Eukaryotic checkpoints are absent in the cell division cycle ofEntamoeba histolytica. *Journal of biosciences*, *27*(6), 567-572.
9. Mehlotra, R. K. (1996). Antioxidant defense mechanisms in parasitic protozoa. *Critical reviews in microbiology*, *22*(4), 295-314.
10. Tekwani, B. L., & Mehlotra, R. K. (1999). Molecular basis of defence against oxidative stress in Entamoeba histolytica and Giardia lamblia. *Microbes and infection*, *1*(5), 385-394.
11. Vassilopoulos, A., Fritz, K. S., Petersen, D. R., & Gius, D. (2011). The human sirtuin family: evolutionary divergences and functions. *Human genomics*, *5*(5), 1-12.
12. Greiss, S., & Gartner, A. (2009). Sirtuin/Sir2 phylogeny, evolutionary considerations and structural conservation. *Molecules and cells*, *28*(5), 407-415.
13. Bitterman, K. J., Anderson, R. M., Cohen, H. Y., Latorre-Esteves, M., & Sinclair, D. A. (2002). Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast sir2 and human SIRT1. *Journal of Biological Chemistry*, *277*(47), 45099-45107.
14. Anderson, R. M., Bitterman, K. J., Wood, J. G., Medvedik, O., & Sinclair, D. A. (2003). Nicotinamide and PNC1 govern lifespan extension by calorie restriction in Saccharomyces cerevisiae. *Nature*, *423*(6936), 181-185.
15. Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K., & Ladurner, A. G. (2005). Splicing regulates NAD metabolite binding to histone macroH2A. *Nature structural & molecular biology*, *12*(7), 624-625.
16. Liou, G. G., Tanny, J. C., Kruger, R. G., Walz, T., & Moazed, D. (2005). Assembly of the SIR complex and its regulation by O-acetyl-ADP-ribose, a product of NAD-dependent histone deacetylation. *Cell*, *121*(4), 515-527.
17. Tong, L., & Denu, J. M. (2010). Function and metabolism of sirtuin metabolite O-acetyl-ADP-ribose. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1804*(8), 1617-1625.
18. Houtkooper, R. H., Pirinen, E., & Auwerx, J. (2012). Sirtuins as regulators of metabolism and healthspan. *Nature reviews Molecular cell biology*, *13*(4), 225-238.
19. Dam, S., & Lohia, A. (2010). Entamoeba histolytica sirtuin EhSir2a deacetylates tubulin and regulates the number of microtubular assemblies during the cell cycle. *Cellular microbiology*, *12*(7), 1002-1014.
20. Merrick, C. J., Dzikowski, R., Imamura, H., Chuang, J., Deitsch, K., & Duraisingh, M. T. (2010). The effect of Plasmodium falciparum Sir2a histone deacetylase on clonal and longitudinal variation in expression of the var family of virulence genes. *International journal for parasitology*, *40*(1), 35-43.
21. Tonkin, C. J., Carret, C. K., Duraisingh, M. T., Voss, T. S., Ralph, S. A., Hommel, M., ... & Speed, T. P. (2009). Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum. *PLoS Biol*, *7*(4), e1000084.
22. Baharia, R. K., Tandon, R., Sharma, T., Suthar, M. K., Das, S., Siddiqi, M. I., ... & Dube, A. (2015). Recombinant NAD-dependent SIR-2 protein of Leishmania donovani: immunobiochemical characterization as a potential vaccine against visceral leishmaniasis. *PLoS Negl Trop Dis*, *9*(3), e0003557.
23. Purkait, B., Singh, R., Wasnik, K., Das, S., Kumar, A., Paine, M., ... & Das, P. (2015). Up-regulation of silent information regulator 2 (Sir2) is associated with amphotericin B resistance in clinical isolates of Leishmania donovani. *Journal of Antimicrobial Chemotherapy*, *70*(5), 1343-1356.
24. Vergnes, B., Gazanion, E., & Grentzinger, T. (2016). Functional divergence of SIR2 orthologs between trypanosomatid parasites. *Molecular and biochemical parasitology*, *207*(2), 96-101.
25. Moreira, D., Rodrigues, V., Abengozar, M., Rivas, L., Rial, E., Laforge, M., ... & da Silva, A. C. (2015). Leishmania infantum modulates host macrophage mitochondrial metabolism by hijacking the SIRT1-AMPK axis. *PLoS Pathog*, *11*(3), e1004684.
26. Silvestre, R., Cordeiro- da- Silva, A., Tavares, J., Sereno, D., & Ouaissi, A. (2006). Leishmania cytosolic silent information regulatory protein 2 deacetylase induces murine B- cell differentiation and in vivo production of specific antibodies. *Immunology*, *119*(4), 529-540.
27. Silvestre, R., Cordeiro-Da-Silva, A., Santarém, N., Vergnes, B., Sereno, D., & Ouaissi, A. (2007). SIR2-deficient Leishmania infantum induces a defined IFN-γ/IL-10 pattern that correlates with

protection. *The Journal of Immunology*, *179*(5), 3161-3170.

28. Vergnes, B., Vanhille, L., Ouaissi, A., & Sereno, D. (2005). Stage-specific antileishmanial activity of an inhibitor of SIR2 histone deacetylase. *Acta tropica*, *94*(2), 107-115.
29. Fessel, M. R., Lira, C. B., Giorgio, S., Ramos, C. H. I., & Cano, M. I. N. (2011). Sir2-Related Protein 1 from Leishmania amazonensis is a glycosylated NAD (+)-dependent deacetylase. *Parasitology*, 1245-1258.
30. Dong, H., Yang, S., Zhao, Q., Han, H., Zhu, S., Zhu, X., ... & Yang, L. (2016). Molecular characterization and protective efficacy of silent information regulator 2A from Eimeria tenella. *Parasites & vectors*, *9*(1), 602.
31. Xie, H., Lei, N., Gong, A. Y., Chen, X. M., & Hu, G. (2014). Cryptosporidium parvum induces SIRT1 expression in host epithelial cells through downregulating let-7i. *Human immunology*, *75*(8), 760-765.
32. Yasukawa, H., & Yagita, K. (2010). Silent information regulator 2 proteins encoded by Cryptosporidium parasites. *Parasitology research*, *107*(3), 707-712.
33. Veiga-Santos, P., Reignault, L. C., Huber, K., Bracher, F., Souza, W. D., & Ulisses De Carvalho, T. M. (2014). Inhibition of NAD+-dependent histone deacetylases (sirtuins) causes growth arrest and activates both apoptosis and autophagy in the pathogenic protozoan Trypanosoma cruzi. *Parasitology*, (6), 814-825.
34. Soares, M. B., Silva, C. V., Bastos, T. M., Guimarães, E. T., Figueira, C. P., Smirlis, D., & Azevedo Jr, W. F. (2012). Anti-Trypanosoma cruzi activity of nicotinamide. *Acta tropica*, *122*(2), 224-229.
35. Alsford, S., Kawahara, T., Isamah, C., & Horn, D. (2007). A sirtuin in the African trypanosome is involved in both DNA repair and telomeric gene silencing but is not required for antigenic variation. *Molecular microbiology*, *63*(3), 724-736.
36. Wang, Y. H., Zheng, G. X., & Li, Y. J. (2016). Giardia duodenalis GlSir2. 2, homolog of SIRT1, is a nuclear-located and NAD+-dependent deacethylase. *Experimental parasitology*, *169*, 28-33.
37. Contreras, L. E., Suárez, A. G., Diaz, G. J., & Ramírez, M. H. (2019). GlSir2. 1 of Giardia lamblia is a NAD+-dependent cytoplasmic deacetylase. *Heliyon*, *5*(4), e01520.
38. Religa, A. A., & Waters, A. P. (2012). Sirtuins of parasitic protozoa: in search of function (s). *Molecular and biochemical parasitology*, *185*(2), 71-88.
39. Hailu, G. S., Robaa, D., Forgione, M., Sippl, W., Rotili, D., & Mai, A. (2017). Lysine deacetylase inhibitors in parasites: past, present, and future perspectives. *Journal of Medicinal Chemistry*, *60*(12), 4780-4804.
40. Scholte, L. L., Mourão, M. M., Pais, F. S. M., Melesina, J., Robaa, D., Volpini, A. C., ... & Nahum, L. A. (2017). Evolutionary relationships among protein lysine deacetylases of parasites causing neglected diseases. *Infection, Genetics and Evolution*, *53*, 175-188.
41. Szklarczyk, D., Santos, A., von Mering, C., Jensen, L. J., Bork, P., & Kuhn, M. (2016). STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic acids research*, *44*(D1), D380-D384.
42. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., ... & Jensen, L. J. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, *47*(D1), D607-D613.
43. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... & Lepore, R. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, *46*(W1), W296-W303.
44. Eswar, N., Webb, B., Marti- Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., ... & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics*, *15*(1), 5-6.
45. Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic acids research*, *40*(W1), W294-W297.
46. Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*(3), 343-350.
47. Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. *Proteins*, *50*, 437-450.
48. Berjanskii, M., Zhou, J., Liang, Y., Lin, G., & Wishart, D. S. (2012). Resolution-by-proxy: a simple measure for assessing and comparing the overall quality of NMR protein structures. *Journal of biomolecular NMR*, *53*(3), 167-180.
49. Wallner, B., & Elofsson, A. (2003). Can correct protein models be identified?. *Protein science*, *12*(5), 1073-1086.
50. Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, *35*(suppl_2), W407-W410.
51. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein science*, *27*(1), 129-134.

52. Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., ... & Kozakov, D. (2017). New additions to the C lus P ro server motivated by CAPRI. *Proteins: Structure, Function, and Bioinformatics*, *85*(3), 435-444.
53. Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., ... & Vajda, S. (2017). The ClusPro web server for protein–protein docking. *Nature protocols*, *12*(2), 255-278.
54. Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Vajda, S. (2013). How good is automated protein docking?. *Proteins: Structure, Function, and Bioinformatics*, *81*(12), 2159-2166.
55. Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., & Hou, T. (2019). HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic acids research*, *47*(W1), W322-W330.
56. Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, *10*(5), 449-461.
57. Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *Journal of chemical information and modeling*. 51(10):2778-2786.
58. Zhang, C., Freddolino, P. L., & Zhang, Y. (2017). COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic acids research*, *45*(W1), W291-W299.
59. Roy, A., Yang, J., & Zhang, Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic acids research*, *40*(W1), W471-W477.
60. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry*, *26*(16), 1701-1718.
61. Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., & van Gunsteren, W. F. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European biophysics journal*, *40*(7), 843-856.
62. Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., ... & Raschellà, G. (2017). Zinc-finger proteins in health and disease. *Cell death discovery*, *3*(1), 1-12.
63. Zhang, W., Xu, C., Bian, C., Tempel, W., Crombet, L., MacKenzie, F., ... & Qi, C. (2011). Crystal structure of the Cys2His2-type zinc finger domain of human DPF2. *Biochemical and biophysical research communications*, *413*(1), 58-61.
64. Gray, K. A., Yates, B., Seal, R. L., Wright, M. W., & Bruford, E. A. (2015). Genenames. org: the HGNC resources in 2015. *Nucleic acids research*, *43*(D1), D1079-D1085.
65. Klug, A. (1999). Zinc finger peptides for the regulation of gene expression. *Journal of molecular biology*, *293*(2), 215-218.
66. Laity, J. H., Lee, B. M., & Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Current opinion in structural biology*, *11*(1), 39-46.
67. Matthews, J. M., & Sunde, M. (2002). Zinc fingers- - folds for many occasions. *IUBMB life*, *54*(6), 351-355.
68. Brown, R. S. (2005). Zinc finger proteins: getting a grip on RNA. *Current opinion in structural biology*, *15*(1), 94-98.
69. Hall, T. M. T. (2005). Multiple modes of RNA recognition by zinc finger proteins. *Current opinion in structural biology*, *15*(3), 367-373.
70. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M., & Mackay, J. P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends in biochemical sciences*, *32*(2), 63-70.