Journal of Genetic Engineering and Biotechnology 23 (2025) 100488

Contents lists available at ScienceDirect

Journal of Genetic Engineering and Biotechnology

journal homepage: www.elsevier.com/locate/jgeb

Comparative proteomics and structure-based approach to unravel the therapeutic drug target of *Theileria* species

Anusha Majumder [1], Jyotisha [1], Fouzia Nasim, Insaf Ahmed Qureshi [*]

Department of Biotechnology & Bioinformatics, School of Life Sciences, University of Hyderabad, Prof. C.R. Rao Road, Hyderabad 500046, India

ARTICLE INFO

Keywords:
Theileria
Transforming species
Comparative proteomics
*Sfi*Ip
Molecular dynamics simulations

ABSTRACT

Theileriosis, caused by protozoan parasites of genus *Theileria*, primarily affects both domestic and wild ruminants. It can lead to significant economic losses in livestock farming due to decreased productivity and high mortality rates in susceptible animals, while treatment measures are not cost-effective. Since most of mechanisms of this disease remain unknown, this study investigates the differences in the mode of pathogenesis between transforming and non-transforming groups through an *in silico* comparative proteomics approach to recognize the key players involved in host cell transformation. Although the major biological processes and molecular functions are almost conserved between the two groups, PEST-motif containing secretory proteins of *Sfi*I, SVSP, and Tash-AT gene families were identified as important candidates with the potential to transform infected host cells. Several members of PEST-motif containing proteins possess signal peptides, nuclear localization signals, and *trans*-membrane helices, further supporting their potential to transform host cells. Additionally, structural analysis helped in the identification of a parasitic protein (*Sfi*Ip) from *Sfi*I family as a plausible drug target. Virtual screening revealed FDA-approved drugs (i.e. atogepant and rimegepant) as promising compounds, showing the highest affinity for *Sfi*Ip during molecular docking. Further studies, including molecular dynamics simulation, principal component analysis, and free energy landscape, suggested that these drug molecules exhibit the stable interaction with protein. Therefore, our research could facilitate the identification and targeting of novel drug candidates that may be further implemented to recognize effective therapeutics against *Theileria* infections.

1. Introduction

Theileriosis causes immense economic loss to the dairy industry and farmers worldwide, particularly impacting domestic livestock, including both large and small ruminants, which results in the death of nearly one million cattle every year only in Africa.[1,2] *Theileria* parasites, causative agents of theileriosis, belong to the order Piroplasmida within the phylum Apicomplexa, and are transmitted by at least four different genera of ticks.[3] There are several known species of *Theileria*, whereas whole genome sequencing data are available for only a few.[4] Among the four species for which whole genome sequencing data are publicly available, *T. annulata* (TA) and *T. parva* (TP) are implicated as the most critical forms of the disease, i.e., tropical theileriosis and east coast fever (ECF), respectively. These species are also referred to as 'transforming species' because they can transform infected host lymphocytes, leading to their indefinite proliferation. *T. orientalis* (TO) results in oriental theileriosis and is classified as a non-transforming species due to its inability to transform host cells, and it mainly multiplies within the infected erythrocytes.[5] *T. equi* (TE) triggers equine piroplasmosis and is regarded as an outlier due to its controversial phylogenetic position, situated between *Babesia bovis* and *Theileria* species.[6] It also lacks the homologs of putative host cell-transforming genes present in TA and TP. Hence, it is imperative to speculate that epidemiological differences and variation in their mode of pathogenesis, along with the clinical

Abbreviations: TA, *Theileria annulata*; TP, *Theileria parva*; TO, *Theileria orientalis*; TE, *Theileria equi*; TM, Trans-membrane; NLS, Nuclear localization signal; MDS, Molecular dynamics simulation; GO, Gene ontology; BP, Biological process; MF, Molecular functions; CC, Cellular components; SDF, Structure-data file; RMSD, Root mean square deviation; Rg, Radius of gyration; RMSF, Root mean square fluctuation; SASA, Solvent accessible surface area; PCA, Principal component analysis; FEL, Free energy landscape.

* Corresponding author.
E-mail address: insaf@uohyd.ac.in (I.A. Qureshi).

[1] These authors contributed equally to this article.

https://doi.org/10.1016/j.jgeb.2025.100488
Received 15 January 2025; Received in revised form 27 March 2025; Accepted 28 March 2025

symptoms have resulted in the speciation of transforming species from the non-transforming group.

Knowledge of the life cycle of a parasite is essential for developing effective disease control strategies and identifying potential drug targets. *Theileria* completes its life cycle in two hosts: the definitive host (tick), where the parasite undergoes sexual reproduction, and the intermediate host (ruminant), on which the asexual life cycle occurs. Among the group of obligate intracellular parasites responsible for diseases in both animals and humans, *Theileria* is unique as the only genus capable of transforming the cells of its mammalian hosts. In contrast to the transforming species, TO lacks host cell proliferation, leading to the enlargement of host cells. Infection and treatment method (ITM) is a vaccination strategy used to control theileriosis, employing either a single parasite isolate or the Muguga cocktail.[7,8] However, ITM has drawbacks such as the disease recurrence, cumbersome storage, and handling requirements, which make it difficult to implement effectively. Simultaneously, several drugs have also been tested against *Theileria* species but they have either proven ineffective or possessed toxicity. Tetracyclines such as chlortetracycline and oxytetracycline were shown to prevent clinical signs of ECF, arrest schizogony in lymphoid cells, and reduce parasitemia.[9] Similarly, a naphthoquinone (menoctone) demonstrated antitheilerial activity, but was soon replaced by buparvaquone (hydroxynaphthoquinone) due to its superior clinical efficacy.[10] Nevertheless, the parasite developed resistance to this drug by creating mutations in the Q_0 quinone-binding site of the mitochondrial cytochrome *b*.[10] Hence, there is a dire need for the identification of new targets and the development of drugs, which are easy to produce, cost-effective with minimal risk factors associated with them.

In this study, we utilized available proteome data to perform an *in silico* comparative analysis of *Theileria* species aiming to uncover differences in their modes of pathogenesis and identify potential drug targets along with their specific inhibitors. In order to identify the proteins of *Theileria* that are responsible for transforming host cells, a systematic functional analysis of the proteomes was carried out. Furthermore, the three-dimensional structure of the target protein (*Sfi*Ip) was predicted using *in silico* approach, followed by an evaluation of its ligand binding sites. Virtual screening of FDA-approved drugs was then performed to identify potential therapeutic candidates. Subsequently, molecular dynamics simulations (MDS) were employed to validate the structural changes induced by the binding of these drugs to the protein of interest.

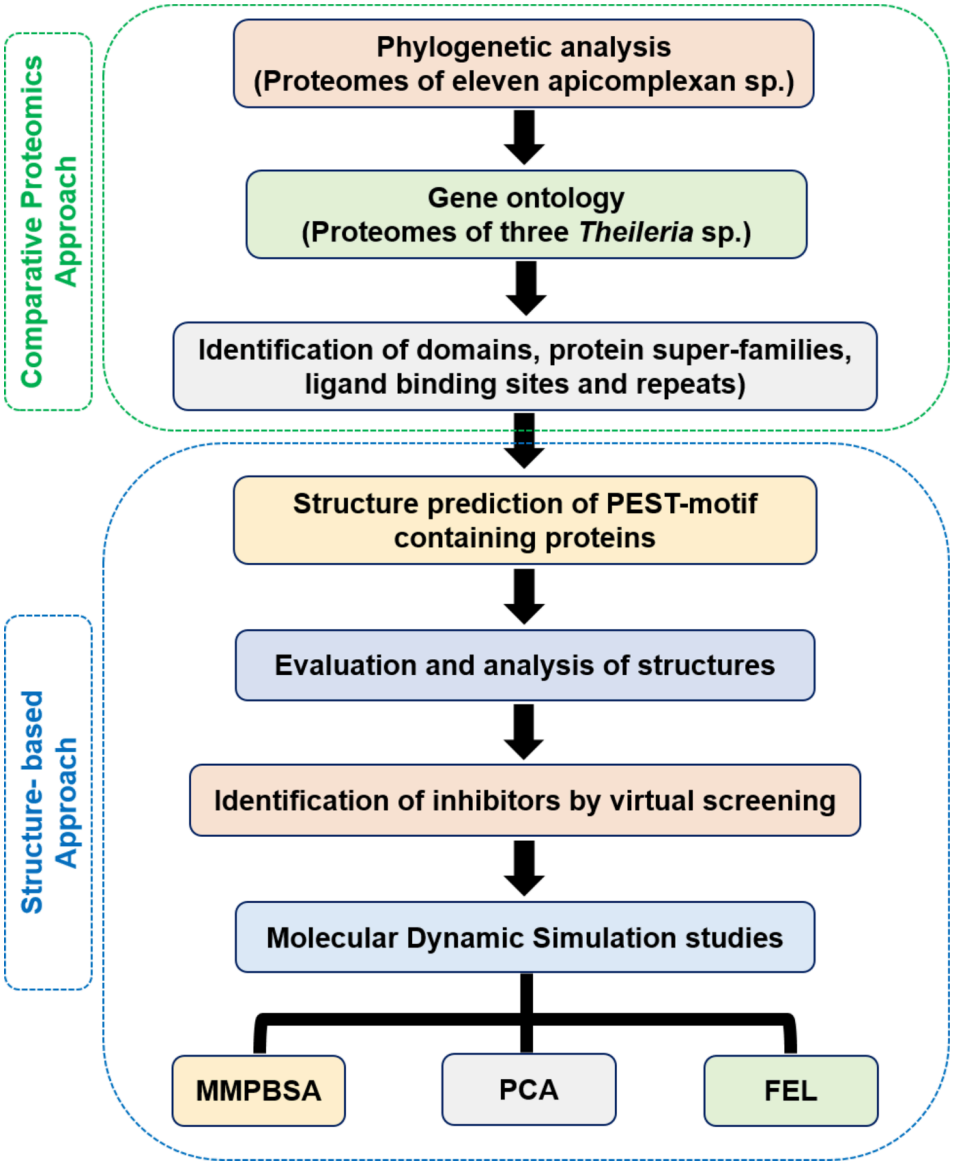

Fig. 1. Schematic representation of the comparative proteome and structure-based analysis for identifying the drug targets and inhibitors of *Theileria* species.

2. Materials and methods

2.1. Species selection

Three species of *Theileria* were chosen from different known species based on information available on their whole genome sequences. Of them, TA causes the deadliest form of disease with the ability to transform host cells, TO does not transform host cells, whereas TE is considered as an outlier. A detailed flowchart of methodology followed in this manuscript is represented in Fig. 1.

2.2. Phylogenetic tree analysis

The proteomes of all the species were obtained from NCBI (ftp://ftp.ncbi.nlm.nih.gov/) in FASTA format for phylogenetic analysis. In total, our dataset included eleven species, namely, four *Theileria* species and other representative members of the apicomplexan family: *Babesia bovis*, *Cyclospora cayetanensis, Cryptosporidium parvum*, *Gregarina niphandrodes, Neospora caninum* Liverpool, *Plasmodium falciparum*, and *Toxoplasma gondii,* which cause diseases such as babesiosis, cyclosporiasis, cryptosporidiosis, neosporosis, malaria, and toxoplasmosis. To generate a phylogenetic tree, OrthoFinder,[11] a command-line-based, fast, and reliable tool was executed that also infers orthogroups and orthologs using STRIDE.[12] The phylogenetic tree obtained was finally visualised in Dendroscope 3.[13]

2.3. Gene ontology of Theileria proteomes

Since a major portion of *Theileria* proteomes is unannotated, Blast2GO from OmicsBox was used to annotate the proteomes of three *Theileria* species.[14] Understanding ontological terms is one of the common ways to interpreting the biological roles of an organism. Gene Ontology (GO) is a set of terms that represent the properties of a gene product and covers the following three main categories: (a) Biological process (BP), which involves through the coordinated actions of multiple molecular activities; (b) Molecular functions (MF), which describes the functions performed by the gene products at the molecular level; and (c) Cellular components (CC) that provides information about localization of a gene product i.e., the site where its functions are carried out. In the present study, a functional analysis module was chosen, wherein, 'DIAMOND CloudBlast'[15] was employed to scan the whole proteomes of TA, TO, and TE. It is a high-throughput alternative program for aligning big sequence data against reference databases such as non-redundant and is 20,000 times faster than the speed of BLAST along with high sensitivity. To analyse the amino acid sequences and identify specific folds of protein domains, motifs, repeats, and conserved sites, CloudInterProScan (https://www.ebi.ac.uk/interpro/search/sequence/) was executed. Furthermore, 'GO Mapping' was utilized to retrieve GO terms associated with the hits acquired from BLAST search (https://blast.ncbi.nlm.nih.gov/). To link potential homologs and domains with their putative roles, mapping relies on available functional annotation in databases such as UniProt (https://www.uniprot.org/) and GO. Finally, 'Blast2GO Annotation' (https://www.blast2go.com/) was implemented to assign the most reliable GO terms to the sequences from the GO pool obtained through the previous step of GO Mapping.

2.4. Presence of GPI-anchor, NLS, signal peptides, and TM-helices in proteins

As these target proteins have unknown functions and lack information on their interactions and the pathways they regulate, the presence of GPI-anchors, nuclear localization signals, and *trans*-membrane helices was correspondingly predicted by PredGPI (https://busca.biocomp.unibo.it/predgpi/), NucPred (https://nucpred.bioinfo.se/nucpred/), and Deep TMHMM (https://dtu.biolib.com/DeepTMHMM) webservers.

2.5. 3D structures of host cell transforming proteins

Since the structures of proteins from host cell transforming gene families (*Sfi*I, SVSP, Tash-AT) have not been reported in PDB database (https://www.rcsb.org), we employed *in silico* approaches to generate their 3D models. First, BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) from NCBI was performed against our query sequences with PDB as the data source. However, since no such homologs were found for any of them, we opted for *ab initio* model building by Robetta (a deep learning-based method, RoseTTAFold) that is based on the knowledge of protein folding (https://robetta.bakerlab.org/). Robetta ranks its predictions by a confidence score, typically ranging from 0 to 1, with a higher C-score indicating greater confidence in the model's quality. On the basis of compactness and the C-score of predicted models, only one structure (*Sfi*Ip) was selected for further analysis. To enhance its structural accuracy, loop refinement was carried out using ModLoop (https://modbase.compbio.ucsf.edu/modloop/). For quality assessment of the selected model, it was subjected to PROCHECK (https://saves.mbi.ucla.edu/) and ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php/) servers.

2.6. Prediction of ligand binding cavities and sites

To approximate the ligand-binding cavities and residues, several web servers were employed. First, CASTp (http://sts.bioe.uic.edu/castp/index.html?2cpk) was implemented to obtain information about the pockets, cavities, and channels present in the protein structures. Next, PrankWeb (https://prankweb.cz/) was executed that uses a machine learning approach to build models based on structural, physicochemical, and evolutionary features. This model was then utilized for the detection of ligand-binding sites on the surface patches, and the ligand-interacting residues were recorded.

2.7. FDA-approved drugs for virtual screening

The 3D structure of *Sfi*Ip was visualized in AutoDock Tools,[16] and prepared for further use as described in previous studies.[17,18] The monomeric state was considered by removing additional chains, repairing missing atoms, and adding polar hydrogen atoms. Kollman charges were then added, charges were distributed equally, and the optimized structure was saved in PDBQT format. Three-dimensional structures of 1984 FDA-approved drugs were obtained from the DrugBank database (https://go.drugbank.com/) in structure-data file (SDF) format as a single file, followed by splitting and energy minimization applying the MMFF94 force field of Open Babel.[19]

2.8. Docking of FDA-approved compounds with SfiIp

Molecular docking was carried out using AutoDock Vina version 1.1.2[20] to assess the binding affinity of the drug molecules to the protein. The saved PDBQT protein file was utilized to incorporate the information regarding ligand-binding site residues applying ADT 1.5.6. The residues of pocket 1, obtained from PrankWeb, were marked to identify the conformational search space, and a grid box with dimensions 54 × 58 × 80 $Å^3$ was specified around the coordinates (19.813, −8.643, −31.621). Virtual screening was executed for the drug molecules individually, using a Perl script employed in the Vina program. All FDA-approved drugs screened were ranked based on their binding affinities (kcal/mol) for the target protein. PyMOL was used for visualization of the *Sfi*Ip-drug complexes after docking (https://pymol.org/2/). For the top 1 % (20 out of 1984) of the virtually screened FDA-approved compounds, their docking scores and the amino acids responsible for interactions with the ligands were analysed using LigPlot + v1.4.5 (https://www.ebi.ac.uk/thornton-srv/software/LigPlus/).

2.9. Molecular dynamics simulation (MDS) studies

To study the conformational flexibility and stability of apo *Sfi*Ip and its drug complexes, MD simulation was carried out in GROMACS version 5.1.4.[21] Topology file of protein was generated through OPLS force field,[22] while the LigParGen web server[23] was executed for generation of GROMACS structure (gro) and topology (itp) file of the ligands. The apo protein or protein–ligand complexes were solvated in a cubic box with the SPC/E water model, and neutralisation of all systems was performed by addition of sodium ions as per earlier studies.[24,25] Steepest descent algorithm was utilised to carry out the energy minimization in 50,000 steps, followed by equilibration of the systems for 100 ps through NVT and NPT ensembles, implementing the leap-frog integrator. Finally, MD simulation was run at 300 K for 100 ns with a time step of 2 fs. To assess the quality, conformational flexibility, and stability of the trajectories, parameters such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), and intermolecular hydrogen bond were correspondingly enumerated applying *gmx rms*, *gmx rmsf*, *gmx gyrate*, *gmx sasa*, and *gmx hbond* built-in commands in GROMACS as mentioned previously.[26] All of these graphs were visualized through GraphPad Prism ver. 8 (https://www.graphpad.com/scientific-software/prism/).

2.10. Principal component analysis (PCA)

PCA in MD simulation studies is essential to understand the simultaneous movement of different parts of a biological macromolecule, including both fast, local fluctuations and slow, collective motions. It significantly reduces the dimensionality of the configuration space while preserving as much variability as possible, facilitating both qualitative and quantitative assessments. To analyse principal component, the *gmx covar* module was executed to build the covariance matrix from the last 40 ns of MD trajectory. The principal components (PCs) are the diagonal eigenvectors of the matrix, representing the protein's correlated motion, whereas the eigenvalues depict the atomic contribution to the motion of protein–ligand complex systems. Additionally, *gmx anaeig* was used to evaluate the eigenvectors of the covariance matrix according to an earlier report.[27]

2.11. Free energy landscape (FEL)

FEL is a valuable tool in bioinformatics for analyzing the folding or aggregation behaviour of protein–ligand complexes following MD simulations, providing insights into the stability of the protein. In this study, the principal component (PC) values were used as order parameters to enumerate the FEL using the '*gmx sham*' module in GROMACS. The FEL was generated based on the entire MD trajectory data. The Gibbs free energy (Gi) for each population bin (Ni) was calculated applying the equation: $Gi = -kB*T*\ln(Ni/N_{max})$, where kB is Boltzmann's constant, T is the temperature (usually 300 K), N_i is the population of bin I, and N_{max} is the population of the most populated bin. A colour model was employed to visualize the energy levels in plot, with red denoting the highest energy level (least stable states) and blue representing the lowest energy level (most stable states).

2.12. Binding free energy calculations

The stability of the protein-drug complexes after MD simulation studies was analysed by enumerating the free binding energy using the g_mmpbsa package[28] integrated with GROMACS. This calculation was performed on every 0.1 ns frame from the last 20 ns of the MD simulation trajectories, as reported previously.[29] The free binding energy calculation abides by the following formulas:

$$\Delta G_{bind} = \Delta G_{complex} - \left(\Delta G_{protein} + \Delta G_{1igand}\right) \quad (1.1)$$

$$\Delta G_{bind,aq} = \Delta H - T\Delta S \sim \Delta E_{MM} + \Delta G_{bind,solv} - T\Delta S \quad (1.2)$$

$$\Delta E_{MM} = \Delta E_{covalent} + \Delta E_{electrostatic} + \Delta E_{VDW} \quad (1.3)$$

$$\Delta E_{covalent} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{torsion} \quad (1.4)$$

$$\Delta G_{bind,solv} = \Delta G_{polar} + \Delta G_{nonpolar} \quad (1.5)$$

ΔG_{bind} represents the binding energy, while $\Delta G_{protein}$, ΔG_{ligand}, and $\Delta G_{complex}$ denote the free energies of the protein, ligand, and complex system, respectively. $\Delta G_{bind,aq}$ is the binding free energy in an aqueous environment; $\Delta G_{bind,solv}$ is the solvation free energy change; ΔE_{MM} is the gas-phase molecular mechanical energy; $T\Delta S$ delineates the conformational entropy change upon binding; and ΔG_{polar} and $\Delta G_{nonpolar}$ depict the polar and non-polar contributions to the solvation free energies.

3. Results

3.1. Transforming Theileria species possess smaller genomes

A comparative study of the selected eleven species (*T. annulata, T. parva, T. orientalis, T. equi, B. bovis*, *C. cayetanensis, C. parvum*, *G. niphandrodes, N. caninum* Liverpool, *P. falciparum*, and *T. gondii*) showed that the average size of the genomes considered was approximately 23.5 Mbp, although variability among them was observed (Table S1). Out of these, *T. gondii* (64.53 Mbp), *N. caninum* (57.5). and *C. cayetanensis* (44.4 Mbp) had relatively larger genomes. Notably, TA, TP, and *B. bovis* possess almost equal genome sizes viz. 8.35, 8.3, and 8.2 Mbp, respectively, while TO (9.1 Mbp) and TE (11.6 Mbp) have comparatively larger genome sizes, which were comparable to that of *C. parvum* (9.11 Mbp). In accordance with the fact that virulent organisms contain small genomes,[30] the host-cell transforming species had a smaller genomic size than the non-host cell transforming species. The GC content of all the species showed considerable variation with an average of 41.1 %. However, *C. cayetanensis*, *G. niphandrodes*, *N. caninum,* and *T. gondii* had higher GC contents (52–55 %), followed by *B. bovis* (41.8 %), *T. orientalis* (41.6 %), and *T. equi* (39.5 %). Among the two transforming species i.e., *T. annulata* and *T. parva* have corresponding GC contents of 32.5 % and 34.1 %, whereas *C. parvum* and *P. falciparum* comprise 30 % and 19 %, respectively.

3.2. Babesia bovis is the last common ancestor of all Theileria species

Phylogenetic relationship analysis revealed the point at which the non-transforming species diverged from the transforming species, establishing an evolutionary link between them (Fig. 2). *T. gondii* and *C. parvum* share a different lineage from the other apicomplexan members and branch into a separate clade, whereas *P. falciparum* is the common ancestor of *Theileria* and *Babesia* species. Additionally, it is clear that the transforming and non-transforming *Theileria* species fall into two distinct clades, with the outlier species in a further separate clade. Therefore, TA and TP are more closely related to each other. It is also worth noting that the last common ancestor of all *Theileria* species is *B. bovis*, which has been reported as a highly pathogenic haemoprotozoan parasite in cattle and is pathologically similar to *P. falciparum* that causes malaria in humans.[31]

3.3. Essential physiological processes are conserved across all Theileria species

The use of gene ontology is a common practice to understand the annotation of a particular gene product. To identify differences in the mode of pathogenesis among *Theileria* species, the whole proteomes of the transforming (TA), non-transforming (TO), and outlier species (TE) were evaluated. BLASTp was performed to find homologous protein sequences against these proteomes and the highest number of hits was

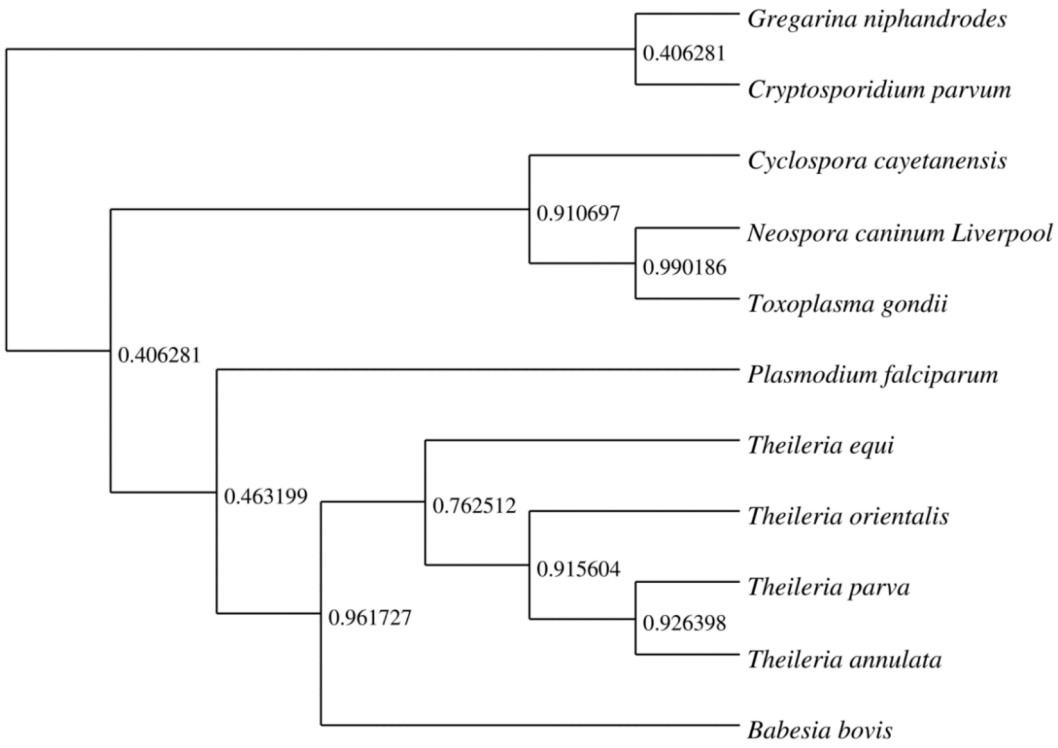

Fig. 2. Phylogenetic relationship of eleven apicomplexan species wherein numbers indicate the Bootstrap values. *P. falciparum* branches out from the clade constituting *Theileria* and *Babesia* species with *B. bovis* being the last common ancestor of all four *Theileria* species.

obtained from other *Theileria* species, *Babesia* species, *Plasmodium* species, and *T. gondii*.

3.3.1. Biological processes (BP)

A biological process is a method of carrying out a genetically-encoded program or module that includes all the necessary steps to achieve a specific biological goal. This process is executed through a set

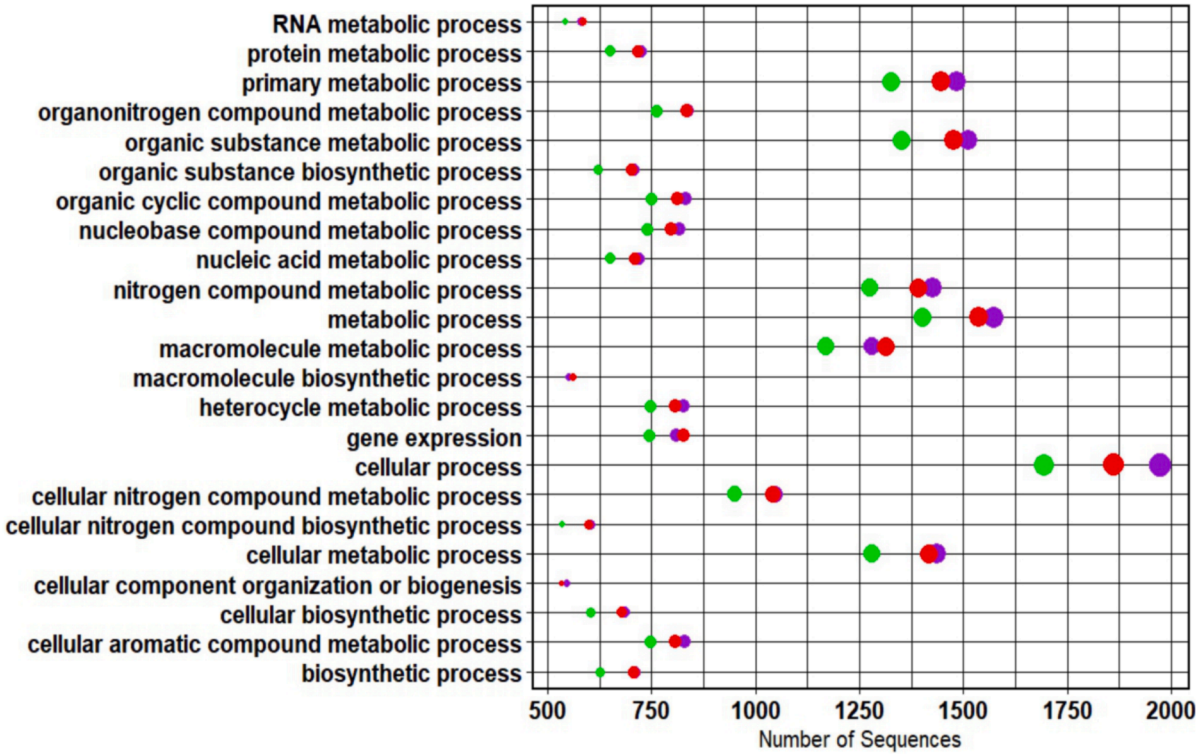

Fig. 3. Biological process distribution of three *Theileria* species. The green, red, and purple circles correspond to *Theileria annulata*, *Theileria orientalis*, and *Theileria equi*. The major biological processes remain mostly conserved in all three species with majority of the proteins participating in various metabolic processes and gene expression. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of functions performed at the molecular level by a particular gene product. It is tightly regulated and is often occurs in a specific temporal sequence. The BP distribution of all three *Theileria* species included processes related to organic substances, primary processes, cellular functions, nitrogen compounds, macromolecules, cellular nitrogen, organonitrogen compounds, metabolic processes, and gene expression (Fig. 3). This suggests that the basic physiological functions of the parasite are similar in the transforming, non-transforming, and outlier species, with majority of the proteins involved in various cellular processes, such as microtubule-based process, cell growth and maintenance, cell communication, and signal transduction. Other enriched BPs observed were nitrogen, organic and organonitrogen compound, and primary metabolic processes etc.

3.3.2. Cellular components (CC)

This ontological term provides information about the location of macromolecules when they are involved in a particular molecular function. CC describes the location of a gene product with respect to either cellular structures or compartments, or the macromolecular complexes of which they are part of. Fig. 4 revealed that a significant proportion of proteins in TA, TO, and TE were membrane-bound (~10 %), indicating their role in host cell invasion. Around 6 % of the proteins were cytoplasmic and might interfere with host signal transduction pathways, thereby manipulating the immune response, and regulating parasite growth and survival. Moreover, nuclear proteins were found to constitute approximately 4 % of the parasite proteomes, playing essential roles in gene expression, DNA replication, and repair.

3.3.3. Molecular functions (MF)

The MF distribution of all three categories of *Theileria* species, as shown in Fig. 5, includes the binding of organic cyclic compounds, heterocyclic compounds, ions, nucleic acids, nucleotides, nucleoside phosphates, and small molecules as well as hydrolase activity. Hence, it can be concluded that the major function of proteins at the molecular level involves either catalysis or binding of cognate ligands, indicating their importance in the pathogenesis of parasites. The catalytic activities encompass the modification of a nucleic acid or a protein such as cyclase, demethylase, hydrolase, isomerase, ligase, lyase, oxidoreductase, and transferase. Binding, in terms of molecular functions, refers to the non-covalent interactions between two molecules, for example, the host cell surface binding activity. However, it should be noted that some proteins fall into more than one category.

3.4. PEST motif is exclusively present in transforming Theileria species

Protein sequences from TA, TO, and TE were analysed using InterProScan to identify domains (Table S2, Fig. 6A), protein superfamilies (Table S3, Fig. 6B), various ligand binding sites (Table S4, Fig. 6C), and repeats (Table S5, Fig. 6D). The over-enriched protein domains were consistent across all species and included the AAA + ATPase domain (ATPases associated with diverse cellular activities), RNA recognition, helicase, and protein kinase domains (Table S2, Fig. 6A). The AAA + ATPase domain participates in several cellular processes such as membrane fusion, proteolysis, and DNA replication. The RNA recognition domain partakes in single-stranded RNA binding, helicase superfamily 1/2, ATP-binding domain, and helicase, C-terminal domain cause separation of double-stranded nucleic acids in an energy-dependent manner, whereas the protein kinase domain phosphorylates proteins. The P-loop NTP hydrolase, which catalyses the hydrolysis of beta-gamma phosphate bond of a bound NTP, is one of the superfamilies that were over-represented in transforming, non-transforming, and outlier species. Other enriched superfamilies included WD40/YVTN repeat-like domain that helps in protein binding; the protein kinase superfamily which is a modulator of several signal transduction processes, and Armadillo-type fold superfamily which allows interaction of large substrates by providing an extensive solvent accessible surface (Table S3, Fig. 6B).

Moreover, differences were observed in the distribution of sites, with the PEST (proline-glutamic acid-serine-threonine) motif being present only in the transforming species and not in the non-transforming species (Table S4, Fig. 6C). The PEST-motif is a hydrophilic stretch of amino acids that is known to cause rapid proteolytic degradation.[32] Additionally, the thirty-three PEST-motif containing proteins in TA considered for further analysis were distributed across three gene families – *Sfi*I, SVSP, and Tash-AT, with 2, 18, and 13 members in each of them, respectively. The highly represented repeats observed in transforming, non-transforming, and outlier species included proteins of unknown function (DUF529, WD40), and tetratricopeptide repeat (TPR) (Table S5, Fig. 6D). The *Theileria*-specific FAINT domain (DUF529) is a repeat of 70 residues long with a conserved aromatic residue in the centre. The presence of the FAINT domain is a characteristic of the macroschizont secretome proteins.[33] The WD40 repeat is a compact motif consisting of approximately 40 amino acids, typically ending with a Trp-Asp (W-D) dipeptide sequence. Proteins containing this motif are commonly involved in signal transduction, protein–protein or protein-DNA interactions, transcription regulation, and cell cycle control. Proteins containing the TPR motif play crucial roles in mediating protein–protein interactions and facilitating the assembly of multi-protein complexes. They are also involved in various cellular processes, such as cell cycle regulation, transcriptional control, protein transport to mitochondria and peroxisomes, neurogenesis, and protein folding.

3.5. GPI-anchors, NLSs, signal peptides, trans-membrane domains of transforming Theileria species

Several proteins from TA that harboured PEST-motifs also possessed signal peptides, i.e., 29 out of 33, which indicate their possibility to be secreted into the cytoplasm of host cell, allowing them to alter signalling pathways (Table S6). Moreover, 20 members of the PEST-motif containing proteins had nuclear localization signals (NLS), which could facilitate their translocation into the host cell nucleus, where they may modulate the expression of important genes, ultimately contributing to host cell transformation. Additionally, 9 out of 33 proteins were found to have *trans*-membrane (TM) helices, suggesting their involvement in host-parasite interactions and the regulation of signal transmission, while none of the proteins contained GPI-anchors (Table S6). These findings align with a previous study in which proteins with signal peptides, NLS, and TM helices were shown to play a key role in host cell transformation.[34] Thus, the differential presence of PEST-motif containing proteins in TA can be associated with the adaptation of transforming *Theileria* species to preferential niches with specific host-parasite interactions.

3.6. Pest-motif containing proteins of transforming Theileria species

3.6.1. SfiI gene family

This sub-telomeric gene family contains an *Sfi*I restriction site and some of its members share sequence similarity with members of the SVSP family[35] with a few members being expressed at the macroschizont stage of *Theileria*.[33] The proteins encoded by members of this gene family are largely hypothetical and require further characterization. *Sfi*I sub-telomeric genes are located adjacent to SVSP genes, and research has shown a significant occurrence of duplication and recombination events within these gene families.[33] One of the members of *Sfi*I gene family, XP_952692.1 (KEGG: TA11400), is characterised by the presence of PEST motif (residues 62–76), multiple FAINT domains (residues 101–156, 199–270, 305–368, 412–466, 522–577, 619–677 and 723–773), and the lentiviral_Tat domain (residues 337–390). This protein consists of 799-residues and also includes a signal peptide from 1 to 20 amino acids.

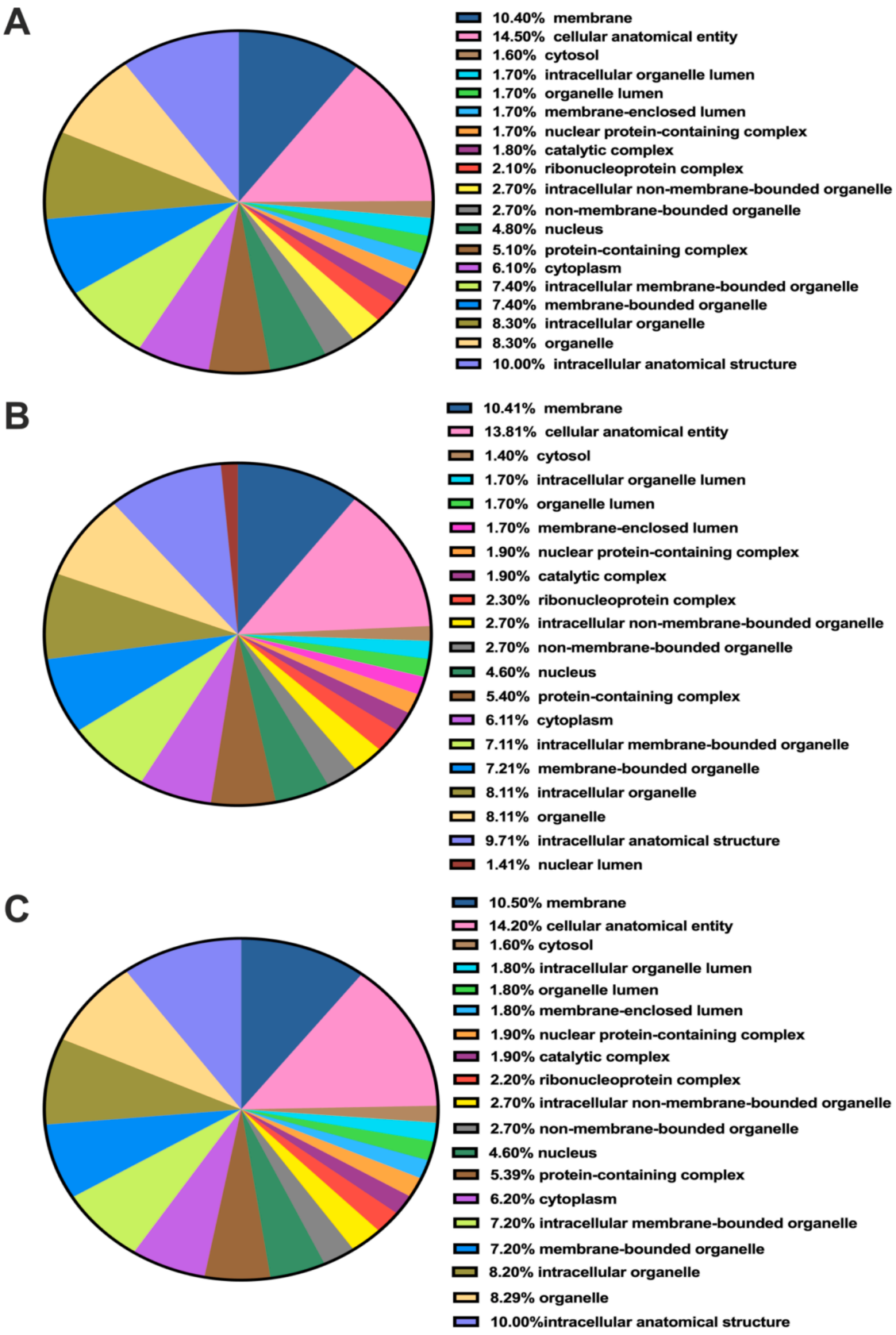

Fig. 4. Cellular components distribution of proteomes of *Theileria annulata* (A), *Theileria orientalis* (B), and *Theileria equi* (C).

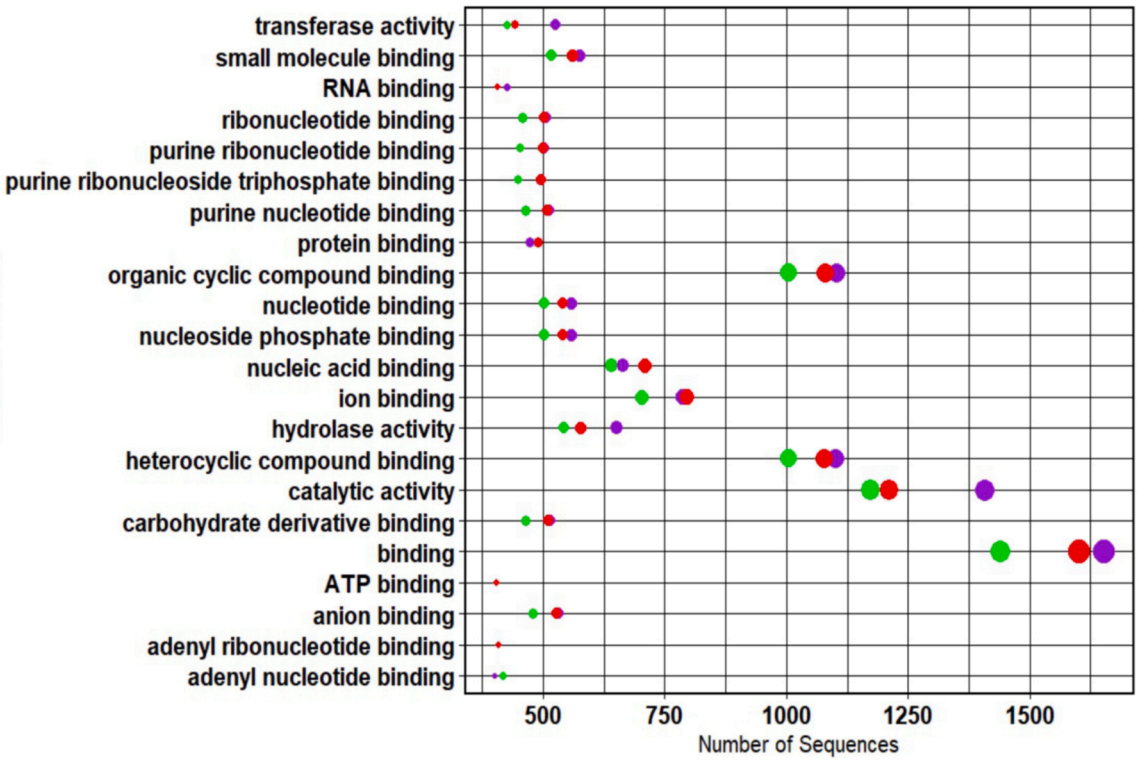

Fig. 5. Molecular function distribution of *Theileria* species. *Theileria annulata*, *Theileria orientalis*, and *Theileria equi* are represented by green, red, and purple circles, respectively. The proteins were predicted to primarily participate in catalytic activities and the binding of their cognate ligands across all three species. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.6.2. *Tash-AT gene family*

Out of the 17 tandemly arrayed Tash-AT genes present in TA, 13 were found to possess the PEST motif. One member of this family, XP_954167.1 (KEGG ID: TA03130), contains a FAINT domain between residues 26–87, while the PEST motif spans amino acids 242–256 and 364–380. The NTP-dependent RNA helicase NPH-II domain is located between residues 242–257. This family has been previously shown to encode for proteins that localize to the nucleus, where they bind to DNA and alter gene expression in the host cells.[36] Notably, 11 out of the 13 candidates were observed to contain an NLS (nucleus localizing sequence), and signal peptides were present in a few others (Table S6).

3.6.3. *SVSP gene family*

It is also a large sub-telomeric multigene family of *Theileria* that encodes macroschizont-specific proteins. With a total of 53 members, 18 of them were characterised by the presence of the PEST motif. One of the member of this family, XP_954704.1 (KEGG id: TA18860) contains the PEST motif (residues 612–628) and the FAINT domain (residues 798–847). The proteins encoded by members of the SVSP gene family typically consist of a short and conserved N-terminal region that usually contains a signal peptide. This sequence is followed by a highly unstructured region rich in glutamine and proline amino acids, and a conserved C-terminal region. Nuclear localization signals were predicted for 9 of the 18 candidates (Table S6), suggesting their potential to be translocated to the infected host cell nucleus.[33] Additionally, a signal peptide was present in all 18 of them (Table S6), indicating the possibility of secretion into the host cell compartment and considered as candidate parasite molecules involved in evasion of a protective immune response.[33]

3.7. *SfiIp contains the least disordered regions*

Since none of the 33 PEST-motif containing protein sequences had any homologs present in the Protein Data Bank (PDB), Robetta was employed to predict 3D structures of these proteins (*Sfi*I, SVSP, and Tash-AT). Among all the PEST-motif containing proteins, *Sfi*Ip displayed the most compact structure in terms of fewer disordered regions, leading to select this model for further analysis. From the five structures of *Sfi*Ip generated by Robetta, the first model with the highest confidence score of 0.73 was chosen for further studies (Fig. 7A). For assessment of the selected model, PROCHECK webserver was used after loop refinement of the structure. Ramachandran plot showed that almost 86 % of residues were in favourable regions, while about 14 % were in additional allowed, 0.3 % residues in generously allowed, and 0.1 % residues in disallowed regions (Fig. 7B). The ERRAT score which represents the overall quality of the protein was found to be 88.378. The ProSA-web server was utilized to evaluate the overall (global) and local quality of the refined protein structure, and the computed Z-score was observed to be −11.4 (Fig. 7C and D). The predicated secondary structural elements revealed presence of 14 beta sheets, 39 beta hairpins, 37 beta bulges, 67 strands, 14 alpha helices, 2 helix-helix interactions, 102 beta turns, and 19 gamma turns in the modelled structure of *Sfi*Ip.

Given that the catalytic activity and ligand binding sites of PEST-motif containing proteins have not been previously studied, we used CASTp and PrankWeb servers to predict the ligand binding pockets and sites, respectively. The first and the second binding cavities in *Sfi*Ip had a surface area of 7485.90 and 318.55 $Å^2$, respectively, with corresponding volumes of 19628.27 and 381.46 $Å^3$. The predicted ligand-binding residues in pocket 1 included Leu631, Glu633, Asn661, Cys662, Glu664, His676, Phe677, Asp678, Phe681, Leu722, Ser724, Thr725, Pro726, and Tyr730. Pocket 2 contained Ile158, Asn159, Ile162, Tyr178, Ile180, Glu187, Glu188, Ile189, Ser193, Gln196, Ile198, Ile230, Ala231, Lys233, Asn236, Ile238, Ala241, and Thr242 as ligand-interacting residues.

3.8. *Screening of FDA-approved drugs against SfiIp protein*

A total of 1984 FDA-approved drug molecules were virtually

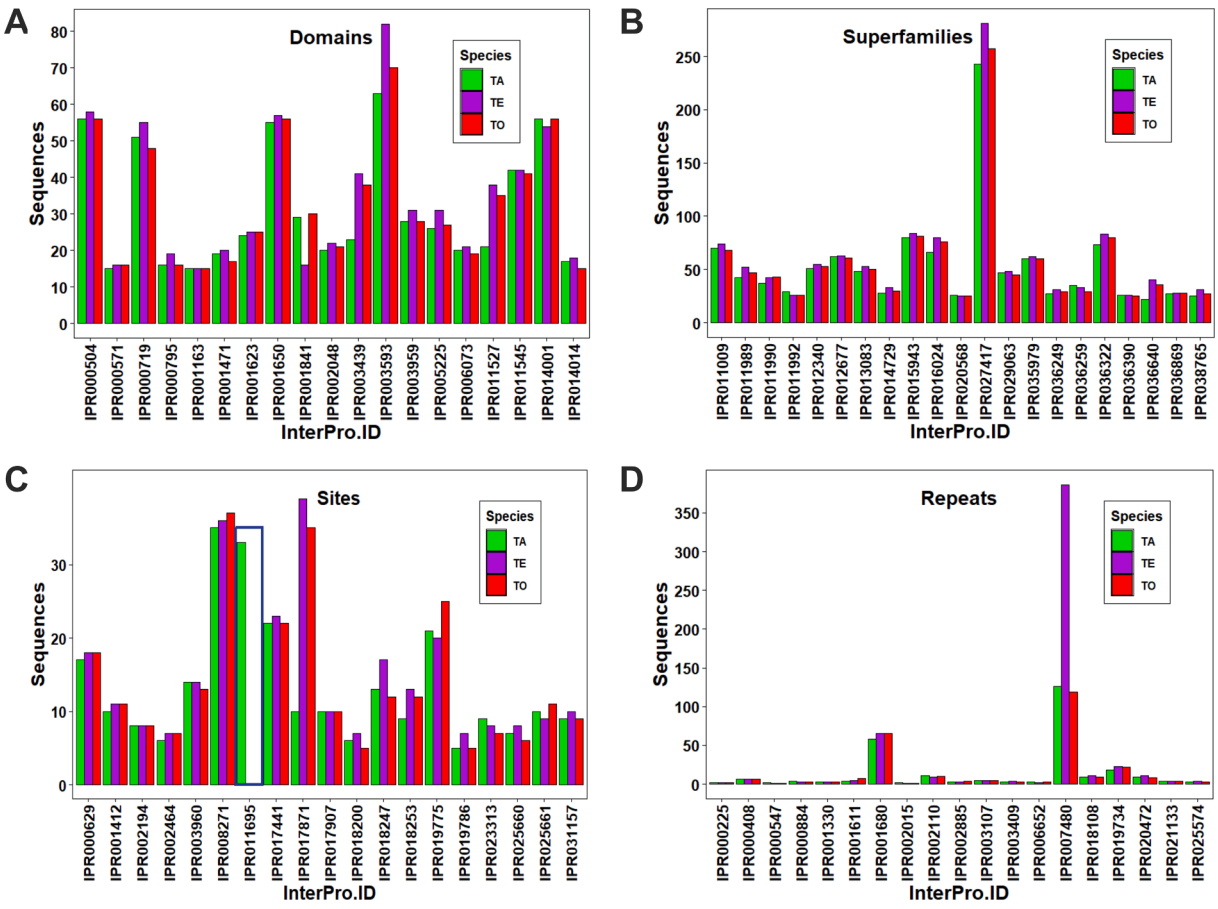

Fig. 6. Distribution of (a) protein domains, (b) superfamilies, (c) sites and (d) repeats in proteomes of *Theileria annulata* (green), *Theileria orientalis* (pale blue), and *Theileria equi* (orange). PEST motif (IPR011695) is characteristic to only the transforming species and absent in other species (*Theileria orientalis* and *Theileria equi*). Notably, the DUF529 repeat (IPR007480) is unique to *Theileria* genus and present in the three species. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

screened against *Sfi*Ip protein (Table S7). All the drugs were classified into three categories based on their binding affinities for *Sfi*Ip viz. −1.1 to −5, −5.1 to −7.9, −8.0 kcal/mol or above containing 143, 1334, and 507 compounds, respectively. The top 1 %, consisting of 20 compounds, were considered for further studies and their docking scores of these 20 drugs were determined. These twenty compounds showed affinities of ≥−9.9 kcal/mol, with 5 of them showing binding affinities of ≥−10.5 kcal/mol. Among them, rimegepant and atogepant demonstrated the highest binding affinities (−10.8 kcal/mol) towards *Sfi*Ip, while capmatinib, dutasteride, and pazopanib depicted corresponding binding affinities of −10.6, −10.6, and −10.5 kcal/mol. Based on the high binding affinity scores of atogepant and rimegepant, these compounds were chosen for further studies. Atogepant, an N-acyl-alpha amino acid derivative, formed several hydrophobic interactions with Glu81, His85, Val88, Ile750, Ala756, Tyr778, Phe782, Leu791, Val794, Lys795, and Asn796 of *Sfi*Ip. Moreover, several H-bonds were also observed involving residues Trp751, Glu752, Thr 754, and Asn755 (Table S7, Fig. 8A). Another studied complex, *Sfi*Ip-rimegepant (an imidazopyridine), did not form any hydrogen bonds but exhibited several hydrophobic interactions with His85, Val88, Thr749, Ile750, Trp751, Glu752, Thr754, Asn755, Ala756, Tyr778, Ile780, Phe782, Tyr789, Leu791, and Val794 of the protein (Table S7, Fig. 8B).

3.9. Simulation studies reveal comparable stability of SfiIp in the presence of ligands

MD simulations help to analyze dynamics of the protein with its ligands in the solvated state; hence, investigation of the stability of the protein-drug complexes relative to the apo protein was executed. The root mean square deviation (RMSD) measures was used to assess the stability of the protein structure by quantifying differences between its initial and final conformations. The apo protein exhibited an average RMSD of 0.80 nm, while the protein complexes with atogepant and rimegepant showed corresponding average RMSD values of 0.69 and 0.95 nm. Notably, the binding of atogepant resulted in the least deviation, suggesting greater structural stability of the protein in complex with atogepant compared to rimegepant (Fig. 9A). The radius of gyration (Rg) determines the overall compactness and conformational stability of protein structures studied during the simulation. A higher Rg value typically indicates less conformational rigidity and a more extended structure, whereas a lower Rg value suggests a more compact and stable structure. Although the apo protein had a Rg of 3.68 nm, among the three protein-drug complexes, binding of rimegepant displayed the most structural compactness with an average Rg of 3.39 nm, while atogepant-*Sfi*Ip demonstrated Rg of 3.51 nm, (Fig. 9B). This suggests that the binding of drug molecules facilitated a more compact structure of *Sfi*Ip.

The root mean square fluctuation (RMSF) measures deviation of each residue from its reference position over time, providing insights into the fluctuation and the degree of flexibility. The atogepant- and rimegepant-*Sfi*Ip complexes exhibited similar RMSF patterns, with average values of 0.35 and 0.36 nm, respectively, compared to an average RMSF of 0.43 nm for the apo protein (Fig. 9C). These values indicate a decrease in backbone fluctuation among the residues of *Sfi*Ip when bound to ligands in comparison to apo form. The Solvent-accessible surface area (SASA) indicates the region of the protein exposed to the solvent. A decrease in

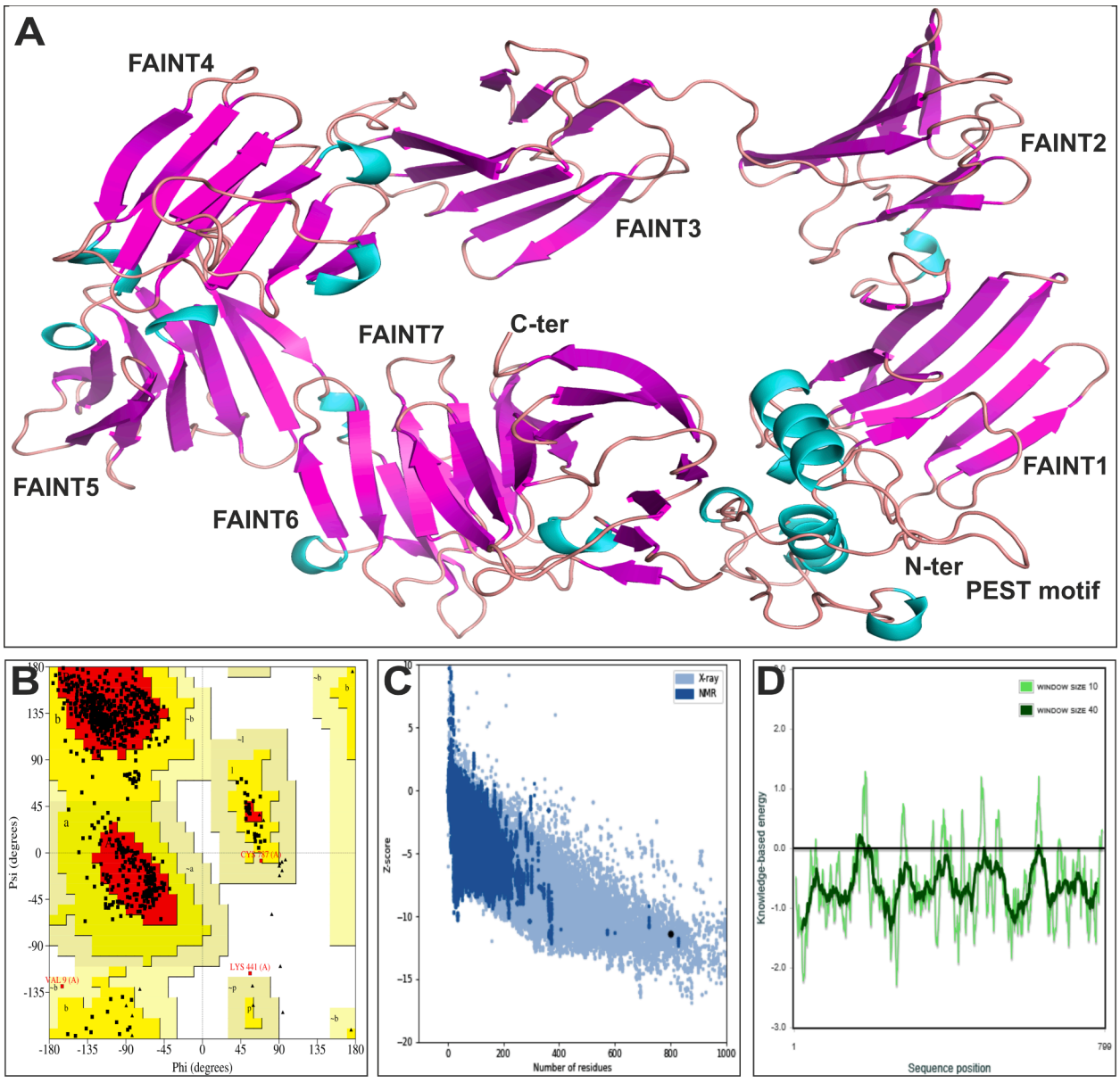

Fig. 7. (A) Modelled three-dimensional structure of *Sfi*Ip, which is a PEST motif-containing protein and encoded by a member of *Sfi*I gene family in *Theileria annulata*. Alpha helices, beta sheets and coils are coloured in cyan, magenta and orange, respectively. (B) Ramachandran plot of *Sfi*Ip obtained through PROCHECK server. (C) Overall and (D) local model quality of refined *Sfi*Ip structure calculated from ProSA webserver. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

SASA values upon ligand binding delineates favourable contraction of the protein, thus increase in hydrophobicity and closing the binding cavity. The apo protein showed SASA value of 475 nm^2, whereas the ligand-bound complexes displayed values of 468 and 474 nm^2, for atogepant, and rimegepant-*Sfi*Ip complexes, respectively, delineating favourable binding of the drugs within a hydrophobic binding pocket of the protein (Fig. 9D). The dynamics of intermolecular hydrogen bond for each complex were also analysed (Fig. 9E). Over the course of the 100 ns MD simulation, distinct patterns of intermolecular hydrogen bonding were observed for all complexes. Notably, the atogepant-*Sfi*Ip complex exhibited a higher number of hydrogen bonds while rimegepant-*Sfi*Ip complex formed only one to two hydrogen bonds throughout the simulation. During the initial 40 ns of simulation, an increase in the number of hydrogen bonds was seen across all systems. However, in the later stages, both complexes stabilized with a constant presence of one or more hydrogen bonds. These findings highlight the differential hydrogen bonding patterns in the protein-drug complexes, with atogepant demonstrating more sustained intermolecular interactions with *Sfi*Ip compared to rimegepant over the course of the simulation.

3.10. Binding of drug molecules reduces phase space occupancy in PCA

PCA (or, essential dynamics) helps in comprehending the collective dynamics of protein systems by reducing the complexity of system's dimensionality. It also provides insights into the ligand induced protein movement by generating eigenvectors. In this study, PCA was applied to the apo protein to compare its overall motion in conformational space, wherein the covariance trace value of apo protein state was 18.87 nm^2 (Fig. 10A). A higher trace value indicates greater flexibility, thus occupying a larger phase space area in the PCA plots. In all the cases, upon ligand binding, the covariance trace values reduce drastically, indicating the overall decreased flexibility of system. The effect of atogepant and rimegepant binding resulted in trace values of 12.30 (Fig. 10B) and 9.98 nm^2 (Fig. 10C), respectively. Therefore, ligand binding caused the

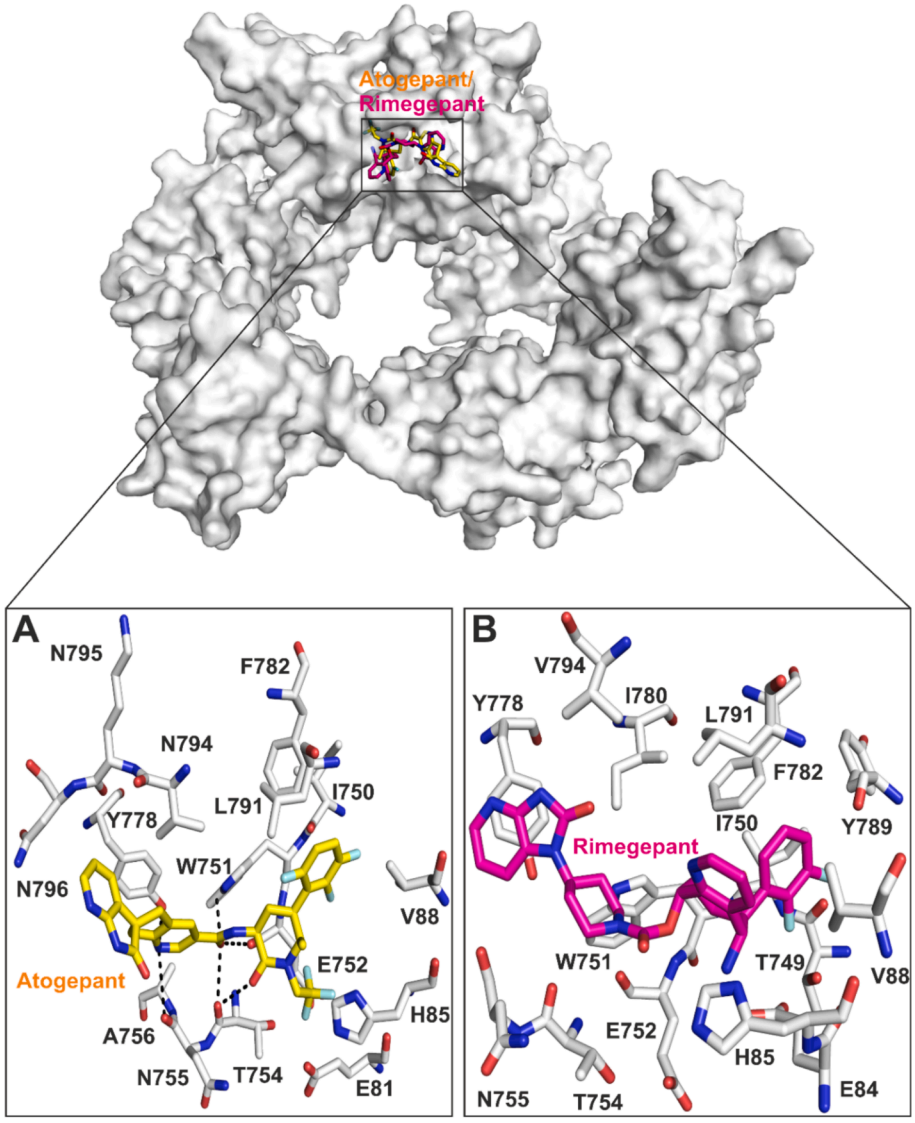

Fig. 8. Interaction analysis of *Sfi*Ip with lead molecules. Docking of *Sfi*Ip was performed with FDA-approved drugs through AutoDock Vina, wherein protein is presented as white coloured surface view, while atogepant, and rimegepant are represented as yellow, and magenta coloured sticks, respectively. The interacting residues of protein with atogepant (A) and rimegepant (B) are shown as white sticks, whereas black dashed lines represent the polar contacts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

clusters to be compressed in conformational space in comparison to the apo protein (Fig. 10D).

3.11. *SfiIp becomes thermodynamically stable in presence of ligands*

The FEL was constructed using the top two principal components (PC1 & PC2) to analyze the effect of ligand binding on conformational redistributions for apo protein and its complex systems (Fig. S1). A strong and stable interaction typically results in a single conformational cluster (global energy minima) depicted by a purple colour in the potential energy map. In contrast, weaker or unstable protein–ligand interactions may lead to multiple energy minima in the FEL, indicating a less stable or fluctuating conformational state. The apo form displayed multiple local energy minima across the landscape, characterized by several clusters, delineating greater conformational flexibility and reduced stability. Conversely, both *Sfi*Ip-atogepant and *Sfi*Ip-rimegepant complexes exhibited fewer and more pronounced energy minima, with highly concentrated low-energy regions marked in blue and purple. Specifically, *Sfi*Ip-atogepant complex demonstrated a dominant global energy minimum, implying a strong and stable interaction. Similarly, the *Sfi*Ip-rimegepant complex showed a slightly broader energy distribution, suggesting stable binding that is marginally less constrained compared to atogepant.

3.12. *In silico analysis indicates strong binding between SfiIp and lead molecules*

To assess the free binding energy of *Sfi*Ip–drug complexes utilizing MM/PBSA method, the trajectories of MDS were extracted and the snapshots were collected at every 100 ps at stable intervals from 80 to 100 ns. The results designated that the two lead compounds – atogepant and rimegepant possessed a corresponding free binding energies of −63.720 and −40.927 kJ/mol. Furthermore, in the context of total interaction energy, van der Waals and electrostatic interactions as well as non-polar solvation energy exerted negatively, whereas polar solvation energy contributed positively to the total free binding energy. Electrostatic energy was calculated to be −14.628 and −32.244 kJ/mol, whereas the polar solvation energy was 54.639 and 100.992 kJ/mol for *Sfi*Ip bound with atogepant and rimegepant, respectively (Table 1). The relative free binding energies of the two protein-drug complexes

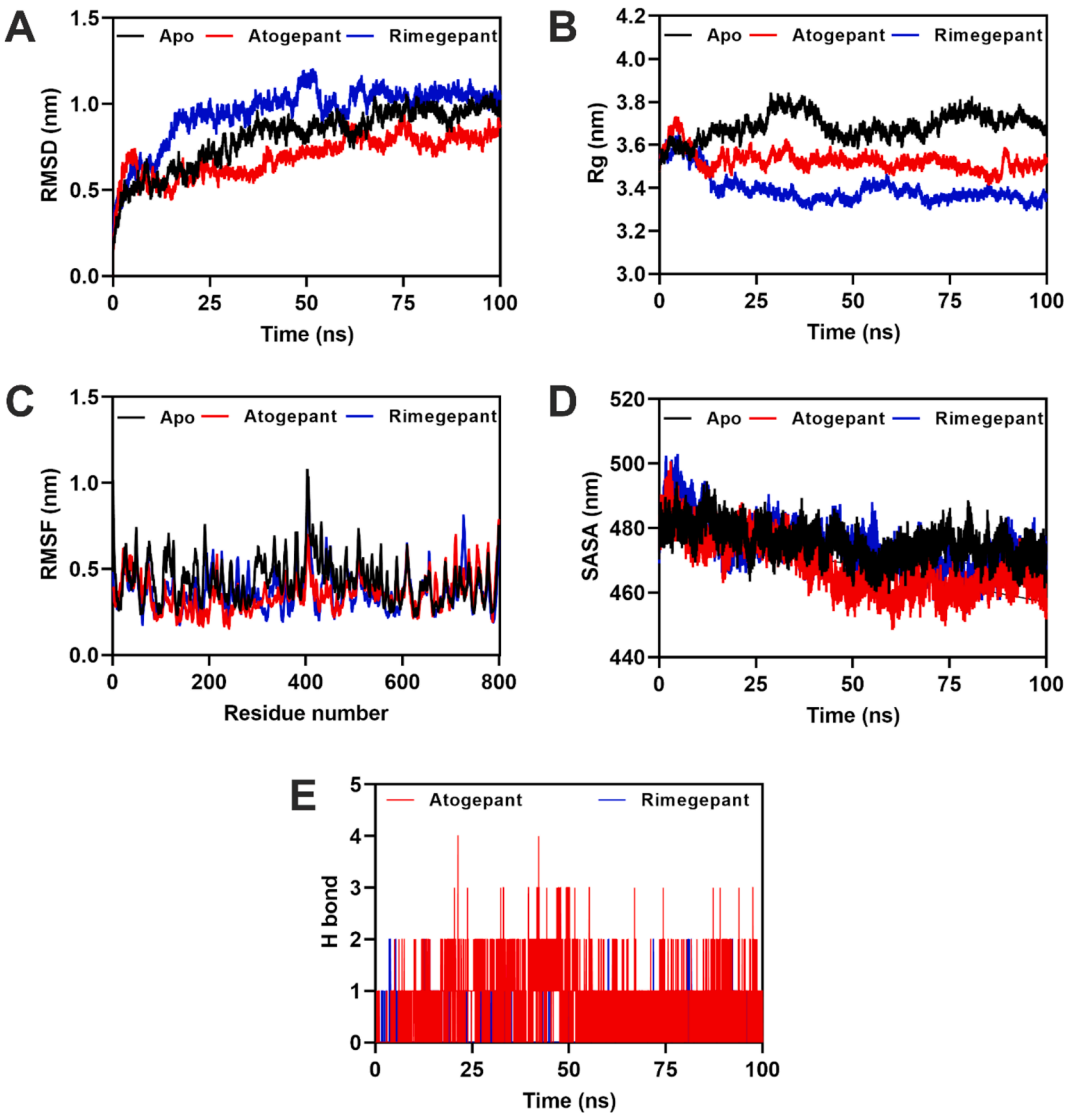

Fig. 9. MDS studies of apo *Sfi*Ip and its complexes with lead molecules. Comparative assessment of RMSD (A), Rg (B), RMSF (C), and SASA (D) of apo, atogepant and rimegepant bound *Sfi*Ip. Illustration of hydrogen bonds formed for *Sfi*Ip complexes (E) during the simulation period of 100 ns.

indicated towards strong binding during the dynamics.

4. Discussion

Theileriosis is a protozoal disease caused by various species of intracellular parasites belonging to the genus *Theileria*. These parasites are transmitted by ticks and primarily affect cattle, leading to significant economic losses in livestock production. The severity of the disease can vary depending on the species of *Theileria*, with transforming species causing the most fatal form of the disease. Treatment of theileriosis involves antiparasitic drugs, but preventative measures such as tick control, vaccination, and quarantine are also critical in the impeding of disease. In addition to its role in establishing and maintaining the infection, host cell transformation by *Theileria* has also been implicated in the development of cancer-like states in infected animals.[37] However, the existing therapeutics used to combat theileriosis have drawbacks, such as limited efficacy, development of drug resistance in the parasites, and side effects that can affect animal health and productivity.[38]

In our study, the host-cell transforming species were seen to have smaller genomic sizes in comparison to the non-host cell transforming species, which is in accordance with the fact that virulent organisms contain small genomes than their non-virulent counterparts.[30] A previous study showed that both the genera of *Theileria* and *Babesia* are closely related within the phylum Apicomplexa, but belong to separate clades or lineages within the order Piroplasmida.[39] Similarly, in our study *B. bovis* was observed to be the last common ancestor of *Theileria* species. Functional anotation revealed that the biological processes, molecular functions, and cellular localization were quite similar across all three *Theileria* species. This suggests that the fundamental physiological activities of the parasite remain consistent across transforming, non-transforming, and outlier species. Moreover, our analysis of *Theileria* proteomes revealed that most proteins were associated with various metabolic processes and gene expression. These findings were consistent with previous studies that characterized the functional catalogues of TA proteins using one-dimensional gel LC/MS, highlighting their involvement in essential biological processes.[40] A significant proportion of these proteins were membrane-bound across TA, TO, and TE, underscoring their role in process of host cell invasion.[41] Occurrence of specific motifs in protein sequences reflect evolutionary conservation and functional significance that often play crucial roles in protein structure, function, and interactions within biological systems.[42] Motif analysis of whole proteomes of TA, TO, and TE revealed the PEST motif to be uniquely conserved in only the transforming species, while they were completely absent from the non-transforming and outlier species. PEST motifs are

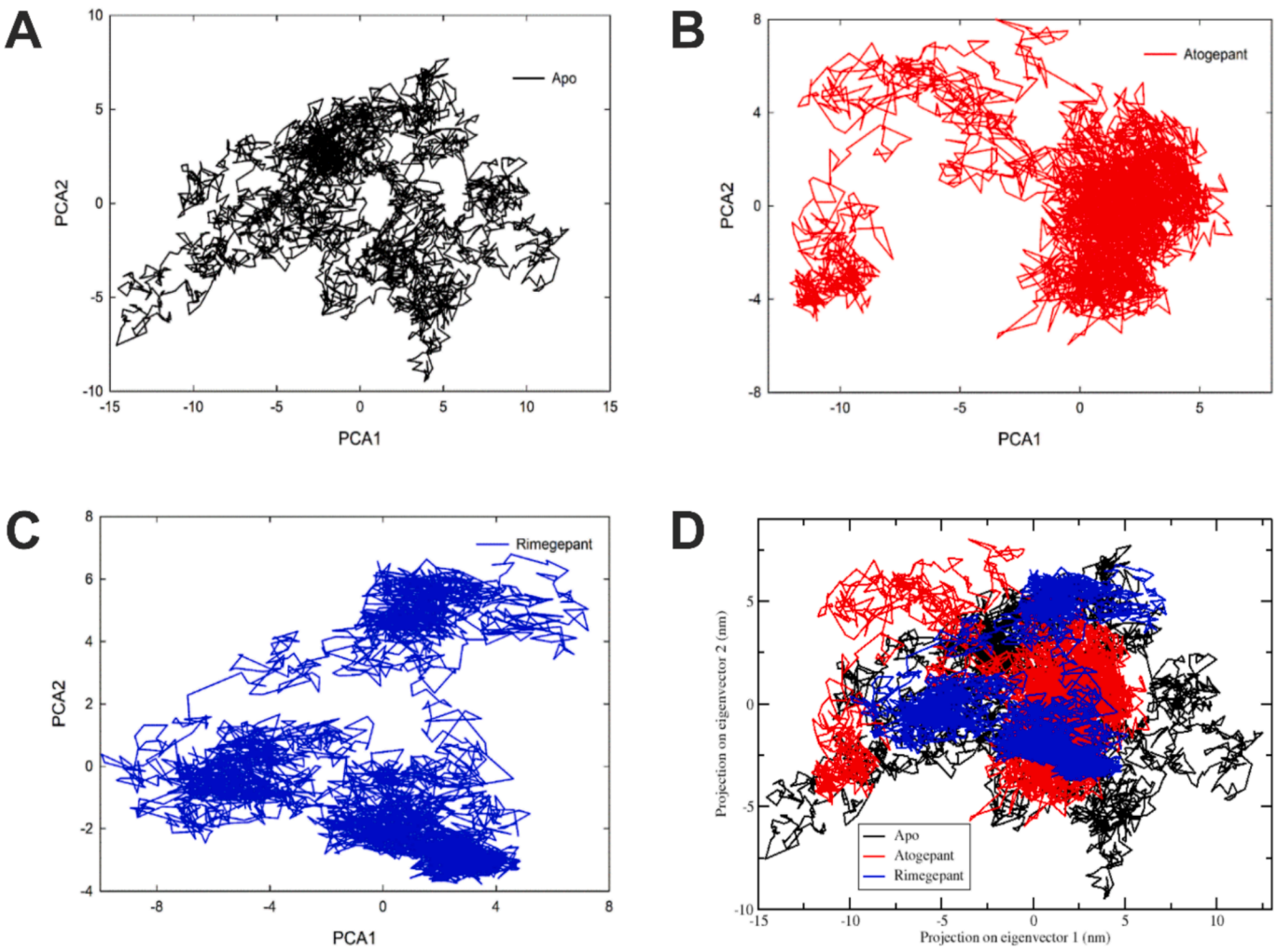

Fig. 10. Principal component analysis (PCA) plots constructed by eigenvector 1 vs eigenvector 2 of apo *Sfi*Ip (A) and its complexes with atogepant (B), and rimegepant (C). The combination of plots for apo (black), atogepant (red) and rimegepant (blue) bound *Sfi*Ip (D). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
MM/PBSA analysis of *Sfi*Ip-drug complexes.

Energy (kJ/mol)	Atogepant-*Sfi*Ip complex	Rimegepant-*Sfi*Ip complex
van der Waal	−92.665	−95.974
Electrostatic	−14.628	−32.244
Polar solvation	54.639	100.992
SASA	−11.067	−13.701
Binding	−63.720	−40.927

typically found in highly regulated proteins with short half-lives, facilitating rapid degradation.[32] Because these proteins need to be tightly controlled, PEST motifs help to ensure their timely degradation when they are no longer needed, preventing excessive or prolonged activity that could disrupt cellular function. *Theileria*-induced transformation relies on continuous activation of host cell signalling pathways,[43] and PEST motifs could help to fine-tune the activity of parasite proteins involved in this process. This might explain why only transforming *Theileria* species have PEST-containing proteins, while non-transforming species do not require such rapid protein regulation. A comparison can be made with *P. falciparum*, where PEST regions are enriched in *Pf*EMP-1, a class of extracellular proteins that mediate host cell adhesion and contribute to malaria pathology.[44] Additionally, the human protein tyrosine phosphatase, which contains a C-terminal PEST motif, has been implicated in oncogenic processes by influencing cell shape and motility.[45] These observations suggest that PEST-mediated proteolysis may be crucial for transformation-related processes in *Theileria*. However, further experimental validation is required to determine their precise functional significance in *Theileria*-mediated transformation and disease progression. Understanding the function of PEST motifs in *Theileria* will help to shed light on parasite-host interactions and identify the potential therapeutic targets for controlling theileriosis.

As suggested by several studies, proteins comprising NLS, TM domains, and signal peptides play significant roles in the pathogenicity of *P. falciparum* (another apicomplexan) by regulating gene expression and manipulating host cells, thereby influencing immune evasion and virulence.[46] For instance, TM domains contribute to invasion and nutrient acquisition by mediating interactions with host cell membranes (e.g., AMA1),[47] signal peptide-comprising proteins are essential for protein secretion and export (e.g., SERA proteins),[48] etc. In line with this, the majority of the PEST-motif containing proteins highlighted in our study were also predicted to possess signal peptides, suggesting their secretory nature and potentially linking them to the adaptation of transforming *Theileria* species to specific host-parasite interactions within preferred niches.

Gene families situated in the sub-telomeric regions of parasite genomes harbour genes that encode for antigens and show high levels of variation in sequence. This variation helps the parasite in evading the host's immune response and facilitates host cell invasion as seen in the

var, *rif*, and *stevor* gene families of *P. falciparum*.[49] Conversely, proteins that facilitate the parasite's adaptation to specific environmental conditions in the host are more likely to remain conserved within the same species. Analogous to these gene families in *P. falciparum*, we observed that PEST-motif containing proteins are majorly distributed in SVSP gene family in *Theileria* along with *Sfi*I and the Tash-AT gene families. In *Theileria* parasites, subtelomeric gene families like Tash-AT contribute to host cell transformation and immune modulation, influencing disease pathogenesis.[34,50] Thus, understanding the roles of these gene families is crucial for interpreting parasite virulence mechanisms and developing targeted interventions to combat apicomplexan infections. Recent studies highlight promising therapies for theileriosis, including histone deacetylase inhibitors (HDACi) like vorinostat and belinostat, which exhibit potent anti-theilerial activity with low host toxicity and effective IC_{50} values.[51] Compounds from the MMV Pathogen Box, such as SAHA, Trichostatin A, and Plumbagin, target pathways like HDAC inhibition and ROS/apoptosis induction.[52] Trifloxystrobin, a broad-spectrum fungicide, effectively reduces *Theileria* parasite load and shows efficacy against buparvaquone-resistant strains.[53]

Despite these advances, challenges persist due to limited understanding of host cell invasion by *Theileria*, a large proportion of the proteins being annotated as hypothetical, and resistance to existing drugs.[34,54] Thus, the identification of both crucial drug targets and novel drug molecules against these targets is required to combat *Theileria* infections effectively. As PEST-containing proteins are often involved in host-pathogen interactions,[44] in the current study, the structure of *Sfi*Ip was predicted along with its ligand-binding sites as it exhibited the most stable and complete conformation among the analyzed PEST-motif-containing proteins. Imidazopyridine derivatives have demonstrated strong anti-trypanosomal activity against *Trypanosoma brucei* and antibacterial effects against various pathogens.[55,56] Similarly, carboxylic acid derivatives have also shown antiparasitic potential. Biarylalkyl carboxylic acid derivatives displayed significant anti-schistosomal activity against *Schistosoma mansoni* at concentrations as low as 10 μM.[57] Furthermore, triazole derivatives of dehydroabietic acid and oleanolic acid, which contain carboxylic acid groups, exhibited antiprotozoal activity against *Leishmania* and *Trypanosoma cruzi*.[58] These findings align with our study, highlighting the potential of derivatives of imidazopyridine (atogepant) and carboxylic acid (atogepant) as promising anti-*Theileria* agents. The complex systems of *Sfi*Ip further showed stability, compactness and reduced flexibility during simulation, delineating stable interactions of the ligands in the binding pocket. Moreover, binding analysis revealed that *Sfi*Ip-lead inhibitor complexes formed highly stable systems with stronger binding affinities, characterized by significantly negative binding energies.

5. Conclusion

The present study highlights key proteins of *Theileria annulata* that have the potential to transform mammalian host leukocytes into immortalized cells thereby mimicking tumour cells and spreading extensively. An *in silico* analysis of annotated proteomes was applied to identify such host cell manipulators, and gene ontology was employed to annotate the majority of unannotated proteins. PEST-motif containing proteins associated with signal peptides, nuclear localization signals, or transmembrane helices were predicted to be highly likely to transform host cells. Moreover, atogepant and rimegepant were proposed as promising anti-theilerial drugs, showing strong interactions with the targeted parasitic protein *Sfi*Ip. Further, *in vitro* validation of target parasitic proteins, screening for binding partners in host cells, followed by downstream analysis will aid in a better understanding of host-pathogen interactions and the underlying mechanisms mediated by their corresponding inhibitors. Our investigation thus provides an *in silico* comparative proteome analysis that can be further implemented to identify effective therapeutics against *Theileria* infections. Furthermore, this study underscores the potential of leveraging structure-based drug design and high-throughput screening for the development of specific inhibitors targeting parasitic proteins. Such efforts could pave the way for the creation of more efficient, selective, and sustainable therapeutic strategies to combat *Theileria*-induced diseases in mammals.

CRediT authorship contribution statement

Anusha Majumder: Writing – original draft, Methodology, Investigation, Formal analysis. **Jyotisha:** Writing – original draft, Formal analysis. **Fouzia Nasim:** Writing – original draft, Formal analysis. **Insaf Ahmed Qureshi:** Conceptualization, Supervision, Resources, Writing – review & editing.

Funding

No funding was received for conducting this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to express their gratitude to the Centre for Modelling Simulation & Design (CMSD), University of Hyderabad, for providing computational resources for the MD simulation studies. Anusha Majumder and Jyotisha are grateful to AICTE and PMRF for their research fellowships. The authors also extend their thanks to Dr. Paresh Sharma, NIAB Hyderabad, for his insightful discussions during the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgeb.2025.100488.

Data availability

The data that support the findings in this study are included within the manuscript and Supplementary information files.

References

1. Bishop R, Musoke A, Morzaria S, Gardner M, Nene V. Theileria: intracellular protozoan parasites of wild and domestic ruminants transmitted by ixodid ticks. *Parasitology*. 2004;129(Suppl):S271–S283. https://doi.org/10.1017/s0031182003004748.
2. Nene V, Kiara H, Lacasta A, Pelle R, Svitek N, Steinaa L. The biology of Theileria parva and control of East Coast fever – current status and future trends. *Ticks Tick Borne Dis*. 2016;7(4):549–564. https://doi.org/10.1016/j.ttbdis.2016.02.001.
3. Gou H, Guan G, Liu A, et al. A DNA barcode for piroplasmea. *Acta Trop*. 2012;124:92–97. https://doi.org/10.1016/j.actatropica.2012.07.001.
4. Wang BH, Du LF, Zhang MZ, et al. Genomic characterization of Theileria luwenshuni strain cheeloo. *Microbiol Spectr*. 2023;11, e0030123. https://doi.org/10.1128/spectrum.00301-23.
5. Clift SJ, Collins NE, Oosthuizen MC, Steyl JCA, Lawrence JA, Mitchell EP. The pathology of pathogenic theileriosis in African wild artiodactyls. *Vet Pathol*. 2020;57:24–48. https://doi.org/10.1177/0300985819879443.
6. Kappmeyer LS, Thiagarajan M, Herndon DR, et al. Comparative genomic analysis and phylogenetic position of Theileria equi. *BMC Genomics*. 2012;13:603. https://doi.org/10.1186/1471-2164-13-603.
7. Radley DE, Brown CGD, Burridge MJ, et al. East coast fever: 1 Chemoprophylactic immunization of cattle against Theileria parva (Muguga) and five theilerial strains. *Vet. Parasitol*. 1975;1:35–41. https://doi.org/10.1016/0304-4017(75)90005-9.
8. Morzaria S, Nene V, Bishop R, Musoke A. Vaccines against Theileria parva. *Ann N Y Acad Sci*. 2000;916:464–473. https://doi.org/10.1111/j.1749-6632.2000.tb05326.x.
9. Neitz WO. Aureomycin in Theileria parva infection. *Nature*. 1953;171:34–35. https://doi.org/10.1038/171034a0.

10. Rashid M, Hayat MH, Zahra N, et al. Systematic review on buparvaquone resistance associated with non-synonymous mutation in drug binding genes site of Theileria annulate. *Vet Parasitol.* 2024;332, 110321. https://doi.org/10.1016/j.vetpar.2024.110321.
11. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y.
12. Emms DM, Kelly S. STRIDE: species tree root inference from gene duplication events. *Mol Biol Evol.* 2017;34:3267–3278. https://doi.org/10.1093/molbev/msx259.
13. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol.* 2012;61:1061–1067. https://doi.org/10.1093/sysbio/sys062.
14. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21:3674–3676. https://doi.org/10.1093/bioinformatics/bti610.
15. Conesa A, Götz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics.* 2008;2008, 619832. https://doi.org/10.1155/2008/619832.
16. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem.* 2009;30:2785–2791. https://doi.org/10.1002/jcc.21256.
17. Narsimulu B, Jakkula P, Qureshi R, Nasim F, Qureshi IA. Inhibition and structural insights of leishmanial glutamyl-tRNA synthetase for designing potent therapeutics. *Int J Biol Macromol.* 2024;254, 127756. https://doi.org/10.1016/j.ijbiomac.2023.127756.
18. Nasim F, Kumar MS, Alvala M, Qureshi IA. Unraveling the peculiarities and development of novel inhibitors of leishmanial arginyl-tRNA synthetase. *FEBS J.* 2024;291:2955–2979. https://doi.org/10.1111/febs.17122.
19. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Cheminform.* 2011;3:33. https://doi.org/10.1186/1758-2946-3-33.
20. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *J Chem Inf Model.* 2021;61:3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.
21. Van Der Spoel A, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem.* 2004;26:1701–1718. https://doi.org/10.1002/jcc.20291.
22. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 1996;118:11225–11236. https://doi.org/10.1021/ja9621760.
23. Dodda LS, Cabeza de Vaca I, Tirado-Rives J, Jorgensen WL. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* 2017;45(W1):W331–W336. https://doi.org/10.1093/nar/gkx312.
24. Kumar J, Jyotisha, Qureshi R, Jagruthi P, Arifuddin M, Qureshi IA. Discovery of 8-hydroxy-2-quinoline carbaldehyde derivatives as inhibitors for M1 aminopeptidase of Leishmania donovani. *Int J Biol Macromol.* 2024;279:135105. https://doi.org/10.1016/j.ijbiomac.2024.135105.
25. Narsimulu B, Qureshi R, Jakkula P, Singh P, Arifuddin M, Qureshi IA. Exploration of seryl tRNA synthetase to identify potent inhibitors against leishmanial parasites. *Int J Biol Macromol.* 2023;237, 124118. https://doi.org/10.1016/j.ijbiomac.2023.124118.
26. Jyotisha, Qureshi R, Qureshi IA. Exploration of membrane-bound ecto-phosphatase to identify potential therapeutic target for leishmaniasis. *Int J Biol Macromol.* 2025;307:141820. https://doi.org/10.1016/j.ijbiomac.2025.141820.
27. Choudhir G, Kumar S, Shahid M, Shamsi A, Islam A. Telomeric RNA quadruplexes as targets for cancer prevention: the therapeutic potential of agonodepsides. *J Genet Eng Biotechnol.* 2025;23, 100454. https://doi.org/10.1016/j.jgeb.2024.100454.
28. Kumari R, Kumar R. Open source drug discovery consortium, Lynn A. g_mmpbsa–a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model.* 2014;54:1951–1962. https://doi.org/10.1021/ci500020m.
29. Singh S, Qureshi IA. Identification of potent inhibitors against chorismate synthase of Toxoplasma gondii using molecular dynamics simulations. *J Mol Graph Model.* 2022;114, 108183. https://doi.org/10.1016/j.jmgm.2022.108183.
30. Sakharkar KR, Dhar PK, Chow VTK. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. *Int J Syst Evol Microbiol.* 2004;54:1937–1941. https://doi.org/10.1099/ijs.0.63090-0.
31. Cuy-Chaparro L, Bohórquez MD, Arévalo-Pinzón G, et al. Babesia Bovis ligand-receptor interaction: AMA-1 contains small regions governing bovine erythrocyte binding. *Int J Mol Sci.* 2021;22:714. https://doi.org/10.3390/ijms22020714.
32. Mitchell D, Bell A. PEST sequences in the malaria parasite Plasmodium falciparum: a genomic study. *Malar J.* 2003;2:16. https://doi.org/10.1186/1475-2875-2-16.
33. Weir W, Karagenç T, Baird M, Tait A, Shiels BR. Evolution and diversity of secretome genes in the apicomplexan parasite Theileria annulata. *BMC Genomics.* 2010;11:42. https://doi.org/10.1186/1471-2164-11-42.
34. Hayashida K, Hara Y, Abe T, et al. Comparative genome analysis of three eukaryotic parasites with differing abilities to transform leukocytes reveals key mediators of Theileria-induced leukocyte transformation. *mBio.* 2012;3:e00204–e00212. https://doi.org/10.1128/mBio.00204-12.
35. Norling M, Bishop RP, Pelle R, et al. The genomes of three stocks comprising the most widely utilized live sporozoite Theileria parva vaccine exhibit very different degrees and patterns of sequence divergence. *BMC Genomics.* 2015;16:729. https://doi.org/10.1186/s12864-015-1910-9.
36. Shiels A, Langsley G, Weir W, Pain A, McKellar S, Dobbelaere D. Alteration of host cell phenotype by Theileria annulata and Theileria parva: mining for manipulators in the parasite genomes. *Int J Parasitol.* 2006;36:9–21. https://doi.org/10.1016/j.ijpara.2005.09.002.
37. Tretina K, Gotia HT, Mann DJ, Silva JC. Theileria-transformed bovine leukocytes have cancer hallmarks. *Trends Parasitol.* 2015;31:306–314. https://doi.org/10.1016/j.pt.2015.04.001.
38. Nene V, Kiara H, Lacasta A, Pelle R, Svitek N, Steinaa L. The biology of Theileria parva and control of East Coast fever – current status and future trends. *Ticks Tick Borne Dis.* 2016;7:549–564. https://doi.org/10.1016/j.ttbdis.2016.02.001.
39. Criado-Fornelio A, Martinez-Marcos A, Buling-Saraña J-C. Molecular studies on Babesia, Theileria and Hepatozoon in southern Europe. Part II. Phylogenetic analysis and evolutionary history. *Vet Parasitol.* 2003;114:173–194. https://doi.org/10.1016/s0304-4017(03)00141-9.
40. Witschi M, Xia D, Sanderson S, Baumgartner M, Wastling JM, Dobbelaere DA. Proteomic analysis of the Theileria annulata schizont. *Int J Parasitol.* 2013;43:173–180. https://doi.org/10.1016/j.ijpara.2012.10.017.
41. Collins CR, Blackman MJ. Apicomplexan AMA1 in host cell invasion: a model at the junction? *Cell Host Microbe.* 2011;10:531–533. https://doi.org/10.1016/j.chom.2011.11.006.
42. Livesay DR, La D. The evolutionary origins and catalytic importance of conserved electrostatic networks within TIM-barrel proteins. *Protein Sci.* 2005;14:1158–1170. https://doi.org/10.1110/ps.041221105.
43. Dobbelaere DA, Küenzi P. The strategies of the Theileria parasite: a new twist in host-pathogen interactions. *Curr Opin Immunol.* 2004;16:524–530. https://doi.org/10.1016/j.coi.2004.05.009.
44. Singh GP, Ganapathi M, Sandhu KS, Dash D. Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins.* 2006;62:309–315. https://doi.org/10.1002/prot.20746.
45. Garton AJ, Burnham MR, Bouton AH, Tonks NK. Association of PTP-PEST with the SH3 domain of p130cas; a novel mechanism of protein tyrosine phosphatase substrate recognition. *Oncogene.* 1997;15:877–885. https://doi.org/10.1038/sj.onc.1201279.
46. De Silva EK, Gehrke AR, Olszewski K, et al. Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci USA.* 2008;105:8393–8398. https://doi.org/10.1073/pnas.0801993105.
47. Alexander DL, Mital J, Ward GE, Bradley P, Boothroyd JC. Identification of the moving junction complex of Toxoplasma gondii: a collaboration between distinct secretory organelles. *PLoS Pathog.* 2005;1:e17. https://doi.org/10.1371/journal.ppat.0010017.
48. Miller SK, Good RT, Drew DR, et al. A subset of Plasmodium falciparum SERA genes are expressed and appear to play an important role in the erythrocytic cycle. *J Biol Chem.* 2002;277:47524–47532. https://doi.org/10.1074/jbc.M206974200.
49. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic variation in Plasmodium falciparum. *Annu Rev Microbiol.* 2008;62:445–470. https://doi.org/10.1146/annurev.micro.61.080706.093134.
50. Schmuckli-Maurer J, Casanova C, Schmied S, et al. Expression analysis of the Theileria parva subtelomere-encoded variable secreted protein gene family. *PLoS One.* 2009;4, e4839. https://doi.org/10.1371/journal.pone.0004839.
51. Barman M, Kamble S, Roy S, et al. Antitheilerial activity of the anticancer histone deacetylase inhibitors. *Front Microbiol.* 2021;12, 759817. https://doi.org/10.3389/fmicb.2021.759817.
52. Kamble S, Singh S, Suresh A, et al. Epidrugs: alternative chemotherapy targeting Theileria annulata schizont stage parasites. *Microbiol Spectr.* 2024;12, e0325823. https://doi.org/10.1128/spectrum.03258-23.
53. Villares M, Lourenço N, Berthelet J, et al. Trifloxystrobin blocks the growth of Theileria parasites and is a promising drug to treat Buparvaquone resistance. *Commun Biol.* 2022;5:1253. https://doi.org/10.1038/s42003-022-03981-x.
54. Villares M, Lourenço N, Ktorza I, et al. Theileria parasites sequester host eIF5A to escape elimination by host-mediated autophagy. *Nat Commun.* 2024;15:2235. https://doi.org/10.1038/s41467-024-45022-7.
55. Silva DG, Junker A, de Melo SMG, et al. Synthesis and structure-activity relationships of imidazopyridine/pyrimidine- and furopyridine-based anti-infective agents against trypanosomiases. *ChemMedChem.* 2021;16:966–975. https://doi.org/10.1002/cmdc.202000616.
56. Hamdi C, Elyoussfi MHA, et al. Synthesis and medicinal applications of imidazopyridine and imidazothiazole derivatives: a review. In: *E3S Web of Conferences.* EDP Sciences; 2024:01014. https://doi.org/10.1051/e3sconf/202452701014.
57. Mäder P, Blohm AS, Quack T, et al. Biarylalkyl carboxylic acid derivatives as novel antischistosomal agents. *ChemMedChem.* 2016;11:1459–1468. https://doi.org/10.1002/cmdc.201600127.
58. Pertino MW, Vega C, Rolón M, Coronel C, Rojas de Arias A, Schmeda-Hirschmann G. Antiprotozoal activity of triazole derivatives of dehydroabietic acid and oleanolic acid. *Molecules.* 2017;22:369. https://doi.org/10.3390/molecules22030369.