STRATHCLYDE INSTITUTE OF PHARMACY & BIOMEDICAL SCIENCES

The Evolutionary and Structural Analysis of Cyclopropane Fatty-Acyl Phospholipid Synthetase in Leishmania Infantum

Author: Amber Minhas

Supervisor: Dr Leighton Pritchard

Signed:  *Date:* 06/03/2022

Word count: 4510

A thesis submitted in partial fulfilment for the degree of BSc (Hons) in **Biomedical Science**. This report is entirely my own work. Any information taken from others has been declared and referenced in the text.

Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow G4 0RE

Table of Contents

Technical Abstract

Leishmania is an intracellular protozoan parasite causing leishmaniasis, the third most important vector-borne parasitic disease leading to 60,000 deaths and infecting up to 1.2 million people worldwide. CfaS (cyclopropane fatty-acyl phospholipid synthetase) is protein which belongs to the SAM methyltransferase family. In this project it was investigated as a possible required protein for the progression of *Leishmania infantum* within its host. This was done by analysing Its structural evolution to identify any crucial residues for the function of CfaS. This study used a combination of bioinformatic techniques including the use of AlphaFold to obtain structures of the CfaS protein as well as InterPro for gathering details of the domains included within the protein. Results showed a number of domains within the CfaS sequence, including part of a Vaccina virus protein for which the structure was further investigated showing similarity with the CfaS structure. A “pocket” binding surface was identified as residues around this structure in both proteins were conserved. Previous evidence showed a similar binding surface/pocket in SAM methyltransferases supporting the theory that proteins from this family would be structurally conserved. CfaS homologues retrieved from UniProt, and their corresponding nucleotide sequences were used to produce a maximum likelihood phylogenetic tree. Using the data from the phylogenetic tree and a BLAST protein search which showed most homologues of CfaS were from the Enterobacteriales class, this suggested that CfaS transferred from bacteria to trypanosomatids through horizontal gene transfer.

238 words

Lay Abstract

Leishmania is a eukaryotic parasite that causes the disease leishmaniasis. It is the third most important vector-borne parasitic disease leading to 60,000 deaths and infecting up to 1.2 million people worldwide. CfaS (cyclopropane fatty-acyl phospholipid synthetase) is a protein that was investigated in this project as a possible required protein for the progression of *Leishmania*

infantum, one of the Leishmania species. Using various bioinformatic techniques, the evolution of CfaS was studied to identify any regions of the protein.
Tools used included AlphaFold to retrieve a predicted 3D protein structure of CfaS and InterPro to gather details about domains in CfaS (a domain is a functional unit of a protein).
Results revealed that part of a Vaccinia virus protein was found as a domain within CfaS and both displayed a similar structure with a “pocket”/binding region. This is possibly due to both proteins being from the same family of methyltransferases, a type of enzyme. Proteins similar to the amino acid sequence of CfaS are known as homologues. These were used in another bioinformatic tool known as Galaxy to produce a phylogenetic tree. This enabled the display of the evolutionary history of CfaS. Finally, using the database BLAST, further homologues of CfaS were found, most of which were from bacterial species. This suggested that CfaS transferred from bacteria to eukaryotes through the movement of genetic material, known as horizontal gene transfer.

228 words

Acknowledgements

Thank you immensely to my supervisor, Dr Leighton Pritchard who has given me plenty of support, advice, and encouragement. Especially grateful for his patience and expert guidance.

Thank you also to the students in my project group.

Introduction

The aim of this project was to investigate the evolutionary and structural analysis of the CfaS protein in *Leishmania infantum* to potentially identify candidate regions that could be plausible drug targets. This was performed using various bioinformatic tools to investigate the sequence evolution and identifying residues that may be required for the function of CfaS.
It was hypothesized that CfaS was involved in horizontal gene transfer from bacteria into the Leishmania species and that proteins such as CfaS which are SAM methyltransferases show conservation across kingdoms.

Leishmania are protozoan parasites which can infect humans as well as other mammalian species. They cause a wide-ranging spectrum of diseases called leishmaniases which is split into 4 forms depending on the particular parasite species:

1) Cutaneous is the most common form of leishmaniases and causes skin lesions.
2) Diffuse cutaneous is a diffuse form of cutaneous leishmaniasis causing diffuse lesions all over the body.
3) Mucocutaneous causes destruction of mucous membranes of the nose, throat and mouth by the formation of lesions in these areas.
4) Visceral is the most severe form of leishmaniases affecting the internal organs, specifically the liver and spleen and is fatal if left untreated (David&Craft, 2009).

Disease incidence is estimated at 2.5 million new cases every year (Bras-Gonçalves et al, 2014). The leishmania genome contains 36 chromosomes with varying sizes of 0.35 to -3 Mb (Wincker et al, 1996).

Leishmania parasites are transmitted as promastigotes, these are flagellated cells and are part of the extracellular stage of the leishmania life cycle. Promastigotes are transmitted to the host via the bite of an infected phlebotomine sand fly (*phlebotominae*) making sand flies the vector for leishmania (Grevelink & Lerner, 1996). Once in the host, promastigotes are broken down, or phagocytosed into the endosomal compartments of primary immune cells such as macrophages where they differentiate into amastigotes. These are non-mobile and aflagellated (without flagellum) and are part of the intercellular stage of the leishmania life cycle. Amastigotes will then continue to infect different organs depending on the leishmania species (Gupta et al, 2013).The life cycle of Leishmania is shown in Figure 1.

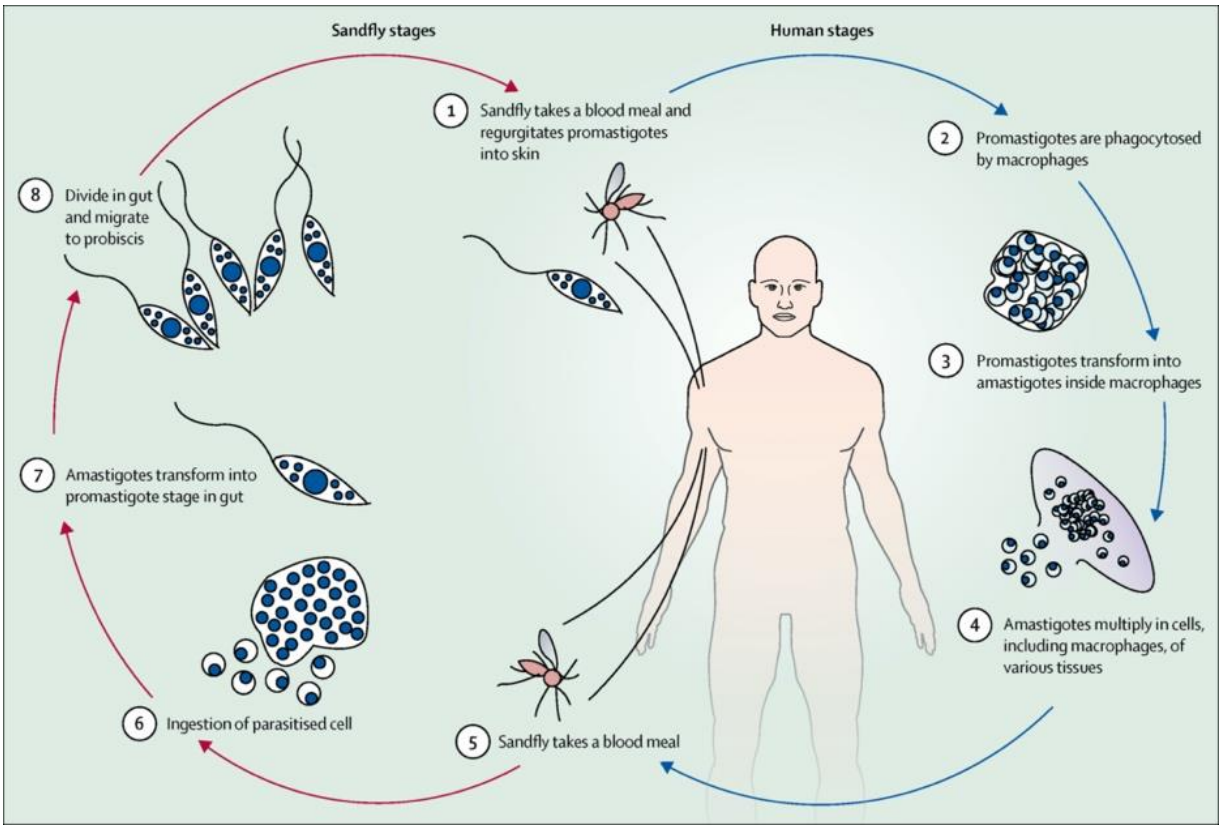

Figure 1: Leishmania life-cycle, showing the transmission of Leishmania to mammalian hosts by sand flies. Reproduced from (Burza et al,2018).

In order for the parasites to spread and establish an infection, Leishmania undergo many morphological and biochemical changes. There are several virulence factors that are required for the infectivity of Leishmania in the host (Table 2).

Virulence factor	Function within Leishmania
Lipophosphoglycan (LPG) (Spath et al,2000)	- Protection against complement - Deactivates macrophage signalling pathway – mitigates host immune response - Aids with macrophage adhesion
GP63 metalloprotease (Olivier et al,2012)	- Prevents excessive triggering of innate immune response - Protects amastigotes from harsh environment in macrophages - Aids with macrophage adhesion
Proteinases (McKerrow et al,2006) (Silva-Almeida et al,2012)	- Involved in migration of parasite through host tissue barriers - Degrade haemoglobin and immunoglobulins
Ecto-Nuleotidases (Peres et al,2018),	- Reduces IFN-γ and increases expression of IL- 10 which in turn decreases immune response - Promotes infection especially within Leishmania Infantum
Heat shock proteins (HSPs) (Prasanna et al,2021)	- Involved in the transformation of promastigote-to-amastigote - Aids in the proliferation of amastigotes within macrophages
Acid phosphatases (ACP) (Singla et al,1992) (Remaley et al,1985)	- Inhibits production of toxic oxygen metabolite by macrophages - Prevent production of superoxide anions which produce free radicals - Aids promastigote virulence
Kinetoplastid Membrane Protein (KMP-11) (Lacerda et al,2012) Elmahallaway et al,2021)	- Assists with attachment to the surface of host cells - Increases expression of IL-10 and reduces NO production, exacerbating

Table 1: Major Leishmania virulence factors and their function.

Leishmania, depending on the particular species has a number of reservoir hosts including humans, rodents and domestic animals. *Leishmania infantum* specifically is the causative agent of visceral leishmaniasis and table 1 shows details about its geographical distribution and hosts (Alemayehu et al, 2017).

Species	Disease form	Geographical location	Hosts
Leishmania Infantum	Visceral leishmaniasis	Middle East and Central Asia to Pakistan, South Europe, Central/South America, Northwest Africa	Foxes, dogs, cats, jackals, humans

Table 2*: Leishmania infantum* background (Alemayehu et al, 2017; Medkour et all, 2019).

Cyclopropane fatty-acyl synthetase (CfaS) is a protein found within *Leishmania infantum* (*L.infantum*)

A crystal structure for CfaS is not available, but its structure has been predicted by AlphaFold2 (Figure 3). AlphaFold2 is an artificial intelligence program that directly predicts a 3D protein structure from its amino acid sequence. It is done using geometric, physical and evolutionary constraints of protein structures (Jumper et al,2021).

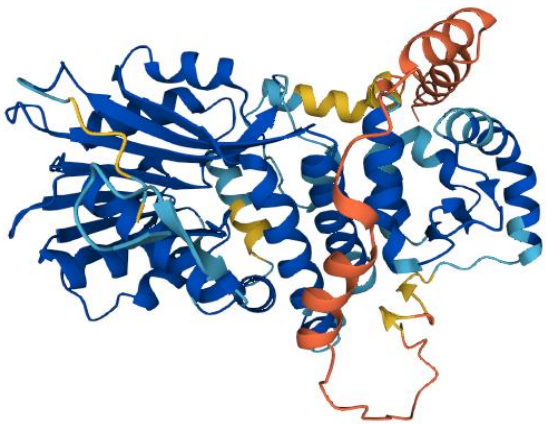

Figure 2: 3D structure of CfaS predicted by AlphaFold

The reverse transcription quantitative real-time PCR (RT-qPCR) technique has been used to analyse CfaS mRNA levels during the life cycle of *L.infantum.* CfaS is encoded by the gene *cfaS* which is found on chromosome 8. Transcripts of *cfaS* were found in promastigote and amastigote parasite forms so it is expressed in both the intracellular and extracellular stages of a parasite life cycle with higher levels of stable mRNA detected in the intracellular amastigotes by two-fold.

CfaS is involved in fatty acid modification in *L.infantum*, It catalyses the transfer of a methylene group from S-adenosyl-L- methionine (SAM or AdoMet) to an unsaturated fatty acid, resulting in a cyclopropane fatty acid (CFA), making it part of the SAM methyltransferase family. The chemical equation showing the formation of these fatty acids is shown in Figure 3.

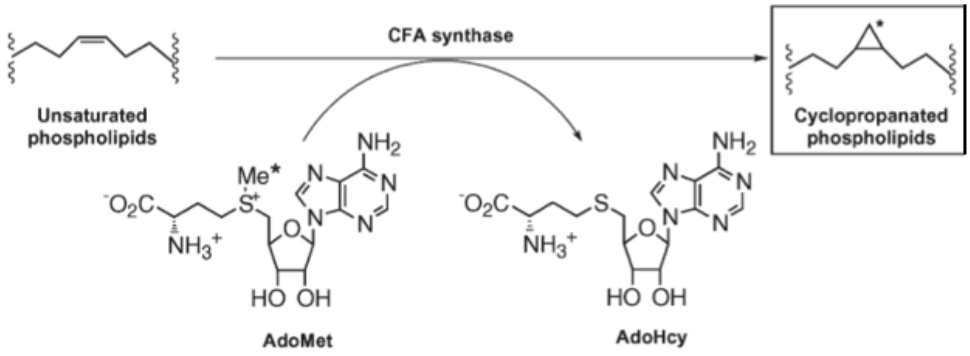

Figure 3: Reaction catalysed by enzyme CfaS (CFA synthase). CfaS uses unsaturated phospholipids and Ado-Met as substrates and Ado-Hcy (S-adenosyl -homocysteine) and cyclopropanated fatty acids are the products of this reaction. Reproduced from (Guianvarc'h et al, 2006).

This *cfaS* gene encodes a 55kDa protein and it shares 48 percent amino acid identity with CfaS-encoding genes of *E.coli* and *Mycobacterium tuberculosis* , cyclopropanated fatty acids have been studied in the most depth in these species. The synthesis of cyclopropane fatty acids have shown an adaptive response to low pH and aeration conditions in *E.coli* (Xu et al, 2017). The addition of cyclopropane fatty acids in the cell membrane of *E.coli* alters lipid composition and membrane proteins thus maintains the cell membranes integrity (Bianco et al,2019).

Although the Leishmania species encode this CfaS protein, Trypanosoma lack this gene (Aquino et al, 2021).

It has also been observed that the loss of the *cfaS* gene in *L.infantum* does not affect promastigote growth or the phagocytosis by macrophages *in vitro* however it does to affect its resistance to oxidative stress. The proline -transporter has also been shown to be defective in the null-parasite (not encoding CfaS protein) suggesting the modification of these phospholipids possibly affects nutrient uptake and virulence in *L.infantum* (Oyola et al, 2012).

Primary sequence alignment was performed to find the S-adenosyl-L-methionine (SAM or AdoMet) binding domain within the CfaS sequence. S-adenosyl-L-methionine is involved in many chemical reactions and after ATP it is the most versatile small molecule ligand (Gana et al, 2013). Using SAM as a cofactor, CfaS produces fatty acids which are essential in Leishmania. S-adenosyl-L-methionine methyltransferases are typically involved in many pathways such as signal transduction, protein repair, biosynthesis, and chromatin regulation (Schubert et al, 2003).

S-adenosyl-L-methionine (SAM) utilising methyltransferases are a large and varied class of methyltransferases which there are approximately 200 known or putative members, they transfer the sulfonium methyl group of SAM to carbon, nitrogen, oxygen and other heteroatom centres on biomolecules (Rosen et al,2016).This methylation of biomolecules using SAM plays a crucial role in many disease processes (Struck et al, 2012). Other methyltransferases such as PRMTs, which are SAM dependant enzymes have been shown to have an important role in Leishmania by regulating RNA-binding protein expression and its function demonstrating their importance (Ferreira et al, 2014)

Along with SAM methyltransferases in *Leishmania infantum* there are other amino acids which are characteristic of this class of enzymes that have been identified as structurally conserved (Oyola et al, 2012).

The majority of SAM-methyltransferase enzymes adopt the Rossman-like fold (Medvedev et al, 2019). A structure of SAM is shown in figure 4.

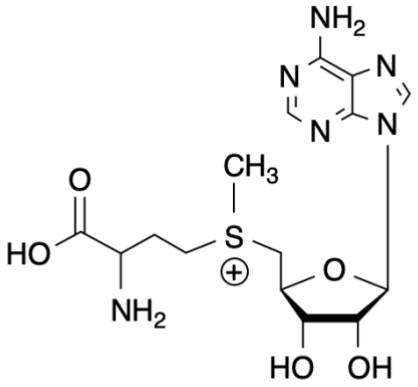

Figure 4: The chemical structure of methyl donor; SAM or AdoMet. (Martin & McMillan,2002)

The mycolic acid cyclopropane synthetase [A0A504Y0L9] from *Leishmania donovani* is a SAM methyltransferase which has 90% amino acid identity with CfaS.

Mycolic acids are long chain branched fatty acids that are found covalently attached to the cell wall and have been shown to be required for the virulence and persistence of mycobacterium tuberculosis and have the potential to be a drug target (Glickman et al, 2000). There are three mycolic acid cyclopropane synthetases known which are PcaA, CmaA1, and CmaA2 and these are responsible for specific modifications of mycolic acid. Genes encoding CmaS have revealed up to 75% identity with other SAM-methyltransferases (Huang et al, 2002). Since CmaS has similar action on mycolic acids as CfaS has on fatty acids, they may have converged on a similar function, this hypothesis was investigated further on.

As well as the sequence being 90% similar, CmaS protein has also been identified to share a domain with CfaS. A functional domain is a region within the protein polypeptide chain that is responsible for a specific role and contributes to the function of the protein.
Besides CmaS, many other methyltransferases were found to share domains within CfaS including ubiE/COQ5 methyltransferase, involved in electron transport (Kurosu et Begari, 2010) and Arginine N-methyltransferase which belongs to the SAM methyltransferase family (Gary et al, 1998).

CmaS and other domains of CfaS were further investigated in this project.

Methodology

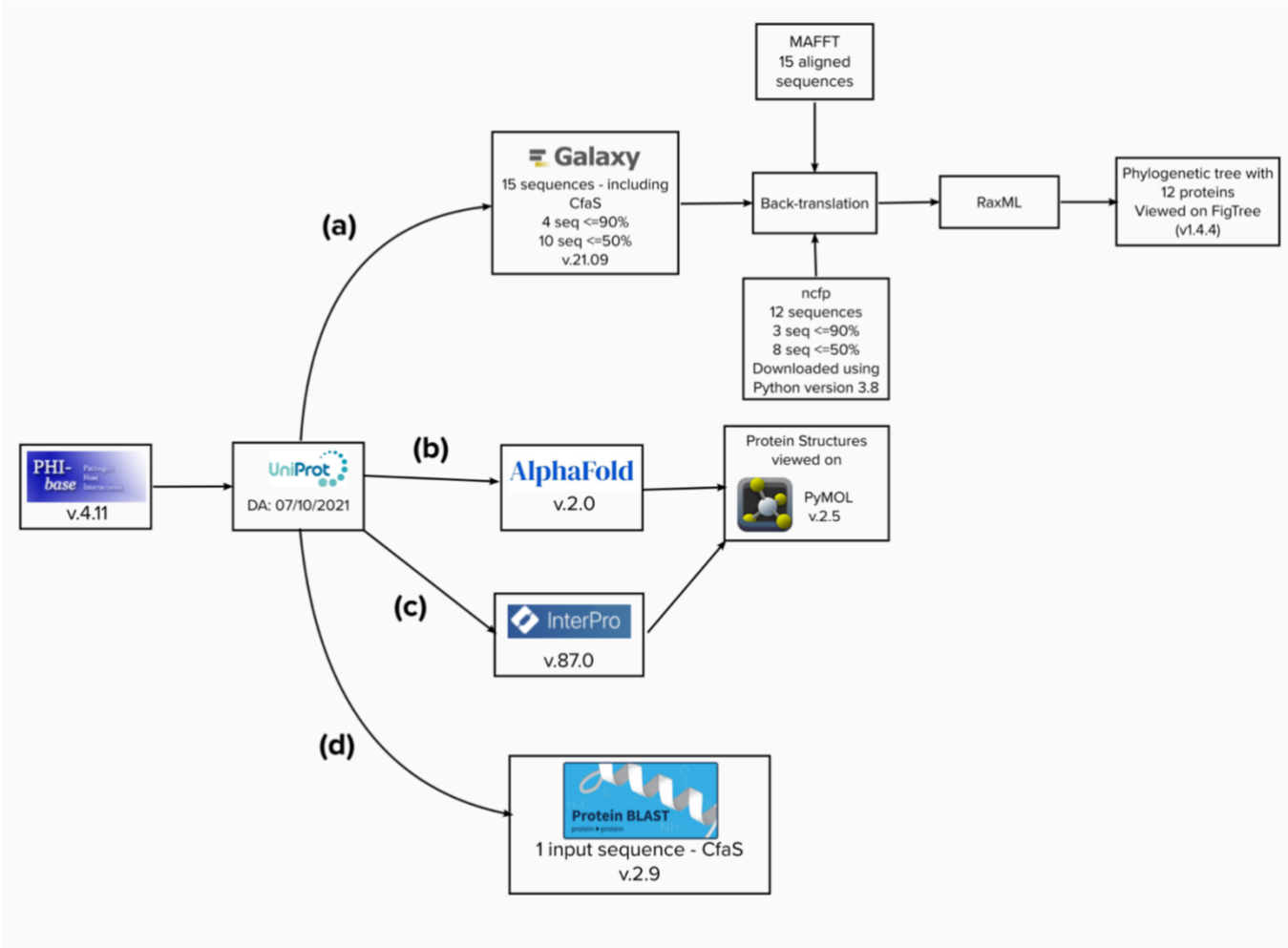

Figure 5: Methodology flow chart.
Databases accessed for (a) Producing a phylogenetic tree using tools within Galaxy and ncfp (b) Visualising 3D protein structures using PyMOL (c) Examining predicted functional domains using InterPro and (d) Investigating evolutionary history of CfaS using BLAST.

To study the evolutionary and structural analysis of CfaS, a combination of bioinformatic techniques were used, described below.

The first step was to access the Pathogen Host Interactions-Base (PHI-Base) record for CfaS, PHI:2643 to view the Uniprot record of CfaS which was identified as “A4HTK3” (Urban et al 2020; Bateman et al,2021).

Producing the Phylogenetic tree (Figure 5(a))

The use of UniProt included details about the proteins function and family as well as 14 homologues of CfaS which all had 50% or 90% sequence similarity.

The FASTA sequence of these homologues along with CfaS were downloaded from the UniProt Knowledgebase (UniProtKB). These 14 sequences along with CfaS, the query protein was uploaded onto MAFFT (Multiple Alignment using Fast Fourier Transform) on Galaxy (Afgan et al,2016) which produced a multiple sequence alignment file containing 15 trimmed and aligned sequences (including CfaS). A corresponding nucleotide sequences file was created using ncfp (https://ncfp.readthedocs.io/en/latest/about.html), only 12 of the 15 sequences successfully converted to nucleotides. Following this, a python script was written on Spyder to rename all the nucleotide sequences in order for them to be read by the back-translate tool (https://github.com/am7861/completed_projects/blob/main/4th_year_project/rename_nt.py).

The renamed nucleotide sequences file and the aligned amino acid sequences from MAFFT were uploaded to Galaxy and put through the “Thread Nucleotides onto a Protein Alignment (back-translation)” tool to produce a codon aware nucleotide alignment.
The subsequent file produced was used as the input file for Randomized Axelerated Maximum Likelihood (RaxML) where 150 bootstrap runs were entered, the algorithm chosen was “Rapid bootstrapping and best ML tree search” and the substitution model was GTR -GAMMA. A phylogenetic tree was the output, created using the maximum likelihood method and consisting of 12 proteins. FigTree was used to visualise the resulting tree.

Visualisation of 3D structures (Figure 5(b and c)).

CfaS

UniProt also displayed a predicted 3D structure of CfaS from the AlphaFold Protein Structure Database.
This CfaS structure was downloaded as a PDB file-type from UniProt and to better visualise and annotate this structure, it was opened on the molecular graphics tool, PyMOL.

CfaS domains from InterPro

The InterPro record (Mitchell et al, 2019) for CfaS was also found on UniProt, recognised by the same ID, “A4HTK3”.

This database was used to identify any predicted functional domains of CfaS, this included the Vaccinia virus protein. This domain was also viewed on PyMOL by downloading it as PDB file on the RCSB Protein Data Bank (Burley et al,2019, v.3.1) identified by '1vp3'. Using the command 'super' (https://pymolwiki.org/index.php/Super) resulted in the best alignment of CfaS and Vaccinia virus protein with the lowest root-mean-square -deviation (RMSD) score, this score shows the degree of similarity between 2 aligned structures. PyMOL was used to further analyse CfaS and the Vaccinia virus protein by colour-coding residues within binding regions.
The sequences of CfaS and Vaccinia virus protein along with homologues of CfaS with at least 50% sequence similarity (from UniProt) were aligned once again using MAFFT resulting in 15 aligned sequences and were uploaded to Jalview (Waterhouse et al, 2009, v.2.11) to identify any conserved patterns in the sequence. Conservation was further studied on the 3D structure of Vaccinia virus protein using ConSurf (Ashkenazy et al, 2016, v.2.42) revealing any functional regions on the proteins.

BLAST search (Figure 5(d))

To retrieve more homologues of CfaS giving an indication whether horizontal gene transfer taking place, a protein BLAST search was performed on the National Center for Biotechnology Information (NCBI) (Johnson et al,2008, v.2.9.0). The CfaS FASTA sequence was used as the input sequence and the non-redundant protein database was searched resulting in proteins similar to the queried sequence. Using this data, a taxonomy table was created.

Default parameters were used for all software unless specified.

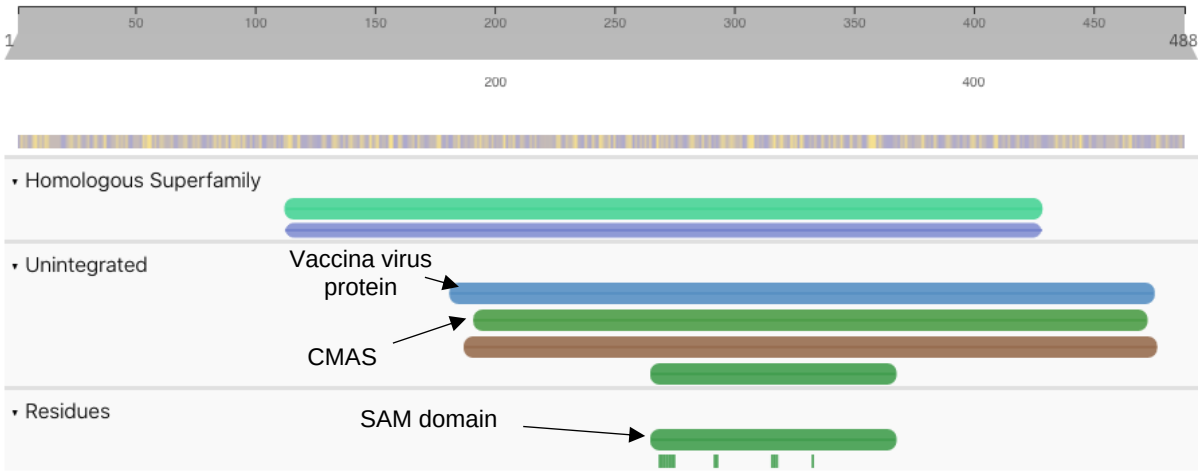

Figure 6: InterPro record (A4HKT3) showing predicted domains of CfaS.

The UniProt record for the PHI-Base protein indicated that InterPro analysis of CfaS (A4HKT3) contained domains S-adenosyl-l-methionine (SAM) domain, Mycolic acid cyclopropane synthetase and the Vaccinia virus protein (VP39) (Figure 6). These domains are consistent with known functions of CfaS, which is they have SAM methyltransferase activity. Analysis into the Vaccinia virus protein is shown further on (page 20).

Taxonomy	Number of Organisms
Trypanosomatidae	12
Leishmaniiae	10
Angomonas deanei	1
Phytomonas sp. Isolate EM1	1
Enterobacteriales	56
Yersiniaceae	46
Pragia	2
Enterobacillus tribolii	1
Morganellaceae	5
Enterbacteriaceae	2

Table 3: Homologues of CfaS using BLAST, searching the non-redundant protein sequences database. Two main families shown in bold.

Performed as described in the methods, BLAST queries of CfaS against non-redundant database identified 100 homologues, 67 different organisms (Table 3).

Twelve proteins from the Trypanosomatidae class are similar to CfaS, with a high query coverage of 83-100% and a percentage identity ranging from 52 - 99% between the twelve proteins. The query coverage is the percentage of the CfaS sequence that is being aligned to the protein in the BLAST database and the percentage identity is the extent to which the aligned sequences are identical.

Ten of the twelve proteins belonged to *Leishmaniiae*, one to *Angomonas deanei* and one to the *Phytomonas* isolate species. The *Angomonas* and *Phytomonas* proteins had a lower query coverage (83%) and percentage identity of 52% and 64% compared to the *Leishmaniiae* proteins which had a query coverage of 93% up to 100% and a percentage identity of 83% up to 99%.

Most of the proteins, a total of 56, that had sequence similarity with CfaS belonged to the Enterobacteriales class. These proteins had a query coverage ranging from 80 – 83% with a lower percentage identity of 47.63 to 49.87%. Within Enterobacteriales, 46 proteins belonged to *Yersiniaceae* family, two to *Pragia*, one to *Enterobacillus tribolii*, five to *Morganellaceae* and two to *Enterobacteriaceae*. Most proteins were from the genus *yersinia* and *serratia.*

While UniProt correlates to some of the homologues found by the BLAST database, the BLAST database included many proteins from the Enterobacteriales class whereas UniProt did not. This is because these bacterial proteins were all under 50% sequence similarity with CfaS.

Since most CfaS homologues are from bacterial organisms, this could possibly suggest that CfaS transferred from bacteria to leishmania through horizontal gene transfer.

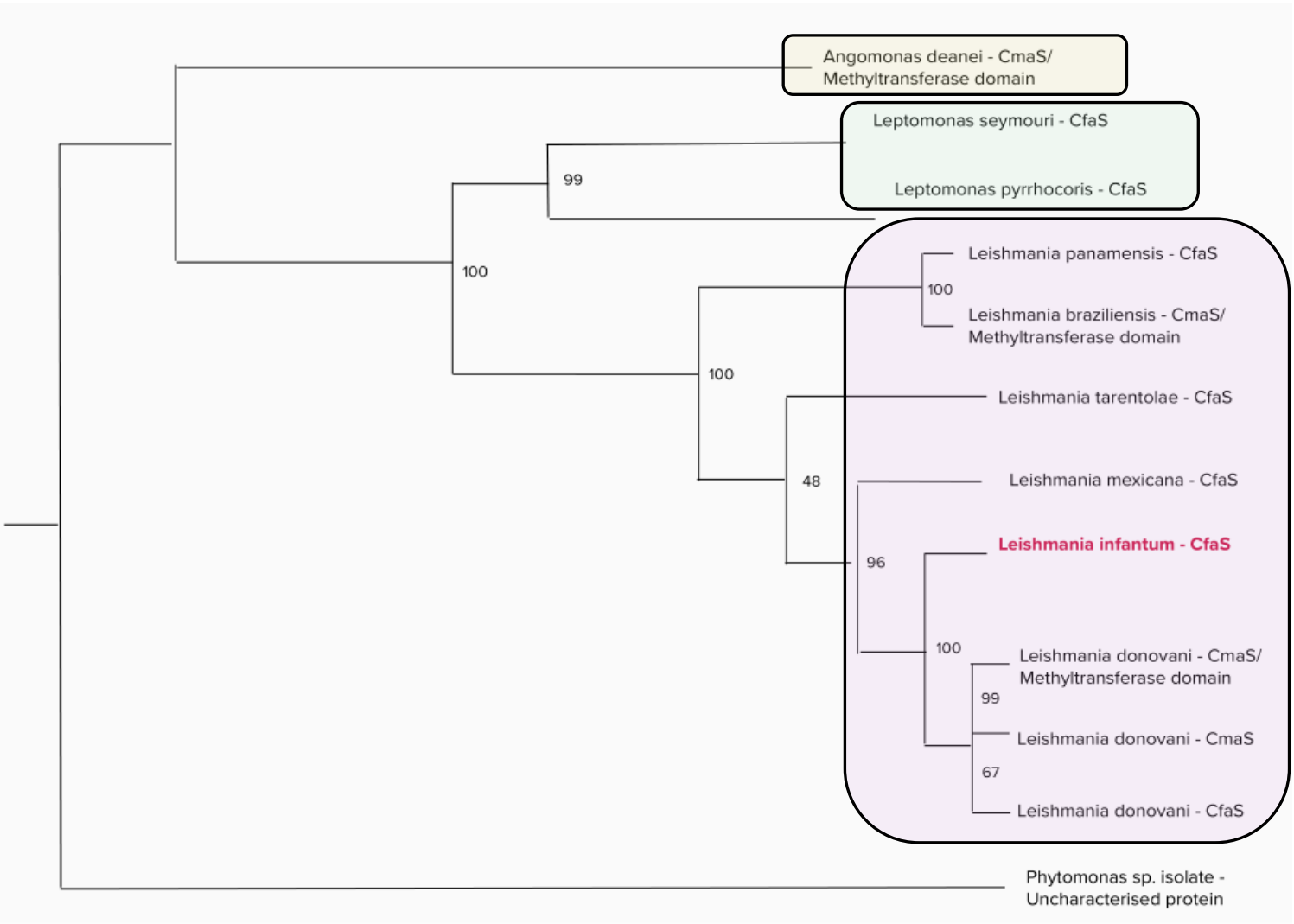

Figure 8: Phylogenetic tree of CfaS and its 50% similar proteins calculated using maximum likelihood on RaxML with 150 bootstrap runs. Colours indicating 3 different sister clades.

As described in methods, CfaS and its 14 homologues (Table 3) along with their corresponding nucleotide sequences were back translated through Galaxy to produce a phylogenetic tree (Figure 8). Only 11 out of the 14 homologues were successful at converting to nucleotides and these along with CfaS made up the phylogenetic tree. Using FigTree, the phylogenetic tree was midpoint rooted and one main clade was determined from the tree, containing three sister clades split into the different genus: *Leishmania*, *Leptomonas* and *Angomonas*.

The largest sister clade with 8 of the 12 proteins belonged to Leishmania, this sister clade contained CfaS in *Leishmania panamensis,* CmaS in *Leishmania braziliensis*, CfaS in *Leishmania*

tarentolae, CfaS in *Leishmania infantum*, and the most closely related group with two CmaS in *Leishmania donovani* and CfaS *in Leishmania donovani*.

The second largest sister clade with 2 of the 12 proteins included CfaS in *Leptomonas seymouri* and CfaS in *Leptomonas pyrrhocoris.*

Finally, the last sister clade contained CmaS in *Angomonas deanei*. Most proteins showed high bootstrap values.

Outside of the clade, an outgroup was identified after mid-rooting the tree containing an uncharacterised protein in the species *Phytomonas*. Specifically looking at CfaS, it only appears to be present in Leishmania and Leptomonas species.

CmaS in *Angomonas deanei and* the uncharacterised protein in *Phytomonas* imply that they are the most recently diverged organisms from a shared common ancestor, determined by the position they are on the tree. All proteins, other than the uncharacterized protein form *Phytomonas* belong to the SAM methyltransferase family.

CmaS has a high protein percentage identity (90%) with CfaS. However, the data from the phylogenetic tree suggests CmaS may have converged on a similar function from CfaS as well as having a high sequence identity.

As stated in Figure 8, no Enterobacteriales are shown on this phylogenetic tree as UniProt doesn’t show any homologues with under 50 % sequence similarity.

Protein name and organism	UniProt entry name
CfaS – Leishmania mexicana	E9AMD4_LEIMU
CmaS – Leishmania donovani	A0A3S5H657_LEIDO
CfaS – Leishmania braziliensis	A4H5C3_LEIBR
CmaS – Angomonas deanei	A0A7G2CBB2_9TRYP
CfaS – Leptomonas pyrrhocoris	A0A0N0VD97_9TRYP
CmaS – Leishmania donovani	A0A504XJ39_LEIDO
CfaS – Leishmania panamensis	A0A088RJ87_9TRYP
CfaS – Leishmania donovani	E9B9G1_LEIDB
Uncharacterised protein – Phytomonas sp. Isolate	W6KGK6_9TRYP
CfaS – Leishmania tarentolae	A0A640KAQ7_LEITA
CfaS – Leptomonas seymouri	A0A0N0P927_LEPSE
CmaS – Leishmania braziliensis	A0A3P3YYW4_LEIBR
CfaS – Leishmania guyanensis	A0A1E1IQ06_LEIGU
CmaS – Leptomonas donovani	A0A504Y0L9_LEIDO

Table 3: Homologues of CfaS, taken from UniProt. 11 of the 14 homologues, were able to be converted into their corresponding nucleotide sequences, ncfp failed to convert 3 (shown in red). These 11 sequences along with CfaS were used in the production of phylogenetic tree.

Further investigation into the Vaccinia virus protein (VP39) was carried out on PyMOL (RCSB PDB accession = 1vp3).

VP39 as a SAM methyltransferase, like CfaS. It is involved in the maturation of both ends of nascent vaccinia transcripts.
VP39 acts on cap 0 structure by methylating (using SAM) the ribose 2′ OH of the first transcribed nucleotide converting the cap 0 form to cap 1 on the 5’ end of RNA. On the 3’ end it acts in poly (A) tail elongation (Hodel et al,1996).
VP39 appears to fold into a core structure with a seven-stranded beta sheet confined around several alpha helices (Figure 9).

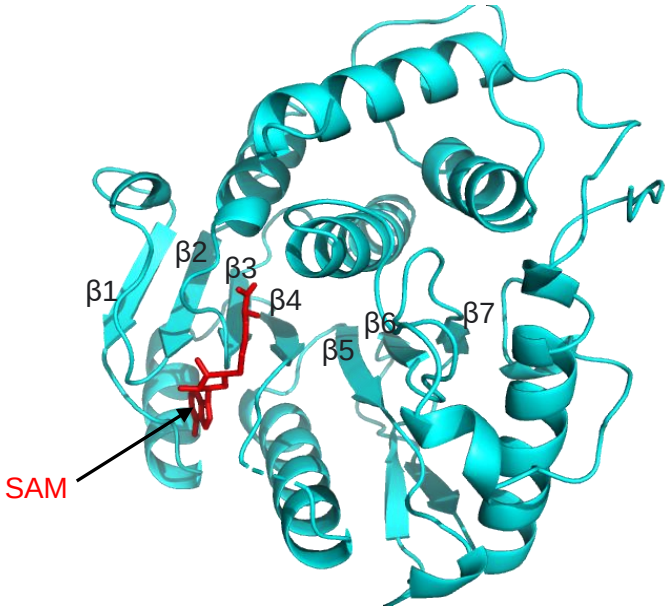

Figure 9: Vaccinia virus protein (VP39) structure displaying seven stranded beta sheet and SAM as the methyl donor. Viewed using PyMOL.

Structures CfaS and VP39 were aligned using the ‘super’ command on PyMOL to obtain an estimate of structural similarity. A number of residues were aligned and compared until a meaningful overlap of the structures was obtained with the lowest RMSD score (Table 4). This alignment is shown in Figure 10.

CfaS residues	VP39 residues	RMSD score (Å)
350-375	121-145	4.707
392-485	150-250	9.950
247-325	39-118	4.340

Table 4: Residues aligned on CfaS with VP39 to identify optimal alignment showing structural similarity

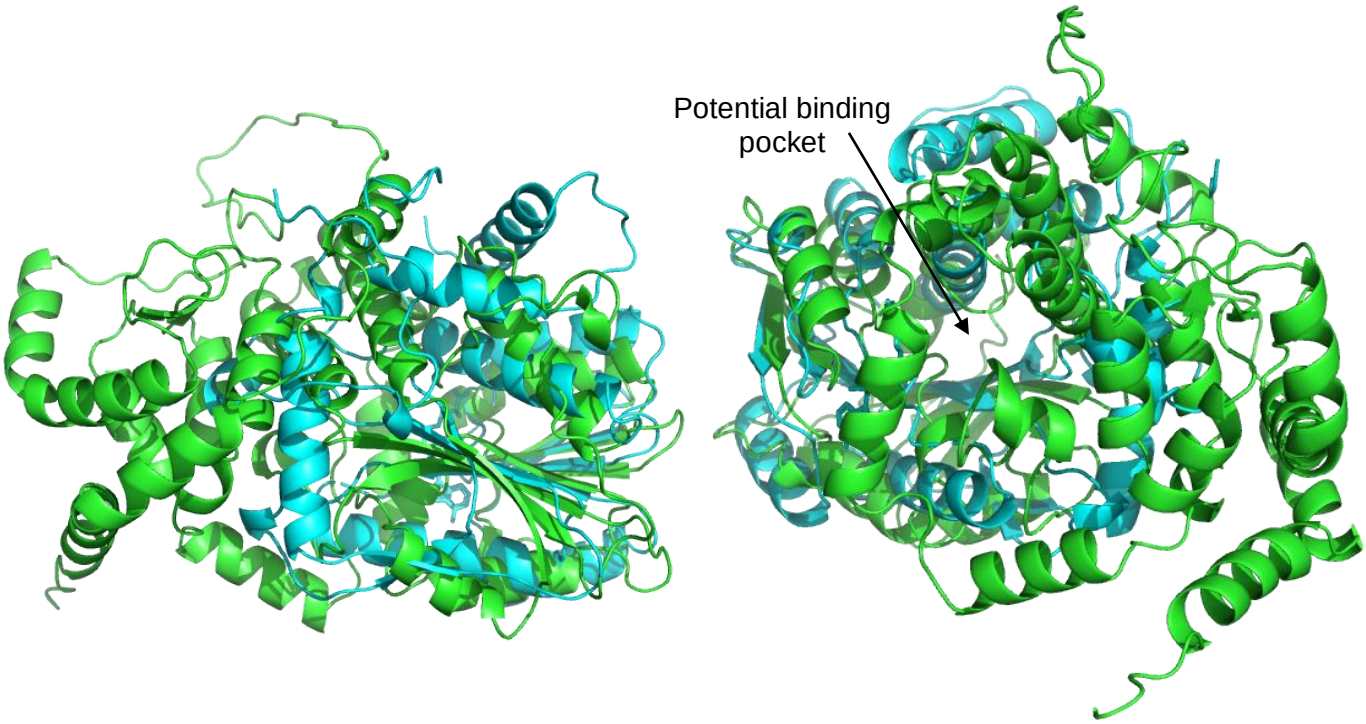

Figure 10: Alignment of CfaS and VP39 at residues 39-118 and 247 – 325. VP39=Cyan, CfaS=Green. Aligned using 'super'.

After aligning CfaS with VP39, a potential binding site/pocket was identified on the proteins, seen by looking directly through the centre of both (right image).

The binding site was then further investigated in VP39, highlighting and colouring all residues within the pocket which looked to be closed at one end and open at the other. The pocket appeared to be in the core of VP39 surrounded by many beta sheets and alpha helixes, but residues selected were specifically found on three beta sheets and an alpha helix parallel to this (Figure 11).

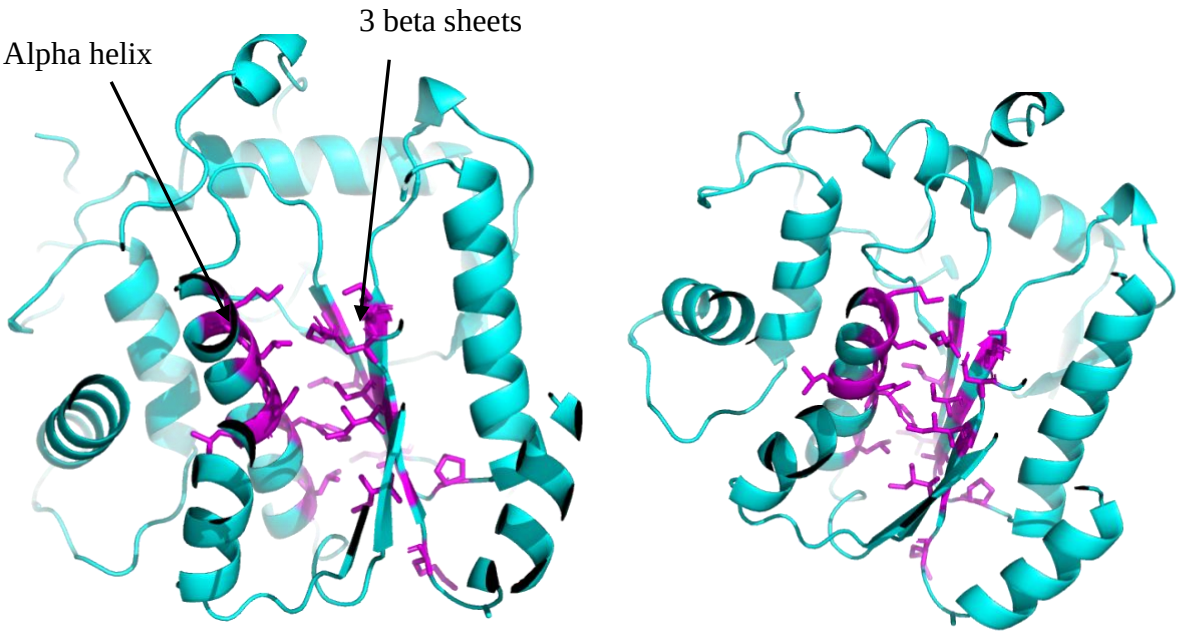

Figure 11: VP39 with binding ‘pocket’ within the structure showed in magenta, viewed on PyMOL.

The ConSurf Server was on used on VP39, CfaS and its homologues to observe any regions of conservation on VP39 structure, the queried protein. A few residues located within the VP39 pocket were highly conserved indicating an evolutionary importance (Figure 12). The surface of VP39 was also viewed also showing conserved residues at the top of this binding region (Figure 13).

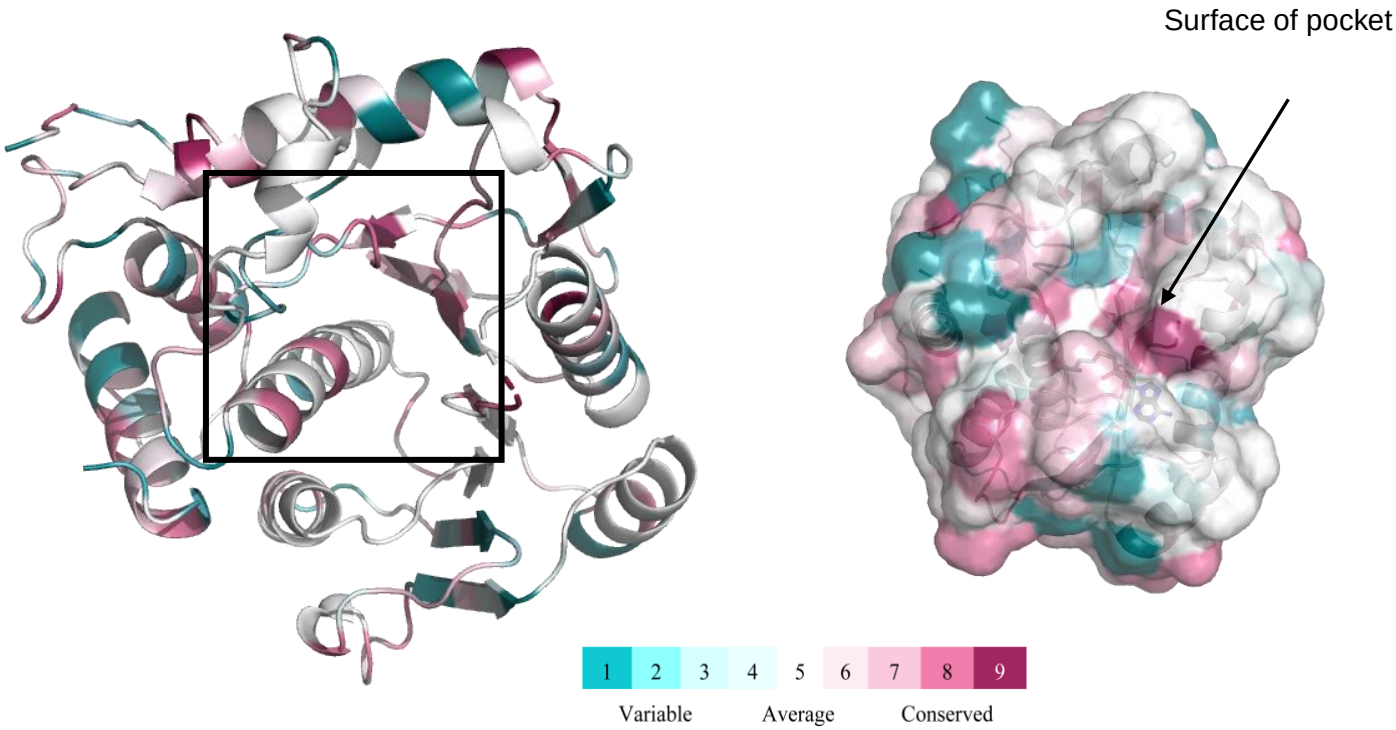

Figure 12: Vaccinia virus protein displaying pocket (left) and surface (right) colour coded using ConSurf and viewed on PyMOL.

The CfaS structure was then investigated to identify if a similar binding pocket was present. Residues were coloured by PyMOL.

Although CfaS protein is a larger protein with 488 resides compared to VP39 with 348, it displays a very structurally similar binding pocket at the bottom half of the protein with several beta sheets and alpha helices parallel to it. While the top half appears to be a tunnel extending from the surface down to the binding pocket (Figure 13).

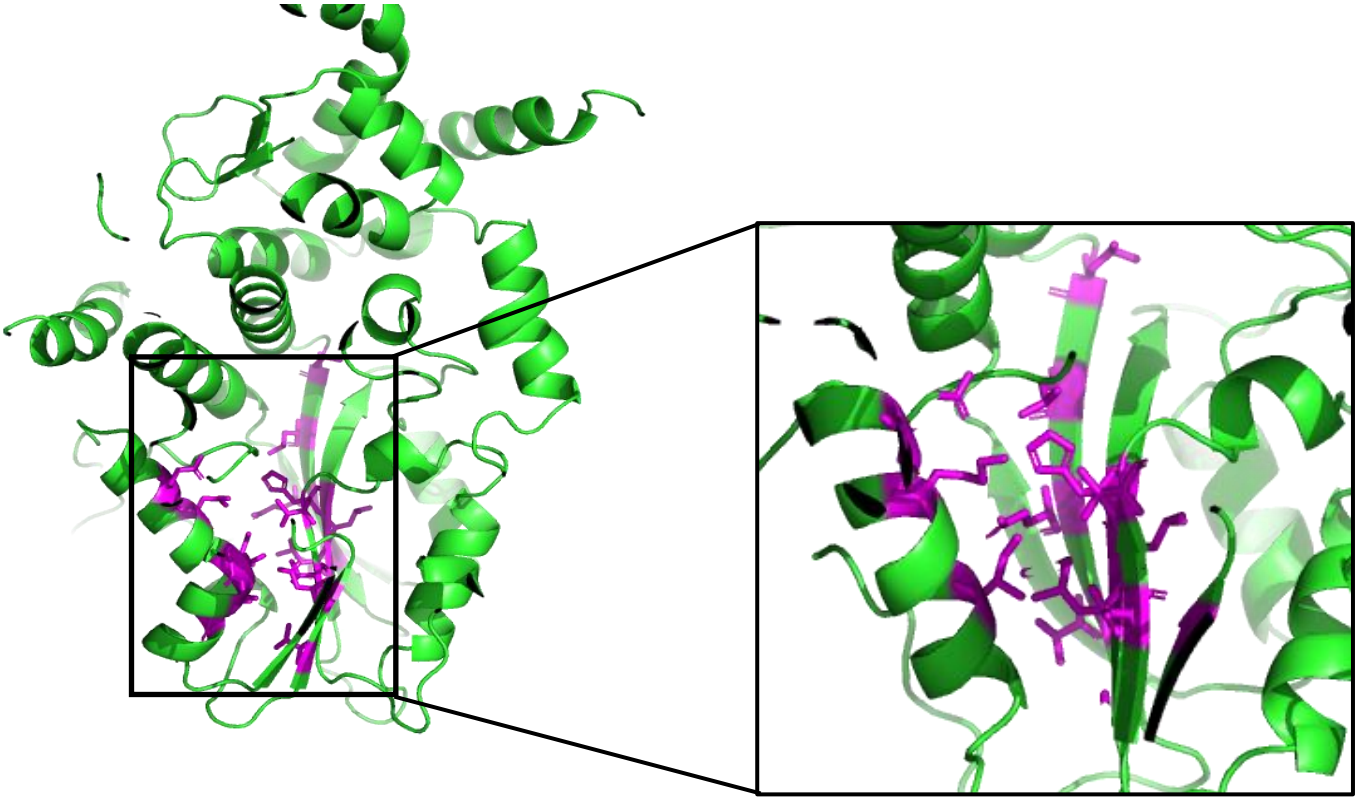

Figure 13: Structure of CfaS, showing binding pocket, viewed on PyMOL

VP39, CfaS and its homologues were then aligned using MAFFT and loaded onto Jalview where any conserved regions along the binding pocket in both CfaS and VP39 were identified when comparing to homologues of CfaS (Figure 14).

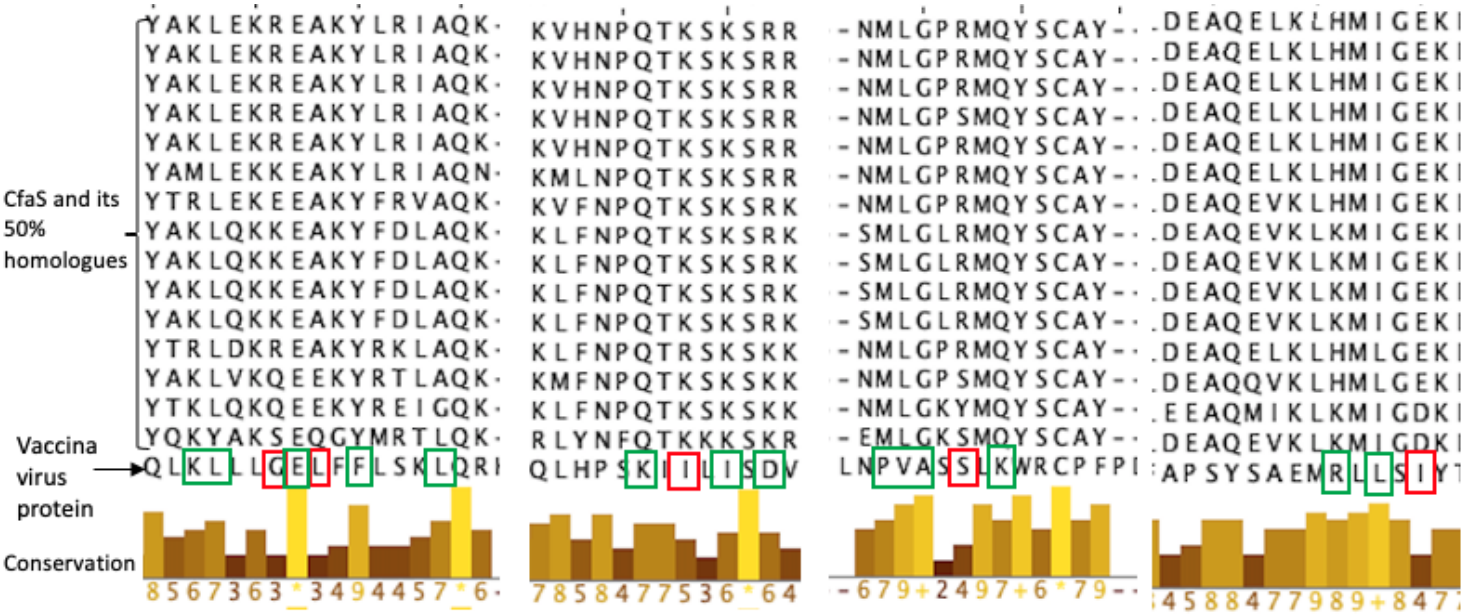

Figure 14: Sequence alignment of VP39, CfaS with >50% sequence identity homologues specifically looking at the residues of VP39. All residues highlighted are amino acids pointing in towards the binding pocket, shown earlier in Figure 11.
Red = low conservation (below 5), green = high conservation (above 5).
Viewed on Jalview.

The sequence alignment shows some variability, with only a few residues showing high conservation. Residues showing conservation included Lys-Leu (41,42bp), Glu(46bp), Phe (49bp),Leu(53bp) Lys (132bp), Ile(136bp), Asp (138bp), Pro-Val-Ala (169,170,171bp), Lys(175bp) and Arg-Leu (209-211bp). The hydrophilic conserved residues are Glu and Asp however most of the conserved residues are hydrophobic suggesting a hydrophobic pocket, this is expected as this binding region points toward the core of the protein.

To compare, the conservation of residues in the binding pocket in CfaS were investigated (Figure 15), using the same sequence alignment in Figure 14.

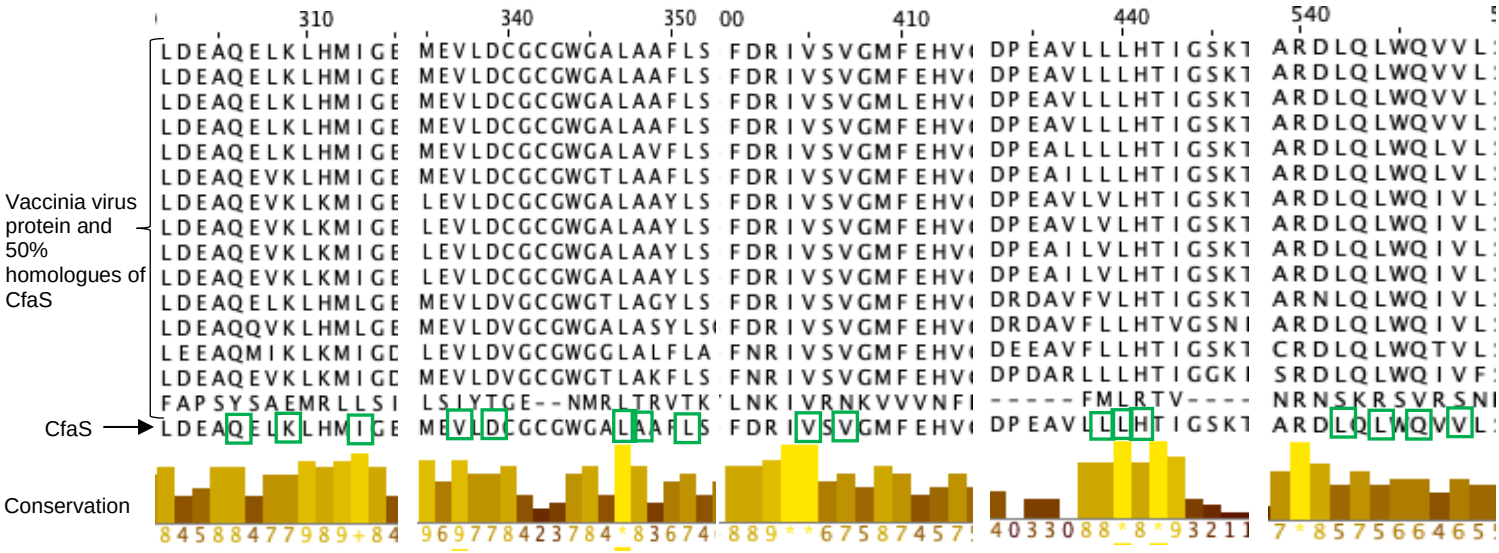

Figure 15: Sequence alignment VP39, CfaS with >50% sequence identity homologues specifically looking at the residues of CfaS. All residues highlighted are amino acids pointing in towards the binding pocket, shown earlier in Figure 14.
Red = low conservation (below 5), green = high conservation (above 5).

The residues surrounding the binding pocket in CfaS are all highly conserved compared to VP39 which showed more variability. Residues showing conservation included Gln (247bp), Lys(250bp), Ile (254bp), Val-Asp(266-268bp), Leu-Ala (276,278), Leu(280bp), Val-Val(331-333), Leu, Leu, His(365,366,367), Leu, Leu, Gln, Val(467-469-471-473). Again, there are some hydrophilic residues present such as His, Lys and Asp but most are hydrophobic.

Since both proteins and CfaS homologues are part of the SAM methyltransferase family, the binding region may possibly be used for a substrate to modify proteins.

Discussion

The purpose of this project was to study the evolutionary and structural analysis of CfaS, this was carried out by investigating the domains and homologues of CfaS to identify if horizontal gene transfer was taking place across species as well as identify any conservation across kingdoms specifically in the SAM methyltransferase family.

The results showed that the Mycolic acid cyclopropane fatty acid synthetase (CmaS) have not only a high protein percentage identity but also seem to have converged on a similar function from the data shown in the phylogenetic tree and previous studies. CmaS modifies mycolic acid and 3 classes of mycolic acids have been found to be produced in mycobacterium tuberculosis according to their modifications at two positions: distal or proximal.
A mycolic acid cyclopropane synthetase, (cmaA1) was found to be one of the proteins modifying these mycolic acids (Verma et al,2020; Yuan et al,1995). This was identified by its high homology to the SAM-dependant methyltransferase, CfaS found in *E.coli* (Wang et al, 1992), and it showed to form cyclopropane rings at the distal region on mycolic acids.
This is similar to the action of CfaS in *Leishmania infantum* as it produces cyclopropane rings but on fatty acids.
This evidence along with the studies showing that cyclopropane rings are a modification found in many bacteria species but limited in eukaryotes suggest the transfer of CfaS from bacterial species to eukaryotes (Sohlenkamp et al,2016; Glickman et al,2000).

A part of CmaS sequence has also been found as a domain within CfaS, its function of modification to mycolic acids may be necessary in Leishmania. The modification of mycolic acids has been shown to play an essential function in Mycobacterium's ability to thrive inside the macrophage as the environment is hostile (Huang et al, 2002; Peterson et al,2019).

It is possible that CfaS has the same function in Leishmania as Leishmania amastigotes also need to withstand environment of macrophages inside the host. As this protein has similar function as CfaS and has been noticed as a potential drug target in *Mycobacterium tuberculosis*, it could also likely be a drug target for *Leishmania infantum*.
CfaS present in bacteria have also shown to provide protection against types of environmental stress such as low oxygen tension and hydrogen peroxide (Jiang et al, 2019).

Horizontal gene transfer was further investigated by the search of the BLAST database that showed many homologues of CfaS belonging to the bacterial species. This correlates with recent studies of trypanosomatids that shows, through evolution, trypanosomatid genomes have modified by the addition of new genes as well as losses by horizontal gene transfer (Maslov et al, 2018).

More than 18 genes were gained, one of which included the involvement of cyclopropane fatty-acyl-phospholipid formation, which is what the protein CfaS forms by transferring a methylene group onto an unsaturated fatty acid (Maslov et al, 2018).

Although cyclopropane fatty acid accounts for only a minor component of the fatty acid content in Leishmania Infantum (Xu et al, 2017) fatty acids are necessary for parasite survival. It's the main energy source during the life cycle stages in particularly Leishmania (Aquino et al, 2021), suggesting a reason for transfer of CfaS into eukaryotic cells.

A previous study (Oyola et al, 2012) has shown that CfaS in Leishmania species is most similar to its bacterial homologues which also correlates to data found using the BLAST database. This data along with the evidence that cyclopropanated fatty acids are rare in eukaryotic cells but are a major plasma membrane component in bacteria further implies the gene *cfaS* was acquired through horizontal gene transfer to Leishmania with secondary loss from *Leishmania major*. This particular Leishmania species doesn't contain CfaS protein and has been hypothesised that it has other mechanisms to compensate for the lack of CfaS and its cyclopropanated fatty acids it produces (Peacock et al, 2007; Okada et al, 2020). However, the expression of a CfaS transgene in *L.major* parasites generates these cyclopropane fatty acids. Which suggests that the substrate for this modification is common to all Leishmania species (Oyola et al, 2012).

Although most of the homologues found by BLAST were bacterial species, the percentage identity was under 50% for all of them compared to the homologues from the trypanosome family that were much greater. This was expected because if horizontal gene transfer occurred from bacteria to eukaryotes, then CfaS would show greater similarity to its homologues within trypansomatids than it would to its bacterial homologues. This can also be inferred from the phylogenetic tree. It displays many CfaS proteins in different Leishmania species, most with high bootstrap values indicating high homology with CfaS in *Leishmania infantum.*

A previous study (Zhang, Beverley, 2019) has shown that MTPs (mannosyltransferase/phosphorylases) that synthesize mannogen (a carbohydrate) are found in Leishmania and Leptomonas species but not in the *Trypanosoma* genus. Phylogenetic analysis from this research showed that the MTP gene possibly was acquired from bacteria, early in the evolution of *Leishmaniiae* by horizontal gene transfer from bacteria. A figure from this study is shown below (Figure 7). Data from BLAST suggests CfaS may have behaved in a similar way.

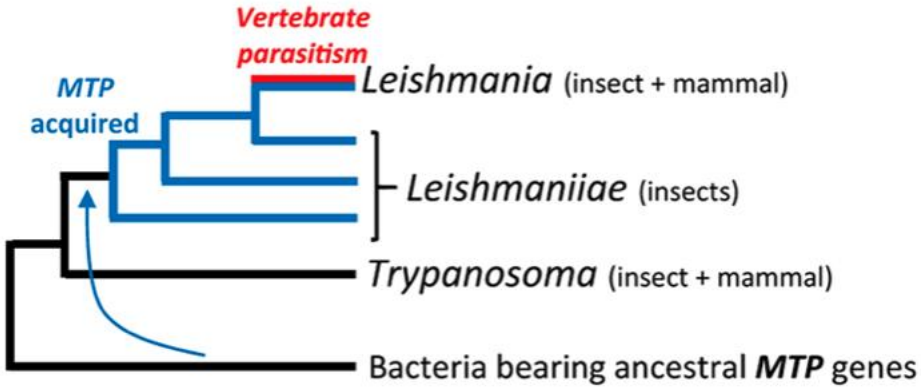

Figure 7: Evolutionary origin of mannogen (Zhang, Beverley,2019; Sernee et al, 2019)

Results from the structural analysis of the Vaccinia virus protein and CfaS displayed structural similarity in both proteins with an indent/pocket. Both proteins are SAM methyltransferases underlining that they use SAM (S-adenosyl-L-methionine) as a cofactor to methylate proteins (Sun et al, 2005).

SAM methyltransferases have a cleft/fold comprised of a seven-stranded beta sheet which is flanked by alpha helices on both sides, resembling the Rossman-like fold that most methyltransferases have (Class I) (Kozbial et al, 2005). The binding site for the co-factor SAM has been shown to be located on the N-terminal part of the B sheet where several active sites have been identified. These aid to stabilize the SAM molecule by an array of van der Walls interactions and hydrogen bonding. It has been demonstrated that SAM can adopt various conformation in the active site (Fyfe et al, 2022). Also, within this cleft, a substrate binding site is found in the C-terminal region of the B sheet. This region amongst these enzymes varies largely in its topology and chemistry. This is most likely due to substrate specificity as SAM methyltransferases methylate many types of distinct proteins and molecules of varying sizes. Depending on the type of SAM methyltransferase. Various substrates are adopted such as carbohydrates, small molecules, RNA, DNA etc.

This binding site appears to be what was found in the Vaccinia virus protein and CfaS. Both proteins have similar structures to other SAM methyltransferases described, specifically they display a beta strand surrounded by alpha helices on each side and a substrate binding domain within this.

Specific studies into the Vaccinia virus protein have showed that SAM binds to the cleft which then deepens to form a hydrophobic pocket, this is possibly where the Vaccinia virus binds to a

substrate and is the pocket identified from the data in this project. The substate that the Vaccinia virus protein has shown to primarily bind to is RNA (Hodel et al,1996).
A comparable structure was also found from studies of CfaS showing a tunnel extending from the surface down to the binding pocket in the central cleft of CfaS which resembled the structure viewed on PyMOL (Sun et al,2021).

The sequence alignment of Vaccinia virus protein, CfaS and homologues of CfaS was performed and specifically looking at the Vaccinia virus binding pocket residues, it showed some conservation but was variable. However, when viewing the residues around CfaS binding pocket, all showed high conservation. The variability of Vaccinia virus protein against the other proteins was expected as generally SAM methyltransferases share a conserved structural fold/pocket but little sequence identity because of the variety of substrates they bind (Martin&McMillan,2002). The reason CfaS showed high conservation across its binding pocket is possibly because all proteins viewed on Jalview other than the Vaccinia virus protein were homologues of CfaS with at least 50% percentage identity. Many of these were CmaS which is also found as a domain in CfaS so it is expected to have high conservation.
Additionally, the homologues were also the same length as CfaS so the binding pocket is likely to be found in the same location along the sequence, unlike the Vaccinia virus protein.

In addition to conservation, the type of amino acids varied in both the binding sites of CfaS and Vaccinia virus. Most residues were hydrophobic, indicting a hydrophobic pocket but hydrophilic residues were also present. As well as varying in sizes, these substrates vary in chemistry, so the amino acids within the substrate binding regions also are expected to vary significantly.

A comparison of Vaccinia virus protein with other SAM methyltransferase revealed that it had evolved from an ancestral fold shared between diverse methyltransferases by the addition of structures necessary for its 5’ and 3’ end-modifying activities. Evidence of its evolution shows that It appears to have tailored its methyltransferase domain instead of acquiring an extra domain purposely for its RNA substrate recognition. Its function is shown to be essential in evading the innate immune response, possibly explaining its presence as a domain in CfaS in *Leishmania infantum* (Hodel et al, 1996; Minnaert et al,2021).

Conclusion

The similar functions of CmaS in mycobacterium and the many bacterial homologues of CfaS leads to the conclusion that horizontal gene transfer of CfaS occurred into the Leishmania species possibly providing a new function allowing Leishmania to survive is host conditions. The structural similarity of the Vaccinia virus protein as well as the functional similarity of CmaS to CfaS supports

the conclusion that there is conservation of SAM methyltransferase proteins across kingdoms. The analysis of CfaS domains as well as the data showing the importance of SAM methyltransferases in the progression on various diseases suggests that CfaS could be used as a plausible drug target. Experiments such as alanine scanning mutagenesis could be performed to further test the function of CfaS.

References

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A. and Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, **44**(W1), pp.W3–W10.

ARHIN, G., 2005. A protein related to the vaccinia virus cap-specific methyltransferase VP39 is involved in cap 4 modification in Trypanosoma brucei. *RNA*.**12**: 53-62

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, **44**(W1), pp.W344–W350.

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Castro, L.G. and Garmiri, P. (2020). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, **49**(D1), pp.D480–D489.

Bianco, C.M., Fröhlich, K.S. and Vanderpool, C.K. (2019). Bacterial Cyclopropane Fatty Acid Synthase mRNA Is Targeted by Activating and Repressing Small RNAs. *Journal of Bacteriology*, **201**(19).

Bras-Gonçalves, R., Petitdidier, E., Pagniez, J., Veyrier, R., Cibrelus, P., Cavaleyra, M., Maquaire, S., Moreaux, J. and Lemesre, J.-L. (2014). Identification and characterization of new Leishmania promastigote surface antigens, LaPSA-38S and LiPSA-50S, as major immunodominant excreted/secreted components of L. amazonensis and L. infantum. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, **24**, pp.1–14.

Burza, S., Croft, S.L. and Boelaert, M. (2018). Leishmaniasis. *The Lancet*, **392**(10151), pp.951–970.

Cheng, X. (1995). Structure and Function of DNA Methyltransferases. *Annual Review of Biophysics and Biomolecular Structure*, **24**(1), pp.293–318.

David, C.V. and Craft, N. (2009). Cutaneous and mucocutaneous leishmaniasis. *Dermatologic Therapy*, **22**(6), pp.491–502.

Elmahallawy, E. and Alkhaldi, A., 2021. Insights into Leishmania Molecules and Their Potential Contribution to the Virulence of the Parasite.*Vet Sci*. **20;**8(2):33

Ferreira, T.R., Alves-Ferreira, E.V.C., Defina, T.P.A., Walrad, P., Papadopoulou, B. and Cruz, A.K. (2014). Altered expression of an RBP-associated arginine methyltransferase 7 in Leishmania major affects parasite infection. *Molecular Microbiology*.

Fyfe, C.D., Bernardo-García, N., Fradale, L., Grimaldi, S., Guillot, A., Brewee, C., Chavas, L.M.G., Legrand, P., Benjdia, A. and Berteau, O. (2022). Crystallographic snapshots of a B12-dependent radical SAM methyltransferase. *Nature*, 602(7896), pp.336–342.

Glickman, M., Cox, J. and Jacobs, W., 2000. A Novel Mycolic Acid Cyclopropane Synthetase Is Required for Cording, Persistence, and Virulence of Mycobacterium tuberculosis. *Mol Cell*. **5**(4):717-27

Guianvarc'h, D., Drujon, T., Leang, T.E., Courtois, F. and Ploux, O. (2006). Identification of new inhibitors of E. coli cyclopropane fatty acid synthase using a colorimetric assay. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1764**(8), pp.1381–1388.

Guja, K.E., Venkataraman, K., Yakubovskaya, E., Shi, H., Mejia, E., Hambardjieva, E., Karzai, A.W. and Garcia-Diaz, M. (2013). Structural basis for S -adenosylmethionine binding and methyltransferase activity by mitochondrial transcription factor B1. *Nucleic Acids Research*, 41(16), pp.7947–7959.

Gupta, G., Oghumu, S. and Satoskar, A.R. (2013). Mechanisms of Immune Evasion in Leishmaniasis. *Advances in Applied Microbiology*, **82**, pp.155–184.

Hanukoglu, I. (2015). Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochemistry and Molecular Biology Education*, **43**(3), pp.206–209.

Hodel, A., Gershon, P., Shi, X. and Quiocho, F., 1996. The 1.85 Å Structure of Vaccinia Protein VP39: A Bifunctional Enzyme That Participates in the Modification of Both mRNA Ends. *Cell*.**85;**2:247-256.

Horning, B., Suciu, R., Ghadiri, D., Ulanovskaya, O., Matthews, M., Lum, K., Backus, K., Brown, S., Rosen, H. and Cravatt, B., 2016. Chemical Proteomic Profiling of Human Methyltransferases. *J Am Chem Soc*. **138**(40): 13335–13343.

Huang, C., Smith, C., Glickman, M., Jacobs, W. and Sacchettini, J., 2002. Crystal Structures of Mycolic Acid Cyclopropane Synthases from Mycobacterium tuberculosis.*J Biol Chem*, **29**;277(13):11559-69

Jiang, X., Duan, Y., Zhou, B., Guo, Q., Wang, H., Hang, X., Zeng, L., Jia, J. and Bi, H. (2019). The Cyclopropane Fatty Acid Synthase Mediates Antibiotic Resistance and Gastric Colonization of Helicobacter pylori. *Journal of Bacteriology*, **201**(20), pp.e00374-19.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36(Web Server), pp.W5–W9.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J. and Back, T. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**(7873), pp.583–589.

Kozbial, P.Z. and Mushegian, A.R. (2005). Natural history of S-adenosylmethionine-binding proteins. *BMC Structural Biology*, **5**(1), p.19.

Kurosu, M. and Begari, E., 2010. *Vitamin K2 in Electron Transport System:* Are Enzymes Involved in Vitamin K2 Biosynthesis Promising Drug Targets?.*Molecules.* 15(3): 1531–1553.

Lacerda, D.I., Cysne-Finkelstein, L., Nunes, M.P., De-Luca, P.M., Genestra, M. da S., Leon, L.L.P., Berrêdo-Pinho, M., Mendonça-Lima, L., Matos, D.C. de S., Medeiros, M.A. and Mendonça, S.C.F. de (2012). Kinetoplastid membrane protein-11 exacerbates infection with Leishmania amazonensis in murine macrophages. *Memorias Do Instituto Oswaldo Cruz*, **107**(2), pp.238–245.

Martin, J., 2002. *SAM (dependent) I AM:* the S-adenosylmethionine-dependent methyltransferase fold. *Current Opinion in Structural Biology*,**3**;1:140

Maslov, D.A., Opperdoes, F.R., Kostygov, A.Y., Hashimi, H., Lukeš, J. and Yurchenko, V. (2019b). Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology*, **146**(1), pp.1–27.

McKerrow, J.H., Caffrey, C., Kelly, B., Loke, P. and Sajid, M. (2006). Proteases in parasitic diseases. *Annual Review of Pathology*, **1,** pp.497–536.

Medkour, H., Davoust, B., Dulieu, F., Maurizi, L., Lamour, T., Marié, J. and Mediannikov, O., 2019. Potential animal reservoirs (dogs and bats) of human visceral leishmaniasis due to Leishmania infantum in French Guiana. *PLoS Negl Trop Dis*, **19;**13(6)

Medvedev, K.E., Kinch, L.N., Schaeffer, R.D. and Grishin, N.V. (2019). Functional analysis of Rossmann-like domains reveals convergent evolution of topology and reaction pathways. *PLOS Computational Biology*, **15**(12), p.e1007569.

Minnaert, A.-K., Vanluchene, H., Verbeke, R., Lentacker, I., De Smedt, S.C., Raemdonck, K., Sanders, N.N. and Remaut, K. (2021). Strategies for controlling the innate immune activity of conventional and self-amplifying mRNA therapeutics: Getting the message across. *Advanced Drug Delivery Reviews*, **176**, p.113900.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.-Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H. and Natale, D.A. (2018). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, **47**(D1), pp.D351–D360.

Okada, S., Taylor, M., Zhou, X.-R., Naim, F., Marshall, D., Blanksby, S.J., Singh, S.P. and Wood, C.C. (2020). Producing Cyclopropane Fatty Acid in Plant Leafy Biomass via Expression of Bacterial and Plant Cyclopropane Fatty Acid Synthases. *Frontiers in Plant Science*, **11**

Olivier, M., Atayde, V.D., Isnard, A., Hassani, K. and Shio, M.T. (2012). Leishmania virulence factors: focus on the metalloprotease GP63. *Microbes and Infection*, **14**(15), pp.1377–1389.

Oyola, S., Evans, K., Smith, T., Smith, B., Hilley, J., Mottram, J., Kaye, P. and Smith, D., 2012. Functional Analysis of Leishmania Cyclopropane Fatty Acid Synthetase.*PLoS ONE*, **7**;12:15

Parreira de Aquino, G., Mendes Gomes, M.A., Köpke Salinas, R. and Laranjeira-Silva, M.F. (2021). Lipid and fatty acid metabolism in trypanosomatids. *Microbial Cell*, **8**(11), pp.262–275.

Peacock, C.S., Seeger, K., Harris, D., Murphy, L., Ruiz, J.C., Quail, M.A., Peters, N., Adlem, E., Tivey, A., Aslett, M., Kerhornou, A., Ivens, A., Fraser, A., Rajandream, M.-A., Carver, T., Norbertczak, H., Chillingworth, T., Hance, Z., Jagels, K. and Moule, S. (2007). Comparative genomic analysis of three Leishmania species that cause diverse human disease. *Nature Genetics*, **39**(7), pp.839–847.

Peres, N.T. de A., Cunha, L.C.S., Barbosa, M.L.A., Santos, M.B., Oliveira, F.A. de, Jesus, A.M.R. de and de Almeida, R.P. (2018). Infection of Human Macrophages by Leishmania infantum Is Influenced by Ecto-Nucleotidases. *Frontiers in Immunology*, **8**.

Peterson, E.J., Bailo, R., Rothchild, A.C., Arrieta-Ortiz, M.L., Kaur, A., Pan, M., Mai, D., Abidi, A.A., Cooper, C., Aderem, A., Bhatt, A. and Baliga, N.S. (2019). Path-seq identifies an essential mycolate remodeling program for mycobacterial host adaptation. *Molecular Systems Biology*, **15**(3).

Prasanna, P. and Upadhyay, A. (2021). Heat Shock Proteins as the Druggable Targets in Leishmaniasis: Promises and Perils. *Infection and Immunity*, **89**(2), pp.e00559-20.

Remaley, A.T., Glew, R.H., Kuhns, D.B., Basford, R.E., Waggoner, A.S., Ernst, L.A. and Pope, M. (1985). Leishmania donovani: Surface membrane acid phosphatase blocks neutrophil oxidative metabolite production. *Experimental Parasitology*, **60**(3), pp.331–341.

Sernee, M.F., Ralton, J.E., Nero, T.L., Sobala, L.F., Kloehn, J., Vieira-Lara, M.A., Cobbold, S.A., Stanton, L., Pires, D.E.V., Hanssen, E., Males, A., Ward, T., Bastidas, L.M., Peet, P.L. van der, Parker, M.W., Ascher, D.B., Williams, S.J., Davies, G.J. and McConville, M.J. (2019). A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in Leishmania Parasites. *Cell Host & Microbe*, **26**(3), pp.385-399.e9.

Schubert, H., Blumenthal, R. and Cheng, X., 2003. Many paths to methyltransfer: a chronicle of convergence. *Cell*.**28**;6:329-335.
Silva-Almeida, M., Pereira, B.A.S., Ribeiro-Guimarães, M.L. and Alves, C.R. (2012). Proteinases as virulence factors in Leishmania spp. infection in mammals. *Parasites & Vectors*, **5**(160), p.160.

Singla, N., Khuller, G.K. and Vinayak, V.K. (1992). Acid phosphatase activity of promastigotes of Leishmania donovani: a marker of virulence. *FEMS microbiology letters*, **73**(3), pp.221–225.

Spath, G.F., Epstein, L., Leader, B., Singer, S.M., Avila, H.A., Turco, S.J. and Beverley, S.M. (2000). Lipophosphoglycan is a virulence factor distinct from related glycoconjugates in the protozoan parasite Leishmania major. *Proceedings of the National Academy of Sciences*, **97**(16), pp.9258–9263.

Sun, Q., Huang, M. and Wei, Y. (2021). Diversity of the reaction mechanisms of SAM-dependent enzymes. *Acta Pharmaceutica Sinica B*, **11**(3), pp.632–650.

Sun, W., Xu, X., Pavlova, M., Edwards, A.M., Joachimiak, A., Savchenko, A. and Christendat, D. (2005). The crystal structure of a novel SAM-dependent methyltransferase PH1915 from Pyrococcus horikoshii. *Protein Science : A Publication of the Protein Society*, **14**(12), pp.3121–3128.

Tang, J., Gary, J., Clarke, S. and Herschman, H., 1998. PRMT 3, a Type I Protein Arginine N-Methyltransferase That Differs from PRMT1 in Its Oligomerization, Subcellular Localization, Substrate Specificity, and Regulation.*JBC*,**273**;27:16935-16945

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y., De Silva, N., Martinez, M.C., Pedro, H., Yates, A.D., Hassani-Pak, K. and Hammond-Kosack, K.E. (2020). PHI-base: the pathogen-host interactions database. *Nucleic Acids Research*, **48**(D1), pp.D613–D620.

Verma, J. and Subbarao, N. (2020). Designing novel inhibitors against cyclopropane mycolic acid synthase 3 (PcaA): targeting dormant state of Mycobacterium tuberculosis. *Journal of Biomolecular Structure and Dynamics*, **39**(17), pp.6339–6354.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25(**9), pp.1189–1191.

Wincker, P., Ravel, C., Blaineau, C., Pages, M., Jauffret, Y., Dedet, J.P. and Bastien, P. (1996). The Leishmania genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Research*, **24**(9), pp.1688–1694.

Xu, W., Mukherjee, S., Ning, Y., Hsu, F.-F. and Zhang, K. (2018). Cyclopropane fatty acid synthesis affects cell shape and acid resistance in Leishmania mexicana. *International Journal for Parasitology*, **48**(3-4), pp.245–256.

Zhang, K. and Beverley, S.M. (2019). Mannogen-ing Central Carbon Metabolism by Leishmania. *Trends in parasitology*, **35**(12), pp.947–949.